Benchmarking Environmental Efficiency of Garment Factories to Understand the Value of Real-Time Environmental Data

by

Jordan Riley Landis
B.S.E Mechanical Engineering and Applied Mechanics,
University of Pennsylvania, 2014

Submitted to MIT Sloan School of Management and the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Mechanical Engineering

In conjunction with the Leaders for Global Operations Program at
The Massachusetts Institute of Technology

May 2020

Signature of Author ______________________________
MIT Department of Mechanical Engineering, MIT Sloan School of Management
May 8, 2020

Certified by ______________________________
Dr. Maria Yang
Thesis Supervisor, Professor of Mechanical Engineering

Certified by ______________________________
Dr. Charles Fine
Thesis Supervisor, Chrysler Leaders for Global Operations Professor of Management

Accepted by ______________________________
Dr. Nicolas Hadjiconstantinou
Chair, Mechanical Engineering Committee on Graduate Students
MIT Department of Mechanical Engineering

Accepted by ______________________________
Maura Herson
Assistant Dean, MBA Program
MIT Sloan School of Management

This page intentionally left blank

.

Benchmarking Environmental Efficiency of Garment Factories to Validate the Requirements for Real-Time Environmental Data

by

Jordan Riley Landis

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering on May 8, 2020 in partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Mechanical Engineering.

Abstract

Li & Fung works with over 10,000 factories distributed across 50 countries to design, produce, and deliver hard- and soft- goods to over 2,000 apparel and consumer goods customers. An increasingly prevalent focus of the industry, driven both by regulation and consumer preferences, is to measure, benchmark, and reduce the overall environmental impact of the supply chain. Currently the measurement mechanisms in place rely on a traditional two-phase approach involving factory self-reporting and verification via independent audits. The scope of this project is to assess the efficacy of currently available measurement data in order to inform the requirements for real-time collected data. This project will be broken into four phases. First, existing industry data sources will be described and evaluated in order to assess data quality, understand requirements, and provide recommendations for future data collection. Second, the features of the data will be analyzed in order to develop an understanding of the underlining relationships. Third, using a set of selected features from the second phase, a predictive clustering algorithm for factory-level resource efficiency will be developed and used to benchmark factories. Finally, an analysis will be performed to evaluate the requirements of real time data and how real-time data could improve the benchmarking tool and future tools and services.

Dr. Maria Yang
Thesis Supervisor, Professor of Mechanical Engineering

Dr. Charles Fine
Thesis Supervisor, Chrysler Leaders for Global Operations Professor of Management

This page intentionally left blank.

Acknowledgements

I would like to take this opportunity to thank everyone who provided guidance and support throughout my internship and research for this thesis. With thanks to…

The LGO Community
For creating a vulnerable and supportive community and constantly inspiring me to reflect deeply on my values and how to leave a positive impact on the world.

My Advisors – Maria Yang & Charles Fine
For your wisdom and willingness to be a sounding board as I developed the scope of my research and worked through the many challenges that came up.

Li & Fung Supervisor – Pamela Mar
For your constant support and guidance during my research and for your commitment to and leadership in sustainability.

Li & Fung Colleagues – JP, Vasu, Chris, Angel, and Jonathan
For keeping things light and fun during the internship. The night hikes, life conversations and long talks about sustainability made my time in Hong Kong particularly meaningful.

Undergrad Friends – Thulani, Brandon, Nick, Rafi & Jorge
For your deep friendship and wisdom that continues to inspire me every day and for the way each of you have always pushed me to live out my values in everything I do.

High School Friends – Ben & Logan
For your long friendship that has helped shape who I am today, especially by making me aware of how important protecting the natural world and environment is and how I can play a part in that.

My Family
For your constant love, encouragement, and wisdom AND for flying across the world to share in the experience of living in Hong Kong.

This page intentionally left blank.

TABLE OF CONTENTS

List of Figure

List of Tables

List of Equations

This page intentionally left blank.

Chapter 1: Introduction

1.1 Project Motivation:

Li & Fung works with over 10,000 factories distributed across 50 countries to design, produce, and deliver hard- and soft- goods to over 2,000 apparel and consumer goods customers [1]. Traditionally, Li & Fung followed a sourcing agent business model essentially acting as the middleman between suppliers, factories, and brands. However, both emerging needs of customers and Li & Fung's understanding of how technology could be used within their operations, have led Li & Fung to pursue a new strategy which to date has focused on innovation, speed and digitalization of the supply chain. Digitalization, specifically, will enable the future growth of the company by transforming Li & Fung from the role of a sourcing agent to a service partner.

A few major trends are driving these emerging customer needs. For example, increased consumer demand for "sustainable" apparel and more stringent environmental regulations, have both led brands to focus on reducing the environmental impact of their supply chains. Brands have primarily focused on developing mechanisms to measure, benchmark, and reduce environmental impact across the supply chain. An example of this is the Sustainable Apparel Coalitions' Higg Factory Environmental Measurement (FEM) module, which provides annually reported factory-level data on environmental impact in order to understand overall impact of the supply chain.

LF's role within the industry is to supplement HIGG and other industry datasets with verification and supplementary data, including real-time data and correlated performance data. To this end, Li & Fung has run a number of pilots focused on data collection. These pilots have ranged from installing wirelessly connected energy sub-meters in factories to developing various applications to track quality defects and factory employee engagement. The energy data, as an example, is streamed to a central platform where it is analyzed and used to drive factory-level reduction. These pilots have successfully demonstrated how data collection can be used to drive factory-level improvements such as reducing energy consumption and increasing first-time pass rates.

However, the underlying assumptions of digitalization, namely that the real-time data can also be used to derive valuable insights across the supply chain have yet to be validated. The requirements for data collection have been primarily focused on end use applications at the factory level. The data required to optimize Li & Fung operations across the complex and geographically dispersed apparel supply chain must meet a different set of standards in order to address challenges such as accuracy, reliability, and interoperability.

1.2 Problem Statement

The underlining assumptions of digitalization as it relates to sustainability of the apparel supply chains have remained largely unvalidated. This research aims to specifically answer the following questions in order to validate the underlying assumptions of current industry digitalization efforts:

1. Given the complex nature of the apparel supply chain, what are the requirements for data collection, data quality, and data processing?
2. How can real-time data be used to improve the sustainability goals of the industry?

To validate these assumptions and answer the outlined questions, the scope of this project will first focus on assessing the efficacy of current environmental and factory data by developing a benchmarking tool for environmental efficiency. This assessment and tool will then be analyzed to understand how real-time factory data could be leveraged to improve the insights provided by Li & Fung. A four-phase approach will be taken to accomplish this:

1. First, existing industry data sources will be described and evaluated in order to assess data quality, develop verification methodologies, and provide recommendations for future data collection.
2. Second, the features of the data will be analyzed in order to develop an understanding of the underlining relationships. This step will focus primarily on selecting the most relevant features for building a predictive model. It will evaluate various machine learning approaches and then report on the most appropriate approach and final set of selected features.
3. Third, using the final set of selected features from the second phase, a clustering algorithm for factory-level resource efficiency will be developed and used to benchmark factories.
4. Finally, an analysis will be performed to evaluate how real-time data could improve the predictive model. This evaluation seeks to evaluate the additional benefit from real-time data and to specify the data requirements for the real-time data.

Chapter 2: Project Background

2.1 Li & Fung:

Li & Fung was established in the early 20th century as a trading startup, based in Guangdong, China. The founders, an English teacher and a local merchant initially focused on exporting Chinese goods to Western markets. By the mid 1930s, the firm grew to be one of the largest import-export firms in the region. After shifting operations to Hong Kong in 1937, the firm expanded into garments, toys, electronics, and hardgoods.

In the 1970s, under the leadership of the third generation of the Fung family, Victor and William Fung took the company public on the Hong Kong stock exchange. The Fung brothers continued to expand the company over the next 30 years through, for example, strategic partnerships with Toys R Us and Circle K. Through the 2000s and early 2010s, Li & Fung continued to grow and expand through a series of acquisitions. By the mid-to-late 2010s significant changes in the market threatened Li & Fung's positioning as a sourcing agent and resulted in multiple restructurings and sell-offs.

The 2017-2019 three-year plan for Li & Fung focused on digitalization, speed and innovation. Li & Fung began investing heavily in digital tools and platforms to meet the changing needs of the market. Today the firm employees over 17,000 people across 230 offices around the world [2].

2.1.1 Fung Academy

Fung Academy is the learning arm of the Fung Group, the majority shareholder of Li & Fung, which seeks to accelerate innovation in the organization. Fung Academy serves the diverse business units across the Fung Group and Li & Fung through a variety of programs, accelerators, and research groups. These services range from talent and capability development programs, consumer and customer insight research, innovation and experimentation accelerators, and supply chain futures research.

Historically Fung Academy has followed a "push" model, developing and incubating new ideas and then working with the business units to find applications for these ideas. As a learning organization, Fung Academy has recognized that it could better serve the business units of Li & Fung if it transitions to a "pull" model. In adapting this new model in the past year, Fung Academy has made significant organizational and operational changes to become a more strategic partner with the rest of Fung Group.

2.1.2 Previous Fung Academy Sustainability Projects

Fung Academy has performed a number of sustainability-related pilots to develop an understanding of data collection and analysis across the supply chain. Starting in 2014, Fung Academy funded a study to understand the energy use throughout their supply chain. Fifteen representative factories were assessed by an independent third party resulting in the identification of 269 energy reduction opportunities, 190 of which were deemed "financially feasible" with an internal rate of return in excess of 15% [3]. The study and follow-up capacity building efforts aimed to encourage factories in the supply chain to invest in these energy reduction opportunities. To aid in the capacity building efforts the study further provided a series of recommendations to Li & Fung. These recommendations included developing targeted approaches based on supply chain segmentation, mapping of broader resource use, and developing an "internet-enabled database" to streamline the process of data collection.

Based on the recommendations of this initial study, in 2015 Fung Academy carried out a research project which aimed to develop a methodology to evaluate the environmental performance of factories in the Li & Fung network. The tool, modeled after established industry best practices for measuring and tracking compliance and aligned with the then-nascent Higg Index suite of tools, served as an "environmental scorecard" for factories. This scorecard and the underlying methodology provided insight into the complexity of accurate data collection and the right quantitative metrics for factory comparison and benchmarks. For example, many factories have on-site dormitories and canteens which use substantial resources (energy, water, wastewater, etc.) to operate. The research showed that the decision to include or exclude this resource use in drawing comparisons is difficult since disaggregation of the collected data is not always feasible [4]. The scorecard, although valuable from a research perspective also proved difficult to manage and implement across the wider network of factories, further highlighting the need for an "internet-enabled database" to drive data collection processes.

In 2016 two additional research projects were conducted to continue to advance the goals of data collection and analysis. The first was a second iteration of the environmental scorecard. Termed the Li & Fung Environmental Tool (LET), this iteration included a mobile phone-based application for collecting and verifying factory-submitted data and a more robust scoring methodology [5]. Full adoption of the tool however was limited by existing business practices and incentives which placed little value on sustainability in sourcing decisions. The data collection process, while streamlined on the Li & Fung side, also required significant time and resources from factories. Li & Fung was unable or unwilling to use limited political capital to apply the pressure necessary to make use of the tool compulsory by factories.

In 2017, Fung Academy developed the third iteration of the environmental performance evaluation which was focused on providing quantitative comparisons between products and between factories rather than a final qualitative score. The data used to develop the quantitative scoring was drawn from both factory surveys (again introducing data quality and verifiability challenges into the analysis) and secondary sources including Life Cycle Analysis (LCA) and government databases. This tool, while more comprehensive than the previous two iterations proved to be cumbersome and difficult to update because of the large number of dependencies on secondary data sources (e.g. research papers rather than databases) [6].

Also, in 2016, Li & Fung introduced a technical audit of factories. The audit, although never rolled out to all factories, provided Li & Fung with an understanding of the main drivers and inhibitors for investment in technology across their supply chain. Further analysis of the audit data concluded that the scope of the technical audit was insufficient to draw any conclusions with regards to technology adoption and factory performance [7].

Following up to this analysis an Industrial Internet of Things (IIoT) pilot was performed at a factory owned and operated by Fung Beauty. Taking the recommendation of the 2014 study, the IIoT platform demonstrated the potential factory-level benefits of an "internet-enable database" for data collection and analysis. This pilot however focused on a segment of Li & Fung's business that had a higher level of technological capabilities than a standard garment factory. Thus, while the pilot project demonstrated the immediate benefit to a beauty factory for implementing an IIoT platform it did not provide any direct insights into how such technologies could also be adapted in low-tech garment factories.

The initial success of the IIoT pilot project, however, led to a few additional IIoT pilots focused on the low-tech garment factory environment. The first, in 2017 was focused on measuring quality performance of factories through a low-tech tool used in place of traditional paper records [8]. This low-tech tool was converted into a mobile application in 2018 [9]. The second was focused on integrating IIoT into factories in order to collect real time energy data. Online platforms and energy-submeters were rolled out in two factories, one in China and another in India. This pilot focused on demonstrating the internal rate of return to a factory for investment in digital solutions for reducing energy consumption [10]. The success of the energy sub-meter pilots was followed by a larger campaign to roll out the platform across the network of factories, especially in India, where the third-party software provider primarily operated. As of mid-2019, despite the case studies showing the relatively short timeframe for return on investment, little traction was made in scaling the system.

2.1.3 Pilot Study of Washing Facilities

In February 2019, as part of the initial efforts of this research, the author worked with five Bangladeshi washing facilities and the Indian third-party software partner from the 2018 project to scale the IIoT platform for energy sub-metering. The goal was to include measurement of water consumption and waste-water discharge in addition to energy consumption. The five washing facilities were identified by a US brand as necessary targets for water-reduction projects. The introduction of the IIoT platform was seen as complementary to the other water-reduction projects that the brand had mandated for the washing facilities to implement. Between February and May 2019, these facilities were engaged for project implementation: an initial technological readiness assessment was performed via a questionnaire to the facilities. After understanding the current state of sub-metering at the facilities, targeted solutions were developed and provided via commercial and technical proposals to the facilities.

The five facilities had varying levels of readiness for the IIoT platform ranging from limited analog sub-meters only in some areas of the factory to more advanced facilities which already had more than 25 digital sensors installed. Some of the facilities also had building management systems and other data collection devices to aid in tracking and trending energy use. Despite the program being classified as a "best practice" and "highly recommended" by the Brand, most facilities were unwilling to invest in the technology at the time of proposal receipt. The industry dynamics and challenges with previous pilots suggests that the status-quo for project success is the nature of the request from the Brand. If a project is compulsory a factory will invest in the implementation of the project; anything less than compulsory will be shelved until such a time as it is considered compulsory.

This pilot provided both new learnings and reinforcement of previous learnings.

1. *Return on investment is not the main driver for investment decisions; garment factory management decision making frameworks are complex and nuanced.*
   First there exists a gap between what factories report as the expected timeframe for Return on Investment (ROI) and the actual timeframe for which they will make an investment. In two surveys, one performed in 2016 as a follow up to the technology audit and the second

by the researcher in 2019, most factories (35% in 2016 [7] and 50% in 2019) reported that the expected timeframe is 1-2 years.

The survey performed by the researcher in 2019 was for a relatively small group of mills (n=52, 37 of whom answered the ROI question) that all supplied raw textiles for a major US retailer. 84% of the respondents also indicated management was willing to invest in new technologies.

Contrary to this data however, in most of the scaling efforts made by Li & Fung there has been considerable push back from factories. For example, the attempted roll-out of IIoT projects in Li & Fung factories in India has yet to result in the development of any projects. This suggests that new approaches to incentivizing factories to invest in technology is required. This aligns with the recommendations from previous projects and research efforts, including the most recent 2018 project which concluded that if Li & Fung and Brands are serious about promoting the use of IIoT they should set up cost sharing programs to help share the initial investment and the cost savings [10].

There are examples of programs that are addressing this gap. One such example is Apparel Impact Initiative (AII) which requires equal financial commitment from Brands and factories for implementation of sustainability projects. The investment of these two stakeholders is then matched by third party investment vehicles such as large banks who have a vested interest in reducing the risk and exposure of their key investments. For example, AII's partnership with Clean by Design focuses on improving textile mills through a ten-step program which includes implementation of resource sub-metering. Strategies like this appear to be gaining traction in the industry and are resulting in a higher number of completed projects [11]. As discussed in 2018 Pulse of the Fashion Industry report, the industry must focus efforts on expanding successful projects rather than continuing to pilot projects that never reach scalability [11].

Some vendors and factories have made sustainability an integral part of their long-term strategy and have thus invested heavily in resource management systems and environmental impact improvement projects. Two specific case studies that highlight this strategy are Aditya Birla and Shahi, both of whom are Indian vendors.

Aditya Birla has taken a strong top-down approach to environmental and social programs. The senior management has articulated a holistic strategy for being the leader in India with regards to environmental and social impact of business. At one factory visited by the researcher, a sustainability team of five people had been established 3 years prior. This team worked closely with maintenance and production teams to develop and execute projects ranging from rainwater harvesting to IIoT platforms for resource use monitoring. The sustainability team was integrated into the factory management and had a voice in all decisions of the organization.

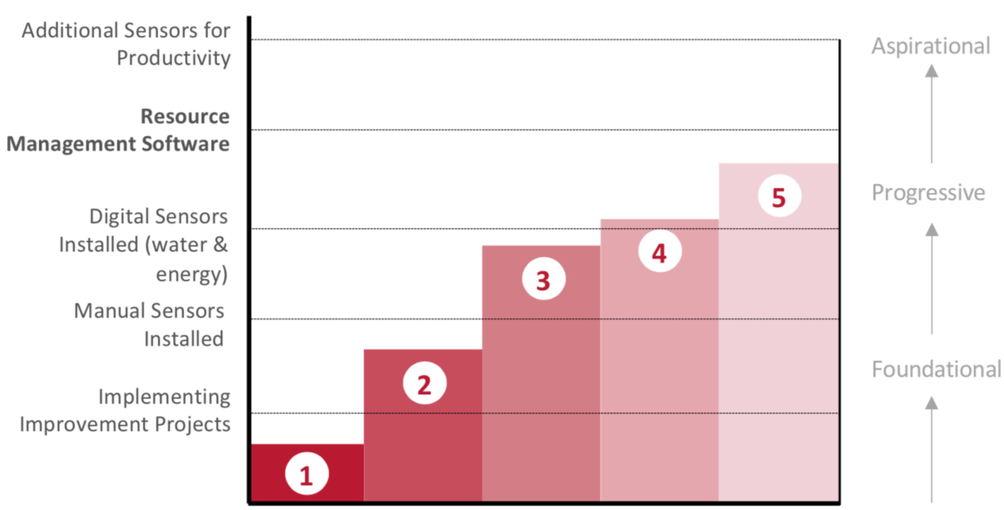
2. *Li & Fung and Brands need to continue to align sourcing strategies with internal sustainability strategies.*

   Academic research is non-conclusive on the business proposition to factories for improving social compliance or reducing environmental impact. For example, an analysis of Li & Fung data shows that compliance ratings do not affect the total dollar value of the annual sales that Li & Fung gives to a factory [10]. This reconfirmed research performed in 2014 which also concluded that there was no correlation between (environmental and social) compliance and "factory success", a proxy for the factory performance [12]. Academic research of broader souring markets suggests that there is a weak relationship between compliance and increased sales. However, this research cannot draw any conclusions on the nature of the relationship. That is to say, the research cannot conclude if the correlation is due to an increased willingness to pay on the part of the buyer or a negligible impact on price for increasing compliance on the part of the seller [13].

Considering Li & Fung's current sourcing process, purchasing decisions are still largely driven by cost and the presence of existing relationships between the sourcing agent and a particular factory. Brands and Li & Fung have the opportunity to institutionalize their sustainability ethos into business practice to further incentivize factories to make the necessary investments and improvements. Brands such as Patagonia and Nike have demonstrated the feasibility of this philosophy in sourcing. However, it has yet at reach wider industry adoption.

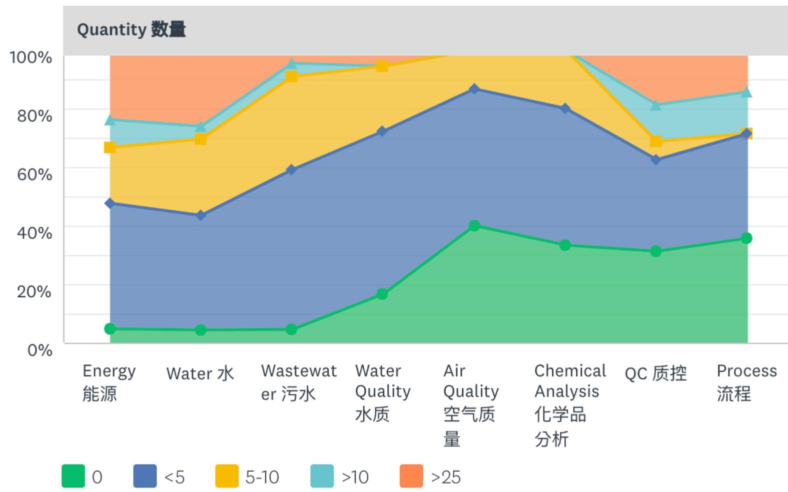
3. *Factories vary greatly in the level of technological readiness; solutions or proposed investments must be tailored to the individual needs of the factories.*
Figure 1 below shows a qualitative comparison of the five washing facilities that were targeted for the IIoT project. As shown in the figure there is a wide gap between the highest performer and the lowest. A one-size fits all technology package will not meet the unique needs of every facility.

Figure 1: Technological Readiness Levels of Bangladeshi Washing Facilities



The survey performed by the researcher for the 52 textile mills provided similar results. For example, 58% (n = 43) of the mills reported that they already had sub-metering in place. Of these 43, 23-25 factories responded to additional questions regarding quantity of sensors (*Figure 2*), sensor type and existing method of data collection (14 mills sent data to online software or Building Management System to be trended vs. 11 mills that manually read the sensors).

Figure 2: Sensor Quantities Installed at Client Mills



Given the small number of surveyed mills and even smaller response rate of the survey this data is not conclusive but does help directionally explain the varying levels of current technological readiness of factories for IIoT solutions. Li & Fung could focus on selecting the appropriate IIoT platform and then giving flexibility to each factory on how to connect existing sensors to the platform.

4. *Li & Fung needs to continue to evolve its strategy around digitalization and develop the resources and capabilities to implement this strategy.*
   The learning from this pilot reaffirmed the recommendations from the previous energy sensor pilots that Li & Fung needs to define their strategy in promoting IIoT [10]. Li & Fung should take an approach aligned with industry best practices for digitization. Mckinsey, for example in their report on digitization in the apparel sourcing industry, reinforced the need for a clear strategy: "digitization is not an end in itself but a key enabler of other priority levers…we see digitization as a core enabler to achieving a step change in performance, transforming to a customer- centric operating model, and creating the transparency that is currently lacking in the global apparel supply chain" [14]. This sentiment is shared by other industry reports such as one produced by the Center of International Manufacturing and University of Cambridge which concludes: "to maximise success, changes to the digital supply chain must reinforce strategic objectives, so there needs to be a clear focus on what organizations are trying to achieve" [15]. By establishing the right strategic goals within the larger framework of "supply chain of the future", Li & Fung can more effectively partner with factories to execute on the strategy.

5. *Li & Fung should focus its efforts on establishing the requirements for data collection and reporting and provide recommendations to factories for platforms and implementation strategies.*
   Given the fragmented and geographically dispersed nature of the apparel supply chain, Li & Fung should shift its focus away from executing one-off IIoT projects and instead focus on defining the requirements for the data collection and reporting from factories. With a clear strategy for digitization the developed requirements can become compulsory. Li & Fung can then shift the focus to connecting 3rd party suppliers, with pre-approved platforms, to

factories. The nature of Li & Fung's role would be limited to providing financial incentive or access to discounted 3rd party services. The technological development of the projects would be left up to the 3rd party and factory.

2.2 Industry Examples of Environmental Performance Measurement

The internal research and pilot projects performed by Li & Fung have paralleled the industry focus on sustainability. The Sustainable Apparel Coalition (SAC) has led a majority of the industry initiatives around measurement and data collection. The SAC Higg platform has also been complimented by a breadth of research in the form of product and process Life Cycle Analyses (LCA).

2.2.1 Higg Factory Environmental Measurement (FEM) Module

The Higg FEM module was developed by the SAC as a tool for brands, factories, and eventually consumers to measure and compare the environmental impact of the garment supply chain. The Higg FEM module is completed on an annual basis by factories through a self-assessment portal. Factories can also elect to have a verified audit of their self-assessment performed, the results of which are also recorded on the same platform. During the self-assessment, the factory provides information on the Environmental Management System (EMS), GHG emissions & energy use, water consumption, wastewater discharge, waste discharge, air quality, and chemicals use. Each category is divided into three levels which build upon each other. Level 1 focuses on reporting impact (e.g. total water use, total energy use, etc.), Level 2 focuses on benchmarking and setting targets for reduction, and level 3 focuses on aspiration goals such as closed loop recycling, engagement with partners, or Scope 3 GHG calculations.

Broadly the comparison methodology for benchmarking a set of facilities is based on a uniform scoring methodology. This methodology does not take into consideration the specific quantitative features of a factory. The scoring is primarily based on whether or not a set of criteria is met for each Level [16]. The only exception to these binary criteria is a set of applicability criteria to differentiate factories-based type of water use, waste-water discharge, air quality system, and chemical usage. However even with these differentiations factories are ultimately given a final score which allows for comparison across all other factories on the Higg platform, regardless of the factory Tier (cut & sew, mill, washing facility, etc.).

The data provided from the surveys do allow for comparisons to be made on a more granular level. For example, by filtering based on "sipfacilitytypesew" = YES, comparisons can be made between two Cut and Sew facilities, normalized based on number of operating days, number of units produced, or square footage of the facility. These analyses provide more meaningful benchmarks for comparing facilities (e.g. facilities that do the same set of processes). The standard practice for the industry to date is to filter on the above mentioned variable for a Cut & Sew (Tier 1) facility and analogous variables for dyeing/printing facilities (Tier 2) and textile mills (Tier 3).

These analyses however are complicated by the wide variety of factory types that exist. The Tier framework that is predominantly used in the industry is a misnomer: one Cut and Sew facility might have a washing facility while other Cut and Sew facilities might outsource all washing. Another facility might be vertically integrated with a mill, a washing facility and a cut & sew facility. Therefore, when making comparisons based on Higg data, it is important to first start by defining the methodology for comparisons between two facilities. A number of parameters exist for the filtering and normalization and these should be tested to determine which provides the most value to the business.

Recent reports from the Higg data analytics partner company, Anthesis, suggest that the 75% of the factories on the platform are now reporting enough information about energy use to estimate GHG emissions [17]. Similarly, trending of the year-on-year data shows that the total number of factories on the platform are expected to increase by almost 2000, from 7,005 in 2017 to 8,938 [17]. The reported success of the platform and the uniform industry buy-in for use of the platform suggest that Higg will continue to be the predominant platform for managing the environmental impact of the industry.

Despite these positive numbers, there are a few shortcomings with the Higg platform, some of which are outlined below and many of which will be discussed in more detail in later chapters:

1. *Not all of the data is verified and some of it lacks accuracy:*
   While guidance exists for filling out the Higg FEM module, there are wide fluctuations in how factories actually record the data. Verified audits are performed. Across all 8,938 factories posting the to the Higg FEM module in 2018, 1560 (17.45%) are expected to be verified [17]. The results from verified audits can be substantially different than the self-assessment.

2. *Question format is structured for ease of factory use rather than data analysis:*
   Factories are given many free response questions which means that the data itself requires additional filtering and interpretation whether manually or through a natural-language processing algorithm. In many cases flexibility around units used or redundancy in questions asked leads to a reduction in data quality.

2.2.2 Kering EP&L

Kering developed an Environmental Profit & Loss account tool to provide "in depth analysis of the resulting impacts a company's activities have on the environment". The tool is built in order to drive business decisions by identifying the largest causes of environmental impacts. The tool is tier-based allowing Kering to evaluate the main drivers for impact across their entire supply chain. The benchmark is broken into six categories: Air Emissions, GHGs, Land Use, Waste, Water Consumption, and Water Pollution [18]. For each category and each tier, the associated economic costs are calculated. This allows for management to understand that, for example, Tier 4 Raw Material Production accounts for the majority of land use impact in the supply chain.
The methodology is based on collecting primary data via surveys to factories and suppliers and then secondary data to quantify the impacts.

2.2.3 RefScale

RefScale is a methodology produced by The Reformation fashion company to provide garment-level comparisons between The Reformation products and other comparable products. The methodology follows standard LCA considerations and draws from primary and secondary data sources [19]. The metrics are communicated as water, energy and waste savings per garment as compared to a similar product from a different brand. The tool was developed to be consumer facing. As such it clearly serves marketing purposes which reinforces the ethos of the brand image. Based on annual reporting from The Reformation the same data also informs setting of environmental targets and overall reduction of their footprint [20].

2.2.4 EIM

The Jeanologia developed software EIM (Environmental Impact Measurement) is a laundry- and garment finisher- facing tool that enables measurement of the environmental impact of finishing processes for the apparel industry. The tool assesses the environmental impact across four dimensions: water, energy, chemicals, and worker health and provides an aggregate score for each process [21]. This scoring metric allows factories and brands to benchmark their processes and the final products to help determine where improvements can be made to reduce the environmental impact. Additionally, the tool is designed to allow designers and brands to understand how their design decisions impacts the environmental costs of the product during the finishing processes [22]. The tool provides an easy to use mobile and online platform.

2.3 Benchmarking Methodologies

The majority of the benchmarking work in the industry has been based on LCAs and methodologies similar to those discussed in Section 2.2. Outside of retail, there has been substantial research into the use of statistical models to benchmark energy efficiency. These include programs such as the EPA's EnergyStar and the Green Building's Council LEED program. The methodologies for these programs employ various regression or scoring techniques to compare a building to other similar buildings within a predefined category (school, office, etc.).

Academic literature has looked at how the methodology behind these programs could be further developed to improve the results of the scoring. For example, Xuefeng Gao, et. al. developed a clustering algorithm to find natural groupings in building data. A clustering algorithm is based on descriptive features within the dataset rather than the type [23]. The clustering approach improved the results of classification and energy prediction over the traditional regression approach based on building type only.

Chapter 3: Data Sources

The data used to develop this research was provided from Li & Fung's internal databases and the Higg FEM platform and included over 60,000 data points for over 700 factories currently on Higg. Relevant data spanning all transactions between 2017 and 2018 were collected from each of these four databases.

3.1 Data Sources

3.1.1 Li & Fung Databases

The Li & Fung datasets includes all relevant information for each order sourced by Li & Fung. The data includes orders from each business unit, for each customer and sourced through each supplier represented within Li & Fung's operations. The data also includes information about the cost, lead time, order quantity, and product description, quality, and compliance (to list a few).

3.1.2 Higg FEM:

The Higg FEM database is managed by Higg and is available by subscription to participating SAC members. The Higg FEM database is currently on the third version and has seen a steady increase in use over the past five years as brands and sourcing agents have started to mandate the use of the platform by all factories. In 2018, over 1000 factories across the Li & Fung network of factories are actively using the platform. Li & Fung also has results from 2017. The majority of the data which falls into seven main categories (EMS, Energy, Water Use, Wastewater, Air Emissions, Waste, Chemicals Management) is self-reported by the factory with the option of a third-party verification of the data. In 2017 X percent of factories were verified. Verification for 2018 data is not due until the end of 2019 and therefore, to date, only X percentage of factories have completed verification. While the database includes over 6000 features, many of these are not relevant or are repetitive and therefore the features that were included in this study were limited to the following:

- Groupid
- Version
- Unverified_assessment
- Verified_assessment
- Sipfacilityprocesses_
- Sipmaterial_
- sipoperatingdays
- sipfulltimeemployees
- siptempemployees
- sitecountry
- sipindustrysector
- sipfacilityannualprodvolquant
- Sipfacilityannualprodvolunits
- Sippermitsreq
- Sippermitsreq_compliance
- Watsources_quant/watersources_units
- Ensource_quant/Ensource_units

- Wstsource_quant/wstsource_units
- Wwtrack
- Method_frequency
- wwtreatment Facility
- airsource_

The dependent variables of interest to this research are derived from a combination of the resource quantity and units features as well as the annual produced volume quantity and unit features. As shown in Section 4.1.2, these features are used to develop normalized resource intensity variables for each factory. The remaining features listed above will be used as independent variables to assist in developing the clustering model.

3.2 Data Categories:

There are four primary categories for the data sources described above. These are: purchase order, verified/audited, and self-reported and streamed data. Within these categories, there are three main data types: numerical, categorical, and ordinal.

3.2.1 Purchase Order Data:

Purchase Order category includes both numerical (e.g. FOB Amount) and categorical (e.g. Factory country) data from the Li & Fung datasets. This dataset and the features captured within are expected to have a high degree of data quality. The process used to generate the data follows strict internal procedures making the measure repeatable. Additionally, these data points are tied directly to financial reporting and contract negotiation, both of which are closely monitored and controlled by current business processes.

3.2.2 Verified/Audited Data:

Verified/Audited category includes data from the Li & Fung dataset and any Higg FEM assessments that are classified as "Verified_assessments". It should be noted that Verified Assessment data does not replace the self-reported data in Higg but provides a point of comparison.

The vast majority of the data points in these data sets are collected by highly trained, third party individuals, increasing the level of independence and ensuring a higher level data quality than self-reported data. Moreover, the specific procedures for generating this data are fine-tuned to ensure reliability. Individuals responsible for generating this data are indoctrinated into the specific programs and undergo periodic training.

However, unlike the data collected in the Purchase Order, which is highly objective in nature, many of the collected data points in the Verified/Audited data sets are subjectively provided. For example, one inspector might ask a question in slightly different manner or chose to probe a little further resulting in discovery of additional information which changes the final recorded data. Additionally, many of the data points in these categories require some degree of estimation which leads to potential deviations between the recorded value and the true value for a data point.

3.2.3 Self-Reported Data:
Self-Reported data includes all three data types from assessments categorized as "Unverified_assessment" within the Higg FEM database. The descriptor "self" in this category refers to factory that reports data on their own facility, operations, and personnel management procedures. As a result of the lack of independence and the inherent subjectivity in some of the measurements, this data is expected to have a lower level of data quality. This is especially relevant to the numerical data points associated with total water and energy usage and total water and waste discharge. These numerical points have variability introduced in two points of data collection: 1) the sensor, sub-meter, or measuring device measure values within some predefined level of accuracy and must be continually calibrated to ensure the data remains reliable and 2) the process and person responsible for transferring the data from the measuring device to the records also introduces possible errors and variance into the data.

3.2.4 Streamed Data
Streamed data is data collected via online sensors and sub-meters. The data quality from streamed data is expected to be higher than self-reported or audit verified/audit data because both the collection and transmission steps of data manufacturing have been operationalized in automated processes. The sensors are still prone to variability based on the specific specifications of the sensor or sub-meter. Ensuring that this variability remains within the expected range given by the calibration requires appropriate preventative maintenance schedules to check calibration and replace or repair sensors as needed.

3.3 Data Features
Features selection is based on the required use cases for each stakeholder along the product value chain. Implicit in this is the need to address interoperability of collection platforms and databases at every node of the value chain.

The following examples help to illustrate this:

In previous Li & Fung IoT pilot projects, the data requirements were developed for the purposes of aiding the pilot factories in trending and reducing their energy consumption. The data output from this platform did not provide the necessary context to allow a sourcing agent or brand to report on GHG emissions or environmental footprint per garment produced. Similarly, the data collected was too granular to present to an end user of the garment.

In many of the research projects carried out by Fung Academy, the source data was collected from a small subset of factories based on a manual email survey (this continues to be the method for much of the compliance data collected outside of regular audits). This data provides little insight to the factory and adds a substantial burden to the already limited resources that factories have available to them.

Both of these examples highlight the challenge of defining data requirements. Moving up the value chain, the data requirements change substantially. To evaluate the data characteristic requirements, the stakeholders along the value chain can be segmented into five main groups. Each group has specific use cases, some of which are well understood and future-state use cases which are not currently known. Table 1 below outlines these five stakeholder groups:

Table 1: Stakeholder Groups

Stakeholder Group	Use Case/Desired Granularity
End User (Consumer)	Product level comparison Brand level comparison
Brand/Fashion Company	GHG/SBTI Targets Product level optimization Network Optimization
Sourcing Agent	Product level optimization Network Optimization
Factory/Supplier	Individual manufacturing line performance Identification of focus areas Driving reduction efforts
3rd Party Auditors	Regulation compliance

As noted in Section 1.2, the first primary goal of this research is to define the data requirements to ensure that digitization efforts meet current and future state needs of the industry. The approach taken by this research is to evaluate the existing data sources to determine the requirements for the future state of data collection and analysis.

3.4 Data Quality Methodology

A major challenge within the contemporary study of supply chain management is measuring and controlling data quality. Data quality if not properly managed has been shown to have significant impacts on both the tangible and intangible aspects of business [24]. Poor data quality, for example, may lead to poor decisions with irreversible consequences for a company. The processes and methodologies for measuring and controlling data quality must be careful operationalized within a business' digitization efforts [24].

The dimensions of Data Quality span two main categories, intrinsic and contextual, the former of which is objective and context-independent [24]. While the process of the data collection for Higg or some Li & Fung data are highly subjective, the collected data can be managed by considering intrinsic data quality measures without consideration to the specific context of collection. This is a simplifying assumption necessary given the difficulty of developing and applying contextual data quality measures to audit data. It is also a mostly realistic assumption since the data collection process is standardized and efforts are made to reduce the contribution of individual inspector or reporter bias in data collection.

Intrinsic data quality is frequently defined along four dimensions which will be explored in detail in the following section: accuracy, timeliness, consistency, and completeness [24]. These four are discussed in detail in Table 2.

Table 2: Intrinsic Data Quality Metrics

Measurement	**Description** [24]	**Quant Metric**	**Applicable Questions**	**Li & Fung Example**
Accuracy (α)	Measure of how closely data reflects the true value	$0.0 \leq \alpha \leq 1.0$ *Proportion of entries that are accurate {1} to the total number of accurate and inaccurate {0} entries for a given feature* *Proportion of features that are accurate {1} to the total number of accurate and inaccurate {0} features for a given entry*	Is the data within an acceptable margin of error from the expected real value?	Factory reports annual water consumption greater than the total volume of water consumption reported for the industry.
Timeliness (λ) [24]	Measure of how frequently data is updated.	$0.0 \leq \lambda \leq 1.0$ $\lambda = \{\max\left[\left(1 - \frac{\text{currency}}{\text{volatility}}\right), 0\right]\}^{s}$ Where: *Currency = (Delivery Time – Input Time) + Age* *Volatility = Shelf life of data, subjectively defined by the use* *s = is sensitivity, user defined*	Is the data a reflection of the relevant period of time?	Higg reports are updated on annual basis vs. sourcing strategies which are set on a 3-year basis.
Consistency (μ)	Measure of how the data is represented	$0.0 \leq \mu \leq 1.0$ *Proportion of entries that are consistent {1} to the total number of consistent and inconsistent {0} entries for a given feature* *Proportion of features that are consistent {1} to the total number of consistent and inconsistent {0} features for a given entry*	Is the format and structure for reporting two data points the same?	The date of a PR falls within the right range of years and is it presented in the right format (e.g. YYYY-MM-DD)
Completeness (χ)	Measure of missing data	$0.0 \leq \chi \leq 1.0$ *Proportion of entries that are complete {1} to the total number of complete and inaccurate {0} entries for a given feature* *Proportion of features that are complete {1} to the total number of complete and incomplete {0} features for a given entry*	Does a single data entry contain response to all of the necessary features? Does a single feature contain responses from all entries?	The number of factories that do not report total energy consumption

Table 2 shows that the metrics can be developed on both a per-entry and per-feature basis. Using the per-entry basis, a new feature, *aggregate data quality score*, could be developed and either used as a dependent variable in development of a system model or as a reporting metric to guide business decisions based on the data for the entry. The per-feature basis will be primarily used in the data-cleaning steps of model development to remove values or features prior to model development.

Thus, the methodology for operationalizing the data quality processes will focus on development of a quality profile whose constituent parts will be determined by these four measurements for each feature. Two methodologies will be developed. First a data quality profile methodology will be developed for the existing data from Li & Fung. Second, a methodology will be proposed and tested for real-time data.

3.4.1 Methodology 1: Existing Data

3.4.1.1 Accuracy

True values are unknown for most of the features in the studied data set. As such, proxies for data accuracy must be developed and validated. Two approaches will be taken to develop an understanding of the data accuracy for each feature. The general approach will be outlined below and applied to every model feature in Section 3.5.

1. *Cross validation*

   There is overlap in the features represented in each data set. In some cases, these features are identical (e.g. site country recorded by Li & Fung and Higg FEM). In other cases, these are structurally the same but might vary in specifics (e.g. number of employees, processes, annual production volumes and garment types produced). Using the union of these databases or the redundant nature of in-data-set features a cross validation can be performed to understand, at least directionally, the accuracy of a given entry or feature. The following cross-validations are used to check for accuracy:

   - Total Annual Production Volume vs. Li & Fung Total Production
   - Total Number of Employees vs. Li & Fung Compliance Employee Count
   - Water Balance (Wastewater Discharge vs. Water Consumption)

   One example will be provided to illustrate the cross-validation methodology for measuring accuracy:

   Total Annual Production Volume vs. Li & Fung Total Production

   The Higg feature, *sipfacilityannualprodvolquant*, provides a measure of the total volume produced by a factory per year. This variable however is defined relative to the operations of each facility. For example, a mill might measure total produced volume as weight (kg) or area (square meters) whereas a cut and sew facility might measure in total cut and sewn garments (Units) or total washed garments (Units). These units pose various challenges for normalizing the data in order to establish meaningful comparisons. For mill data, without additional information on the specific material type (e.g. density of the fabric) there is no way to convert between kilograms and square meters. Similarly, on the cut and sew side since each facility performs different operations a completed unit may vary greatly across the factories.

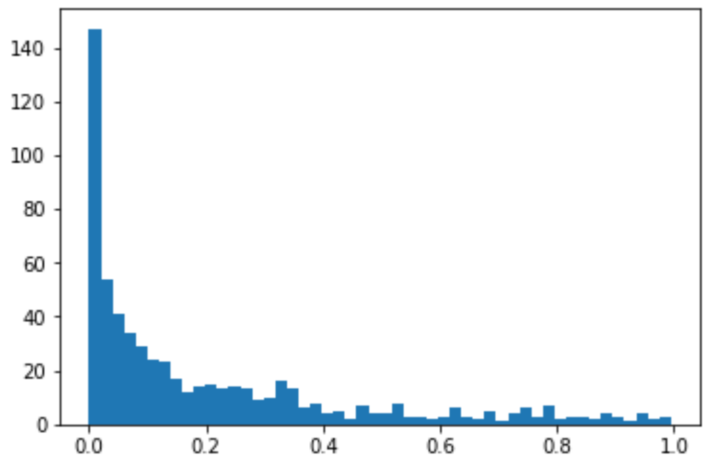
   The Higg guidance on the unit of measurement for the quantity is that it must be consistent with the units made in establishing the baseline and reduction targets. For example, if the baseline is measured in kW-hr/Units produced then the units recorded in total annual production needs to also be in the same Unit.

   Thus a more complex approach must be applied to verify the accuracy of the Higg feature, *sipfacilityannualprodvolquant*. The Li & Fung dataset provides an annual production volume quantity variable which encompasses all units ordered and delivered to Li & Fung. Using the percentage of a factory's production that is purchased by Li & Fung, an estimate of the accuracy of data points can be assessed.

   Figure 3 shows the results of the percentage calculation for the 2018 Higg Unverified Data. The following observations are made from these results:

- Of the 660 observations, 621 fall between 0-100% as expected.
- The remaining 39 observations are greater than 100%. This suggests that for these 39 observations, the production values reported by Higg are either inaccurate (36 factories) or equal to zero (3 factories)
- Of the 621 observations between 0-100%, ~222 represent less than 5% and ~70 are less than 0.05%.
- Looking at the order of magnitude represented by the Li & Fung Orders and the self-reported total production. Only 6 of the 621 factories reported greater than 2 x10^8 total units produced in a year. For these 6, Li & Fung represented less than 0.5%.

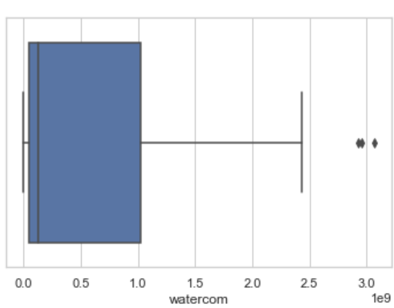
Figure 3: Li & Fung Order Quantity as a Percentage of Total Annual Production (N = 621)



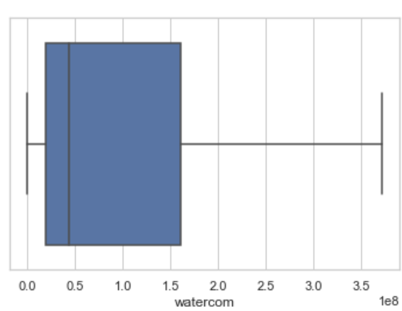
2. *Statistical analysis*
   The second methodology to measure accuracy is to perform a statistical analysis of the data and use statistics (standard deviation and mean) as a proxy for accuracy.

   a. *Statistical analysis of the distribution standard deviation*
      To verify future data points, the spread of these aggregated resource datasets were analyzed to understand the distribution. As an example, Figure 4 shows a box and whisker plot for the 2018 and 2017 verified data.

Figure 4: Box & Whisker Plot of Verified Total Water Consumption



2018 Verified Data (n = 27)



2017 Verified Data (n = 126)

Each feature is transformed on a log scale which was done since most of the distributions are heavily skewed to the left. Then any entries outside of 3 standard deviations from the mean are assumed to be inaccurate values and are scored as such.

b. *A Tukey test is performed on the means of the "true" data and the collected data.* The first method assumed the dataset was the entire population (e.g. 2018 Unverified Entries). A second statistical analysis is to look at the difference between the means of the Verified and Unverified data for a given year. This approach leverages the higher expected accuracy of the verified data given that the collection process is more independent. This verified data is used as a proxy to determine the true value of the statistics, e.g. mean, of the data.

   A Tukey test is a statistical analysis to test the hypothesis that the means of two different samples are statistically significant. This test is similar to a Student t-test but accounts of the standard errors of the data.

3.4.1.2 Timeliness

Timeliness as defined in Table 2 is a measurement of the periodicity of data collection. However, in application, the governing equation for timeliness includes non-intrinsic features of the raw data such as the time of delivery and a sensitivity parameter, *s*, which is use case dependent. Ballou, et. al bifurcate timeliness into two components: *currency* and *volatility*.

1. *Currency* is a function three dates: the original input time which is the time stamp at which the data was received into the system (e.g. date Higg FEM report is uploaded), the delivery date which is the time stamp at which the data was sent to the user (e.g. when the data was used to make a decision), and the age which is the time duration between when the corresponding physical event that led to the creation of the data occurred and when the data was sent to and received by the system. To illustrate this, an example of Higg annual average water intensity data is provided below:
   - Occurrence: water is used over an annual basis and the annual water utility bill is sent to the factory with a date stamp of 31 December 2018.
   - Input Time: on 1 March 2019 the factory records the annual water use into the Higg FEM module
   - Delivery Time: on 1 June 2019 Li & Fung retrieves this data and uses it to compare the water use of the factory to a benchmark.

   The Timeliness for the water measurement is then calculated as follows:

   Currency = (1 June 2019 – 1 March 2019) + (31 December 2018 – 1 March 2019) → 5 months

2. *Volatility* is a function of the shelf life of the data. The determination of the shelf life is a subjective exercise based on the expected frequency of changes in data and the impact these changes have on the decisions based on the data. As an example, the volatility of Higg annual average water intensity data is likely on the order of 6 months since this is the average timeframe for implementing an improvement project that would result in the reduction of water intensity [25].

Timeliness is a function of both currency and volatility and measures if the frequency of data updates is aligned with the time horizon for using the data in decision making. It is more informative to measure timeliness on an absolute scale and therefore, Ballou et. al. postulated a methodology to measure effectiveness on a scale from 0 to 1 as demonstrated in the below example:

$$\lambda = \{\max\left[\left(1-\frac{\text{currency}}{\text{volatility}}\right),0\right]\}^{s} \rightarrow \lambda = \{\max\left[\left(1-\frac{5\text{ months}}{6\text{ months}}\right),0\right]\}^{s} \rightarrow \lambda = \{0.16667\}^{s} \rightarrow \lambda = 0.1667$$

S = 1, assuming that the timeliness is not affected by the ratio between currency and volatility

A timeliness value close to or equal to 0 suggests that the data is does not meet the timeliness needs of the data use cases. A timeliness value close to or equal to 1 suggests that the data meets the timeliness needs.

Application of the timeliness measurement will be discussed in more detail in Section 7.1.1.1 when considering the incorporation of real-time data into the developed model.

3.4.1.3 Consistency
Academic literature makes a distinction between intra- and inter-relation as it pertains to data consistency. Intra-relation is a measurement of how much the data fits into a range of potential values whereas inter-relation is a measurement of how closely the data conforms to a defined format for the feature [24].

1. *Intra-relation*

   To measure intra-relation consistency, acceptable ranges must be understood for each numeric feature and acceptable values must be understood for each categorical variable. Higg guidance provides acceptable values for categorical data. For example, the feature *sipindustrysector*, must fall into one of seven categories of which one is "other". Any responses outside of these seven responses would be classified as inconsistent and would require either correction or removal from the model.

   Quantitative ranges are not provided by Higg or LF for numeric features. Therefore, a research approach was applied to develop the appropriate ranges. Industry and academic literature provided insight into the potential order of magnitude for each valuable. Based on this and comparisons across the databases (e.g. cross validation as discussed in Section 3.4.1.1), final ranges, as detailed in Table 3 were established.

2. *Inter-relation*

   The inter-relation of the data in the various datasets can be easily reviewed by comparing the formats of each entry of a specific data feature. Any inconsistencies in the recorded format would be an instance of inconsistent data as measured by inter-relation.

   For example, when importing the Li & Fung data there are multiple features that follow a date format. Some of the entries in these features are recorded in a YYYY-DD-MM format while others are in a YY-DD-MM format. One format must be selected and consistently used. Any deviation from this selected format would be an inconsistent data point.

3.4.1.4 Completeness
Completeness is a measure of what data is missing from a specific feature or entry. The process of measuring completeness is a straightforward process of identifying any empty, Null, or Nan data points in the data set.

For example, for the Higg dataset some factories do not track and thus cannot record the total annual waste produced. For these factories the value of *wastecom* will be equal to 'Nan' or equivalent depending on how a missing value is recorded. The completeness score for *wastecom* would be the total number of records with a recorded value for *wastecom* divided by the total number of records.

3.4.2 Methodology 2: Real Time Data
Li & Fung does not currently have a developed network of "connected" factories, however as discussed in Section 2.1 and Section 7.2 the long term strategy is to build out this capability across the network. As such, an adaption of the proposed data quality assessment methodology must be developed for the collection of real time data.

This methodology will be primarily based on the work of Benjamin Hazen, et. al. who proposed the application of a statistical process control (SPC) approach which built on Total Data Quality Management cycle developed by Wang, et. al [24]. The primary contribution of Hazen, et. al. was to develop a novel solution for measuring and improving data quality in "real-time" at the point of creation rather than after the data has been created.

Applying this approach, there are two nodes within the studied supply chain where data is created. The first is at the factory level, where sensors, distributed across the factory, collect disaggregated real-time data for resource use. The second is at the Li & Fung level where aggregated factory-level data is collected from many factories. To develop this methodology real-time data, collected by IoT sensors at three different factories for a finite period of time, will be used. These datasets will be treated as "live" and thus partitioned into historic data (to establish the control limits) and new (incoming) data (to demonstrate the real-time quality measurement process).

Table 3: Intra-Relation Quantitative Consistency Ranges & Justifications

Range	Upper Limit	Lower Limit	Upper Limit Explanation	Lower Limit Explanation
Water Usage (Liters)	1x10^9	1 x10^3	Most academic research on a country basis puts the order of magnitude for total industrial water demand in the range of billions of cubic meters (trillions of liters, or 1x10^12 Liters). For example, textile industry water demand in Bangladesh is estimated to be 1.5 Billion Cubic Meters in 2014 [26]. Similarly, in India the total Industrial Water Use 15 Billion Cubic Meters [27]. Therefore, considering that no one factory represents more than 0.1% of the total industrial or textile industry in their respective country, a reasonable cut-off for the maximum value of water usage is on the order of 1x10^9 liters maximum.	The lower limit would apply to a facility that only requires domestic water (drinking, sanitation). The WHO estimates that a minimum of 7.5 liters per capita of water is required [28]. Assuming that at least one third of this is used during the work day, the facility will provide at least 2.5 liters per person per operating day. The minimum number (excluding zero) of man days from the Higg dataset (e.g. sipoperatingdays x (sipfulltimeemployees + siptempemployees) is 1425-man days. Given 1425-man days times 2.5 liter per capita, a realistic minimum number of liters per facility is on the order of 1 x 10 ^3 liters.
Energy Usage (W-hr)	4x10^11	0	Most academic research puts the approximate energy consumption for cut & sew facilities on the order of 1x10^7 [29]. Similarly, in Li & Fung's research around energy consumption, the maximum annual energy use was 5x10^10 [3]. On a macro scale, the global energy use for the textile industry in 2008 was estimated to be 1x10^15 [30], therefore following similar logic to what was applied for water usage, no single factory should be consuming more than 0.01% of to global energy use in the industry. This puts the value around 1x10^11. Finally, after reviewing the Higg data the final value of 4x10^11 was selected and used.	Theoretically an operating plant must use some energy. However, a factory might only use a small amount of power and therefore no lower limit was imposed.
Wastewater Discharge (Liters)	1 x 10^9	0	Wastewater output cannot be more than the maximum water usage value.	Factories are expected to discharge some wastewater; however, it is possible that all water is treated and recycled continuously.
Waste Discharge (kg)	1 x 10^8	0	Waste discharge must be less than a fraction of the total annual production volume. 50% of the production volume is used.	Theoretically an operating plant must create and dispose of some waste. However, a factory might not track waste and therefore a lower limit is not imposed.
Total Annual Production Volume (Units)	2 x 10^8	0	60 billion kilograms of textiles were produced in 2008 [30]. Assuming an average weight of 0.2 kg per unit the total number of units produced per year should be on the order of 3 x10^11. Again,	Theoretically an operating plant must produce product. However, for the purposes of filtering the data, as long as a positive number is provided it is deemed acceptable.

			assuming that no one factory will produce more than 0.1% of the global production and comparing the values provided by Higg, the final upper limit of 2 x 10^8 was selected.	
Total Employees	1 x 10^4	0	A factory in general is located in single location. The maximum number of people should not be more than 10,000.	Theoretically an operating plant must produce product. However, for the purposes of filtering the data, as long as a positive number is provided it is deemed acceptable.
Total Operating Days	365	0	The upper limit of number of working days is 365 as dictated by a standard calendar.	The lower limit of number of working days is greater than 0 as dictated by a standard calendar.

Below is a description of the three plants and the measurements taken.

- Factory 1 (Tier 1 - Textile Mill): Energy use for four spinning units measured over 22 days at 1-hour increments.
- Factory 2 (Tier 2 - Washing Facility): Water (one point of measurement) and Energy (12 points of measurement) use for the facility measured over 59 days at 1-hour increments.
- Factory 3 (Tier 3 - Cut & Sew Facility): Incoming and discharged Water (measured at 10 locations) and Energy (measured at 12 locations) use for the facility measured over 2 months at 1-day increments.

3.4.2.1 Factory Implementation

The following illustrative example will be limited to a measure of intra-relation consistency of the data at each of the factories.

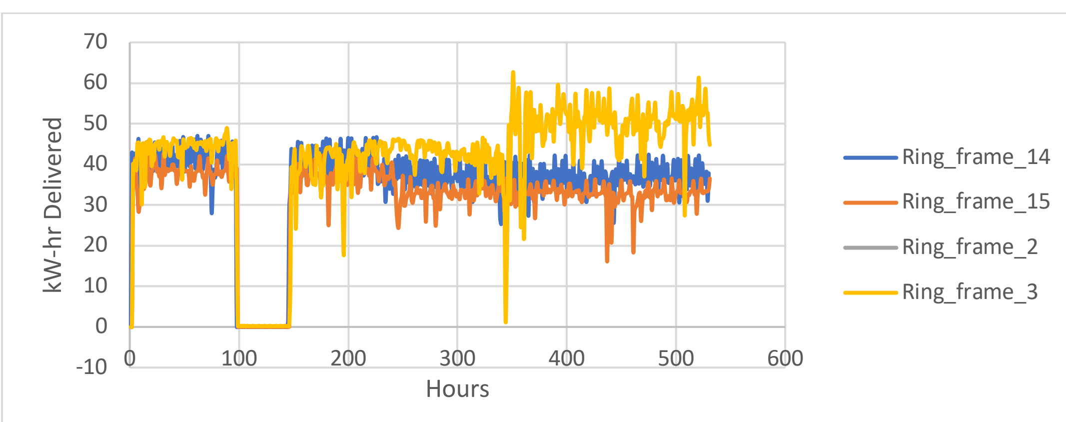
At the factory level, the implementation of statistical process control (SPC) to measure and control data quality is applied using a Bernoulli CUSUM control chart as recommended by Hazen et. al and derived from Reynolds et. al. [24]. First the historic data (X) is used to establish the Upper (UCL) and Lower Control Limits (LCL).

Figure 5 shows the kW-hr delivered to the four spinning units at Factory 1. The intra-relation consistency measurement for the collected data will be based on the allowable range of energy draw from each spinning machine and the expected accuracy of the energy sub-meter. For the purposes of this illustrative example, the consistency is defined by a binary variable based on a range defined by the mean and standard deviation of the historic data set (Equation 1). In this example, given the period of between 100-150 hours when none of the units were operating, the average and standard deviation are just based on operating time periods between hours 0 and 300 excluding hours 100-150.

Equation 1: Real-Time Consistency Equation

$$C_i = \begin{cases} 1, if\ x_i\ > \bar{X} + 0.5\sigma_X \\ 0, if\ x_i\ \leq \bar{X} + 0.5\sigma_X \end{cases}$$

Figure 5: Real-Time kW-hr Delivered to Spinning Units (Factory 1)



A Bernoulli CUSUM chart is defined by the values of p_0 (in control cumulative proportion of consistent data) and p_1 (detectable value) and H_b (the control limit). For example, a value of $p_0 = 0.01$ and a value of $p_1 = 0.025$ will be used in the following example for each of the three example plants. The control limit is set based on a variety of other factors not specifically considered in this example. Instead the value of H_b is set at 8 to illustrate how the control chart will be analyzed during data creation. Figure 6 and Figure 7 provides examples of this SPC.

Figure 6: CUSUM Ring 14 Consistency SPC

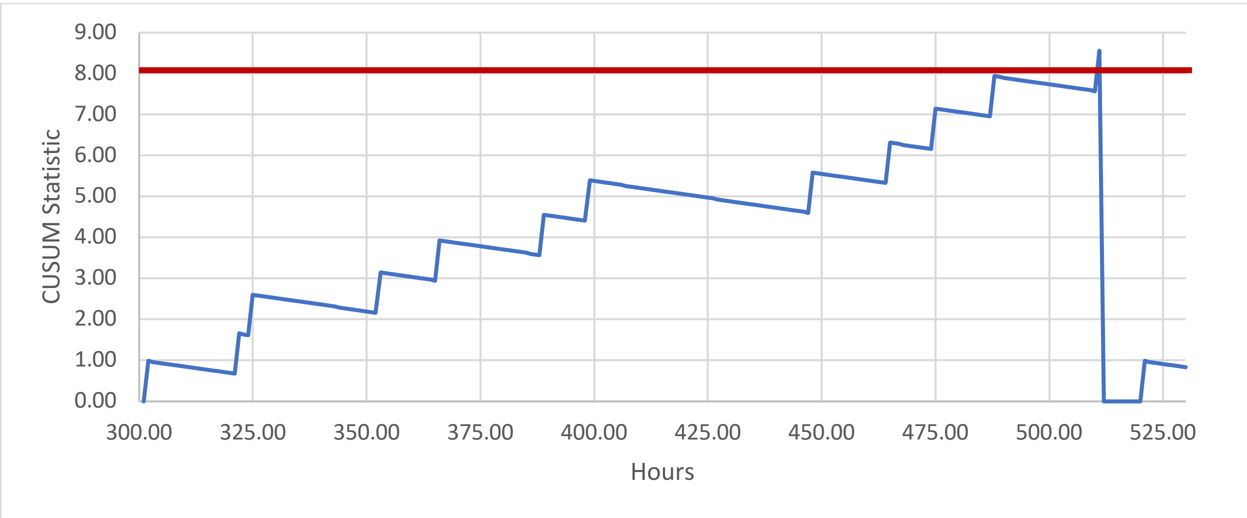
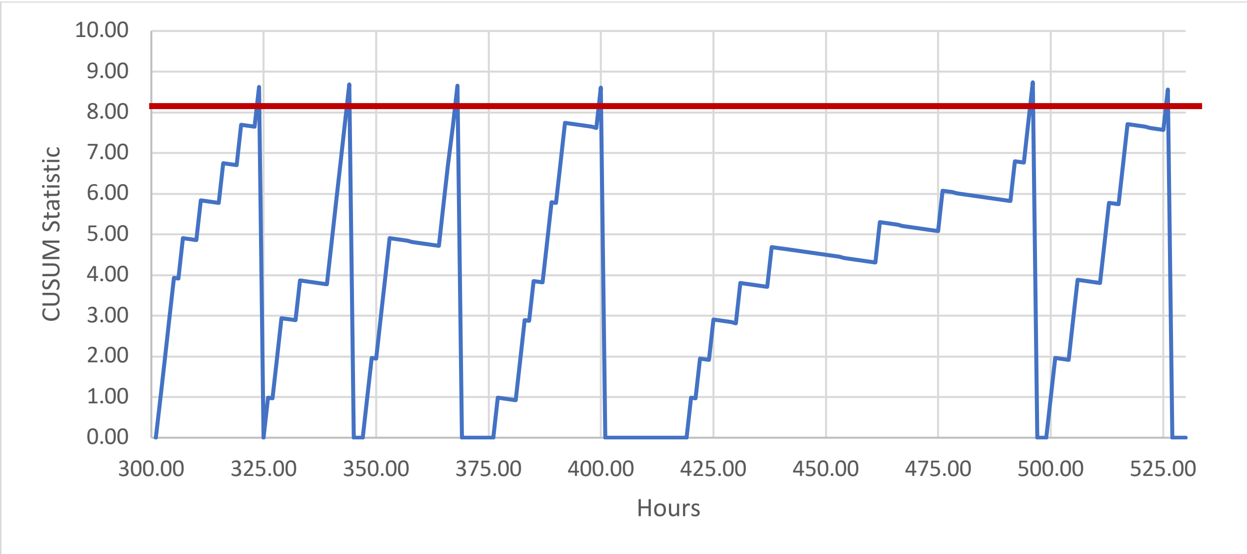


Figure 7: CUSUM Ring 2 Consistency SPC



As each real-time data point is created in the system, the consistency is measured in accordance with Equation 1and measured as a binary {0,1}. In parallel, a cumulative sum (CUSUM) is calculated and used to trend and control the overall consistency measurement of the data. The time to detect an error in the system is set by additional factors (not specifically discussed here) that can be adjusted to ensure the appropriate threshold for error detection. This threshold must be optimized to reduce false error rates and while also providing a quick response time to errors.

As shown in Figure 6, an error is detected after 200 hours of operation. When the control limit is reached the sensors must be checked and any issues identified and corrected. The system then resets back to zero.

Ring 2, as shown in Figure 7, has an error within the first 25 hours of operation. After the control limit is reached, the system automatically resets to zero, however because no steps are taken to remedy the issue on the data collection side the control limit continues to be reached on a periodic basis (~25 hours). The system appears to sell correct itself at hour 425 before moving back into the same periodic failure frequency.

These two examples illustrate a simple tool that could be implemented by the factory to control data quality during data collection. The software used for data collection will certainly also trend overall resource use and provide flags to the operators if there is a spike in resource use. There will be an expected correlation between data quality error rates and resource use error rates in the system monitoring. However, measuring and controlling data quality in real time will also ensure that issues related to system calibration such as consistency, accuracy and completeness are tracked and remedied.

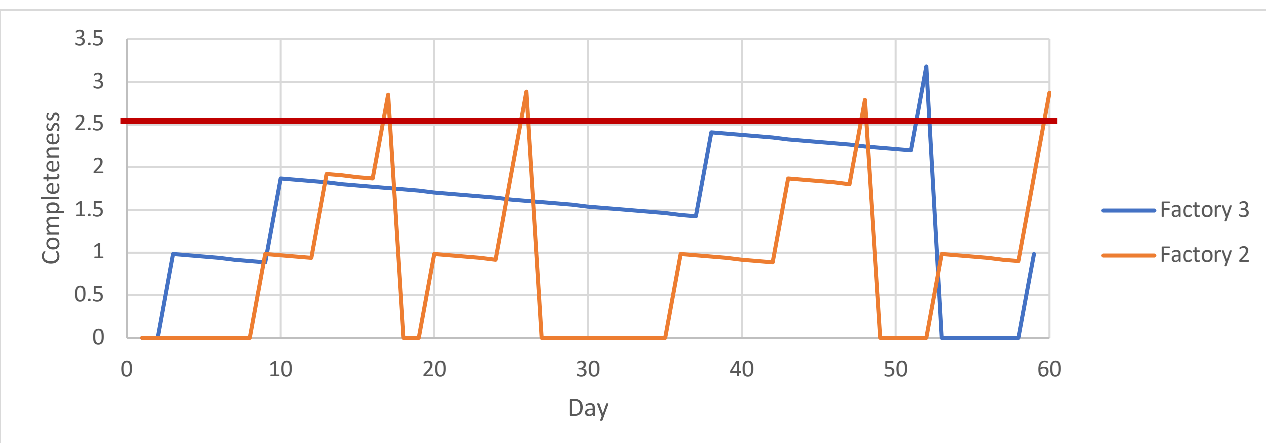
3.4.2.2 Li & Fung Implementation

Similar to data quality control at the factory level, Li & Fung must implement processes to monitor and control incoming real-time data quality levels.

A simulation of the data transmission from Factories 2 & 3 will be used to illustrate how the SPC could be used by Li & Fung to control real-time data quality. The factory data will be aggregated and sent to Li & Fung on a daily basis and the simulation will measure and control data completeness for both Factories.

Figure 8 shows the SPC for completeness of water data from Factories 2 & 3:

Figure 8: Completeness Measurement SPC for Li & Fung Factories



3.4.3 Data Quality Scoring

The methodologies outlined in Section 3.4.1 were applied to the four databases used in this study. Table 4 provides average scores for each of the four measurements across each of the dataset. *Note that with the exception of Completeness, these aggregated scores only include the features discussed in Section 3.1 for 2018 datasets and is not a comprehensive score for all features in any of the databases. It also excludes any factory that reported annual production in units other than 'Unit (piece or pair)'.*

Table 4: Average Quality Scores by Dataset

Database	Accuracy**	Timeliness	Consistency*	Completeness	Completeness (all)
Higg FEM***	0.98	--	0.94	0.85	0.24
LF Data 1***	--	--	0.99	0.99	--
LF Data 2	1.00	--	1.00	0.99	--
LF Data 3	--	--	--	--	0.48

Note: Completeness was scored with higgdataframe.csv not HiggCombined.csv. This means that some initial manipulation was performed prior to scoring the data on these four measurements.
** Only intra-relation is evaluated and scored.*
*** Because cross-validation is used to score accuracy, this measurement is done after merging the various dataframes. Thus, filtering is performed before*
**** All measurements are made only on numeric or string features. Since one-hot categorical features are either 1 or 0 for all entries.*

Table 5 provides representative scores for each of the four measurements for some of the more critical features included in the model.

Table 5: Feature Level Quality Scores

Feature	Accuracy	Timeliness	Consistency	Completeness
watercom	0.98	--	0.53	0.89
energycom	0.98	--	0.94	0.90
wastewatercom	0.99	--	0.68	0.48
wastecom	0.99	--	1.00	0.94
sipfacilityannualprodvolquant	0.94	--	0.92	0.99
siptempemployees	1.00	--	0.99	0.12
sipfulltimeemployees	0.99	--	0.98	0.99
sipoperatingdays	0.99	--	0.96	0.98
sitecity	--	--	0.99	0.99
factoryname	--	--	0.98	1.00
avgleadtime	--	--	0.99	0.99
avgunitcost*	--	--	1.00	1.00
orderqty*	--	--	1.00	1.00
totalfob*	--	--	1.00	1.00
totalorderqty*	--	--	1.00	1.00

** Assumed to be 1.00 since used in cross validation*

3.5 Data Cleaning & Preparing

Both datasets were imported into a Python script where they were formatted and merged together. A common dictionary of factory names allows the Higg dataframe to be merged to the LF dataframes. The Higg *groupid* is the unique identifier for each factory in the Higg database, which could be matched with a similar identifier from the Li & Fung dataframe. A fuzzy natural language algorithm was used to match the *groupid* feature from the two different systems to provide a complete dataset for all factories that exist within both systems.

Prior to merging the datasets, the Higg dataset was divided into four subsets to avoid duplicate instances of the same *groupid* when merging with the LF data. These four subsets included the following: 2018 Verified, 2018 Unverified, 2017 Verified, and 2017 Unverified.

3.5.1 Identifying and Correcting Inconsistent and Incomplete Data Points

The methodologies described above were applied to the final combined datasets to identify the values that needed to be removed or corrected. Table 6 thru Table 8 provide a summary of the results for the dataset that is the focus of all future analysis, 2018 Unverified.

Table 6: Incomplete Data Corrections

Feature	Qty per Feature	Action
sipoperatingdays	3	Impute the Mean
sipfulltimeemployees	1	Remove
siptempemployees	584	Keep - not a required variable
sipfacilityannualprodvolquant	3	Remove
watercom	95	Remove
energycom	83	Remove
wastewatercom	223	Keep - not analyzing at this time
wastecom	33	Keep - not analyzing at this time

Table 7: Inconsistent Data Corrections

Feature	Value	Action
sipoperatingdays	380	Impute Mean
sipoperatingdays	520	Impute Mean
sipoperatingdays	-291	Correct Sign
sipoperatingdays	3100	Divide by 10
sipfulltimeemployees	324492	Remove - Outside Range
sipfulltimeemployees	-1400	Correct Sign
sipfacilityannualprodvolquant	452000000	Remove - Outside Range
sipfacilityannualprodvolquant	2000000000	Remove - Outside Range
sipfacilityannualprodvolquant	665756868	Remove - Outside Range
sipfacilityannualprodvolquant	800000000	Remove - Outside Range
sipfacilityannualprodvolquant	220000000	Remove - Outside Range
sipfacilityannualprodvolquant	1500000000	Remove - Outside Range
watercom	6.7125E+10	Remove - Outside Range
watercom	1788580000	Remove - Outside Range
watercom	1.3669E+10	Remove - Outside Range
watercom	3588732000	Remove - Outside Range
watercom	1145364000	Remove - Outside Range
watercom	1029384000	Remove - Outside Range

watercom	1033677000	Remove - Outside Range
watercom	1400000000	Remove - Outside Range
watercom	3452382000	Remove - Outside Range
watercom	1074991000	Remove - Outside Range
watercom	1872000000	Remove - Outside Range
watercom	1122978000	Remove - Outside Range
watercom	1386159000	Remove - Outside Range
energycom	1.96E+20	Remove - Outside Range
wastewatercom	1100000000	Remove - Outside Range
wastewatercom	4.2053E+10	Remove - Outside Range
wastewatercom	1.4971E+11	Remove - Outside Range
wastewatercom	1.783E+10	Remove - Outside Range
wastewatercom	5.8581E+10	Remove - Outside Range
wastewatercom	1439653100	Remove - Outside Range
wastewatercom	3200000000	Remove - Outside Range
wastewatercom	5.843E+10	Remove - Outside Range
wastewatercom	3683162000	Remove - Outside Range
wastewatercom	1120000000	Remove - Outside Range
wastewatercom	3920325000	Remove - Outside Range
wastewatercom	6069780000	Remove - Outside Range
wastewatercom	1350000000	Remove - Outside Range
wastewatercom	1.4575E+10	Remove - Outside Range
wastewatercom	1080580000	Remove - Outside Range
avgleadtime	-1.2	Remove - Outside Range

Table 8: Accuracy (Cross Checking) Data Corrections

Feature	**Value**	**Action**
sipfacilityannualprodvolquant	821.134	Remove - Failed Verification
sipfacilityannualprodvolquant	0	Remove - Failed Verification
sipfacilityannualprodvolquant	1148472	Remove - Failed Verification
sipfacilityannualprodvolquant	144000	Remove - Failed Verification
sipfacilityannualprodvolquant	600000	Remove - Failed Verification
sipfacilityannualprodvolquant	4815000	Remove - Failed Verification
sipfacilityannualprodvolquant	11000000	Remove - Failed Verification
sipfacilityannualprodvolquant	580000	Remove - Failed Verification
sipfacilityannualprodvolquant	970390	Remove - Failed Verification
sipfacilityannualprodvolquant	7000000	Remove - Failed Verification
sipfacilityannualprodvolquant	12000000	Remove - Failed Verification
sipfacilityannualprodvolquant	4719294	Remove - Failed Verification

sipfacilityannualprodvolquant	6481488	Remove - Failed Verification
sipfacilityannualprodvolquant	0	Remove - Failed Verification
sipfacilityannualprodvolquant	350000	Remove - Failed Verification
sipfacilityannualprodvolquant	278400	Remove - Failed Verification
sipfacilityannualprodvolquant	180000	Remove - Failed Verification
sipfacilityannualprodvolquant	25	Remove - Failed Verification
sipfacilityannualprodvolquant	609367	Remove - Failed Verification
sipfacilityannualprodvolquant	7200000	Remove - Failed Verification
sipfacilityannualprodvolquant	700000	Remove - Failed Verification
sipfacilityannualprodvolquant	700000	Remove - Failed Verification
sipfacilityannualprodvolquant	2687831	Remove - Failed Verification
sipfacilityannualprodvolquant	2374192	Remove - Failed Verification
sipfacilityannualprodvolquant	500	Remove - Failed Verification
sipfacilityannualprodvolquant	1380000	Remove - Failed Verification
sipfacilityannualprodvolquant	1000000	Remove - Failed Verification
sipfacilityannualprodvolquant	4800000	Remove - Failed Verification
sipfacilityannualprodvolquant	700000	Remove - Failed Verification
sipfacilityannualprodvolquant	1108006	Remove - Failed Verification
sipfacilityannualprodvolquant	4217691	Remove - Failed Verification
sipfacilityannualprodvolquant	1338728	Remove - Failed Verification
sipfacilityannualprodvolquant	6056376	Remove - Failed Verification
sipfacilityannualprodvolquant	9600	Remove - Failed Verification
sipfacilityannualprodvolquant	1248000	Remove - Failed Verification
sipfacilityannualprodvolquant	15436	Remove - Failed Verification
sipfacilityannualprodvolquant	150000	Remove - Failed Verification
sipfacilityannualprodvolquant	0	Remove - Failed Verification
sipfacilityannualprodvolquant	1500000	Remove - Failed Verification

3.5.2 Imputing Values for Incomplete or Inconsistent data points

As shown in Table 6 and Table 7, inconsistent or incomplete data points were evaluated, and an action was taken to resolve the error directly in the data set. After correcting these errors, the final dataset was reduced from 660 entries down to 475 entries.

A check can then be performed to verify that the features were corrected as expected. Table 9 shows the final Completeness and Consistency measurements for the final dataset.

Table 9: Final Feature Quality Scores (Consistency & Completeness)

Feature	Completeness	Consistency
sipoperatingdays	1.00	1.00
sipfulltimeemployees	1.00	1.00

siptempemployees	0.12	1.00
sipfacilityannualprodvolquant	1.00	1.00
watercom	1.00	1.00
energycom	1.00	1.00
wastewatercom	0.71	1.00
wastecom	0.97	1.00

3.5.3 Measuring and Improving Overall Accuracy

Accuracy is measured with the proxies defined in Section 3.4.1.1. The initial 2018 Unverified dataset is evaluated, and the results are shown in Table 10:

Table 10: Pre- and Post- Correction Accuracy Scores

Feature	Accuracy (pre-correction)	Accuracy (post-correction)
watercom	0.98	0.98
energycom	0.98	0.98
wastewatercom	0.99	0.99
wastecom	0.99	0.99
siptempemployees	1.00	1.00
sipfulltimeemployees	0.99	0.99
sipoperatingdays	0.99	0.99
sipfacilityannualprodvolquant	0.94	1.00

As shown in the table the majority of features had a high level of accuracy prior to removing the inconsistent and incomplete entries. The accuracy remains high but does not substantially change for any of the features with the exception of *sipfacilityannualprodvolquant.*

Sipoperatingdays has less than 1.00 accuracy but after closer evaluation all values fall between 0-365 and thus the reason it scores less then 1.00 is because some values still fall out of +/- 3 Standard Deviations from the mean of the data. No corrections are required for this.

Sipfacilityannualprodvolquant now has 1.00 accuracy. This improvement shows the efficacy of the verification method used.

Comparative analysis was performed using the Water Balance verification method. In this method the total recorded water discharge is compared to the total water consumption. Any entry that reflects a higher value for water discharge than water consumption is assumed to be inaccurate and removed. This would result in the removal of an additional 38 entries. However, the removal of these entries actually reduces the accuracy of the *watercom* feature by 1% and has a negligible impact on other features. These results suggest that this verification method does not significantly improve the overall data quality of the dataset and therefore it is not implemented going forward.

Overall the accuracy is at a high and acceptable level. Since the deviation from 1.00 is a reflection of the underlying feature distributions and not due to objective errors no further filtering will be performed.

3.5.4 One-Hot Categorical Variables

Given the nature of initial data import, it was not feasible to perform quality verification on each categorical feature during initial merging. For example, there are four main features of the LF dataset that are used to categorize a given order. When importing the data from the LF dataset the orders were merged together based on factory. As a result, the order level data was lost. Rather than perform a quality assessment of the order level data, the quality assessment was then performed on the aggregated factory-level data. This process of aggregating categorical features involved converting them from descriptive strings of text to binary sub-categories. This process is called "one-hotting" and results in a new feature for every possible sub-category under the main feature. Each new subcategory becomes itself a new feature of the dataset represented as a column of either {0,1} for each factory if the factory produces garments that fit into the specific category.

This same process was applied to the categorical features from Higg and the other LF databases. After One-hotting all applicable categorical variables, the final dataset has a total of 434 categorical binary features.

3.5.5 Standardization of Numeric Features

Many forms of comparative analysis and machine learning algorithms use the Euclidian distance between two points to form comparisons or group similar data. The absolute magnitude of the distribution of data is less important than the relative spread of the data. A typical approach to address the different magnitudes of features is to scale each feature to a standard scale based on the mean and standard deviation of each feature. Specifically, the feature is transformed by in accordance with Equation 2:

Equation 2: Standardization Equation

$$F(X) = \frac{X - \mu}{\sigma}$$

Note that different machine learning algorithms are sensitive to standardization in different ways. Standardization is performed after the data is divided into testing and training sets (if applicable) to avoid introducing any new or additional information into the dataset during building of the model.

3.6 Exploratory Analysis

3.6.1 Correlation Maps

It was hypothesized that larger facilities should use more resources. What constitutes "larger facilities" is a qualitative assessment that is not easily measured from the available data. Total employees and total annual production are used as proxies for facility size. It is expected that there should be a correlation between resource use and these two features. Figure 9 provides a correlation heat map of all of 13 numeric features of the combined initial dataset.

Evaluating the heat map the following confirmatory observations are made:

- There is a strong correlation (~0.7) between *sipfacilityannualprodvolquant* and *sipfulltimeemployees.*
- There is a strong correlation (~0.75) between *watercom* and *wastewatercom.* Again, this is expected and aligns with the water balance verification methodology discussed in Section 3.5.3.
- Weaker correlations (~0.2-0.4) exist between the combined resource features. Again, this aligns with the intuition that a larger facility should use more of every resource.

Evaluating further the strength of the relationships between the independent variables (e.g. resource use) and the potential dependent variables, another heat map showing the features that rank in the top 10 highest correlation with *watercom* are shown in Figure 10. Whereas Figure 9 only included numeric features, Figure 10 considers all 456 features in the combined dataset. A few observations are made:

- As expected, water consumption is correlated to various processes performed at the facility. It is found that amongst process features, the highest correlation with *watercom* is *sipfacilityprocessess_Washing*. Although this relationship is weak (~0.41) it does confirm the intuition and increase the confidence in the data.
- There is a correlation (~0.3) between the existence of a wastewater treatment facility, *wwtreatment*, and water usage. A similar correlation exists between wastewater discharge, *wastewatercom*, and *wwtreatment* (~0.29) as expected based on the intuition of factory processes.
- Analysis of the main correlations with *energycom* and *wasterwatercom* show similar results. The main interesting takeaway from the top 10 *wasterwatercom* correlations is that *sitecountry_Bangladesh* and *equipmentqty* are both in the top 10 with a correlation of ~0.25.
- Also, important to note is the features that are not highly correlated. For example, intuition would suggest that facilities that have larger pieces of equipment (boilers, AC, etc.) should have higher energy use. Based on the data the correlations are quite weak: *equipmentqty ~0.147* and *boilersize* ~0.151.

Water, Wastewater, and Energy all have similar top 10 correlated features. An analysis of *wastecom* is shown in Figure 11. As shown in Figure 11, the main features correlated with the magnitude of *wastecom* are product related. Again, this aligns with the intuition that waste is more dependent on certain products than it is on specific manufacturing processes. It is also noted that the top correlated products are hardgoods rather than apparel. This again aligns with the intuition since waste or scraps from hardgoods are generally for higher density materials than soft-goods or apparel.

This further highlights the need to properly cluster factories to ensure that comparisons, especially on the waste feature, are made between "similar" factories where "similar" refers to specific processes and products produced.

It is also observed that there is a perfect correlation between a few of the *prodcatdesc* features. This is an indication that during feature selection, sparse or highly correlated features can be combined to reduce the dimensions of the dataset.

Figure 9: Correlation Heat Map for Numeric Features

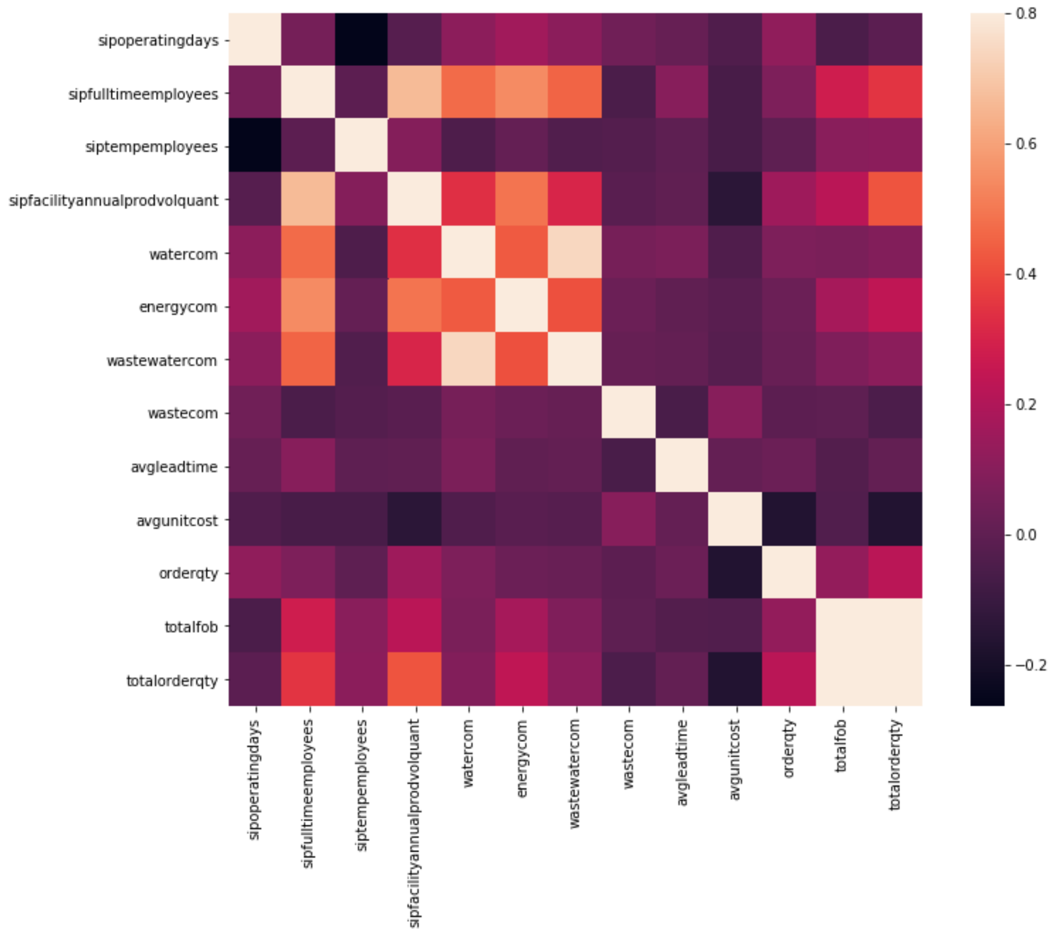


Figure 10: Watercom Top 10 Correlated Feature Heat Map

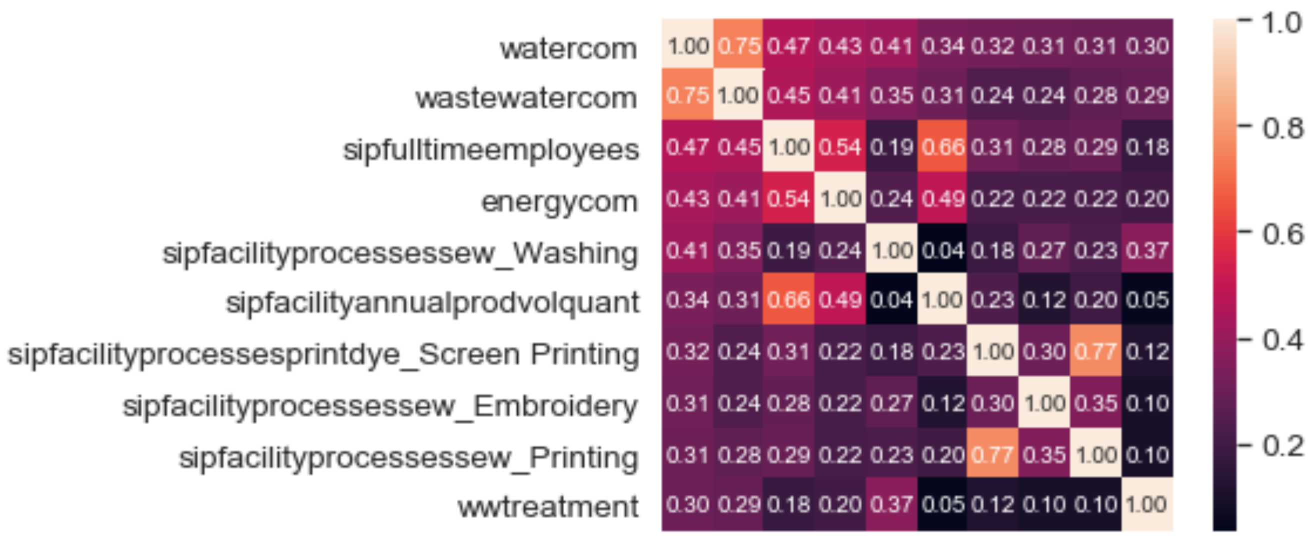
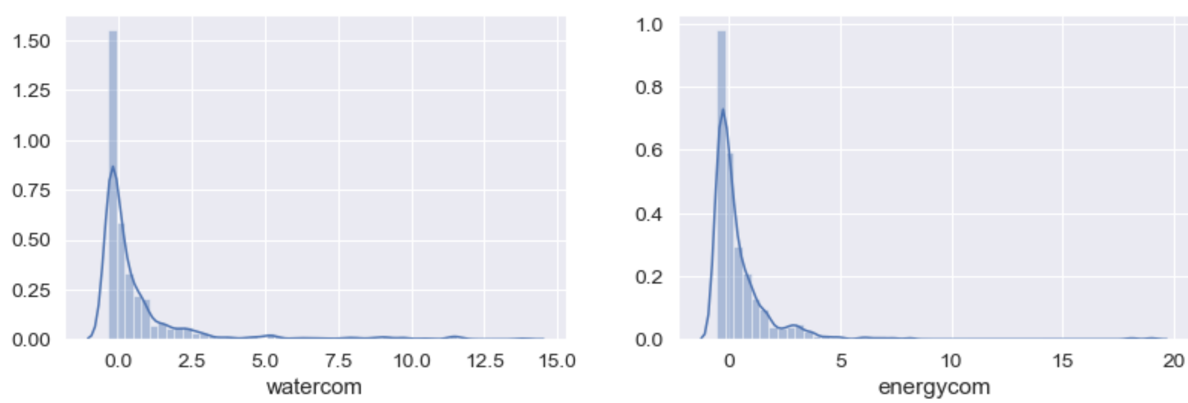


Figure 11: Wastecome Top 10 Correlated Feature Heat Map

3.6.2 Distribution Analysis

The distributions for the Water and Energy are shown in Figure 12. As shown in these figures the distributions are left-skewed as expected.

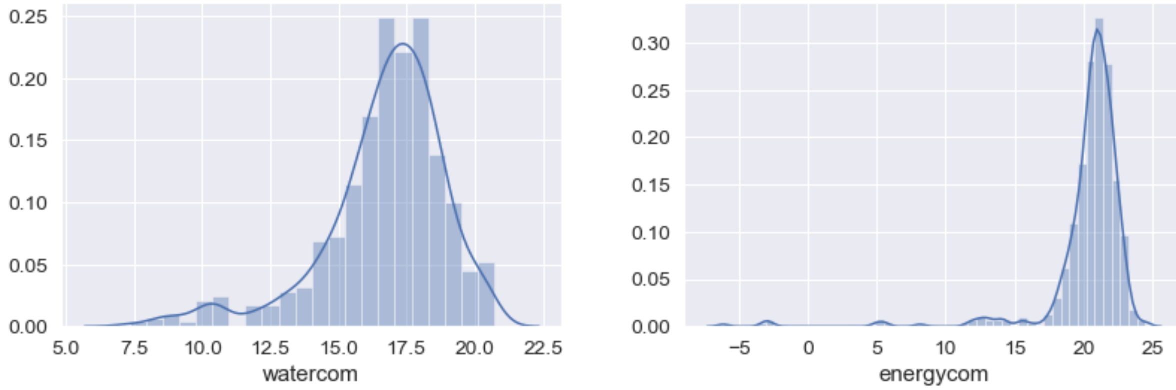
Figure 12: Distributions of Watercom and Energycom



Water (original) Energy (original)

Left-skewed distributions can be transformed via a Logarithmic transformation (based on the non-standardized data) and then tested for normality. Figure 13 show the transformed distributions and a QQ-plot to visually test for normality. As shown in these two figures, the dataset does not meet the standard of normality even after a log-transformation.

Figure 13: Log-Transformed Distributions for Watercom and Energycom



Water (transformed) Energy (transformed)

Based on the exploratory analysis the data does not conform to a normal distribution. Further investigation will need to be performed, depending on the specific algorithms used, to ensure that the underlying statistical assumptions are met.

Chapter 4: Feature Creation & Selection

4.1 Feature Creation Methodology

The data sources explored in Chapter 3 provide a variety of potentially meaningful parameters for predicting environmental performance of nodes across the garment supply chain. The purpose of the feature creation step is to create new features that combine one or more related features into a single interpretable feature. For example, the Higg dataset includes two features related to employment: number of fulltime employees and number of part time employees. Combining these features into one feature for total employees will provide a more meaningful comparison across all factories.

The methodologies employed in feature creation vary for each new feature however a few general principles apply to each:

- A new feature can be created based on a physical relationship between two or more other features
- A new feature can be created based on a learned collinearity between two or more other features. Creating a new feature based on collinearity might lead to loss of information but also serves to reduce the dimensionality of the dataset.
- A new feature can be created if there is a generally accepted metric or normalization methodology which makes a set of features more interpretable across the dataset or in physical space.

4.1.1 Combined Resources

The total resource quantities for water, energy, wastewater, and waste were aggregated from the Higg database. The aggregation process included converting each value to a common set of units based on the provided unit of measurement for each datapoint. The result was a new set of features notated as 'XXXcom' where the XXX refers to the specific resource. This was done for all of the 2017 and 2018, verified and unverified data points.

4.1.2 Resource Intensities

When developing energy benchmarks, the standard practice across different industries and methodologies is to compare an intensity or specific consumption. For example, the US Energy Star methodology benchmarks buildings based on the building Energy Use Intensity (EUI) which is defined as Energy Use per Area [23]. A similar method is applied in most academic research around energy use in the textile and garment industry. For example, specific energy consumption or sometimes referred to as energy intensity is defined as energy consumption per produced unit of output [29].

This same method was applied to all of the resources used and tracked via Higg. As discussed in Section 4.1.1, total consumption was calculated by converting each of the various inputs to a common set of units and then aggregating the totals for energy, water, wastewater and waste. Based on the created aggregate resource use features and the total production feature, *sipfacilityannualprodvolquant*, the intensities can be calculated as follows:

> *$RI = RC/TP$, where RC is the resource consumption and TP is the total production over the same period.*

Using this formal definition, the independent variables for the study were developed as follows:

- *Energyint* equal to the Total Annual Energy Consumption (W-hr)/Total Annual Production (Units)
- *Waterint* equal to the Total Annual Water Consumption (liters)/Total Annual Production (Units)
- *Wastewaterint* equal to the Total Annual Wastewater Discharge (liters)/Total Annual Production (Units)
- *Wasteint* equal to the Total Annual Waste (kg)/Total Annual production (Units)

4.1.3 Combined Employees

Two features exist within the Higg dataset that both measure employee count. These two features, *sipfulltimeeployees* and *siptempemployees*, when combined provide a standardized measurement of the total number of employees. The assumption made in the creation of this new feature, *totalemployees*, is that *siptempemployees* are contractually temporary employees but work full time the entire year.

4.1.4 Permit Quantity & Compliance

Data from Higg was used to account for the presence of and compliance with regulation. The Higg database includes a set of questions regarding required permits for a variety of processes including wastewater discharge, air quality, and chemical usage. For each type of permit two questions are asked: 1) is a permit required by regional regulations and 2) does your site have the required permit.

The "Yes/No" responses from each of these questions were used to derive two new features which provide directionally accurate indication of the role of regulation in the facility operations:

- Permit Quantity, *permitqty*: the sum total of permits that a site maintains
- Permit Compliance, *permitcompliance*: a binary variable that equals True if the site possesses the number of required permits and False if the site does not.

4.1.5 Proxy for Data Quality

A significant challenge associated with data collection and analysis across complex supply chains such as the apparel supply chain is understanding the quality level of the data used in decision making [24]. As discussed in Chapter 3, the data sources used in this analysis come from a variety of sources with different degrees of data quality. A detailed analysis of data quality was provided in Section 3.4.3 and it was shown that the final dataset used in the analysis has a sufficiently high level of data quality. This analysis however was done on a per feature and per total dataset basis and does not reflect the expected data quality of the remaining features.

The following methodology was employed to create a proxy data quality measurement for the resource data collected.

- For each resource, the Higg dataset includes a series of questions regarding the method and frequency for tracking the data.
- The tracking method must fall into one of six categories (Meters, Weighed, Invoices, Estimated, Estimates, Unknown). Similarly, the frequency must fall into one of seven

categories (Continuously, Daily, Weekly, Monthly, Bi-Monthly, Quarterly, and Annually).
- These method and frequency categories are ordinal variables with a higher weight placed on more objective collection methods and more frequent data collection.
- A quantitative scoring was developed to convert each ordinal category into a numeric value as shown in Table 11.

Table 11 Scoring Metric for Data Quality Proxy Feature

Category	Ordinal Value
Meters	5
Weighed	5
Invoices	4
Estimated	1
Estimates	1
Unknown	0

Category	Ordinal Value
Continuously	5
Daily	4
Weekly	3
Monthly	2
Bi-Monthly	1
Quarterly	1
Annually	1

- Based on these weights, each applicable feature for each factory was converted into a weighted score. For example, if electricity was metered continuously it would receive the highest score (10) but if the same meter was only read weekly it would receive a lower score (8). This reflects the prevalent industry assumption that more frequent data collection from a source with minimal human subjectively introduced will lead to the highest level of data quality.
- The new created feature, *resourcedataquality*, is the average of all of the measuring method/frequency scores.

These features will be used for two purposes:

1. Test whether there is a correlation between forms of data tracking and collection (e.g. streamed vs. bills vs. estimates) and the accuracy of the recorded data. This will involve considering the pre-processed dataset and will be explored in more detail in Chapter 6 as part of the discussion on IoT. Also check whether there is any correlation between advanced resource tracking and year on year percent reductions in use.

2. Provide an in-sample estimate on the degree of reliability for an individual factory. As a simplified proxy, it is unknown if this feature will weigh into the final scoring and clustering of the factories.

4.1.6 Number of Large Pieces of Equipment

The number of pieces of large equipment is a proxy for how industrial the setting of the factory is. Boilers, Compressors, Air Conditioning, and Refrigeration are all heavy energy users and based in academic research are generally the main consumers of electricity for garment factories [29, 30].

The Higg database provides a series of questions related to the presence of large equipment (Boilers, Generators, Engines, Ovens, Heating, Refrigeration, and Air Conditioning) as they relate to air pollution. These (Yes/No) questions were converted into binary features. Each site either has or does not have each piece of equipment. The only exception is boilers. If a factory has a boiler, they must also specify the size of the boiler {Small: Less than 50 MW, Medium: 50-300 MW, Large: more than 300 MW}. This series of questions was consolidated into two new features: *equipmentqty* and *boilersize*. *Equipmentqty* is a sum of the binary columns for each type of equipment that a factory identifies as using. *Boilersize* is an ordinal variable between 0-3 for each of the three sizes specified.

4.1.7 Employee Efficiency

Employee efficiency is the ratio of units produced, *sipfacilityannualprodvolquant*, to total employees, *totalemployees*. These two features have a correlation around 0.66. Combining them reduces the dimensionality of the overall dataset. The combined feature also provides a more normalized view of both of the features. For example, a highly automated factory and a highly manual factory might have similar annual production volumes but a very different employee count. Comparing the annual production volume as a basis for clustering would not account for the employees.

This however is speculative on the part of the researcher. Additional analysis will be performed to understand if using this as a proxy feature provides better results for the developed algorithm. Since there is a high degree of correlation between the two variables, in general an increase in annual production is followed by a proportional increase in total employee count and thus the only function of this variable, seemingly, might be to reduce dimensionality.

4.1.8 Li & Fung Database

The raw data from the LF database were purchase order-level records. During the import process the purchase-order level details were consolidated and aggregated to create a factory-level picture of the data. The following five features were created based on the consolidation and aggregation process of data import.

Other academic literature, including a recently published article that also used a LF dataset have highlighted the influence of lead time and cost on the performance of factories. These features were created to build a profile of the quality grade of garments made (high end vs. fast fashion), comparative margins, and time pressure that the factory operates under. These features will also be used, as discussed in Chapter 7, in the development of the business tools.

- *Average Lead Time*
- *Average Cost per Unit*
- *Average Order Qty*
- *Total Order Qty*
- *Total FOB*

4.2 Feature Selection

The process of feature selection is important in any machine learning application [31]. There are various approaches to feature selection, each which has different tradeoffs for the final model. For this application, feature selection was performed by applying both deterministic filtering and statistical analysis to the dataset to remove features that did not meet specific criteria such as: variance threshold, importance thresholds, and collinearity thresholds. Each of these will be developed below and then results will be provided in the subsequent section.

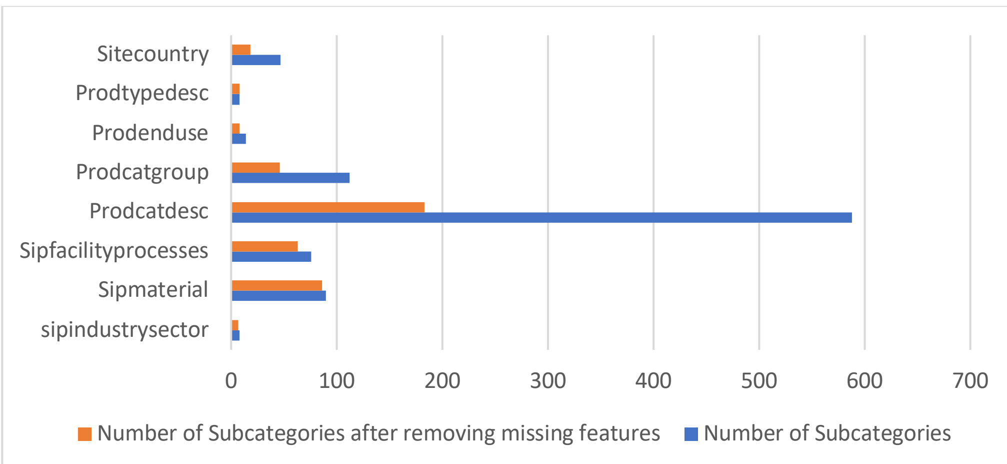
4.2.1 Initial Evaluation

Prior to feature selection the dataset had a total of 982 features in the dataset. The majority of these features were created during the one-hot processes to convert categorical variables to binary variables.

For example, there are a total of eight categorical features that have been converted to one-hot features as described in the Figure 14. The majority of the high dimensionality of the dataset comes from these one-hot features.

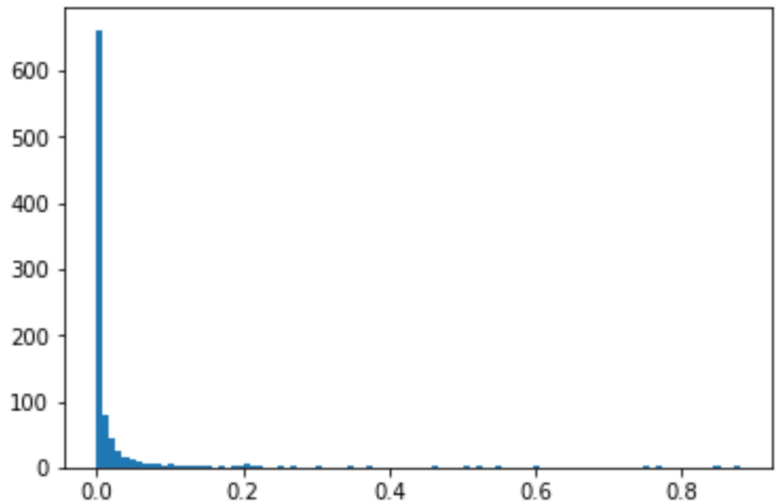
Prior to removing features, it is important to gain an understanding of the existing distribution of features, especially one-hot categorical features. A closer evaluation of these features reveals that the large majority of the one-hot features apply to less than 1% of the total dataset.

Figure 14: Distribution of One Hot Features



As shown in Figure 15, of the initial 982 features, over 700 are applicable to less than 1% of the population. Further investigation shows that 422 of the features are not represented by any of the factory population for the 2018 data.

Figure 15: Distribution of One Hot Feature Relevance



Comparing both 2017 and 2018 data there are 397 features that are not represented by any of the factories during either year. The reason these features exist is because they are represented in the LF database for specific orders. However, the factories that the orders were issued to are not currently using the Higg FEM module to record environmental data. As such these features are not represented in the combined dataset and can be removed. There is another subset of features that are represented in less than 1% of total factories.

4.2.2 Feature Selection Methodology

4.2.2.1 Removing Incorrect Features

During the creation of one-hot features, some features were created by categorical descriptions that have no physical meaning. For example, many of the Higg categories had incorrect values such as "Nan" encoded in the entry. This value was then converted to a new one-hot feature. Since the one-hot feature carries no meaning to the model or dataset it should be removed.

Two passes were performed at different stages of the dataset creation.

1. The first pass occurred during merging of the datasets. After matching the LF datasets with the Higg dataset a significant number of entries were lost because the two datasets did not have identical factory lists (e.g. some factories only existed in one of the datasets). This loss of entries also resulted in many one-hot features that were no longer applicable to any of the remaining factories. Therefore, during this phase the threshold for acceptable variance was set at a very high value, equivalent to any feature that did not have at least one applicable entry. The value, for the datasets used in this research was 0.998. This first pass reduced the number of features from 982 down to 456.

2. The second pass was performed after all data quality processing was performed and prior to performing any additional model development. This second pass included removal of two types of one-hot features: those with a suffix "_Nan" and any features that had less than one applicable entry. The _Nan features were the biproduct of the one-hot process in Python and provide no physical meaning to the dataset. This pass removed 56 features that did not

meet the variance threshold and an additional 16 "_Nan" features. The final number of features after this pass is 384.

Finally, it was decided to remove the Higg "Score" variables since these are calculated values based on an internal Higg methodology and are not a representation of the physical factory. Removing these further reduces the overall feature dimension to 376.

4.2.2.2 Test for Collinearity between Features

A test for collinearity between the independent features was performed. This analysis was performed on all one-hot variables to understand if there was any existing redundancy between two one-hot features. For example, the LF dataset included four descriptive features for classifying an order. In most cases, there was dependence between these four feature sets (e.g. the majority of denim products were also classified as bottoms). Therefore, to understand and limit mutual information between features (without losing the predictive power of the overall dataset), collinearity tests were performed to understand what variables were shared high degrees of mutual dependence.

As an example, for a correlation threshold of 0.7, there were 102 features identified as being collinear. In this methodology, the features that have a collinear value above the threshold will be identified and then one of the two features will be removed. In the example of a threshold equal to 0.70, 51 features would be removed further reducing the dimensionality of the dataset from 376 features to 325.

4.2.2.3 Stacked Ensemble Method for Select Features

The final methodology to select features is an iterative approach to determine which features contribute the most to predicting resource use. This process involves the following 4 steps:

1. k ensemble machine learning algorithms are developed. These machine learning algorithms combine multiple machine learning algorithms in an attempt to improve overall performance of the combined algorithm. For the purposes of this research, the following algorithms were used: random forest, gradient boosting, and adaptive boosting (e.g. AdaBoost).

2. Each k algorithm is used to develop n x m models, where n is the number of resources considered in the final clustering algorithm (in this case $n = 4$) and m is the number of iterations that the model is run for each resource.

3. For each of the n x m x k models the relative importances of the final features selected in the model are stored in a new array. These importances are then averaged over the n x m x k models.

4. The final features are selected based on a normalized importance value of each features derived from the scoring of each of the n x m x k models. Specifically, a cumulative sum of the importances is developed and any feature importance that does not contribute to a specific threshold (e.g. 95% of total feature importance) is removed.

4.2.3 Feature Selection Application

Each of the approaches discussed above was employed to both analyze the existing relationships between features and then using the learned knowledge to reduce the overall dimensionality of the dataset.

In applying the developed methodology, there are a number of tuning parameters that can be adjusted to improve the overall results of the feature selection process. These tuning parameters are outlined below:

- Collinearity Threshold
- Cumulative Importance Threshold
- Variance Threshold
- Number of iterations for identifying zero importance
- Ensemble method hyper parameters

4.2.3.1 Metrics

In order to guide the parameter tuning a number of metrics were selected for use in assessing and comparing various models. Based on the initial explanatory analysis performed in Section 3.6, it is not expected that any of the models developed will score highly on the metrics used. However, these metrics will be directionally accurate. For example, if a model consistently has an R^2 value below 0, then the model should not be used in feature selection.

These metrics include the following, each of which is averaged for all *n* resources used in the model selection process.

- R^2: coefficient of variance, R^2, is a measure of the explained variance.
- Mean Squared Error: average squared difference between the predicted and actual value
- Mean Absolute Error: average absolute difference between the predicted and actual value
- Dimensionality: number of features

4.2.3.2 Establishing Baseline

Prior to tuning any parameters, a baseline model will be established with initial values for each parameter as outlined in Table 12. Additionally, a seed (random_state = 5) for the random number generator will be set to ensure comparability across different iterations.

Table 12: Baseline Stacked Ensemble Model Parameter Values

Parameter	**Value**
Collinearity Threshold	0.90
Cumulative Importance Threshold	0.99
Variance Threshold (min number of applicable entries)	2
Number of iterations for identifying zero importance	2
Ensemble method hyper parameters	None

Based on these initial parameters, the baseline model performs as outlined in Table 13:

Table 13: Baseline Performance of Stacked Ensemble Model

	R^2	MSE	MAE	Dimensionality
waterint	0.247	1154.083	19.360	153
energyint	0.210	942822.340	495.559	
wastewaterint	0.106	953.419	16.772	
wasteint	-0.083	63.928	2.150	
average	0.120	236248.442	133.460	

Given the number of possible tuning parameters, solving the resulting optimization problem for correct parameter values is highly complex. First, a seed is again set for splitting of the data into test and train sets. The same seed will be used for each iteration of parameter tuning. For larger number of iterations for identifying zero importance, a string of seeds will be used consistent (random_state = [1 ,40, 32, 81, 9]).

A set of heuristics will be used to set the values for some of the parameters. These steps are detailed in Table 14. Prior to using these heuristics, each parameter is evaluated to develop an understanding of the likely effect it will have on the model:

- First the Collinearity and Variance Thresholds are used to reduce the dimensionality of the dataset and performance of the model. If for example, the Collinearity and Variance Thresholds are set too low then a large number of potentially important features will be removed prior to building the stacked ensemble model. This might result in a decrease in model performance.
- The Importance Threshold is expected to have minimal effect on the performance of the model but will significantly reduce the dimensionality of the model.
- The Number of Iterations is expected to increase the performance of the model up to some value.

Table 14: Heuristic Approach to Stacked Ensemble Model Optimizing

Step	Collinearity Threshold	Importance Threshold	Variance Threshold	No of Iterations	R^2	MSE	MAE	Dimensionality
Baseline	0.90	0.99	2	2	0.120	236248.44	133.46	153
Step 1, set Collinearity Threshold to 0.8	0.80	0.99	2	2	0.115	236647.85	132.967	140
Step 2, set Collinearity Threshold to 0.95	0.95	0.99	2	2	0.123	238609.11	132.602	150
Step 3, set Collinearity Threshold to 0.99	0.99	0.99	2	2	0.123	238609.11	132.602	150
Step 4, set Importance Threshold to 0.90	0.95	0.90	2	2	0.123	238609.11	132.602	87
Step 5, set Importance Threshold to 0.8	0.80	0.80	2	2	0.123	238609.11	132.602	63
Step 6, set Importance Threshold to 0.5	0.50	0.50	2	2	0.123	238609.11	132.602	38
Step 7, set Variance Threshold to 5	0.95	0.99	5	2	0.115	236902.30	134.003	135
Step 8, set Variance Threshold to 1	0.95	0.99	1	2	0.123	238209.08	131.908	150
Step 9, set Number of Iterations to 3	0.95	0.99	2	3	0.160	184836.12	126.059	153
Step 10, set Number of Iterations to 5	0.95	0.99	2	5	0.179	221176.66	128.031	160

** Grey highlights indicate final "optimized" value for each threshold*

4.3 Feature Selection Results

Based on the above discussed feature selection methodology, the final selected parameters and included features (top 15 importance) are shown in Figure 16 and Figure 17. Table 15 provides the final detailed model performance for final tuned model from Table 14.

Table 15: Final Stacked Ensemble Model Performance

Resource	R^2	MSE	MAE	Dimensionality
waterint	0.264	845.925	17.325	160
energyint	0.301	883129.202	478.162	
wastewaterint	0.210	676.100	14.653	
wasteint	-0.060	55.401	1.986	
average	0.179	221176.657	128.031	

Figure 16: Cumulative Feature Importance for Final Stacked Ensemble Model

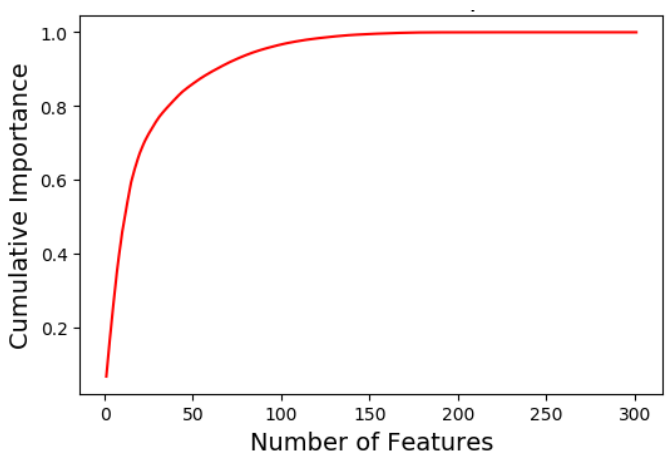
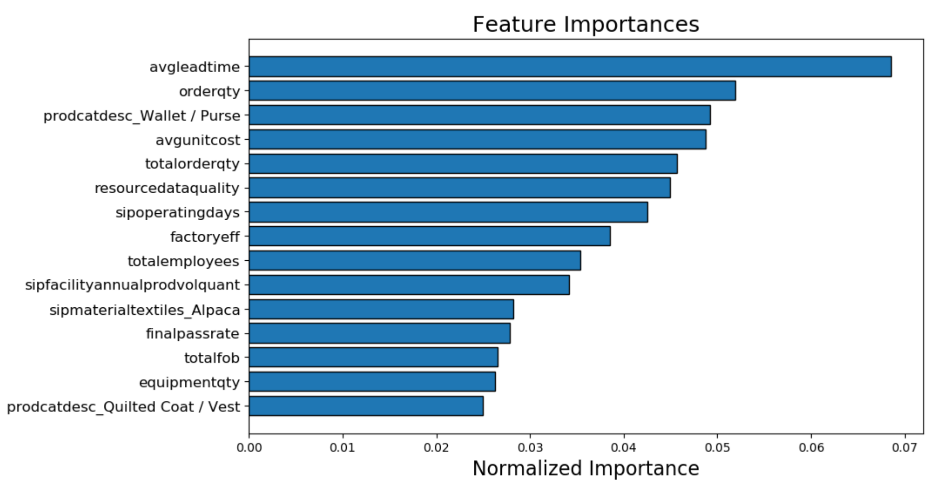


Figure 17: Top 15 Features based on Importance



Chapter 5: Clustering

5.1 Introduction to Current Tier Clustering

The apparel industry, as discussed in Chapter 2, generally partitions nodes along the supply chain into Tiers. Each Tier represents a general set of processes ranging from raw material, yarn formation, textile formation, textile finishing, and cut and sew. The division of these processes into tiers is further guided by the perceived level of direct or indirect control a brand has on the specific supplier at each node. Although the Tiers are often thought of as following a linear set of processes up the supply chain, the reality is that a single piece of fabric might often move between Tiers in a non-linear, or recursive manner. The most commonly used nomenclature to divide these processes into Tiers is as follows:

1. *Tier 1 - Cut & Sew facilities:* Cut and Sew facilities are responsible for the final garmenting process. They convert finished textiles into apparel though a set of labor-intensive, manual cutting and sewing processes. Brands and sourcing agents such as Li & Fung have direct relationships with this Tier and can influence environmental and social practices of the facilities.

2. *Tier 2 – Dying, Finishing, and Trim facilities*: This Tier is often a catch all for processes that are generally outsourced by a Cut & Sew facility and also not part of the Tier 3 processes. Most commonly these facilities receive raw textiles from a Tier 3 facility and perform the dying, washing, and any other finishing processes required to convert the raw textile into the desired finish material. These facilities often also perform a set of processes that occur after the cut and sew step. For example, many finished goods require a final wash phase. This wash phase might be classified under Tier 2 and outsourced by the cut & sew supplier to a different facility.

3. *Tier 3 – Mills:* This Tier is generally thought of as the set of suppliers responsible for textile formation, but often in reality suppliers are vertically integrated and perform yarn formation and textile formation in the same facilities. This Tier is the furthest removed from the Brands and Sourcing Agents however it is also generally the Tier with the largest environmental impact.

Chapter 2 highlighted a specific example of why the classification of factories as a specific Tier is not necessarily informative from the perspective of understanding how a facility is performing. Most facilities do not fall into a specific Tier because they are either vertically integrated (which is often the case for facilities in India) or because they perform a majority of processes that are generally classified as Tier 2.

This chapter will provide an overview of the Tier breakdown of the studied facilities and then develop a set of metrics for properly measuring the validity of a classification methodology. It will conclude with a comparison of these metrics for the developed clustering methodology and the Tier based classification.

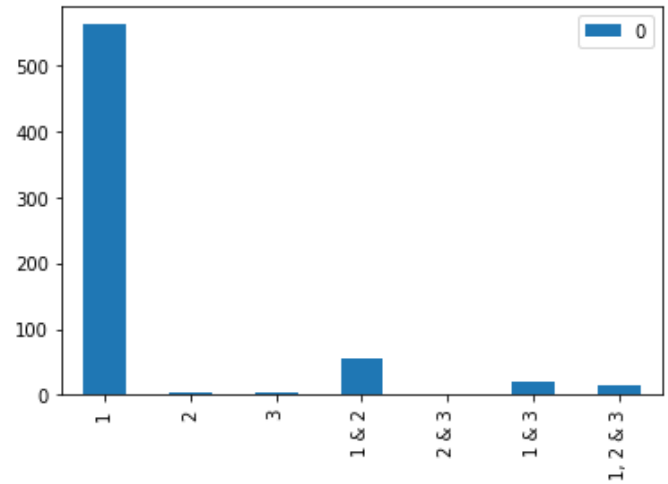
5.1.1 Analysis of Dataset based on Tiers

The Higg dataset has a set of features, *sipfacilitytype_*, that are used to classify a facility into the above described Tiers. The most general approach, and the one that will be used as the baseline for validating the clustering methodology will be to use these features to classify the factories and then analyze each Tier based on the metrics for comparison discussed in Section 5.2.

In analyzing the Higg dataset the first observation was that many factories were classified as more than one Tier. Thus, the intial baseline is to classify each factory into one of the following 7 possible classifications: Tier 1, Tier 2, Tier 3, Tier 1 & 2, Tier 2 & 3, Tier 1 & 3, or Tier 1, 2, & 3. Figure 18 below shows a breakdown of how the original, unfiltered set of factories from 2018 are classified:

Figure 18: Histogram and Table of Tier Breakdown (pre-filtered)

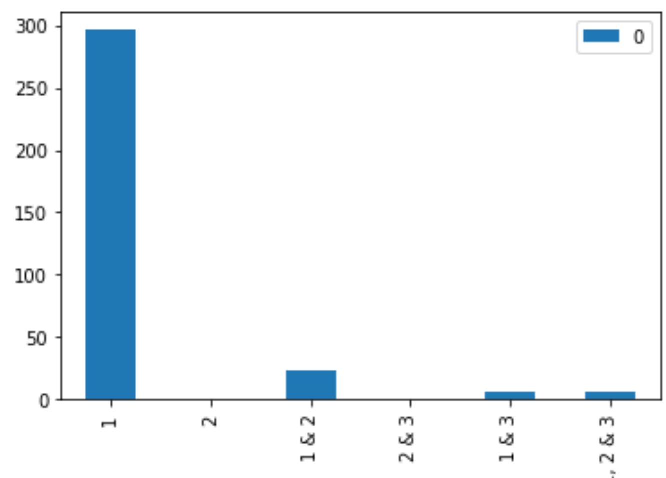
Tier	Percentage
Tier 1	85.2%
Tier 2	0.4%
Tier 3	0.6%
Tier 1&2	8.3%
Tier 1&3	2.9%
Tier 2&3	0.2%
Tier 1,2&3	2.4%



After filtering, the final set of factories that will have been included in the clustering model are distributed as shown in Figure 19:

Figure 19: Histogram and Table of Tier Breakdown (post-filtered)

Tier	Percentage
Tier 1	88.1%
Tier 2	0.3%
Tier 3	0.0%
Tier 1&2	7.1%
Tier 1&3	2%
Tier 2&3	0.3%
Tier 1,2&3	2%



Further analysis was performed to understand the distributions of resource intensity across each of the seven categories.

The expectation is that different tiers should perform differently on environmental performance. Academic research, life cycle analyses, and industry knowledge all point to Mills for example as being the most water and energy intensive part of the manufacturing process.

The underlying assumption is that a Tier based categorization provides the correct basis for comparisons. To understand this, Table 16 below provides the key statistics for each of the 6 Tiers represented in the final filtered datasets.

A few general conclusions can be drawn from Table 16. First as shown visually in Figure 18 and Figure 19, the majority of the data represents Tier 1 facilities. This aligns with current practices within the industry where most brands and sourcing agents work directly with Tier 1 facilities and then Tier 1 facilities contact out the Tier 2 and 3 facilities. Given the small sample size for Tier 2 & 3 facilities, the statistics presented in Table 16 cannot be considered conclusive.

Second, assuming that the small sample size for the Tier 2 & 3 facilities still represent the overall distribution of the Tier 2 & 3 facilities (this is not a statistically sound assumption), with the exception of Waste Intensity (which does not have the necessary granularity to draw any conclusions), it appears that the mean intensity value for each of the remaining three resources tends to be lower for a Tier 1 facility versus the others.

Table 16: Key Distribution Statistics for Tier Based Clusters

	Water Intensity			
	Tier 1, 2 & 3	Tier 1 & 2	Tier 1 & 3	Tier 1
count	7.00	24.00	7.00	296.00
mean	32.04	44.63	7.35	21.47
std	35.43	63.91	12.29	63.59
min	4.26	1.79	1.03	0.00
25%	12.85	6.12	1.66	2.24
50%	21.64	21.03	2.30	6.19
75%	32.85	52.33	4.96	19.50
max	106.96	266.77	34.84	969.51

	Energy Intensity			
	Tier 1, 2 & 3	Tier 1 & 2	Tier 1 & 3	Tier 1
count	7.00	24.00	7.00	296.00
mean	779.85	908.74	579.81	686.09
std	763.32	1569.48	343.10	1391.04
min	5.02	0.00	153.44	0.00
25%	42.64	142.96	335.94	132.47
50%	1006.82	325.77	541.74	278.47
75%	1198.85	834.16	806.86	654.89
max	1964.12	7326.33	1077.88	15411.00

	Wastewater Intensity			
	Tier 1, 2 & 3	Tier 1 & 2	Tier 1 & 3	Tier 1
count	7.00	24.00	7.00	296.00
mean	20.45	33.96	2.29	16.83
std	15.72	49.29	1.81	54.33
min	8.52	0.01	0.13	0.00
25%	11.03	3.93	1.22	1.09
50%	16.76	12.36	2.30	4.47
75%	20.87	39.59	2.76	12.99
max	54.10	172.11	5.61	788.57

	Waste Intensity			
	Tier 1, 2 & 3	Tier 1 & 2	Tier 1 & 3	Tier 1
count	7.00	24.00	7.00	296.00
mean	0.03	0.49	0.03	0.85
std	0.03	2.23	0.06	5.18
min	0.00	0.00	0.00	0.00
25%	0.00	0.01	0.01	0.00
50%	0.02	0.03	0.01	0.01
75%	0.04	0.05	0.02	0.04
max	0.08	10.98	0.16	68.34

Table 16 can be visualized in Figure 21. As shown in Figure 21 the mean of the distributions for non-Tier 1 factories is generally to the right of the mean value for Tier 1 factories.

However, based on the scatter plots no conclusion can be made about the distribution of one Tier vs. another Tier.

To test this further, a Tukey Test can be performed between the difference of means for each of the Tiers for each of the resources are shown in Figure 20.

Figure 20: Tukey Test Results for Tier Clustering

Water Intensity Tukey Test

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
1	1 & 2	23.1615	-14.9542	61.2772	False
1	1 & 3	-14.1219	-82.7981	54.5544	False
1	1, 2 & 3	10.5713	-58.105	79.2476	False
1	2	-19.5458	-199.4381	160.3466	False
1	2 & 3	-18.8796	-198.7719	161.0128	False
1 & 2	1 & 3	-37.2834	-114.4282	39.8614	False
1 & 2	1, 2 & 3	-12.5902	-89.735	64.5546	False
1 & 2	2	-42.7073	-225.9998	140.5852	False
1 & 2	2 & 3	-42.0411	-225.3336	141.2514	False
1 & 3	1, 2 & 3	24.6932	-71.3013	120.6877	False
1 & 3	2	-5.4239	-197.4129	186.5651	False
1 & 3	2 & 3	-4.7577	-196.7467	187.2312	False
1, 2 & 3	2	-30.1171	-222.1061	161.8719	False
1, 2 & 3	2 & 3	-29.4509	-221.4399	162.5381	False
2	2 & 3	0.6662	-253.3114	254.6437	False

Energy Intensity Tukey Test

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
1	1 & 2	222.6515	-619.0752	1064.3782	False
1	1 & 3	-106.2829	-1622.8933	1410.3275	False
1	1, 2 & 3	93.7545	-1422.8559	1610.3649	False
1	2	-601.6618	-4574.3087	3370.9851	False
1	2 & 3	-360.4188	-4333.0657	3612.228	False
1 & 2	1 & 3	-328.9344	-2032.5586	1374.6898	False
1 & 2	1, 2 & 3	-128.897	-1832.5212	1574.7272	False
1 & 2	2	-824.3133	-4872.0474	3223.4209	False
1 & 2	2 & 3	-583.0703	-4630.8045	3464.6638	False
1 & 3	1, 2 & 3	200.0374	-1919.8538	2319.9286	False
1 & 3	2	-495.3789	-4735.1613	3744.4035	False
1 & 3	2 & 3	-254.136	-4493.9184	3985.6464	False
1, 2 & 3	2	-695.4163	-4935.1987	3544.3661	False
1, 2 & 3	2 & 3	-454.1734	-4693.9558	3785.609	False
2	2 & 3	241.2429	-5367.462	5849.9479	False

Wastewater Intensity Tukey Test

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
1	1 & 2	17.1316	-15.1346	49.3979	False
1	1 & 3	-14.5419	-72.6787	43.595	False
1	1, 2 & 3	3.627	-54.5099	61.7639	False
1	2	-14.9832	-167.2684	137.302	False
1	2 & 3	-14.2958	-166.5809	137.9894	False
1 & 2	1 & 3	-31.6735	-96.9792	33.6323	False
1 & 2	1, 2 & 3	-13.5046	-78.8104	51.8011	False
1 & 2	2	-32.1148	-187.2784	123.0487	False
1 & 2	2 & 3	-31.4274	-186.5909	123.7361	False
1 & 3	1, 2 & 3	18.1689	-63.0938	99.4316	False
1 & 3	2	-0.4413	-162.9667	162.0841	False
1 & 3	2 & 3	0.2461	-162.2793	162.7715	False
1, 2 & 3	2	-18.6102	-181.1356	143.9152	False
1, 2 & 3	2 & 3	-17.9228	-180.4482	144.6026	False
2	2 & 3	0.6874	-214.3135	215.6883	False

Waste Intensity Tukey Test

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
1	1 & 2	-0.3542	-3.3581	2.6498	False
1	1 & 3	-0.8169	-6.2293	4.5956	False
1	1, 2 & 3	-0.8231	-6.2355	4.5894	False
1	2	-0.7311	-14.9086	13.4464	False
1	2 & 3	-0.5818	-14.7594	13.5957	False
1 & 2	1 & 3	-0.4627	-6.5426	5.6172	False
1 & 2	1, 2 & 3	-0.4689	-6.5488	5.611	False
1 & 2	2	-0.3769	-14.8224	14.0686	False
1 & 2	2 & 3	-0.2276	-14.6732	14.2179	False
1 & 3	1, 2 & 3	-0.0062	-7.5716	7.5593	False
1 & 3	2	0.0858	-15.0451	15.2167	False
1 & 3	2 & 3	0.235	-14.8959	15.3659	False
1, 2 & 3	2	0.092	-15.0389	15.2229	False
1, 2 & 3	2 & 3	0.2412	-14.8897	15.3721	False
2	2 & 3	0.1493	-19.867	20.1655	False

Finally, to help further visualize the entire feature space, a Principle Component Analysis (PCA) is performed to reduce the dimension space to two principle components.

Figure *22* visualizes the PCA components with color-coded Tiers.

Based on the results from the Tukey test and visualizations it appears that the available data does not have sufficient evidence to support the granular case studies and Life Cycle Analyses (LCAs) that suggest significant differences between resource use of different Tiers. This is likely due to the limited information encoded in the *sipfacilitytype* set of features. While ascribing to the Tier classification, it does not provide the necessarily granularity to capture the effects of different processes, material types, equipment, etc. that affects the overall environmental performance of a factory.

Figure 21: Scatter & Distribution Plots for Resource Intensities

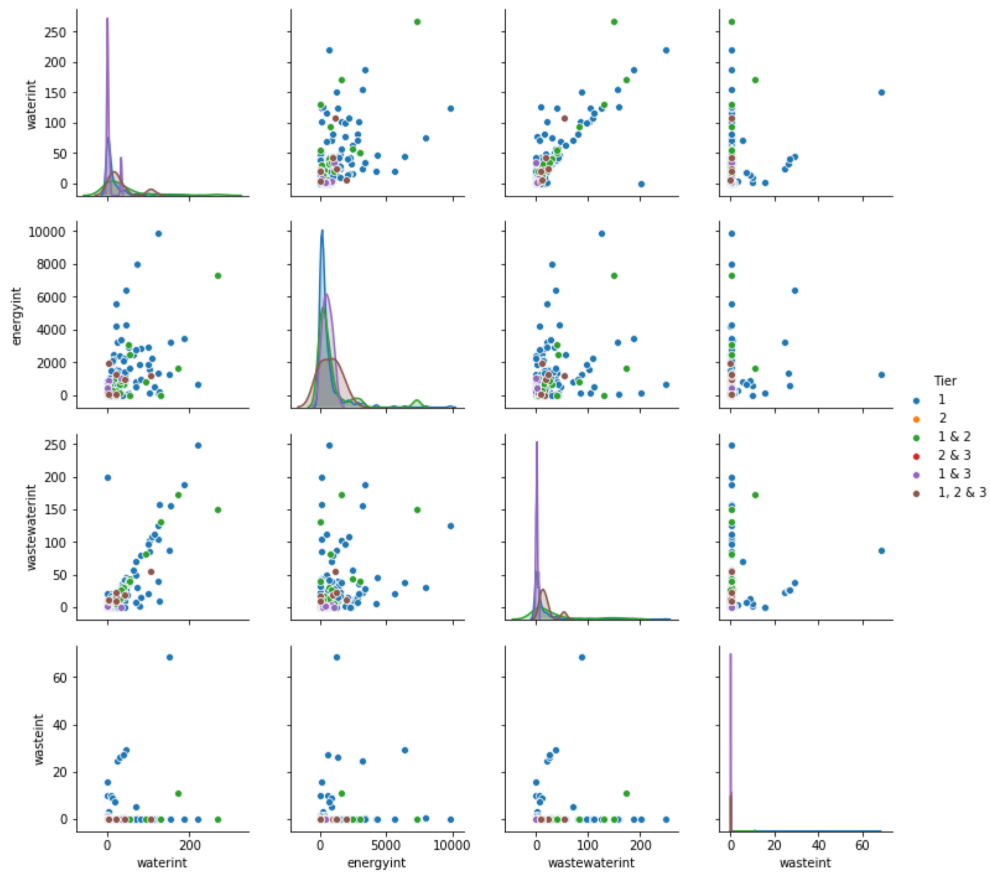
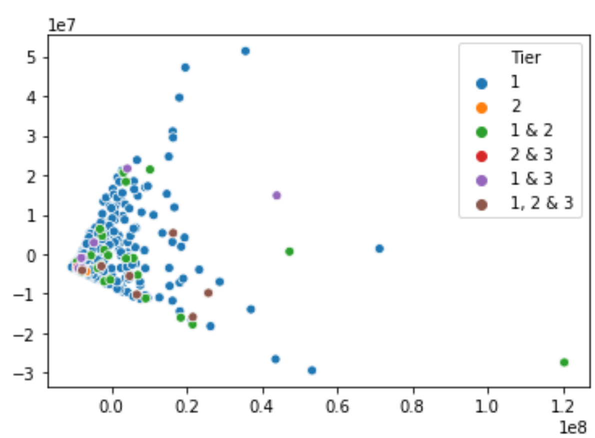


Figure 22: PCA Scatter Plot of Tier Clusters



5.2 Metrics for Comparison

The metrics used to compare the tested clustering algorithms to the baseline Tier classification should account for performance along multiple dimensions. The goal of clustering is to group "similar" factories to enable appropriate comparisons to be made. The end use of these comparisons will be to evaluate supply chain performance and improvement, guide sourcing decisions to reduce environmental impact, and benchmark factories to optimize the overall network. As such, the metrics for optimization cannot just focus on reducing the variance between resource intensity. Nor can it just focus on grouping factories that perform the same set of processes. The ideal grouping will be optimized on both of these dimensions to allow for business tools to be developed.

1. *Cluster "Goodness" Measures:* five measures will be used to assess the overall variance, separation, and density of the clusters.
   a. *Silhouette Score:* Measure of how well defined a cluster is based on the distance between points within a cluster and the distance between points and all other points in the next nearest cluster.

   b. *Calinski and Harabasz Score*: Measure of variance between clusters based on the ratio of between-cluster dispersion to within-cluster dispersion.

   c. *Davies-Bouldin Score*: Measure of similarity and separation based on distances between points and the centroid and the distances between centroids.

   d. *Mutual Information Score:* Measures the similarity of two labels of the same set of data. This will be used to compare two clustering methodologies to understand agreement.

   e. *Tukey Test:* A statistical analysis to test the hypothesis that the means of two different clusters are statistically significant.

2. *Mutual Feature Measures:* This measure will compare specific factories within the same cluster to understand how many similar one-hot features they share and the percent decrease in standard deviation between the clusters and the overall dataset. For each factory in a given cluster, the normalized number will be calculated and used to score the degree to which the cluster groups "like" factories.

   The specific measurement is calculated as follows:

   Average proportional similarity between binary features: The proportion of "Yes" and "No" for each feature on a per cluster basis is calculated and then the weighted average feature proportion is calculated for all features across all clusters in a specific methodology. Finally, the average of the total feature space is calculated. The theoretical highest possible score is 1; the lowest score is 0.5 since the equation will always select the larger proportion (e.g. whichever is larger "yes" or "no").

   Average percent decrease in standard deviation of numerical features: The standard deviation of each numerical feature for a given cluster is calculated and then the weighted average is taken for each feature across all clusters. Finally, the average of all features is calculated. The possible range is between 0 and 1.

Overall Mutual Feature Measure: The average of the previous two measurements is taken to provide an overall measurement of mutuality for the clustering.

The weighted average is calculated based on the total number of factories in a given cluster. This weight penalizes smaller clusters which would likely have higher proportions and smaller standard deviations. For example, if a cluster of one is created then the standard deviation would be zero and the proportion would also be one.

5.2.1 Metrics Applied to Tier Classification

Table 17 below provides the results for the Tier based classification. These values serve as the baseline for testing the relative improvement of the clustering algorithm.

Table 17: Tier Clustering Scoring

Metric	**Score**
Silhouette Score (higher is better)	-0.467
Calinski and Harabasz Score (higher is better)	2.697
Davies-Bouldin Score (lower is better)	4.538
Tukey Test	See Section 3.4.1.1
Mutual Feature Measures	0.086

5.3 Clustering Methodologies

Two families of clustering algorithms were assessed to understand which performed best based on the metrics defined above. Most clustering algorithms depend on Euclidean distances between numerical features. Given that the final dataset includes both numerical and categorical features, the applied families of algorithms were selected based on their ability to include both data types.

The two clustering methodologies that will be explored are K-means and ROCK. K-means further has a sub-set of methodologies that can incorporate categorical data. These include K-modes (categorical only) and K-prototypes (mixed type).

5.3.1 K-means

K-means is one of the most common unsupervised machine learning approaches to finding natural clusters in large datasets. The algorithm seeks to assign each observation or entry into k exclusive clusters defined by the cluster centroid.

K-means is based on Euclidean distances between features and thus for this application, given the large number of categorical features, K-means is limited in its application scope. Huang adapted the K-means clustering algorithm to categorical data, called K-modes. The final algorithm, K-prototypes, can find clusters in mixed-type datasets [32].

A Python package, *kmodes*, [33] based on the K-modes algorithm was developed and available online. This package was used to develop a clustering algorithm for the data set. Models of each of the three implementations of the standard K-means approach, K-means, K-modes, and K-prototypes were developed in Python with this package. The results of these various clusters are compared in Table 18.

5.3.2 ROCK

The ROCK algorithm, *RObust Clustering using linKs*, is a derivative algorithm developed by Guha, et. al to address the shortcomings of the K-means algorithm. ROCK builds clusters based on *links* between datapoints [34]. Rather than relying on Euclidean distances between points, ROCK relies on assessing the number of *common neighbors* from a global rather than just local perspective. This notion of links is especially helpful in assessing the similarity of data points based on categorical variables.

An implementation of the ROCK algorithm is available in the PyClustering library [35]. This package was used to perform a ROCK clustering on the dataset, the results of which were compared to the K-means approaches discussed above.

5.4 Model Validation

The results of various clustering methods are shown in Table 18. Each clustering method is compared across the four metrics and to the baseline *Random* and *Tier* approach to clustering. The *Random* clustering approach used a random number generator in Python to create three clusters of the data. As can be seen below, the *Random* clusters actually perform better than the *Tier* approach in 2 of the 4 metrics.

Table 18: Scoring for Each Clustering Model

	Number of Clusters	Mutual Features	Silhouette	Davies Bouldin	Calinski Harabaz
Tier	5	0.086	-0.467	4.538	2.697
Random	3	0.187	-0.091	12.192	1.469
Kproto	2	0.277	-0.041	2.806	13.981
Kmeans	2	0.169	0.561	1.156	139.447
Kmodes	2	0.187	-0.085	3.468	7.207
Kproto	3	0.313	-0.048	3.393	28.602
Kmeans	3	0.292	0.015	3.199	69.566
Kmodes	3	0.194	-0.075	4.117	7.138
Kproto	4	0.313	-0.127	5.128	40.719
Kmeans	4	0.241	-0.043	1.900	50.570
Kmodes	4	0.180	-0.106	22.614	2.919
Rock	5	0.161	-0.032	1.516	53.043
Rock	3	0.117	-0.078	1.325	47.929

The final comparison is a Tukey test comparing the difference in means between each cluster for each of the four resource use parameters. For Kproto4, this means are provided in Table 19. The results of the Tukey tests are presented in Table 20 for the top four clustering models. As shown in the Tukey results, in most cases there is not a statistically significant difference between the means of the various clusters. Kproto4 has statistically significant differences between groups for 50% of the *energyint* means and for 1 of the *wasteint* means.

Table 19: Means for Resource Intensity for each Cluster

Kproto4	Qty in each cluster	Means			
		waterint	energyint	wastewaterint	wasteint
0	198	27.17	655.38	20.98	0.13
1	30	4.95	218.55	3.79	0.012
2	90	18.17	708.2	14.41	2.23
3	15	34.05	2073.31	24.49	2.46

Table 20: Tukey Results for Selected Clusterings

Kmeans2 Tukey Test

```
tukey test for kmeans2 waterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower   upper  reject
----------------------------------------------
  0      1    -18.6416 -37.8428 0.5597 False
----------------------------------------------
tukey test for kmeans2 energyint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
=================================================
group1 group2  meandiff  lower    upper    reject
-------------------------------------------------
  0      1    -451.4508 -872.287 -30.6145  True
-------------------------------------------------
tukey test for kmeans2 wastewaterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower   upper  reject
----------------------------------------------
  0      1    -15.1739 -31.4185 1.0707 False
----------------------------------------------
tukey test for kmeans2 wasteint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
=============================================
group1 group2 meandiff  lower  upper  reject
---------------------------------------------
  0      1     -0.672  -2.1843 0.8404 False
---------------------------------------------
```

Kproto3 Tukey Test

```
tukey test for kproto3 waterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower    upper  reject
----------------------------------------------
  0      1     16.95   -6.5676  40.4677 False
  0      2     9.4802  -16.2219 35.1822 False
  1      2    -7.4699  -25.9596 11.0199 False
----------------------------------------------
tukey test for kproto3 energyint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
=================================================
group1 group2 meandiff   lower     upper   reject
-------------------------------------------------
  0      1    373.6303 -141.8356  889.0961 False
  0      2    460.1429 -103.2005 1023.4862 False
  1      2    86.5126  -318.7497  491.7749 False
-------------------------------------------------
tukey test for kproto3 wastewaterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower    upper  reject
----------------------------------------------
  0      1    15.2239  -4.6575  35.1052 False
  0      2     9.7275  -12.0005 31.4554 False
  1      2    -5.4964  -21.1272 10.1344 False
----------------------------------------------
tukey test for kproto3 wasteint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
============================================
group1 group2 meandiff  lower  upper  reject
--------------------------------------------
  0      1     0.1139  -1.6954 1.9233 False
  0      2     2.4263   0.4489 4.4037  True
  1      2     2.3124   0.8899 3.7349  True
--------------------------------------------
```

Kproto4 Tukey Test – Part A

```
tukey test for kproto4 waterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower    upper  reject
----------------------------------------------
  0      1    -22.219  -53.9313  9.4932 False
  0      2    -8.9927  -29.5703 11.5849 False
  0      3     6.8895  -36.458    50.237 False
  1      2    13.2263  -20.8978 47.3504 False
  1      3    29.1085  -22.0776 80.2947 False
  2      3    15.8822  -29.2597 61.0241 False
----------------------------------------------
tukey test for kproto4 energyint
 Multiple Comparison of Means - Tukey HSD,FWER=0.05
===================================================
group1 group2  meandiff    lower      upper   reject
---------------------------------------------------
  0      1    -436.8303 -1117.2321  243.5715 False
  0      2     52.8158   -388.6871  494.3188 False
  0      3    1417.9348   487.8922 2347.9774  True
  1      2     489.6461  -242.5037  1221.796 False
  1      3    1854.7651   756.5402 2952.9899  True
  2      3    1365.1189   396.5757 2333.6622  True
```

Kproto4o Tukey Test – Part 5

```
tukey test for kproto4 wastewaterint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
==============================================
group1 group2 meandiff  lower    upper  reject
----------------------------------------------
  0      1    -17.193  -44.0498  9.6638 False
  0      2    -6.5745  -24.0015 10.8524 False
  0      3     3.5137  -33.1968 40.2243 False
  1      2    10.6185  -18.2809 39.5179 False
  1      3    20.7068  -22.6423 64.0558 False
  2      3    10.0883  -28.142  48.3186 False
----------------------------------------------
tukey test for kproto4 wasteint
Multiple Comparison of Means - Tukey HSD,FWER=0.05
============================================
group1 group2 meandiff  lower  upper  reject
--------------------------------------------
  0      1    -0.1098  -2.5606 2.3411 False
  0      2     2.1091   0.5188 3.6994  True
  0      3     2.3297  -1.0203 5.6798 False
  1      2     2.2189  -0.4183 4.8561 False
  1      3     2.4395  -1.5163 6.3954 False
  2      3     0.2206  -3.2681 3.7093 False
--------------------------------------------
```

Ultimately, the conclusion from this analysis shows that there are statistically significant ways to group the factories to provide a more useful basis for comparing the factories on environmental performance.

5.5 Development of Benchmarking Tool

The results of the clustering analysis were used to develop an internal benchmarking tool. Each factory is classified into a specific cluster and then the z-score of the environment performance is calculated for each factory for each environmental parameter. The z-score is simply the individual consumption (energy, water, wastewater, waste) minus the cluster average divided by the cluster standard deviation.

With the standardized performance of each factory, the results can be visualized to enable use in decision making. The goal of the benchmarking tool is to provide a streamlined approach to factoring environmental performance into business decisions. The tool was developed to serve three major stakeholder groups:

- Merchandisers: Merchandisers make the purchasing decisions for each order. In most cases merchandisers have strong relationships with a subset of the Li & Fung factory network based on years of experience. Purchasing decisions are currently made by each individual merchandiser based on their preferred factories and price considerations. To allow Li & Fung to begin to optimize the network for environmental performance, the benchmarking tool will provide an environmental performance comparison between selected factories. This will allow merchandisers to begin to make decisions based on cost and environmental performance.

- Account Managers: Account Managers (AM) are a relatively new function within Li & Fung. The AM role fills the gap that currently exists between client management and merchandisers. The AM interfaces directly with the client. Li & Fung plays a role in helping clients develop their long-term strategy for factory engagements. This strategy, for many clients, now includes meeting specific long-term strategies for environmental performance. The benchmarking tool will provide clients with an understanding of how their network of factories compares to competitors' factory networks or a how their network of factories is performing overall in order to develop a long-term strategy for shifting their factory network.

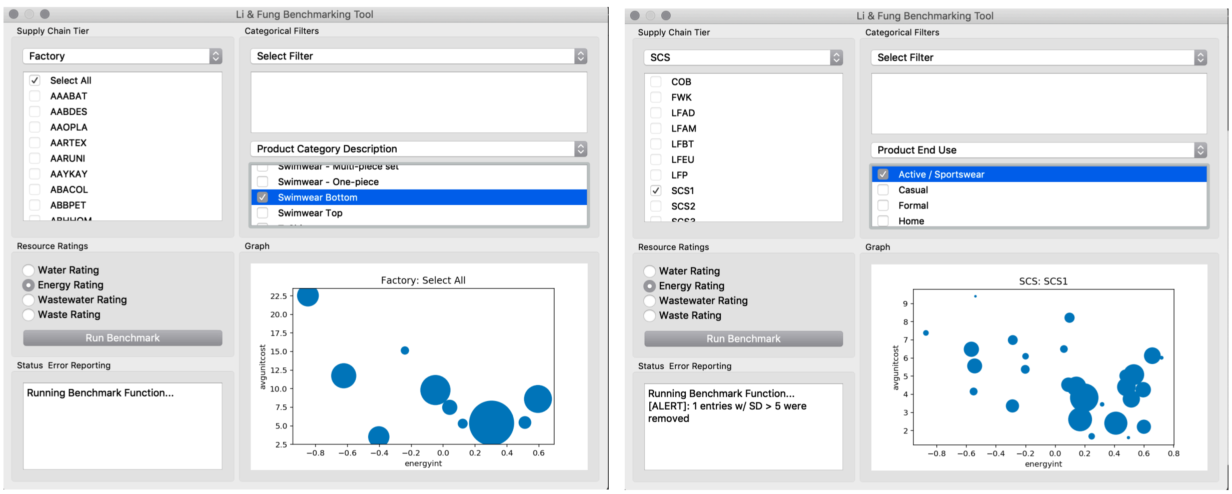
- Vendors: A key new role that Li & Fung has begun to play in the industry is to help develop a stronger supplier network. This has involved providing various services to factories including a WorkerApp (to help improve working conditions for employees), sponsoring the HERproject (to empower women in apparel supply chains) and the End of Line QC app. The benchmarking tool will be used to provide feedback to vendors (groups of factories) and individual factories on their environmental performance relative to their peer groups.

Given the needs of the above three major stakeholder groups, the benchmarking tool was developed to allow for filtering and visualization of the dataset. There are a few major filtering categories:

- Supply Chain Tier (Client, Factory, and Supply Chain Solution division within Li & Fung)
- Resource Rating (Water, Energy, Wastewater, Waste)
- Categorical Filters (Country, Product Category, Product End Use, etc.)

A first iteration of the tool was developed to demonstrate the potential use case of the tool within Li & Fung's current business processes. Figure 23 below shows the final tool with examples for each stakeholder group. The visualization shows the performance of a subset of factories. On the x-axis is the standardized resource score for the factory and on the y-axis is the standardized average price point of a product from the factory. The size of the bubble is proportional to the amount of product procured through the specific factory in the previous year.

Figure 23: Benchmarking Tool for Various Use Cases



Chapter 6: Analysis of Incorporating Real-Time Data

6.1 Testing Existing Hypotheses for Real-Time Data

As discussed in Section 4.1.5, a feature was created to provide an ordinal measurement of the data collection method and frequency. This section will outline and test existing hypothesis regarding the use of real-time data in the industry. The goal is to develop a working understanding of the current state which will then be used to develop a sensitivity analysis with real-time data for the model developed in Chapter 5.

It should be further noted that the created feature, *resourcedataquality*, is an aggregated average value, between 0 and 10, for each factory, where a score of 10 means that all of the resource data is collected with real-time sensors, a score of 1 means that all of the resource data is collected on an annual basis from estimates, and a score of 0 means that no resource data is collected.

6.1.1 Hypothesis 1: Real time data improves data quality

Testing Methodology: An analysis of the difference between the data quality scores of facilities that meet a minimum threshold for data collection/frequency and facilities that do not meet this threshold.

To test this hypothesis, the original dataset was divided by year collected (2018 or 2017) and verified or unverified. These four datasets were then further divided based on a minimum resource data measurement and frequency score (between 0 and 10). A threshold of 8 was used as this represents factories that have a majority of sensor data collected on at least a weekly basis.

Table 21 presents the scores for a threshold of 8:

Table 21: Data Quality Scores for Sensor and non-Sensor Data Sources

Database	Accuracy	Consistency	Completeness
2017 Unverified (above 8)	0.973	0.824	0.146
2017 Unverified (below 8)	0.982	0.852	0.140
2018 Unverified (above 8)	0.977	0.815	0.143
2018 Unverified (below 8)	0.985	0.883	0.139

Based on Table 21, on all measurements there is a negligible difference in quality scores. Furthermore, the factories that have lower *resourcedataquality* scores on average have a higher overall data quality score on accuracy and consistency. This suggests, that based on the available data from Higg, real-time data collected from sensors does not necessarily increase overall data quality scores.

It should be noted that there are some nuances to the data quality score that make this comparison less than appropriate for drawing actionable conclusions. For example, the *resourcedataquality* variable only applies to resource data whereas the data quality score is based on the overall factory entry in Higg.

To gain a better understanding of a direct comparison between data quality and the *resourcedataquality* feature another analysis was performed where only the combined resource features are included in the data quality assessment. Table 22 below provides the results of this analysis:

Table 22: Data Quality Scores (Resource Consumption Only)

Database	Accuracy	Consistency	Completeness
2017 Unverified (above 8)	0.980	0.961	0.839
2017 Unverified (below 8)	0.982	0.957	0.812
2018 Unverified (above 8)	0.981	0.926	0.911
2018 Unverified (below 8)	0.987	0.959	0.863

Based on Table 22 there is again negligible difference, on an aggregated basis, between the factories that use real-time sensors and those that do not. The only conclusive observation is that real-time sensors lead to more complete data. One potential explanation for this might be that larger factories, potentially outside of the three standard deviations for resource use currently guiding the accuracy measurement, might be more likely to adapt technology but are penalized because they exceed the limits set by consistency and accuracy. Additional outlier analysis would need to be performed to verify this.

Conclusion: Based on the currently available industry data there is no evidence that sensors and real-time data collection improves the data quality of the generated data. This conclusion further illustrates the need for procedures and standardization of data collection practices for real-time IoT networks.

6.1.2 Hypothesis 2: Real-time data improves management of resource use

Testing Methodology: the following two questions will be answered to test this methodology:

1. Is there a correlation between data collection methodologies and year-on-year reduction in resource use?

   To answer this question, the 2017 and 2018 datasets were filtered to remove outliers (following the same methodology employed in Section 3.5). Once filtered, resource intensities were calculated, and the two datasets were merged to only include common factory *groupid* features for which resource use was recorded for both years. The difference between the 2017 and 2018 intensities were calculated and the average of the *resourcedataquality* feature was also calculated. The final dataset included a relatively small number of entries (n = 72).

   Pairwise scatter plots (Figure 24) were then made for the five features and a correlation table (Table 23) was generated to understand if any relationship exists between the use of real-time sensor data collection methods and reduced year-on-year resource use.

Figure 24: Pairwise Scatter Plot for Resource Intensities and resourcedataquality Feature

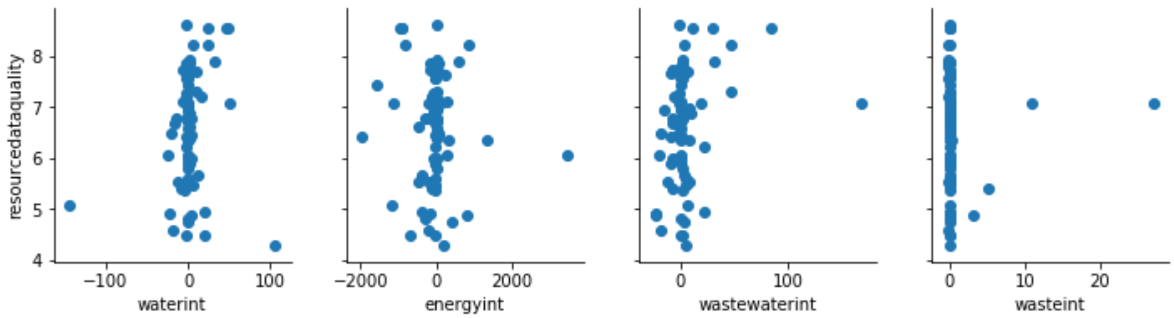


Table 23: Correlation Values for Resource Intensities and resourcedataquality Feature

	waterint	energyint	wastewaterint	wasteint
resourcedataquality	0.17219963	-0.0898337	0.26408167	0.02244894

As shown in the scatter plots and correlation table, the correlation between *resourcedataquality* and year on year improvements in resource intensity are weak at best.

This analysis provides a few takeaways:

- There is still insufficient data to conclude that the use of improved data collection techniques results in better tracking, trending and year on year reduction of resource intensities.
- The evidence that does exist shows that a higher *resourcedataquality* is actually positively correlated (although weakly) with increased water, wastewater, and waste intensity. There is a weak negative correlation between energy intensity increase and *resourcedataquality*.
- There may be confounding variables not adequately represented in the data that contribute to the changes in resource use intensity.

2. Are Higg Level 2 scores, on average higher for factories that have more advanced data collection methodologies?

   Higg Level 2 scores relate to establishing a baseline and then recording year on year improvements against that baseline. In theory, establishing a baseline first requires accurate measurements of resource use and then a method to trend, analyze and implement solutions to reduce resource use year on year. The premise of this test is that those factories that have both established a baseline and improved on that baseline year on year should also be the same factories that have improved their measurement and trending programs to ensure that they are tracking and reducing their resource use year on year.

   To test this, the Level 2 scores for each factory from 2018 and 2017 were analyzed against the *resourcedataquality* scores to understand what, if any correlation, exists between the two variables. Figure 25 show the relationship between the two features for 2017 and 2018. The aggregated correlation data is show in Table 24.

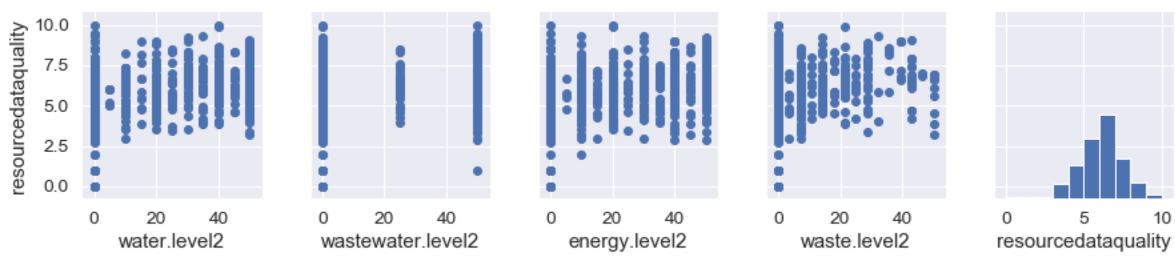
As shown in Table 24 and Figure 25 there appears to be weak correlations between these two features, with the greatest relationship between energy and water use. This is expected since the main focus of the industry is on energy and water. However, the relatively weak correlation again suggests that there is still insufficient evidence from industry data that real time data collection is strongly correlated to factory-level management practices around resource tracking, trending and reduction.

Table 24: Level 2 Scores and resourcedataquality Feature Correlations

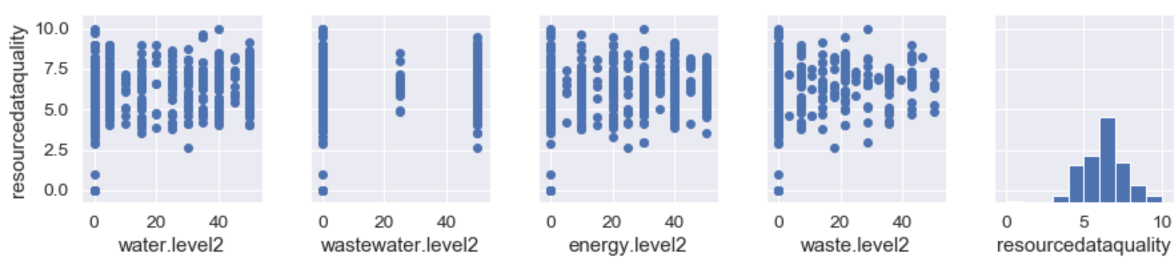
	resourcedataquality (2017)	resourcedataquality (2018)
water.level2	0.232	0.317
wastewater.level2	0.177	0.186
energy.level2	0.216	0.276
waste.level2	0.147	0.213

Figure 25: Pairwise Scatter Plot for Level 2 Scores and resourcedataquality Feature

2018 data:



2017 Data



6.1.3 Hypothesis 3: Real time data improves transparency and traceability [14].

This hypothesis cannot be adequately tested with the current data available to the researcher. This is a hypothesis that should be tested further in the future. Data from Li & Fung's End of Line QC app or WorkerApp might both provide data necessary to test this hypothesis.

6.1.4 Hypothesis 4: Sensors can optimize factory scheduling and reduce manufacturing disruptions [14, 15].

This hypothesis cannot be adequately tested with the current data available to the researcher. This is a hypothesis that should be tested further in the future. Data from Li & Fung's End of Line QC app or WorkerApp might both provide data necessary to test this hypothesis.

The above analysis included testing 2 of 4 possible hypothesis related to the use of real-time sensor data to improve data quality or reduce resource use in the industry. Based on the data currently available to the researcher there is little evidence that the current adoption of real-time sensors across the industry is correlated to a reduction in resource use or improvement in data quality. This analysis was based on an aggregation of data from the industry and therefore the conclusions should not be misconstrued to represent each factory in the network or industry. There are likely confounding variables that influence the conclusions made in this section. However, the lack of strong correlations reemphasizes the importance of developing a more robust set of requirements for real-time data collection.

This analysis is based on the current state of industry practices. With more robust standards around data collection and reporting, it is expected that the theoretical benefits of real-time data can be fully realized in the industry and will become an important tool to achievement of sustainability goals across the industry.

6.2 Reducing the Introduction of Variance

Section 3.4 introduced two aspects of data quality: intrinsic and contextual however only the former was formally discussed. We turn our attention to the latter, contextual, to understand how IIoT data could help improve the data quality on this dimension.

Significant variance and uncertainty are introduced into the data because the instructions for data collection allow for subjective interpretation. The example previously used to illustrate this subjectivity is the decision to include or exclude resource use for an on-site dormitory or canteen. A factory can choose to include or exclude resource use for these facilities, likely depending on the corporate structure for overseeing these additional assets and services. However, since Higg provides no information on the available facilities on-site, the relevant contextual data is not available for use in understanding the provided, objective data points.

Data generated from an IIoT system could reduce this variability and uncertainty if properly managed at the enterprise level. First, data collected from a connected IIoT network of factories must conform to a uniform set of requirements, specifically capturing both physical and data dimensions of the system.

6.2.1 Physical Requirements:

1. *Sensor at all input and output echelons:* a minimum requirement of sensors at all input and all output points of resource use should be imposed on any factory providing data to the enterprise level system.

   Certainly, factories will benefit from more granular level monitoring at different points throughout the manufacturing process, however for the enterprise level analysis, the data must allow for a resource balance calculation to be performed to validate the intrinsic quality of the data.

2. *A mapping of the physical sensor location to the facility layout:* the mapping of sensor locations provides an additional level of verification at the enterprise level that the streamed data captures the expected values. This physical mapping will also provide insight as to whether the facility includes dormitories, canteens, or other ancillary facilities that also consume resources.

6.2.2 Data Requirements:

1. *Data Reporting to Enterprise:* In reviewing current examples of IIoT systems already actively collecting real-time data across a subset of the factories in the network, it is clear that a standardize approach for reporting needs to be developed and utilized across then network. For example, Figure 26 shows examples of the Reports generated from various factory IIoT systems. Not only is the format inconsistent across the different sites, but the content itself varies.

Figure 26: Examples Real-Time Data Reports from IIoT Systems

Summary																												
Date	01-Feb-19	02-Feb-19	03-Feb-19	04-Feb-19	05-Feb-19	06-Feb-19	07-Feb-19	08-Feb-19	09-Feb-19	10-Feb-19	11-Feb-19	12-Feb-19	13-Feb-19	14-Feb-19	15-Feb-19	16-Feb-19	17-Feb-19	18-Feb-19	19-Feb-19	20-Feb-19	21-Feb-19	22-Feb-19	23-Feb-19	24-Feb-19	25-Feb-19	26-Feb-19	27-Feb-19	28-Feb-19
Source : Energy Meters	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH
LT_Incomer																												
DG1_250KVA																												
DG2_500KVA																												
DG3_125KVA																												
Total																												
Distribution : Energy Meters	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH	KWH
First_Floor_Panel_1																												
First_Floor_Panel_2																												
Ground_Floor																												
Utility_fd																												
Fusing																												
Air_Duck_Panel																												
Cutting_Panel																												
Borewell_Elec																												
Total																												
Difference																												

DG1_250KVA	DG LOG SHEET		
DG Set Details			
D.G SET CAPACITY	250 KVA	ENGINE MAKE	CUMMINS
D.G SETTYPE	NA	ENGINE CAPACITY	NA
D.G SET MODEL	NIC 855-G-FFC	ALTERNATOR MAKE	NA

Report Period:
Report Prepared By:
Report Prepared At:

Date	No. of Times	Time		Hour Meter Readings		Load Run Hours	No Load Run Hours	Avg Voltage			Avg Load in Amps			Frequency in HZ	kWh Readings		Total Units Generated	Diesel Readings		Diesel Consumed Ltrs	Rs.	kWh / Ltr	Rs.	Remarks
		START	END	START	END			R	Y	B	R	Y	B		START	END		START	END		Cost Per Unit	Units Per Ltr	Running Cost	

Daily	WATER REPORT	Mar-19

Report Period: 01-Mar-2019 To 31-Mar-2019
Report Prepared By: Report Prepared At: 01-04-2019 10:06:55

Sump				Staff_Hand_Wash				STP_flush_GND_FF				Gardening			
Date	Start Reading	End Reading	Litres	Date	Start Reading	End Reading	Litres	Date	Start Reading	End Reading	Litres	Date	Start Reading	End Reading	Litres

Key	Name:Suffix	Trend Definitions Used
Point_1:	AHU.SREC	15 minutes
Point_2:	CANT1.SREC	15 minutes
Point_3:	CANT2.SREC	15 minutes
Point_4:	CHL1.SREC	15 minutes
Point_5:	CHL2.SREC	15 minutes
Point_6:	D2MAIN.SREC	15 minutes
Point_7:	DBOUT.SREC	15 minutes
Point_8:	FCU.SREC	15 minutes
Point_9:	LDB.SREC	15 minutes
Point_10:	PDB.SREC	15 minutes
Point_11:	Pump.SREC	15 minutes
Point_12:	SOLAR.SREC	15 minutes
Point_13:	SOLAR2.WH.FWD	1 hour
Point_14:	STP.SREC	15 minutes
Point_15:	E.LIGHT.SREC	15 minutes
Point_16:	E.MACHINE.SREC	15 minutes
Time Interval:	1 Hours	
Date Range:	3/10/2019 00:00:00 - 5/8/2019 23:59:59	
Report Timings:	All Hours	

					Meter Destructed					Meter Destructed							
<>Date	Time	Point_1	Point_2	Point_3	Point_4	Point_5	Point_6	Point_7	Point_8	Point_9	Point_10	Point_11	Point_12	Point_13	Point_14	Point_15	Point_16

One of the main areas of variation in data recording is for power generated on site through the use of diesel or gas generators. Some sites monitor the real time energy generated by the diesel generator (DG), some monitor the real-time usage of diesel, and others monitor both. Comparing energy usage across these different sites requires an assumed generator efficiency coefficient which varies greatly with different diesel generator models.

Additionally, recording only the energy output of the diesel generator provides enough information to determine the energy efficiency of the factory but not enough to determine the specific GHG emissions since. Thus, in establishing an enterprise system for real-time data collection, both the format and required data features must be standardized across all network factories. The most common approach to this is to develop a data dictionary which standardizes the format and content of the collected data.

2. *In situ Quality Monitoring*: the in situ quality monitoring developed in Section 3.4.2 should be used to ensure that the real-time data collected and provided to the enterprise platform is of a high level of quality. Following the methodology developed, outlier management would be done at the source reducing the need for loss of data during later data analysis.

6.3 Improvements to the Benchmarking Model & Future Tools

With the proper implementation of the improved reporting standards outlined in Section 6.2 and later in Section 7.1 and the collection of real-time data, the benchmarking model could be greatly improved and additional tools could be developed to support the sustainability needs of the industry.

The benchmarking model is currently a static tool that can be updated on an annual basis with Higg Data. Given that most improvement or resource-use-reduction projects are 3-6-month efforts, annual updates to this tool are aligned with the realized improvement across the network.

However, there are a number of key shortcomings of the current tool that would be greatly improved with real-time data updates:

1. *Annual review of performance is not frequent enough to establish a continuous improvement culture across the industry.*
   While annual reporting of resource use is an important first step for improving transparency across the industry, it does not necessary lead to immediate improvements at a factory level. Developing a culture of continuous improvement across the industry is necessary to ensuring meaningful improvements from an environmental perspective. As has been shown time and time again across a variety of industries, moving from mediocrity to excellence is difficult but necessary to ensuring long term sustainable competitive advantage. Factories that focus on reducing waste across their value-stream will not only be more competitive from a sustainability perspective but will also be able to provide competitive pricing and better jobs. The use of real-time data in the benchmarking tool can provide immediate insights into which factories are not only performing well but are also improving.

2. *Real-time updates will allow Li & Fung to identify problem areas immediately and shift resources to support improvement*
   Building on the above point, real-time data in the benchmarking tool will allow Li & Fung to become nimbler in its sourcing strategies. Rather than shifting sourcing on an annual or semi-annual basis, Li & Fung will be able to understand how various factories are performing day-to-day or month-to-month. Much of the resource data can be used to extrapolate on factory performance to understand current capacity, potential labor challenges, or localized exogenous factors that might increase the risk of additional

sourcing from a particular factory. In the vein of continuous improvement, real-time benchmarking will allow Li & Fung to identify these risks and then take a proactive approach to helping the factory resolve an issue in order to ensure continued environmental performance and improvement.

3. *Trends can be identified to understand how exogenous factors (weather, geopolitical, macro-economic trends, etc.) affect the environmental performance of factories.*
   In addition to improvements of the benchmarking tool, the real time data collected can be trended over time to identify issues at the factory level on a real-time basis. For example, hourly, daily or weekly data could be collected on the enterprise platform and then trended over time for each factory. The data could then be used to identify potential disruptions at the factory level that might affect environmental performance. For example, spikes in energy usage at a factory could be indicators of excessive overtime or overcapacity of a specific factory. Sufficient baseline data needs to be collected as a training set for developing the algorithms.

4. *Real-time leaderboards could be used to build new incentive structures throughout the industry to drive further improvement.*
   As discussed in Section 5.5, a key function of the benchmarking tool is to provide a feedback loop to vendors on how their performance stacks up to their competitors. Improving the transparency of environmental performance across the industry can help raise the standards for all vendors and factories. With real-time data this functionality could be greatly improved by providing real-time leaderboards ranking performance of all factories in the network. Careful consideration would need to be placed on how this is communicated and utilized to avoid incentivizing dishonest behavior, but the net affect could be to provide the backwards transparency necessary to continue incentivizing factories to drive improvement across their operations.

Chapter 7: Recommendations & Conclusions

7.1 Recommendations for Improving Data Collection, Quality and Processing

This study has focused on a specific benchmarking use case for environmental data in order to explore the future state requirements for data collection, data quality, and data processing. Chapter 3: explored these three dimensions for a specific industry data set. The purpose of this section is to apply the learnings from Chapter 3: to recommendations on future data collection, data quality, and data processing within the industry.

7.1.1 Li & Fung

7.1.1.1 Timeliness Framework

A main focus for Li & Fung is understanding the Timeliness dimension of data quality. A useful framework for guiding the development of a plan for digitalization, as it fits into the larger strategic goals of the organization, is to understand the real requirements for data timeliness. Chapter 6 outlined potential considerations for incorporating real-time data into the benchmarking tool. These discussions however were meant to be exemplative rather than prescriptive. The point of exploring the model sensitivity to real time data was to outline the primary considerations Li & Fung should use in developing a roadmap to digitalization.

Both components of Timeliness require a deep business understanding of the desired use cases for the data. Volatility, or the shelf-life of the data, is dependent on the time horizon for use in business decisions. For example, if Li & Fung decides that sustainability as it is broadly defined will become a key consideration on all sourcing activities then it may become necessary that tools such as the factory benchmark developed in Chapter 5 are updated on a monthly or weekly basis to provide the latest information to merchandisers. In this scenario, the shelf-life of the data would be only a few days. Similarly, if Li & Fung wanted to use real-time energy data as a proxy for capacity within the network, hourly data might be required to inform each purchasing decision on a daily basis. Alternatively, if the end use was simply to inform 6-month or 1-year sourcing strategies that needed to be aligned with high level sustainability KPIs then the use of the annual Higg data might be sufficient to meet current needs.

The Timeliness framework is an important starting point to minimize the tendency to throw technological solutions at problems that could be solved through simpler and less expensive means. What Li & Fung needs to focus on is developing a clear corporate strategy around the "supply chain of the future" and then aligning the development of the necessary resources and capabilities, in some cases which may include technology, to execute on this strategy.

Once the strategy is defined and the resources and capabilities are further outlined then the right solution can be selected based on the derived requirements for data timeliness.

7.1.1.2 Improved Verification Methodology

Li & Fung is in a unique industry position to continue to develop and refine the methodology and baselines for verification of factory data. As discussed in Section 7.2, when considering IIoT solutions the requirements for calibration and hardware accuracy specifications should be clearly defined and enforced throughout the network. However, specifying requirements is only the first order method of ensuring data quality.

As discussed in Section 3.4, data accuracy can be verified through various other processes. Li & Fung should continue to find ways to increase data accuracy through verifications processes. Verified data will become more important to the industry as the shift to environmental reporting through protocols such as GHG and SBTi become the industry standard.

GHG reporting is in a nascent phase with most companies using academic research and industry estimates to establish baselines and targets. While this is sufficient to address the improvements that could be classified as "low hanging fruit" in colloquial terms, as the required improvements advance to more complex and specific areas increased data quality and more specific data collection techniques will be required. Especially in reporting Scope 3 emissions, it will become more important that the data used to establish new baselines, new targets, and most importantly to drive insights into where to invest additional resources is verifiable and of a high quality.

7.1.2 Higg FEM

This research provides a useful framework for analysis of Higg data and lessons learned for future development of the platform.

1. *Reduced variability in responses*
   Free responses in data collection provide additional level of granularity for specific factories. However free responses are also not easily analyzed on an aggregate level. The tradeoff between these two perspectives must be carefully considered. Given the limited English proficiency of many of the platform users, free responses questions often lack the specificity necessary to properly interpret the data. As such, Higg could consider reducing the number of free responses questions or providing more specific and less free-form options for response to questions that must remain free form. These could be automatically generated and selected by the specific user.

2. *Built in Verification processes*
   The data processing performed by the researcher was aligned with the expectations for any data analysis project. However, given the wide range of factories that are currently using this platform post-collection data verification is often inaccurate and results in removal of data points that appear to be incorrectly recorded.

   Higg could incorporate secondary and automated checks during data entry. For example, following the process in Section 3.4.1.3, for verifying data consistency, ranges could be established for the majority of numeric data points collected. If a value outside the range is provided then an error message could be given to the individual entering the data. This would help limit the amount of data that needs to be disregarded during data analysis due to errors with units, order of magnitude, or data entry that could easily be avoided.
   A second order verification process that could be implemented would be to establish logic between the questions on the platform. This logic could allow there verification to be further tailored to the specific type of facility. For example, since ranges for water use at a mill versus a cut & sew facility vary widely, the users responses in the *sipfacilityprocess* section of the data collection process could inform the guidance given in the Level 1 water use section.

3. *Increased granularity of data to allow for normalization*
   Given the complexity of the industry a degree of flexibility must included in the platform. However, much of the numeric data currently collected on the platform lacks the appropriate context necessary to allow for appropriate comparisons across the network. For example, the allowable units for *sipfacilityannualproductionvolqty* range from *kg* to *square yards* to *units*. These various units are not comparable without additional context regarding average weight of a unit or density of the material.

   Since the industry practice for comparing factories is on a resource-intensity basis, the lack of consistency between units results in a loss of data that limits the insights able to be derived from the data.

4. *Consistent missing value records*
   The platform filters questions based on applicability, previous answers, etc. This results in a significant number of incomplete data points throughout the final dataset. Rather than coding these points as incomplete the platform could perform filtering, based on the applicability logic, and code questions that are unanswered differently than questions that were not applicable to the specific user.

5. *Additional Data Formatting Improvements*
   The output data files provided by the platform could conform to a standards for data formatting and reporting. Coding "Yes/No" as binaries, converting all incomplete cells to a consistent NaN or Null, removing the ambiguity between 0 and Null, and labeling each feature as a specific data type (categorial, ordinal, numeric, etc.) would all help support efforts to analyze the data and reduce loss of data due to poor quality and reporting procedures.

   Government datasets, such as the US Department of Energy Building energy use dataset [36], can be used as a template to build a cleaner output CSV file from the platform along with better guidance for the users.

6. *Reduction of redundancy (processes, etc.)*
   Similar information is requested at different parts of the data collection process. For example, processes are specified more than one time on the platform increasing the likelihood of inconstant data entry by a single user. Auto-populate or an automated cross-check could be performed during data entry to ensure that the individual entering the data does not provide inconsistent responses to similar questions.

   Removing redundancy would also help reduce the effort required from the individuals entering the data.

7. *Inclusion of additional data points related to facility size*
   Since most of the current use cases for the data fall into one of two categories: benchmarking and reporting, it would be useful to include additional information to allow more accurate comparisons between facilities. The researcher used a clustering technique to find patterns between different factories that would improve the benchmarking criteria, however this effort could be further supported through collection of a few additional data points that tend to be better indicators of environmental performance. These would include,

for example, details on heavy equipment (to supplement the limited list in the Air Quality section), details on total square meters, details on on-site facilities such as dormitories or canteens. All of these examples have been shown in previous research to have a significant influence on resource use. Requesting this data in the general section would further improve the usefulness of the data set.

8. *Improved guidance and instructions*
   Much variability is introduced into the dataset because there appears to be a lack of consistency between reporting practices across factories. The example used throughout this paper is the inclusion or exclusion of resources consumed by ancillary facilities such as dormitories and canteens. Higg guidance can be improved to clarify what is in and out of scope for reporting quantitative data to the platform.

7.2 IIoT Requirements

7.2.1 Hardware Requirements

Hardware selection will necessarily vary across the network. To ensure quality data at the enterprise level it is important that the hardware selected 1) supports a standard communication protocol and 2) meets a minimum set of technical specifications.

7.2.1.1 Communication Protocol & Gateway

Communication protocol refers to a set of standardized rules for communication between hardware devices. The most commonly used protocol for sensors is Modbus. All sensors should ideally have a Modbus RTU or Modbus TCP/IP port for purposes of data collection and transmission.

Some digital sensors do not have a Modbus capable port but do have a pulse output. The pulse output is an electrical signal transmitted from the sensor that is proportional to a flow rate (e.g. water or electricity). The pulse output must be translated into a Modbus output via a "transformer". The "transformer" converts the electrical pulse signal into a Modbus compatible electrical signal.

Generally, a Modbus-enable sensor is connected to the platform or cloud through a gateway. Gateways have different communication protocols and often different platforms are only compatible with specific gateways. Developing a robust set of requirements for integration of sensors is necessary to reduce the high costs of one-off integration of thousands of different sensors.

7.2.1.2 Technical Specifications

Technical specifications should be imposed on all hardware reporting data to the enterprise platform. The specific requirements will vary by sensor type and should include details on accuracy, sampling rate, and display type. Table 25 provides an example technical specification for a Turbine Water Meter.

Table 25: Example Technical Specification for Turbine Water Meter

Indicator Specifications		Specifications	Mandatory Specification
1	Enclosure	Metal/ABS, Weather proof IP 65 min	YES
2	Parameters measured	Flow Rate and Totalised Flow Volume	YES
3	Input Supply	230 V AC or 24 V DC	
4	Display Range for Flow Rate	0 - 999.9 m^3/hr or 0 - 9999 Ltrs/hr	
5	Display Units for Flow Rate	Ltrs/hr or m^3/hr; user settable	
6	Display Range for Flow Volume	0 - 999999.9 m3 or 0 - 9999999 Ltrs	
7	Display Units for Flow Volume	Ltrs or m3; user settable	
8	Flow Rate Sampling	Minimum every 3s or better	YES
9	Accuracy	+/- 2% FSD or better	YES
10	Display Type	Alphanumeric LCD	YES
11	Linearity	+/- 2%	
12	Communication Output	Modbus RTU or Modbus TCP/IP	YES

Sensor Specifications		Specifications	Mandatory Specification
1	Type	Full Bore Turbine	YES
2	Fluid Name	Water	YES
3	Operating Pressure	Upto 5 Kg/cm^2	
4	Operating Temperature	60^o C max	
5	Impeller Material	Polypropylene / SS	

7.2.2 Enterprise Platform Requirements

A variety of frameworks exist for selecting an enterprise platform. Table 26 below outlines some key considerations for selecting the appropriate platform.

Table 26: Key Attributes and Considerations for Enterprise IIoT Platform

Attribute	Considerations
Integration	The approach of the platform provider for integration is key to ensuring the success of scaling a decentralized IIoT system across the apparel industry.
Connectivity	Understanding the communication protocol requirements of the platform is important to ensuring that a variety of sensors can be easily connected. Some platforms use open-sourced communication protocols while others rely on proprietary protocols which can significantly increase the cost of connecting additional sensors to the gateways or gateways to the platform.
Edge or Cloud	Some platforms can be hosted locally on a computer, on the cloud, or on the gateway itself. Edge computing capabilities enable much of the data analysis to be performed in real time at the sensor or gateway. Some platform providers have their own custom gateways tha [30] [30]t fully integrate this capability.
Customization	Some platforms are built for customization. For example, Tulip allows easy modification of applications for use on the shop floor.
Analytics	Some platforms come with built in data analytics tools including pre-trained machine learning packages and ability to integrate with other common ML/AI packages.

7.3 Business Implementation

Section 5.5 outlined three potential stakeholder groups that could benefit from a benchmarking tool. The benchmarking tool was developed as a prototype to prove out conceptually the potential use cases of data currently available within the Li & Fung network. A variety of future use cases could be developed to continue to harness this data to drive meaningful improvements across the industry and to monetize a variety of new services.

As an example, Li & Fung can begin to consider their sustainability functions as a service opportunity to stakeholders across the industry. Other firms are already beginning to play in this space with the development of Environmental Social and Governance (ESG) reporting platforms enabled by IIoT. A recent example of this is the San Diego startup, Measurabl, that has developed a platform for the real estate industry. The platform allows real estate owners to connect their utility accounts and building sensors directly to a platform for automated ESG report generation.

Given the large industry focus on reporting through protocols such as GHG and SBTi, providing one-stop-shop solutions for developing these reports would be a highly valuable service for Li & Fung to provide.

7.4 Conclusion

Understanding currently available data is an important starting point to developing the right framework and strategy for scaling an IIoT platform across the wider industry. This research developed an analysis of what data is currently available, the appropriate approach for assessing and controlling the quality of this and future data collected in real-time from an IIoT platform, and potential use cases for this data. Finally, this research provided recommendations on how real-time data could be integrated into similar data analytics tools to help accelerate the rate environmental impact reduction across the industry.

Chapter 8: References

[1] Li & Fung, "Supply Chain Sustainability," [Online]. Available: https://www.lifung.com/our-strategy/our-sustainability-strategy/supply-chain-sustainability/. [Accessed 19 3 2020].

[2] Li & Fung, "About Li & Fung Our History," [Online]. Available: https://www.lifung.com/about-lf/our-purpose/our-history/. [Accessed 15 12 2019].

[3] Fung Group & Energenz, "Outcomes Report: Supply Chain Energy Management Capacity Building Program v2.0," Li & Fung, Hong Kong, 2014.

[4] E. D. Koepsell, "Environmental Sustainability Assessment Tool for Factories," Massachusetts Institute of Technology, Cambridge, 2016.

[5] A. Hashim, "Leveraging Data for increased Sustainability of Products & Factories," Massachusetts Institute of Technology, Cambridge, 2017.

[6] D. Hillstrom, "Multi-Tier Supply Chain Assessment of Garment Environmental Sustainability," Massachusetts Institute of Technology, Cambridge, 2018.

[7] Z. J. Stauber, "Technology Adoption in Consumer Goods Manufacturing in Asian, Low-cost Sourcing Countries," Massachusetts Institute of Technology, Cambridge, 2017.

[8] B. D. Jennings, "Leveraging Smart System Design to Collect and Analyze Factory Production Data," Massachusetts Institute of Technology, Cambridge, 2018.

[9] K. Hsu, "Using Factory-Level Digital Tools to Improve Quality and Productivity in Garment Factories," Massachusetts Institute of Technology, Cambridge, 2019.

[10] J. P. Talampas, "Electricity Sensors for Resource Efficiency and Supply Chain Visibility in Factories," Massachusetts Institute of Technology, Cambridge, 2019.

[11] M. Lehmann, S. Tarneberg, T. Tochtermann, C. Chalmer, J. Eder-Hansen, D. J. F. Seara, S. Boger, C. Hase, V. V. Berlepsch and S. Deichmann, "Pulse of the Fashion Industry," Global Fashion Agenda and The Boston Consulting Group, Boston, 2018.

[12] A. T. Chan, "A Predictive Approach for Identifying High Performance Factories," Massachusetts Institute of Technology, 2015.

[13] G. Distelhorst and R. M. Locke, "Does Compliance Pay? Social Standards and Firm-level Trade," *American Journal of Political Science,* vol. 62, no. 3, pp. 695-711, 2018.

[14] A. Berg, S. Hedrich, T. Lange, K.-H. Magnus and B. Mathews, "The apparel sourcing caravan's next stop: Digitization," McKinsey Apparel, Fashion & Luxury Group, 2017.

[15] J. Singh Srai, P. Christodoulou and E. Settanni, "Next Generation Supply Chains: Making the Right Decisions About Digitization," University of Cambridge Institute for Manufacturing, Cambridge, 2017.

[16] Higg Co, "Higg FEM Scoring System Guidance," [Online]. Available: https://apparelcoalition.zendesk.com/hc/en-us/articles/115002450511-Higg-FEM-Scoring-System-Guidance-. [Accessed 15 12 2019].

[17] E. Kocherovsky, "Higg Facility Module Adoption & Performance Trends," Anthesis, Presentation, 2019.

[18] Kering, "Environmental Profit & Loss (EP&L) 2018 Group Results," 2018. [Online]. Available: https://keringcorporate.dam.kering.com/m/4cf9e7e0187fd328/original/https-keringcorporate-dam-kering-com-media-.pdf. [Accessed 15 8 2019].

[19] The Reformation, "RefScale Methodology," [Online]. Available: https://thereformation-weblinc.netdna-ssl.com/media/W1siZiIsIjIwMTgvMDkvMTkvMTkvMDUvMDUvYjNlYWQyYWUtNTU3OS00YzVkLTg0ZTEtOTM1NjJjMjBhYzAzL1VwZGF0ZWQgUmVmU2NhbGUgTWV0aG9kb2xvZ3kucGRmIl1d/Updated%20RefScale%20Methodology.pdf?sha=832bb6ee07fb26f8. [Accessed 15 08 2019].

[20] The Reformation, "Sustainability Report Q1 - January-March 2019," The Reformation, 2019. [Online]. Available: https://www.thereformation.com/pages/sustainability-report-q1. [Accessed 15 08 2019].

[21] Jeanologia, "EIM Environmental Impact Software," Jeanologia, 2019. [Online]. Available: https://www.jeanologia.com/portfolio/eim-environmental-impact-software/. [Accessed 15 08 2019].

[22] Jeanologia, "Environmental Impact Measuring Software," Jeanologia, 2019. [Online]. Available: https://eim2.jeanologia.com/v2/login. [Accessed 15 08 2019].

[23] X. Gao and A. Malkawi, "A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm," *Energy and Buildings*, vol. 84, no. December 2014, pp. 607-616, 2014.

[24] B. T. Hazen, C. A. Boone, J. D. Ezell and L. A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," *International Journal of Production Economics*, vol. 154, no. August 2014, pp. 72-80 , 2014.

[25] D. Ballou, R. Wang, H. Pazer and T. Giri Kumar, "Modeling Information Manufacturing Systems to Determine Information Product Quality," *Management Science*, vol. 44, no. 4, p. 462, 1998.

[26] T. Sagris and J. Abbott, "An Analysis of Industrial Water Use in Bangladesh with a Focus on the Textile and Leather Industries," 2030 Water Resources Group, Washington, DC, 2015.

[27] A. Prakash, M. Sharma and J. Chourey, "Water in India: Situation and Prospects," UNICEF, FAO, SaciWATERs, New Delhi, 2013.

[28] World Health Organization (WHO), "Water Sanitation Hygiene," WHO, [Online]. Available: https://www.who.int/water_sanitation_health/emergencies/qa/emergencies_qa5/en/. [Accessed 15 08 2019].

[29] A. Cay, "Energy Consumption and Energy Saving Potential in Clothing Industry," *Energy*, vol. 159, no. September 2019, pp. 74-85, 2018.

[30] M. I. Islam, "Energy Consumption Determinants for Apparel Sewing Operations: An Approach to Environmental Sustainability," Kansas State University, Manhattan, 2016.

[31] I. Guyon and A. Elisseef, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, no. March 2003, pp. 1157-1182, 2003.

[32] Z. Huang, "Clustering large data sets with mixed numeric and categorical values," in *The First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Singapore, 1997.

[33] N. d. Vos, "Github Kmodes," 2016. [Online]. Available: https://github.com/nicodv/kmodes. [Accessed 15 08 2019].

[34] S. Guha, R. Rastogi and K. Shim, "ROCK: a robust clustering algorithm for categorical attributes," in *Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337)*, Sydney, 1999.

[35] A. Novikov, "PyClustering: Data Mining Library," *Journal of Open Source Software*, vol. 4, no. 36, pp. 1230, https://doi.org/10.21105/joss.01230, 2019.

[36] EIA, "Commerical Buildings Energy Consumption Survey," EIA, 2012. [Online]. Available: https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata. [Accessed 15 08 2019].