Improving Online Demand Forecast using Novel Features in Website Data: A Case Study at Zara

by

Lila Fridley

S.B., Chemical Engineering
Massachusetts Institute of Technology, 2013

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

Master of Business Administration
and
Master of Science in Civil Engineering and

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

June 2018

© 2018 Lila Fridley. All Rights Reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author .... Signature redacted
MIT Sloan School of Management
Department of Civil Engineering
May 11, 2018

Certified by .... Signature redacted ....
Georgia Perakis
William F. Pounds Professor of Management Science, MIT Sloan School of Management
Thesis Advisor

Certified by .... Signature redacted ....
Patrick Jaillet
Dugald C. Jackson Professor of Electrical Engineering and Computer Science, MIT
Thesis Supervisor

Accepted by............ Signature redacted ....
Jesse Kroll
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Accepted by............ Signature redacted ....
Maura Herson
Director of MBA Program MIT Sloan School of Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 07 2018
LIBRARIES
ARCHIVES

The author wishes to acknowledge the Leaders for Global Operations Program for its support of this work.

Improving Online Demand Forecast using Novel Features in Website Data: A Case Study at Zara

by

Lila Fridley

Submitted to MIT Sloan School of Management on May 11, 2018 in Partial Fulfillment of the requirements for the Degree of Master of Business Administration and Master of Science in Civil Engineering

Abstract

The challenge of improving retail inventory customer service level while reducing costs is common across many retailers. This problem is typically addressed through efficient supply chain operations. This thesis discusses the development of new methodologies to predict e-commerce consumer demand for seasonal, short life-cycle articles. The new methodology incorporates novel data to predict demand of existing products through a bottom-up point forecast at the color and location level. It addresses the widely observed challenge of forecasting censored demand during a stock out.

Zara introduces thousands of new items each season across over 2100 stores in 93 markets worldwide [1]. The Zara Distribution team is responsible for allocating inventory to each physical and e-commerce store. In line with Zara's quick to retail strategy, Distribution is flexible and responsive in forecasting store demand, with new styles arriving in stores twice per week [1]. The company is interested in improving the demand forecast by leveraging the novel e-commerce data that has become available since the launch of Zara.com in 2010 [2].

The results of this thesis demonstrate that the addition of new data to a linear regression model reduces prediction error by an average of 16% for e-commerce articles experiencing censored demand during a stock out, in comparison to traditional methods. Expanding the scope to all e-commerce articles, this thesis demonstrates that incorporating easily accessible web data yields an additional 2% error reduction on average for all articles on a color and location basis. Traditional methods to improve demand prediction have not before leveraged the expansive availability of e-commerce data, and this research presents a novel solution to the fashion forecasting challenge. This thesis project may additionally be used as a case-study for companies using subscriptions or an analogous tracking tool, as well as novel data features, in a user-friendly and implementable demand forecast model.

Thesis Supervisor: Georgia Perakis
William F. Pounds Professor
Sloan School of Management Science, MIT

Thesis Supervisor: Patrick Jaillet
Dugald C. Jackson Professor
Department of Electrical Engineering and Computer Science, MIT

This page intentionally left blank.

Acknowledgements

During the second half of 2016, I was honored to have the opportunity to work at Zara (Inditex S.A.) HQ in A Coruña, Spain. This experience was only possible due to the longstanding collaboration between Inditex and MIT, which enabled me dive deeply into this research. For that I am grateful to MIT, the Leaders for Global Operations Program, and the faculty and staff who continue to foster such collaboration with companies like Zara.

I would like to thank my supervisors and project mentors at Zara: Ane Insausti Altuna, Jose Luis Goñi, and Manuel Fontenla Cadavid. Everyone I had the pleasure of working with provided me with countless hours of patience, detailed explanations of inventory management practices, and algorithmic and data analysis support. I appreciate the teamwork, feedback, and constant drive for improvement in our models and results. To everyone else whom I worked with or who provided me with travel recommendations and local restaurant suggestions: thank you for your kindness and support during my six months at Zara. I feel like I achieved near-local expertise of Galicia.

I would like to thank my MIT research advisors: Professor Georgia Perakis and Professor Patrick Jaillet. My collaboration with each enabled me to more deeply understand demand forecasting, retail operations, data analyses and statistical methods. Thank you for your patience, long hours of reading emails, conference calls, and helpful guidance throughout this process.

Finally, I would like to thank my family, friends, and LGO community who offered me sage advice and thoughtful words of wisdom. I am grateful for having so many people in my life who continue to support me on my ever-changing journey.

This page intentionally left blank.

Table of Contents

Summary of Figures

This page intentionally left blank.

Summary of Tables

This page intentionally left blank.

Chapter 1

Introduction

The global fashion industry was estimated to be worth USD$2.4 trillion in 2016. Despite a challenging year in fashion sales in 2016, the industry has continued to grow annually at 5.5% in the past decade and continues to be one of the largest industries in the world. Overall growth has been fueled by three dominating trends: affordable luxury wear, value fashion, and athletic/leisurewear [3]. Within the value fashion segment, fast fashion showed strong growth over the past decade. Fast Fashion is defined in literature by the complementary implementation of 1) quick response, and 2) enhanced design [4]. This segment includes brands like H&M, Mango, and Uniqlo.

This thesis researches the operations at Zara, a leading fashion retailer. Zara is the flagship brand of Inditex S.A., representing 65% of group sales, and is one of the largest fashion companies in the world. Inditex operates in nearly 100 distinct markets with nearly 7,500 stores worldwide. In 2016, Inditex earned USD$28 billion in revenue, and continued to grow sales throughout 2017 [5].

Many brands in the Inditex group, including Zara, focus on providing customers with the latest fashion trends at affordable prices. To achieve this goal, they implement a lean operations strategy through an agile supply chain and inventory distribution process. Furthermore, Zara is highly aware of changing customer demand preferences. They respond quickly to trends with new designs, fast delivery, and specific inventory distribution to optimize delivering the right articles to the right location at the right time [1]. This operational efficiency is critical for fulfilling their strategic objective of meeting changing customer preferences, a challenge intrinsic to the fashion industry.

1.1 Research Motivation

Zara launched its online store, Zara.com, approximately 10 years ago [2]. In this time, sales through Zara.com have grown dramatically, and Zara has adapted rapidly to meet demand in this new sales channel. For example, Zara built distribution warehouses around the world to fulfill online orders within a few days. Furthermore, customer engagement has continued to grow with online and social media presence, and Zara has explored methods to improve operations through these new tools.

This thesis has two key objectives in improving demand forecast accuracy. First, this research aims to improve the point-forecast on an article and location basis after a period of censored demand during a stock out. Second, it aims to explore new methods to predict e-commerce demand using available web data.

Consumer retailers are eager to improve demand prediction at an article-and-location level. Improving prediction accuracy enables retailers to achieve the same customer service level with lower inventory and distribution costs. This is a widespread challenge across the fashion industry, which experiences short life-cycle articles and rapid evolution of consumer preferences.

To measure the success of new methods to achieve the goals described above, this work compares prediction error on an article-and-location basis. Linear models were trained on previous datasets, and prediction accuracy was tested on out-of-sample training data. This method is typical for distribution operations improvement projects. The dataset explored offers two novel categories of information that were explored with the goal to improve e-commerce demand prediction, corresponding to the two research objectives. To address the first goal of improving the demand prediction during a stock-out, this research presents the first analysis of using opt-in customer "subscriptions" as a feature in demand forecasting. The hypothesis was that this type of information would effectively un-censor demand during stock-outs. To address the second goal of incorporating widely available e-commerce data into the demand forecast, this research aggregates several data sources and explores the relationship between the observed features and future demand. The hypothesis was that website data like article position and category would provide new information to improve the demand forecast.

New models are presented in this thesis which explore both categories of novel data, with the intention of incorporating data to improve the e-commerce demand prediction. These models range from simple linear regression techniques (favorable for ease of implementation and usability), to logarithmic transformation and logistic regression. All models are analyzed in depth for prediction accuracy and robustness using out of sample test data.

Figure 1 illustrates several key features investigated in this study. First, Figure 1-Left shows an example of an article available for sale on Zara.com. In this image, *Subscriptions* are available for customers to "click" the envelop and enter a contact address to receive notification of replenishment. Prior to this research, Zara had not yet identified a systematic way to incorporate this Subscription information into the demand prediction model. Second, Figure 1-Right is an example of the article display structure on the Zara website. The blue arrow points to indicate article Category. Website article category is investigated in this thesis as a possible feature for improving demand prediction.

FIGURE 1: FEATURES STUDIED: Subscriptions & Structure on Zara.com website [6]

To achieve these goals and objectives, the solution was designed to meet the following four tightly knit Business and Technical Needs: Implementable, Communicable, Scalable, and Robust. These four characteristics are illustrated in Table 1, below. Each characteristic was identified as a critical factor for the research project success, and new models were evaluated on their ability to fulfill these needs.

TABLE 1: Alignment of research goals in the context of business & technical needs

	Business Need	**Technical Need**

Implementable	New model should be in a format that is easy to implement	New model integrates existing e-commerce data with existing distribution data
Communicable	New model should be simple to explain to non-technical teams	New model follows simple and logic-based assumptions
Scalable	New model should be usable on thousands of articles	New model generates prediction with short computation time & stable solution output
Robust	New model should improve prediction error for many articles with small demand, not just small percent of articles with large demand	New model error analysis should demonstrate that in the 90^{th} percentile case, the new prediction is better than the baseline

1.2 Research Approach

The proposed investigation was divided into three phases, as shown in Figure 3 below. These phases included: Problem Scoping, Research & Analysis, and Implementation.

The modelling process for this research was adapted from Tennent and Friend's Guide to Business Modelling. This guide provides the framework to designing stages of research and model development within a business context. Using this framework, the first phase is to define the business question and the scope of inputs and outputs required to address the challenge. The second phase is the core of the quantitative analysis: to collect data, design the input features, train the model, and test the model with out-of-sample data. The third phase is to share the findings with the stakeholders and implement the new model [7].

In the context of this thesis, each phase was critical to the overall success of the project. The first phase was essential to achieve three preliminary milestones. First, the project scope was defined to improve the demand forecast prediction for replenishment and e-commerce sales only. Second, prioritization was established between the two project goals: the top priority was given to incorporating Subscriptions, with a secondary goal of investigating and adding other e-commerce features. Third, this phase achieved initial alignment across business units; it was critical to establish unified goals given that the project ultimately affected both Distribution and Zara.com teams.

The second phase was iterative, as data collection and analysis uncovered new findings and new approaches were tested throughout the project. Feature engineering and model development aimed to address the four key Business and Technical Needs of the project: Implementable, Communicable, Scalable, and Robust. The model development split the data into training and test sets, in order to train linear and logistic regressions on observed data. The analysis compared how the models performed on out-of-sample data to predict sales demand. Statistical metrics of error, such as R^2 and WMAPE were used to identify which models best fit the available data and demonstrated the best out of sample performance in predicting new data points.

Finally, the third phase enabled a successful hand-off of the final solution. The new model was presented to the wider Distribution and Zara.com stakeholders, and a specification guideline was written to provide a detailed description of the data analysis, exception cases, and recommended path forward.

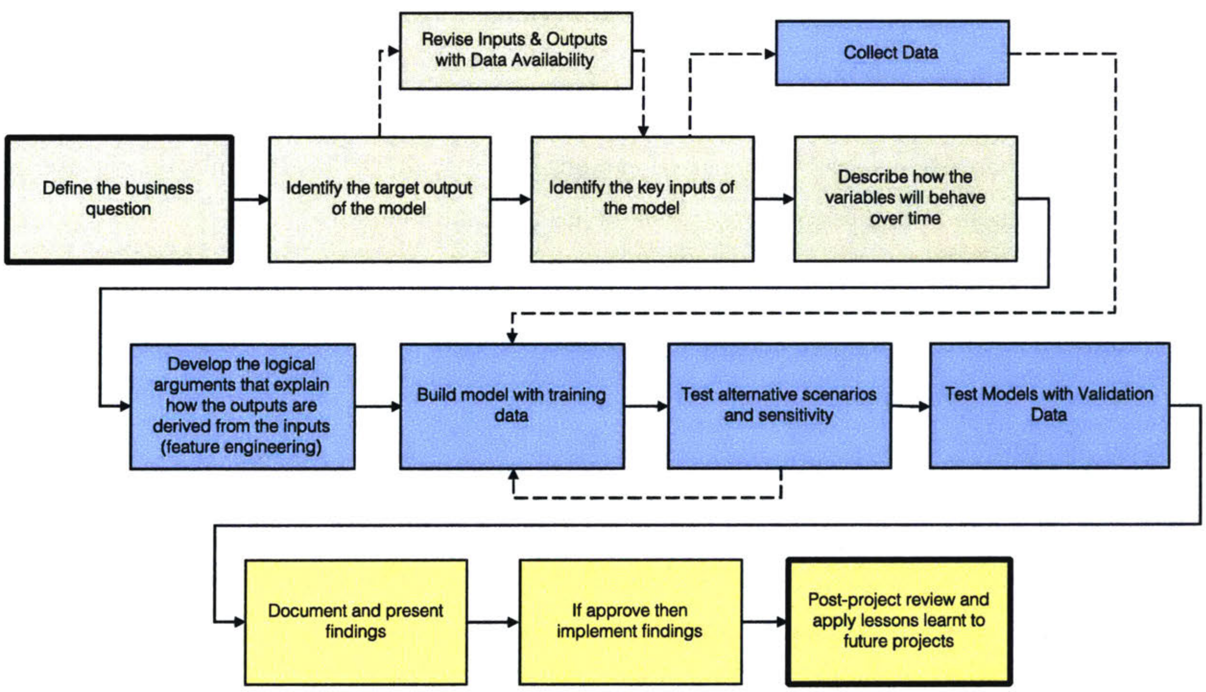

FIGURE 2: BUSINESS MODELLING PROCESS: Adapted from Tennent's Guide to Business Modeling [11]; process is divided into three phases: phase 1 in grey, phase 2 in blue, and phase 3 in yellow.

1.3 Thesis Contributions

The contributions of this thesis are the following: first, we investigate linear and logistic regression models for fashion retailers; second, we develop a systematic method to incorporate user subscriptions into the demand forecast; third, we present a method for using time available for sale as an input to demand modelling; fourth, we discuss the importance of novel web features in demand forecasting.

Linear and logistic regression models: This research focuses on linear and logistic regression models because they fulfill the needs of fashion retailers who prefer simple, implementable, robust, and scalable solutions. They are built using readily available e-commerce data and adhere to the widely adopted fashion forecasting method of using previous sales history to predict future demand. This research identifies the best performing baseline model (where Demand is a function of observed sales only) and yields 3% WMAPE prediction improvement versus a baseline regression.

Systematic incorporation of subscriptions: This research identifies a reliable method to improve the demand prediction accuracy (WMAPE reduction of 16%) for articles with censored demand. We accomplish this by using e-commerce tracking information such as user subscriptions and combine a trained clustered linear model with a decision tree to select the optimal sub-model. This improvement captures two challenging forecasting scenarios: first, it improves the ability to predict a future increase in sales after a period of stock out; second, the model improves the ability to predict a decline in sales after censored demand has been fulfilled.

Time available for sale as model feature: This work demonstrates that incorporating time available for sale ("article age") improves the out of sample WMAPE by 2% across all articles. This is achieved by performing a linear regression of observed sales while clustering articles based on similar age groups.

Use of novel web features in demand forecasting: Finally, this research discusses other web features which may be used to predict e-commerce demand. This discussion highlights methods that were investigated but did not yield prediction improvements in the context of Zara. However, these features show promising correlation with e-commerce demand, and may be investigated in future research.

The findings of this thesis have demonstrated methods to improve demand forecast accuracy using available retail data and may be used as a case study for other fashion retailers looking to improve e-commerce demand prediction.

1.4 Thesis Overview

This thesis is organized into chapters to facilitate ease of review according to the following outline:

Chapter 1 Introduces the thesis. It provides the research motivation, investigation approach, and thesis contributions.

Chapter 2 Reviews the relevant literature for this research. It provides a background on the fashion industry and insights into demand forecasting and the challenges facing fashion retailers.

Chapter 3 Provides an overview of Inventory Planning and introduces the Replenishment distribution decision on which this research is focused.

Chapter 4 Describes the data used in this analysis, and discusses the methods used to integrate disparate sources, aggregate features, and clean the data into a tidy set.

Chapter 5 Reviews statistical forecast methods and introduces the linear regression techniques used to develop the novel models presented in this thesis. Results for each model are presented.

Chapter 6 Discusses the robustness and reliability of each methodology explored in Chapter 5, and addresses the original hypotheses made about predicting future demand. Implementation recommendations for Zara and new areas for further research are provided.

This page intentionally left blank.

Chapter 2

Literature Review

Literature were reviewed to understand previous research in fashion retail and demand forecasting in the fashion industry. The fashion industry evolved dramatically throughout the 20th century and reached another significant turning point in the 21st century with the introduction of omni-channel retailing and e-commerce. These change present significant challenges to fashion retailers, and a new opportunity for operations research. The literature presented discuss fashion, demand forecasting, and statistical methods for evaluating data modelling.

2.1 Traditional Fashion Life Cycles

Traditional consumer behavior of 20th century fashion was presented in a series of pioneering works by George Sproles throughout the 1970s and early 1980s. He presented conceptual frameworks, analyses of consumer behavior, and perspectives on marketing retail fashion. These works continue to be among the most cited works in the retail fashion industry and are useful in defining the traditional fashion industry today.

Sproles presents the fashion life cycle as a temporary cyclical phenomenon, with stages of: (1) concept introduction and adoption (fashion leaders), (2) increasing public acceptance of style (growth), (3) mass conformity (maturation), and (4) decline to obsolescence. He presents "short run acceptance" of styles in the context of months to years, which is typical for the traditional fashion retailers. Furthermore, he argued that "styles lasting only a short period… of weeks or months" should be termed "fads." At the heart of Sproles' research was an analysis of the forces behind fashion trends [8].

Scholars and analysts in the fashion industry contest which forces lead to public acceptance and mass conformity in the fashion life cycle. Sproles presented four scenarios: Upper Class Theory, Mass Market Theory, Subcultural Leadership Theory, and Theory of Collective Selection. These theories compare the perceived effects of consumer behavior, fashion leaders, and cultural influences that lead to fashion maturation. At the time of his publishing (1981), it was unclear if mass market consumers had any substantial influence on the direction of fashion trends and design [8]. The immense success of fashion brands like Uniqlo, Inditex, and H&M in the late 20th century provide examples to support the theory that mass-market consumption can drive fashion trends.

During this pivotal era in fashion history, many companies were experimenting with a new concept which would later be termed *fast fashion*. This new way of operating and appealing to customers was predicated on the Mass Market Theory. Since then, the past thirty years continued growth demonstrates that consumers can – and do – drive the latest trends and fads.

2.2 Fast Fashion

As described in Chapter 1, fast fashion is defined in literature by the complementary implementation of 1) quick response, and 2) enhanced design [4]. Companies have shortened lead times and incorporated consumer feedback, resulting in a new fashion life cycle outside of the traditional retail methods described by Sproles. The business strategy behind this success was observed in the early 21st century in major global retailers such as H&M, Mango, Uniqlo, and Zara, in a category known as *fast fashion*. One of the first early demonstrated successes of this strategy was the first Inditex store in the mid 1970s. Key to the operating model was shortening lead times and incorporating direct consumer feedback into the supply chain – both distinct elements of *quick response* and *enhanced design* [9].

Quick response is characterized by short production and distribution lead times, enabling adaptability to unpredictable and highly variable demand. Enhanced design is characterized by developing and iterating on trendy designs such that the retailer keeps up with – and anticipates – consumer preferences. Delivering on the fast fashion business strategy, both elements must synchronously work together. For example, in order to leverage the complementary benefits of these two strategies, Zara historically employed in-house designers, used local labor, and expedited shipping [4]. Furthermore, the 2010 Inditex Annual Report stated that "fast effective customer service through store personnel and other enabled communication channels [was] one of the tenants of [Inditex]" [2].

Achieving *quick response* requires a thorough alignment throughout business operations. For example, if a firm's production operations are flexible, they maintain the ability to more closely align supply and demand by adjusting inventory volumes after learning new information about real demand. However, short lead times and *just in time* manufacturing are only part of the operations required to fulfill highly uncertain demand in fashion retail. Fashion companies must leverage *accurate response* through improved demand forecasting. For example, Zara has chosen to forecast at the item level, which "includes all the sizes and colors of a given garment – rather than using collections" [10]. Operationally, this allows them to design, produce, and distribute each item individually rather than grouping thousands of articles together. Instead, items can be classified within the "fashion triangle" (Figure 3) which help determines operational decisions. For example, a basic t-shirt that is produced each season may be ordered with long lead time, versus a trendy fashion item that is produced with very short lead time [10].

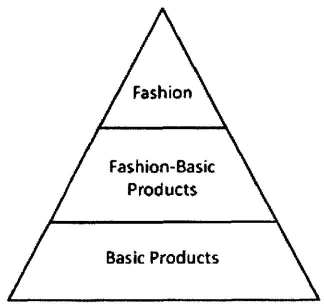

FIGURE 3: THE FASHION TRIANGLE [10]

Achieving *enhanced design* requires retailers to be diligent about staying on top of consumer trends. At Japan-based Uniqlo, the team achieves this by leveraging "today's increasing digital world to communicate directly with customers and quickly transform their desires into actual products" [11]. Furthermore, design decisions are a group effort. According to their website, Uniqlo merchandisers closely monitor the sales of each item to determine whether to increase or decrease production and shipment volumes within a single season [11]. Similarly, Inditex aims to incorporate consumer feedback into the design process. According to their website, Inditex's founding approach was to "shrink the gap between fashion creation and the customer, bringing customers closer to the products they want, all at an affordable price" [11].

Uniting the complementary effects of *quick response* and *enhanced design* was a completely novel concept in the late 20th century. One of the earliest documented businesses pursuing this goal was presented in the pioneering Sport Obermeyer case study. In this study, Fisher et al. challenged traditional production planning within a fashion company. Prior to Fisher's work, Sport Obermeyer made production decisions months in advance – based on false assumptions of predictable and stable demand. The study presented a new approach which leveraged new demand information, dynamically changing production decisions using signals from the marketplace. Such real-time information, using the complementary operation practices of *quick response, enhanced design,* and *accurate response*, enables the operational foundation required to deliver on the fast fashion business strategy [12].

2.3 Sales Forecasting for Fashion Retailing

Available literature has documented many challenges that face demand forecasting for the fashion retailer. These challenges may be categorized to better understand difficult problem of developing an improved demand prediction, as shown in Table 2 below. Each categorical issue was considered in the context of this thesis's study of demand forecasting.

TABLE 2: CHALLENGES FACING THE RETAIL FASHION FORECASTER: This project aimed to overcome these three key challenges in developing an improved forecasting method for Zara.com.

	Challenge	Description
1	Data Aggregation	Should the retailer forecast demand using a bottom-up or top-down approach? [13]
2	Mathematical Approach	Should the retailer employ statistical methods or artificial intelligence when developing a forecast model? [14]
3	Demand Censoring	How should the retailer account for censored demand during a stock out event? [15]

2.3.1 Data Aggregation

In a 2011 study, Williams & Waller performed an analytical study to determine the tradeoffs between bottom-up (BU) and top-down (TD) approaches to demand forecasting. Their study considered four scenarios: first, comparing the BU and TD methods if retail Point-of-Sale (POS) data was not available (applicable to vendors in the supply chain who do not interface with the end consumer), and comparing the BU and TD methods when retail POS data was available (applicable to retailers who interface with the end consumer, or to vendors who have access to this data). The analysis using POS data was most relevant in the context of this thesis [13].

The study demonstrated that bottom-up forecasts based on POS data have a significantly lower error metric than do the top-down forecasts for the same dataset. This finding was a surprise because it showed a counter-intuitive outcome versus the expected gains due to risk pooling that are made by top-down approach. However, Williams & Waller showed that risk pooling effects are offset by new error introduced in estimations required to disaggregate the aggregated forecast. Therefore, they recommend using a bottom-up approach to demand forecasting for retail scenarios where point of sale data is available [13]. This method is in line with Zara's baseline forecasting methods and confirms that the baseline approach is the best method for their scenario.

2.3.2 Mathematical Approach: Statistical Methods versus Artificial Intelligence

Choosing the analytical technique for forecasting that is best suited for the business context is a challenging decision for the analyst. Liu et al. wrote a comprehensive literature review of modern techniques for fashion sales forecasting in 2013 [14]. Their work provides a useful framework for understanding types of forecasting models available for demand predictions, as well as which models are popularly employed in the fashion industry.

The Mathematical Approach embodies a wide range of challenges, especially as forecasting applies to fashion retail. In particular, fashion is especially volatile, consumer tastes change rapidly, product life cycles are short, and external factors like seasonality, weather, marketing, and social media influence buying decisions. Given the variety and volume of factors that may affect consumer purchase behavior, the forecaster is faced with vast quantities of data and a challenge deciphering where to begin analysis and how to structure the model [14].

Despite the many challenges, demand forecasting is the important first step to proper retail inventory management. It is critical for supply chain success, and triggers sequential operations like due date management, production planning, pricing, and customer service level management. To provide guidance on approaching the Demand Forecast question, Liu et al. compared the benefits and drawbacks of three techniques popularly researched in Operations Management: "Traditional" Statistical Methods, Artificial Intelligence (AI) Methods, and Hybrid Approaches. They observed that advanced techniques like AI have demonstrated better accuracy than traditional time series methods, but require longer data history and more computational power [14].

Traditional statistical methods are still widely used in retail today. These methods include: linear regression, moving average, weighted average, exponential smoothing, exponential smoothing with trend, double exponential smoothing, and Bayesian analysis. The benefits of statistical forecasting are that they are simple, easy to implement, and have extremely fast computation time. A slightly more advanced improvement to these methods include classification techniques, which simulate AI clustering methods using prescribed classes like product family. Such practices have shown improved precision across the sample population pool. Despite the benefits, the statistical methods have drawbacks. They require human inputs and "expert" knowledge to choose the right tools and how narrow scope of statistical analysis. Furthermore, statistical methods have not been the focus of Operations Research for nearly fifteen years, during which the academic literature has studied AI and Machine Learning methods [14].

AI methods include: artificial neural networks (ANN), fuzzy logic models, and evolutionary networks and models (or "extreme learning machine," ELM). The benefits of AI forecasting are that they have demonstrated higher accuracy in fashion retail sales forecasting versus statistical methods. The drawbacks are that these methods can be very time intensive for computation, and some machine models are unstable, meaning that they may generate different outcomes in each simulation. In the context of demand forecasting, this may be undesirable to the forecaster who is looking for a stable prediction to use in business decisions [14].

Hybrid methods employ a mix of statistical and AI models to meet the needs of the fashion retail forecaster. For example, hybrid methods may use AI methods to identify critical factors to use in a statistical method rather than relying on human judgement with limited data. Another example is using AI methods to cluster and classify articles based on *impact factor*, which improves statistical forecasting eliminates the human factor in classifying articles based on limited data [14].

2.3.3 Demand Censoring

The demand forecaster faces the challenge of analyzing imperfect data, which often includes periods of stock-out. During these periods, the true consumer purchase behavior is *censored*: unsatisfied customers are unable to purchase desired items due to a stock out. Ignoring this effect understates the actual customer demand. If this understated observed demand is used to predict future demand, the retailer will always forecast an underestimate [15].

Traditional forecast methods leverage point-of-sale (POS) data, which has been demonstrated to provide retailers more accurate information on the true underlying demand and stock relationship. Further research and retail experience show that Bayesian updating – wherein the observed no-sale data during a stock-out is corrected-for in the forecast method – improves the demand estimation and reduces prediction error. This method was used as a baseline for this thesis research. However, the challenge of handling censored demand still introduces variability and prediction error, and remains an opportunity for improvement beyond Bayesian updating [15].

2.4 Evaluation Methods

Forecast evaluation is the topic of textbooks, research papers, and coursework in Operations Research and Statistics. Selecting applicable and representative metrics is critical for evaluating mathematical methods in demand forecasting.

This thesis focuses primarily on the evaluation of linear regression methods. To compare the performance of each regression, the following key parameters were analyzed:

- **R^2** coefficient of determination is a statistical measure of how well the regression line approximates the real data points
- **Root Mean Square Error (RMSE) is** the measure of differences between the values predicted by the model and the observed values
- **Weighted Error Measurement (WMAPE)** is the average measure of differences between the values predicted by the model and the observed values, weighted to the magnitude of each data point. WMAPE was used in this thesis to evaluate and screen models.
- **Percent Under Prediction** is the ratio of number of values predicted by the model which were less than the observed value, versus the total number of data points

The above performance metrics were studied for each model on the relevant dataset(s). Metrics were calculated on the average basis (across the entire sample population) and analyzed for extreme cases (e.g. maximum, 90^{th} percentile, minimum).

2.5 Incorporating e-commerce features

Previous MIT LGO theses have studied demand forecasting and e-commerce. Of particular relevance is the foundational work presented on predictive analytics for replenishment distribution decisions [16]. That research supported using website data to improve the demand prediction for online sales. Over 30 features were surveyed model development. These ranged in category, weight, and complexity, and enabled a rapid analysis of impactful feature selection for model advancement. Table 3 highlights key features which demonstrated possible significance to improving demand prediction for brick and mortar and e-commerce sales.

TABLE 3: SUMMARY OF FEATURES PREVIOUSLY INVESTIGATED: Previous LGO theses identified the following features that demonstrate high correlation and possible significance to predicting future demand [16].

Replenishment Shipment: B&M Feature Importance				
	E-commerce Data Only		w/ Subscriptions	
Rank	Feature Name	% MSE Increase	Feature Name	% MSE Increase
1	Vm	0.4018	Vm	0.5846
2	Subfamily	0.0544	Subfamily	0.1389
3	Stores	0.0506	Net Cart	0.0951
4	Net Cart	0.0410	Article Age	0.0824
5	Article Age	0.0492	Rank	0.0823
6	Article Age (Ecomm)	0.0479	Article Age (Ecomm)	0.0796
7	Ecomm Sales (MCC)	0.0273	Stores	0.0659
8	Avg. Best Position	0.0243	Ecomm Sales (MCC)	0.0538
9	Region	0.0193	Region	0.0467
10	Rank	0.0180	Avg. Best Position	0.0367
11	Price	0.0155	Total Capacity	0.0312
12	Avg. Categories	0.0101	Avg. Instock (MCC)	0.0311
13	Total Capacity	0.0072	Price	0.0252
14	Avg. Instock (MCC)	0.0056	Subscriptions (MCC)	0.0252
15	Days New	0.0048	Avg. Categories	0.0198
16	Mall	0.0013	Days New	0.0175
17			Mall	0.0043

While previous work has begun the investigation of the types of features that may be useful in e-commerce demand forecasting, there were no studies that showed *which* features yield the best improvement nor *how to incorporate* them systematically into a reliable forecast model. Based on these two research gaps, the open question of *how to use novel web features to improve demand forecasting?* was the research opportunity for this thesis.

This research aims to build upon previous studies showing that there is a correlation between website data and future demand. It uses data from Zara.com sales and website to build a comprehensive dataset with key features. These features are analyzed for importance in demand prediction. The most important features are combined in statistical models to predict online sales. The research demonstrates new methods to utilize these data, as well as future opportunities for further research using the identified data and higher order modelling tools.

Chapter 3

Logistics Operations

Logistics Operations is a topic of significant research in the business and academic communities. Within logistics operations are Inventory Planning and Inventory Distribution. Like many fashion retailers, Zara executes logistics operations on an article-and-location basis. This means that the logistics team must decide how much of each article to purchase, allocate, and distribute to each location on a rolling basis. Given the rapid pace of change in the fashion industry, these decisions introduce significant challenges and areas of opportunity for improvement.

3.1 Demand Forecasting for Inventory Planning

In supply chain management, "demand forecasts form the basis of all supply chain planning" [17]. The fashion industry is widely known as unpredictable and challenging to forecast, similar to other fast-moving retail items like consumer electronics and high-tech products. To handle inventory management for a wide variety of products, product planning occurs in several phases, often known as product "cycles." These cycles separate business decisions at different stages of the supply chain, such that the product buying decision (how much product to order from the supplier) is separate from the distribution decision (how many units to ship to stores and e-commerce warehouses). However, the demand forecast ultimately drives all aspects of the supply chain, and is critical for purchase, planning, and distribution decisions [17].

The demand forecast for product distribution is the focus of this study. Fashion product life cycles are short given the rapid introduction of new articles and seasonality of demand. Therefore, products quickly move between the stages of the product life cycle (market development, growth, maturity, and decline) [18]. To manage the distribution throughout a product's lifetime, Zara uses a two-period operating model. The first period, called "Initial Period," refers to the first shipment from the Distribution Center to the stores and warehouses. This Initial Shipment covers demand for the first seven days that the article is available for sale. The second period, or "Replenishment Period," refers to each subsequent shipment from the Distribution Center to the stores and warehouses [16]. The Initial and Replenishment periods vary in two important ways: Forecast Method and Shipment Volume Determination. Table 4 below describes these differences.

TABLE 4: COMPARISON BETWEEN INITIAL AND REPLENISHMENT: Zara.com segments articles based on "Initial" introduction of new articles, and "Replenishment" of existing articles on the website. This segmentation enables different distribution practices based on article shipment type [16].

	Initial Period *Initial shipments of product*	Replenishment Period *All shipments after Initial*
Demand Forecast	Does not include article sales history	Based on observed sales of the article being forecasted.
Shipment Volume Decision	Based on selection of initial coverage period	Based on selection of replenishment review period & desired coverage period

This section discussed the importance of the demand forecast for making business decisions in the supply chain. The forecast is especially challenging in high-turnover, rapidly changing industries. This challenge, as it relates to product replenishment, is the focus of this thesis. The next section, "Replenishment: Inventory control policy" discusses how the Demand Forecast is used in supply chain inventory distribution.

3.2 Replenishment: Inventory control policy

As Section 3.1 discusses above, the demand forecast is used throughout all aspects of the supply chain. This chapter discusses how the forecast is used in inventory control policies to determine shipment volumes.

An inventory control policy is a "procedure that helps to define how much and when to order" inventory [19]. The inventory planner has four parameter that she may use to affect this decision:

1) Replenishment interval (cycle time)
2) Order quantity
3) Re-order level (minimum inventory)
4) Target inventory level

Some firms may choose to fix certain parameters constant: or example, some electronics retailers have been cited using a one-week replenishment interval and always order on Fridays to meet the next week's demand. The combination of fixed and variable parameters are documented in literature as *periodic review systems* and *continuous review systems*. The four resulting policies referenced in literature are shown below in Figure 4 [19].

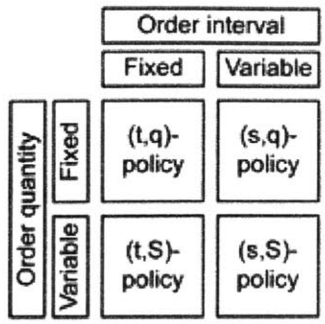

FIGURE 4: INVENTORY CONTROL POLICIES: Static and dynamic policies to ship inventory to satisfy downstream demand [19]

Once the firm determines its Inventory Review Policy, the supply planner may use the fixed parameters to determine the undefined parameters. In a fixed review period policy, which is common in fast moving retail systems, the shipment volume must be determined using forecasted demand and coverage time. The shipment volume can be expressed as:

$$Shipment\ Volume = Forecasted\ Demand * Coverage\ Time$$

The two variables are Forecasted Demand and Coverage Time. The demand forecast is the subject of this thesis, and Chapter 5.1 explores mathematical forecasting methods in detail.

These two variables are influenced by six key components:

- The systematic component (base demand)
- The trend (expected gradual change in demand)
- The seasonality (expected step-changes in demand)

Influences Demand Forecast (*See Ch. 5*)

- The random component (variability in demand)
- The cycle time (time between review)
- The lead time & variability in lead time (shipping times)

Influences Coverage Period

To determine the coverage period, the inventory planner must account for the three components identified above: random variability, cycle time, and shipment lead time. This can be expressed using the equation below, where the variability in demand and variability in lead time are combined in the "safety period" component [19].

$$Coverage = \ Cycle\ Time + Lead\ Time + Safety\ Period$$

The safety period covers several factors. First, the safety period covers the systematic variability in lead time and demand. Second, safety period covers the unknown variability due to imperfect demand forecasting. Therefore, improvements to the demand forecast have the ability to reduce the required safety stock associated with each shipment. The demand forecast accuracy has a direct impact on inventory levels in the supply chain [20].

As seen in the two equations presented above (Shipment Volume and Coverage), the Demand Forecast is deeply embedded in inventory planning. First, the demand forecast is used to determine the expected consumer demand over the cycle time and shipment lead time. Second, the forecast prediction accuracy is used to determine the safety period for inventory levels. Based on these factors, it is demonstrated that the demand forecast is a critical parameter for operationally efficient supply chains.

This section explained inventory control policies for supply chain planning and discussed how the demand forecast is used to make informed business decisions. These concepts were used as the foundation for this research thesis. The research presented here discusses methods to improve the demand forecast using replenishment shipment data from Zara.com. This research has the potential therefore to improve operations by reducing error in demand prediction, thereby leading to higher customer service levels and lower inventory costs to the retailer. Demand Forecast for making informed business decisions that impact the inventory position throughout the Zara supply chain.

Chapter 4

Data Description

Model development requires a reliable and rich data source to provide sufficient information to draw conclusions and predict behavior of un-seen data points. This research leveraged data from two departments within Zara and used administrative databases where possible. This process ensured that the highest quality data was used and minimized the occurrence of questionable exception cases.

4.1 Data Sources & Structure

Data for this thesis were collected from raw source data from the Zara Distribution and Zara.com databases. These databases were provided in an accessible format, without pre-filtering or trimming. Three original sources of data were integrated to build the dataset used for analysis: 1) Sales history, 2) Subscriptions history, 3) Inventory, 4) Website Structure.

The original source data were screened for the following criteria:

Department: The study focused on one department of Zara products. Within this department, the sale behavior of articles was studied for one year, and one season (e.g. Winter or Summer). Narrowing the data scope in by these criteria achieved two goals: first, it mirrored the way Zara has historically categorized articles for inventory planning purposes; second, it limited the data volume to a manageable size.

Geography: Key regions (countries or groups of countries) were studied that that are large and representative of Zara sales. These geographies were selected because they represent either (1) the most complex to model and (2) the most common distribution model. Examples of retail distribution are shown in Figure 5 below.

Pricing: The data did not include any "special prices" or items with online discount. The dates studied screened out 4 periods of store-wide discounts (Black Friday, Zara.com Spring Sale, Zara Winter Sale, or Zara Summer Sale). Discount days dramatically affect sales: section 4.3 "Filtering" demonstrates the powerful impact of these events and describes how they were addressed in this research.

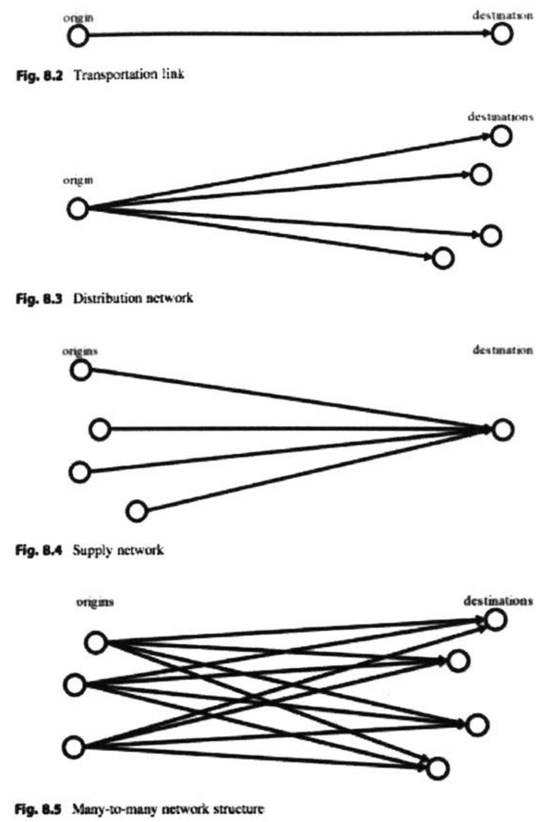

FIGURE 5: EXAMPLE DISTRIBUTION STRUCTURES: Generic distribution structures are shown above, excerpted from *Global Supply Chains and Operations Management* [21]. This project studied data from two regions that used a combination of both simple and complex distribution structures. The simple structure approaches the top "Transportation link" example, and the complex structure approaches the "Many-to-many network structure" example above.

A dataset was built using available data from discrete databases within Zara. Building a retail dataset has been studied in literature. A typical dataset development process and data structure is shown in Figure 6 below. The research for this thesis used a similar approach.

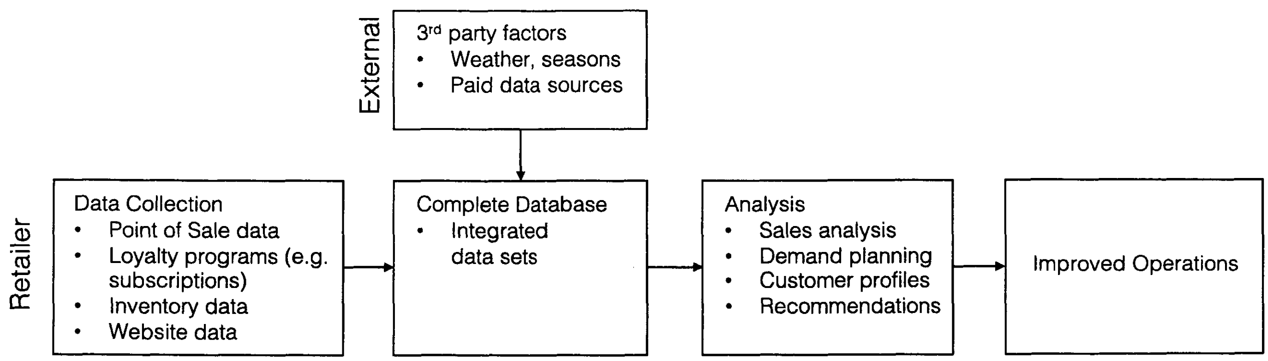

FIGURE 6: DATA INTEGRATION & STRUCTURE: Current practices in retail include developing datasets from disparate internal and external sources for supply chain analysis. Schematic is adapted from Tengberg, 2013 [22].

The dataset used in this research included data on: transaction history, subscription history, inventory history, and website layout. These disparate datasets were integrated to form a complete database, which was used for analysis. The data integration required identifying unique keys, such as *date* and *SKU*. Original sales records included information on Customer ID, Date, SKU, and size identifiers. This data was transformed to calculate observed sales per day. Similarly, the original Subscriptions records were available based on unique Customer ID, Subscription Date, Notification Date, and SKU identifiers. This data was transformed to calculate "Active Subscriptions" for each SKU. Next, inventory records were available for each Date, SKU and size, and was transformed to calculate average inventory availability across sizes for Date. Finally, the transformed Sales, Subscriptions, and Inventory datasets were integrated to build the complete analysis data for each geography.

The data were inspected and screened to ensure that the compiled dataset was accurate and representative to be valid in this study. The prepared datasets were inspected for outliers, and common exception cases were either removed or adjusted using existing logic-based rules that Zara uses for Replenishment Forecasting. For example, this logic included:

- Dataset was limited to articles which had been available for sale more than n days, to fall within the Replenishment Period
- Items with no sales throughout the season (but available stock) were assumed to be un-listed on the website, and were removed from analysis
- If an item was stocked-out for longer than m days, only the period without stock-out was used to calculate demand forecast

Finally, the dataset was split into unique events for articles with Subscriptions and unique events for articles without Subscriptions. The hypothesis was that Subscriptions were an important feature in forecasting Demand, and therefore the demand of those articles should be modelled independently from the rest of the items.

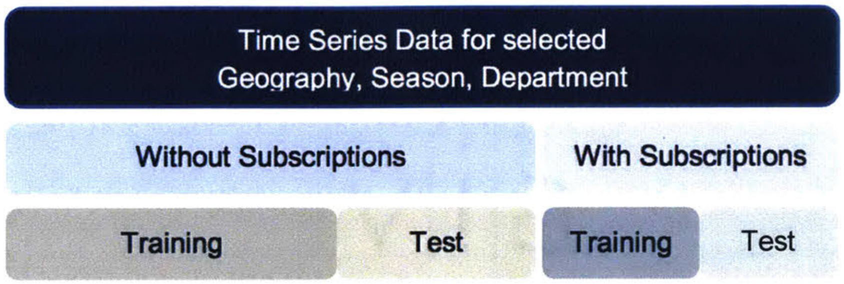

FIGURE 7: DATASET SEGMENTATION: Data were segmented as indicated for statistical analysis and Modeling

A summary of available data points of interest, after all filtering was applied, provided in Table 5 below. These values indicate the number of available rows in the full datasets studied.

TABLE 5: SUMMARY OF DATA: Total available data in selected Geography and Selling Seasons, after filtering based on discussed criteria

	Geography 1 e-commerce	Geography 2 e-commerce
SKU	4,700	6,200
Dates	124	124
Total Rows	336,000	253,000

4.2 Data Filtering

The data were filtered down to study two key geographies. These geographies were selected as representative for proof-of-concept of the challenges and opportunities for improving the e-commerce demand prediction. Source data were integrated and aggregated for each geography separately, and prediction models were developed independently. To prepare data for analysis, the datasets were split into Training & Test sets. The training set was used to identify feature correlation, feature importance, and to train models. The Test set was used to test the models and to select the best performing models for out-of-sample improvement.

As was supported by demand modelling literature, the dataset was filtered on three criteria: promotions, seasonality, and holidays. These factors have been shown to influence customer demand [23]. Within the dataset in this research, it was evident that price promotion and holidays had a significant impact on consumer buying behavior. For example, Figure 8 below shows the total unit sales by day for one region in the season studied. The abrupt peak corresponds to a single-day discount event. Consumer purchases more than quadrupled on that day, and subsequent days observed below-average sales. To minimize the effect of discount pricing, the data were filtered to only include dates up to one week before known major price promotion events on Zara.com.

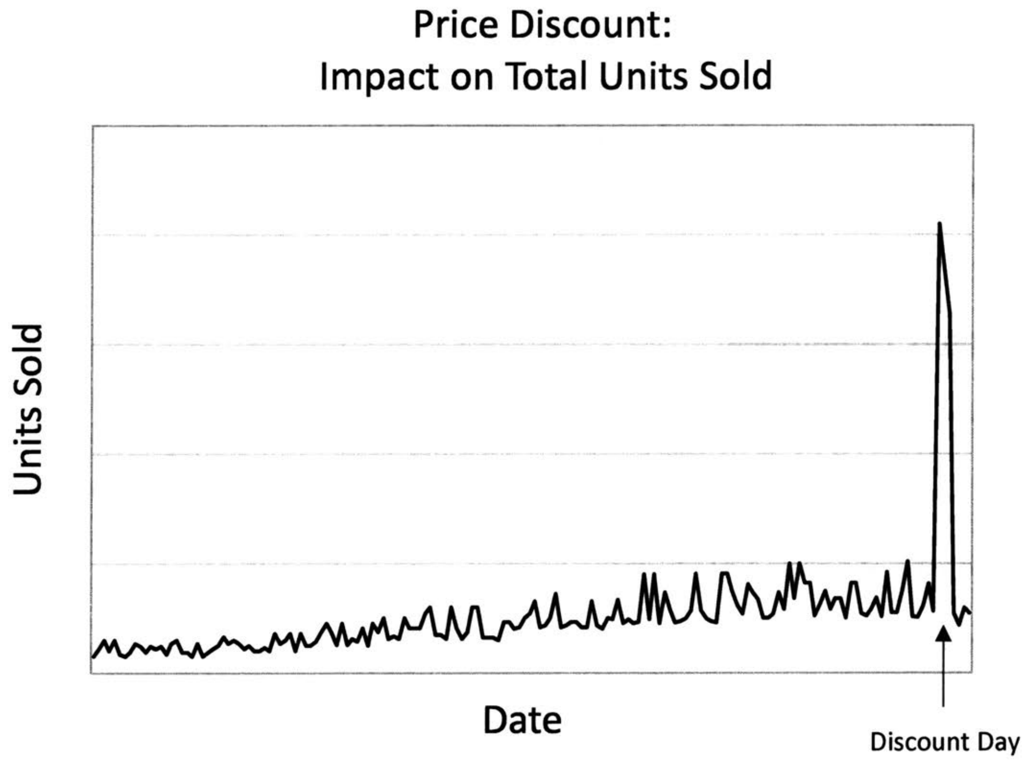

FIGURE 8: DISCOUNT PRICING EFFECT ON SALES: A single day of discounts has an extreme effect on consumer buying decisions, and results in a demand peak that was ultimately screened from the analysis in this project.

4.3 Data Feature Description

Previous research was leveraged to narrow the scope of feature engineering. Feature analysis and model development were closely interrelated, and often performed in tandem. This enabled flexibility in engineering and problem-solving, as new features were rapidly and dynamically explored using Training Data.

Like the thesis as a whole, new data features were categorically split between Subscriptions and Other e-Commerce Features. The project was structured in these discrete avenues because of the clear differences in the phase of analysis. To accomplish the first project goal, incorporating Subscriptions into the replenishment demand forecast, much less feature engineering and experimentation as required. Previous research had already proven that Subscriptions were a meaningful, statistically relevant feature [16]. This project iterated on that finding in order to identify the best way to incorporate Current vs. Past Subscriptions, Conversions, and Duration of Stock-out (days with Subscriptions). The second project goal, investigating other e-commerce features to improve the demand prediction for online sales, was a more experimental research project. This project focused on features that had been previously highlighted as statistically relevant, but still required significantly more iteration, development, and testing of these variables for use in a robust prediction model.

4.4 Notation & Definitions

SKU = Specific article notation (does not differentiate size)

t = A single time period where t = 0 represents the date that the demand forecast is prepared

v = Observed sales of a given SKU & size

$\mathbf{D}^{R}$ = Predicted average demand for the x days after forecast date [units per day]

$\mathbf{D}^{OB}$ = Observed average demand for the y days after forecast date [units per day]

DPV = Possible Sales Day: Binary variable to indicate if an SKU-size was available on a specific date

$\mathbf{W}_t^{OB}$ = Observed average e-commerce sales through Zara.com during the time period of interest. Aggregated at the SKU level (all sizes) for each region of interest. Time periods available include cumulative since the product was first introduced on Zara.com and for the previous week beginning the day before forecasting is performed. It is calculated as:

$$W_t^{OB} = \frac{\sum v}{\sum DPV}$$

Subscription Email tracking service for stocked-out articles, where customers may enter their email address to receive a notification of replenishment.

Notification Automated email from Zara.com to subscribed customers, sent when a stocked-out item is replenished & available for sale.

Subscription Features

Active Subscriptions The cumulative number of distinct email addresses signed up to receive notification for out-of-stock SKU beginning up to p days before replenishment. After replenishment & subscribed customers receive Notification, the number of Active Subscriptions equals zero.

Converted Sales from Subscription Sales of a previously stocked-out SKU to a customer who had signed up for Subscription, sold within q days of Replenishment Notification

Average Conversion Rate The ratio of Converted Sales from Subscriptions versus the total number of email subscribers prior to Notification

Duration of Stock-out Cumulative number of days of SKU stock-out during which customers may subscribe to receive replenishment notification

Peak Day Observed date after Notification event wherein SKU sales spike due to conversions

E-commerce Features

Garment Cluster: Pre-defined featured pulled directly from the Administrative Article Database as a Garment Cluster code. Incorporated to model as the average weighted average sales across garment cluster relevant to SKU being forecasted.

Abbreviated Notation of Garment Cluster Average Sales: $\bar{W}_{t,Garment\ Cluster}$

Article Age The number of days between the initial sale offering of the SKU on Zara.com and the forecasting date, calculated as *ForecastDate – InitialDate.* This represents the number of days an SKU has been available for sale online

Position on Website SKU display position on the forecast date, only identified if featured in the New In section of the website.

New In Binary variable to indicate if SKU is featured in New In (1 if SKU appears in New In on Forecast Date, 0 if not).

Days in New In Cumulative number of days SKU has been featured in the New In section of the website

Average Sales in Physical Stores Observed weighted average sales across all physical stores in geography relevant to Zara.com market being analyzed [units per day]

Abbreviated Notation of Average Sales in Physical Stores: $\overline{W}_{t,MCC,physical}$

One day change in Sales Observed difference in SKU sales between two previous days of sale: positive value indicates increase in sales between t-2 and t-1, and negative value indicates decrease in sales between t-2 and t-1.

4.5 Feature Covariance

Feature Covariance is a widely used tool in statistical modelling to identify data features which are highly correlated to each other. In statistical modelling, it is foremost desirable to have features which are correlated with the variable being predicted (in this case, Demand). Second, it is desirable to have features which are not highly correlated with the other prediction factors.

This thesis used feature covariance to narrow the scope of features used in model development. When compared across the entire dataset, this research reduced the feature set down to factors which demonstrated high correlation to Demand and did not have high covariance with other prediction features. In the case where multiple features correlated with Demand and with each other, judgement was used to identify if the features were offering directly redundant information, or if they should be evaluated with further analysis.

As shown below in Figure 9, website features were tested for covariance. In the lower right-hand quadrant, three variables are all highly correlated with values of 0.79, 0.83, and 0.94. These features are all different aggregations representing observed sales of an article-location pair. It was an intuitive finding that all three features showed extremely high correlation. However, only two of these features demonstrated significant correlation with future Demand (covariance of 0.2 and 0.21, versus 0.05 for the third aggregation method). Given that there was no statistical difference between covariance of 0.21 and 0.2, the simplest representation of observed sales history was selected to include in the demand models studied. This was in line with historical practices at Zara and validated the baseline prediction methodology.

A second group of features emerged as highly covariant in the upper right-hand quadrant of Figure 9. These three features are all methods to represent the duration of a stock out and available Subscriptions information. These features demonstrated covariance of 0.62, 0.53, and 0.63. However, of these three features, only one showed significant correlation to future demand. This was the cumulative number of subscriptions observed the past 7 days and showed a covariance with future article-location demand of 0.08. Based on this result, the past week cumulative subscriptions were selected for model development.

Another notable discovery was the inverse correlation between Article Age, category New In, and article Position on the website. Feature correlation identified that Article Age was negatively correlated with New In and Position: the fewer days it was available, the stronger correlation it had with being in a good position and being in New In. Furthermore, Article Age demonstrated a strong negative correlation with future sales (-0.45), and New In demonstrated a mild correlation with future sales (0.21). Meanwhile, Position was highly positively correlated with New In (0.47), indicating that new articles were more likely to be prominently displayed in a good position. These findings make intuitive sense, but provided critical insight to the Zara team, who previously thought that all features would provide new information to the model. Instead, this covariance analysis indicates that only one of Position, New In, or Article Age should be used in demand model development.

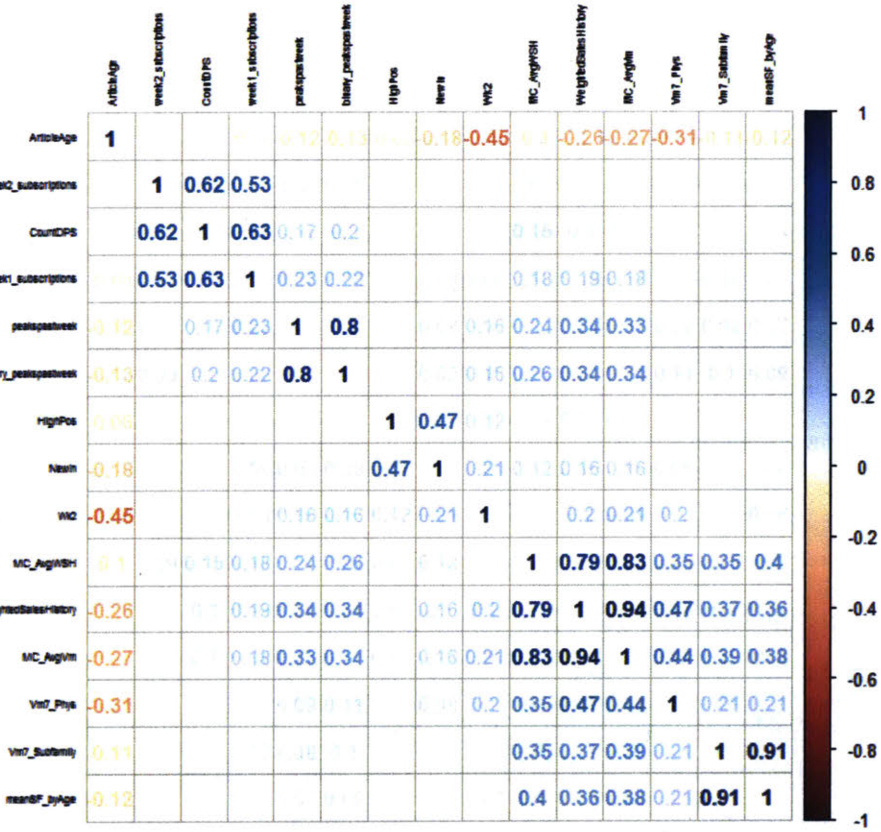

FIGURE 9: COVARIANCE MATRIX: Covariance indicated for features of interest in the available dataset

4.6 Feature Importance

Feature importance is a method to rank the relative impact of data features when building a new model. This method enables the researcher to focus model improvements on features which have the most impact on model accuracy. In this research, feature importance was determined iteratively. A list of possible features was generated through review of prior research, inspection of the dataset, and interview conversations with business analysts in Zara Distribution and Zara.com. The identified features were constructed in the dataset and tested for covariance. If a feature was determined to be sufficiently independent of other variables, and demonstrated correlation to Actual Customer Demand, it was explored during model development. This analysis identified the following features, indicated in Table 6, as promising the highest relevance for predicting actual customer demand of the dataset studied.

TABLE 6: SUMMARY OF RELEVANT FEATURES: Features shown below were Identified as relevant to forecasting consumer demand and were later explored in model development

Subscriptions Features	**Other e-Commerce Features**
Active Subscriptions	Garment Cluster
Converted Sales from Subscription	Article Age
Average Conversion Rate	Position on Website
Duration of Stock-out	In New In (Binary)
Peak Day for Item with Subscriptions	Days in New In
	Average Sales in Physical Stores
	One day change in Sales

This page intentionally left blank.

Chapter 5

Model Development

Retail forecasting demand prediction methods vary widely in complexity, from naïve (for example, future sales equal observed sales) to complex AI (as discussed in Chapter 2). This research aims to develop a model that complements the existing Zara forecasting methods. To achieve this goal, it uses observed datasets to train statistical methods such as linear and logistic regression. This chapter includes a description of the types of statistical methods that were formulated in the development of an improved demand prediction. The descriptions are arranged in increasing complexity, from most simplistic model to the most advanced.

Previous studies have evaluated the formulation of demand forecast for the purpose of short-lead time prediction and replenishment [24]. This study aims to further reduce prediction error by incorporating new data and by testing model performance using the best-performing methods from the earlier study. Of the five standard statistical methods presented below, this study focused primarily on adaptations of method 3: predictions based on Weighted Moving Average Sales.

At Zara, the demand forecast is reviewed, discussed, and adjusted by many different stakeholders. The Distribution Team develops the forecast and has considered higher order sophisticated modelling. Some advanced AI methods require significant monitoring and evaluation from a data analytics team to ensure optimal performance. In considering the tradeoffs between model improvement and complexity, and it is common to evaluate new models by comparing the performance versus a baseline model. This thesis uses this evaluation method, and compares each new model versus the baseline model, and where necessary, evaluates any additional complexity introduced.

5.1 Description of Statistical Forecast Methods

General Notation

y_t Sales during period t [sales/period]

$\hat{y}_{t+1}$ Predicted sales during the forecast period [sales/period]

n Number of previous period(s) to use in lagging average of previous sales [periods]

$\boldsymbol{\omega_i}$ Linear weighting factor, where $\sum_i \omega_i = 1$

Description of General Statistical Methods:

This thesis only studied replenishment period, and therefore all forecast date-article datum had at least seven sales days of history to build a prediction. To achieve the underlying goals of *implementable* and *communicable*, this research aimed to build upon the existing forecast method. Models explored iterated on a variety of statistical methods, described below.

1. **Naïve:** The most basic time-series forecast can be made from the available information about previous sales. It can be described as a "random walk" where the next period sales are predicted by the previous period sales.

$$\hat{y}_{t+1} = y_t$$

2. **Simple Moving Average:** is a modification to the Naïve forecast. It approaches the cumulative average when n includes all available periods. In this scenario, the slope approaches zero. The moving average approaches the Naïve forecast as an extreme short-term average when n = 1, and the forecast behavior approaches a random walk.

$$\hat{y}_{t+1} = \frac{1}{n} \sum_{i=t-n}^{t} y_i$$

3. **Weighted Moving Average:** is a modification to the Simple Moving Average forecast. The average sales during previous periods of defined length *n, m, p, etc.* are combined with linear weights ω_i. It approaches the Simple Moving Average when $\omega_1 = 1$.

$$\hat{y}_{t+1} = \omega_1 \frac{1}{n} \sum_{i=t-n}^{t} y_i + \omega_2 \frac{1}{m} \sum_{i=t-m}^{t-n} y_i + \omega_3 \frac{1}{p} \sum_{i=t-p}^{t-m} y_i + \cdots$$

4. **Holt Single Exponential Smoothing** is distinct from the aforementioned Naïve, Moving Average, and Weighted Moving Average statistical forecasts. Similar to the Weighted Moving Average, this method weights newer observations more heavily than older observations. However, the weights decrease exponentially as they age. In this model, α is the exponential smoothing parameter. High values ($\alpha \rightarrow 1$) tends toward a volatile prediction, approaching the Naïve method. Low values ($\alpha \rightarrow 0$) tends toward a stable prediction, approaching the cumulative or moving average methods.

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1-\alpha) \cdot \hat{y}_{t-1}$$

5. **Holt Exponential Smoothing with Linear Trend** is an adaption of the Holt Single Exponential Smoothing Method and assumes a linear trend. This forecast is a combination of the latest estimate for the level (l_t), calculated using simple exponential smoothing with the α parameter, and the trend (b_t), calculated using the exponential smoothing β parameter.

$$\hat{y}_{t+1} = l_t + hb_t$$
$$where\ l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) = \text{level}$$
$$h = time\ between\ last\ obervation\ \&\ forecast$$
$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} = \text{trend}$$

5.2 Description of Models Evaluated

Models M1 – M3: Exploring the relationship between Predicted Sales, D_{SKU}^{R} and Observed Average Sales, Vm_{MCC}^{OB}

- **M1** is the biased forecast, where the M0 Weighted Average Sales History is fit to the true demand profile using linear regression with intercept of zero and best-fit slope equal to β_1. This is the closest approximation to the baseline forecast method and will hereto be referred to as "Baseline."

$$D_{MCC}^{R} = \beta_1\ (\omega_1 \mathrm{W}_{\mathrm{SKU},1}^{\mathrm{OB}} + \omega_2 \mathrm{W}_{\mathrm{SKU},2}^{\mathrm{OB}} + \omega_3 \mathrm{W}_{\mathrm{SKU},3}^{\mathrm{OB}}) \qquad \textbf{(M1)}$$

- **M2** is a biased forecast of the logarithmic transformation of the variables. Previous research demonstrated the log normal distribution was a suitable approximation to the demand profile. Therefore, the Weighted Average Sales were transformed and fit to a new biased forecast with intercept of zero and best-fit slope of β_1.

An important feature of average sales (observed and predicted) is that the range of values may include zero and less than one values (for cases with zero or very few sales). To address this issue in the log-normal model (where ln(0) = infinity; ln(0<X<1) = negative), D_{SKU}^{R} and W_{SKU}^{OB} were transformed to $D_{SKU}^{R}{}'$ and $W_{SKU}^{OB}{}'$ as shown:

$$D^{R}_{MCC}{}' = D^{R}_{MCC} + 1$$

$$Vm^{OB}_{MCC}{}' = Vm^{OB}_{SKU} + 1$$

$$\ln(D^{R}_{MCC}{}') = \beta_1 \ln(\omega_1 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},1} + \omega_2 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},2} + \omega_3 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},3})' \qquad \textbf{(M2)}$$

Error analysis for M2 was performed after returning the predicted output to the untransformed variables.

Model M3: Incorporating Article Information

Previous analyses studied the effects of including categorical linear features in the demand forecast model. In that research, Item Type, and Buyer (Woman, Teen, Basics) were tested as individual indicators and resulted in marginal model improvement. However, when they were combined to develop a cross-categorical clustering model, the forecast prediction improved significantly.

This thesis set out to investigate whether categorical information could be included in the new forecasting tools. Data limitations were identified and explained why earlier findings had not been adopted from previous studies. Instead, forecast clusters were created using specified ranges of observed average sales (W).

- **M3** is a linear regression of Weighted Average Sales with intercept equal to zero, where β_{W} is a vector of coefficients depending on observed Weighted Average Sales.

$$D^{R}_{MCC} = \beta_{Vm}\ (\omega_1 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},1} + \omega_2 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},2} + \omega_3 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},3}) \qquad \textbf{(M3)}$$

As shown in Figure 10 below, the Baseline forecast method performed equally to the Logarithmic Transformation (M2). The clustered regression, with unique coefficients calculated for each specified range of W performed slightly better than the Baseline prediction baseline. Based on these results, the Clustering technique was selected as the best performing model of the most basic regression scenario and was used for testing with additional features included in the model.

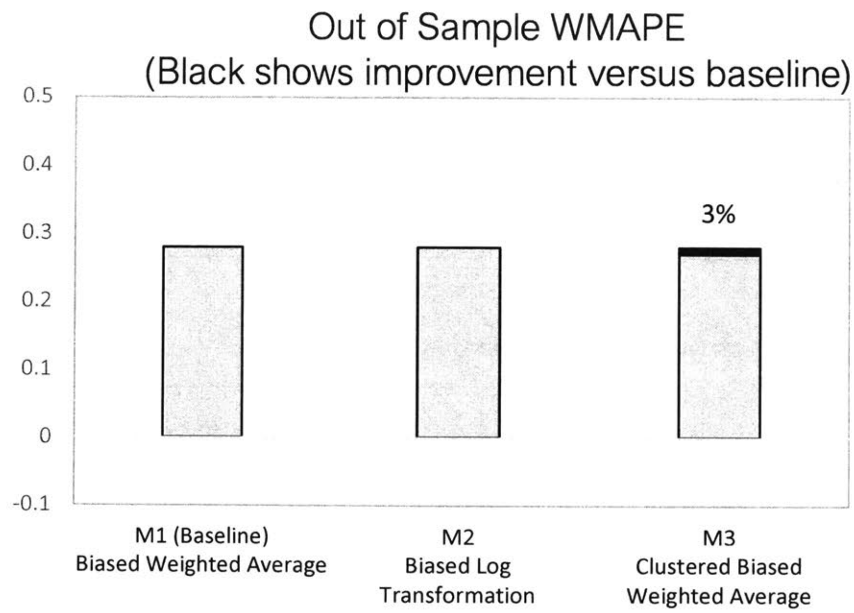

FIGURE 10: SUMMARY OF PERFORMANCE FOR M1, M2, AND M3. M2 shows no improvement versus the Baseline prediction method (M1). M3 shows slight improvement versus the baseline prediction. Results presented for single test geography, with all relevant data included.

Based on the relative WMAPE performance of models M1, M2, and M3, further analysis was studied to compare M1 to M3 These results are shown in Tables 7 and 8 below, which compare the regression results for the two models. The regression was performed using Training data for a single geographic region during one full selling season of articles.

The detailed regression analysis shown in Table 7 shows that the regression coefficient is highly significant for each unique cluster of W (with t-value >> 1), and the magnitude of coefficient varies significantly depending on cluster. This suggests that there is, in fact, new information to be gleaned by clustering the model on the magnitude of observed weighted average sales, W. Furthermore, Table 8 shows the R^2, Adjusted R^2, and RMSE results for M1 versus M3. These results indicate that M3 performs better than the baseline model for all metrics tested and are in line with the observed out of sample WMAPE.

TABLE 7: REGRESSION ANALYSIS FOR MODELS M1 AND M3. Regressions were performed using the full dataset including article-dates with and without active subscriptions.

M1: Baseline	**M3: Clustered Regression**

Var.	β	Std. Error	t-value	Pr(>|t|)	Var.	β	Std. Error	t-value	Pr(>|t|)
W	0.8417	0.0009191	915.8	<2e-16	W_clusterA	1.044	0.011046	94.56	<2e-16
					W_clusterB	0.9624	0.003029	317.76	<2e-16
					W_clusterC	0.8989	0.001629	551.89	<2e-16
					W_clusterD	0.8412	0.001967	450.6	<2e-16
					W_clusterE	0.7595	0.001506	504.4	<2e-16

TABLE 8: MODEL FIT RESULTS FOR MODELS M1 AND M3

	M1	M3
R^2	0.7354	0.73857
Adj R^2	0.8153	0.8211
RMSE	5.186	5.10279
OOS WMAPE	0.280	0.271

Models M4 – M8: Incorporating Article Subscriptions

Previous research has shown that censored demand during a stock out, in the form of Subscriptions, are a valuable indicator for predicting future demand. This was corroborated through the feature analysis and covariance analysis described in Chapter 4. Quantitatively, there were two goals of incorporating subscriptions in the analysis for this project. These goals were based on hypotheses formed after analysis of consumer behavior patterns after a stock-out and subscription event.

The first goal of incorporating subscriptions was to predict a future increase in demand immediately following a stock-out. This would be an improvement versus the baseline model, which is blind to the censored demand and under-predicts the demand after replenishment. The hypothesis was that including the number of subscribers to a given article would provide information about the demand immediately following replenishment. This was based on the observed purchasing behavior where previously-stocked-out articles would receive a short demand peak after replenishment.

The second goal of incorporating subscriptions was to predict a subsequent *decline* in demand after the pent-up subscriptions had been fulfilled. The hypothesis for this was that if a subscriber intended to purchase the stocked-out item, s/he would do so immediately after receiving the replenishment notification. Therefore, the hypothesis stated that after a short demand spike corresponding to fulfilled subscription orders, the future demand would decline. This was based on observed sales behavior where stocked-out items gathered pent-up demand in the form of subscriptions and upon replenishment, demand for these articles peaked noticeably before rapidly declining.

FIGURE 11: DEMAND PROFILE WITHOUT (LEFT) AND WITH (RIGHT) SUBSCRIPTIONS: Left - Demand profile for an article that did not experience stock-out. Right - Demand profile for an article that experienced stock-out and subscriptions. After notifications, it was observed that demand rapidly responded.

Predicting an increase in sales after subscription period

Models M4 through M6 were developed and tested using the subset of data with non-zero Active Subscriptions. Since this data subset was significantly smaller and with distinct demand profiles versus the full dataset, the models were evaluated versus a new baseline error rate. This baseline was calculated using the baseline model (M2) trained on the relevant data subset (SKU-dates with Active Subscriptions).

- **M4** is a linear regression with two features: the Weighted Average Sales (like M2), and the number of Active Subscriptions for the SKU being forecasted.

$$D_{MCC}^{R} = \beta_1(\omega_1 \mathrm{W}_{\mathrm{SKU},1}^{\mathrm{OB}} + \omega_2 \mathrm{W}_{\mathrm{SKU},2}^{\mathrm{OB}} + \omega_3 \mathrm{W}_{\mathrm{SKU},3}^{\mathrm{OB}}) + \beta_2 S_{MCC} \qquad \textbf{(M4)}$$

- **M5** is a linear regression with two features: the Weighted Average Sales and the number of Active Subscriptions for the SKU being forecasted (like M4). It varies from M4 in that $\beta_{i,w}$ is a vector of coefficients depending on observed Weighted Average Sales.

$$D_{MCC}^{R} = \beta_{1,Vm}(\omega_1 \mathrm{W}_{\mathrm{SKU},1}^{\mathrm{OB}} + \omega_2 \mathrm{W}_{\mathrm{SKU},2}^{\mathrm{OB}} + \omega_3 \mathrm{W}_{\mathrm{SKU},3}^{\mathrm{OB}}) + \beta_2 S_{MCC} \qquad \textbf{(M5)}$$

- **M6** is a linear regression with two features: the clustered Weighted Average Sales and the number of Active Subscriptions for the SKU being forecasted (as in M5). It varies from M5 in that the coefficient for Subscriptions, $\beta_{2,W}$, is also a vector of coefficients depending on observed Weighted Average Sales.

$$D_{MCC}^{R} = \beta_{1,Vm}(\omega_1 \mathrm{W}_{\mathrm{SKU},1}^{\mathrm{OB}} + \omega_2 \mathrm{W}_{\mathrm{SKU},2}^{\mathrm{OB}} + \omega_3 \mathrm{W}_{\mathrm{SKU},3}^{\mathrm{OB}}) + \beta_{2,Vm} S_{MCC} \qquad \textbf{(M6)}$$

As shown in Figure12 below, all three models with Subscriptions included as a regression feature (M4 through M6) performed better than the Baseline forecast method. The best performing model was M6, in which a vector of linear coefficients was fit to each of the two model features (Weighted Sales Average and Subscriptions) based on the value of Weighted Average Sales at the time of the forecast. Table 7 below provides an example output of regression coefficients for each model (real numerical values have been changed to protect data privacy). Based on these results, the double-clustering method for both Weighted Average Sales and Active Subscriptions was selected as the best performing model for predicting a future increase in demand due to subscriptions.

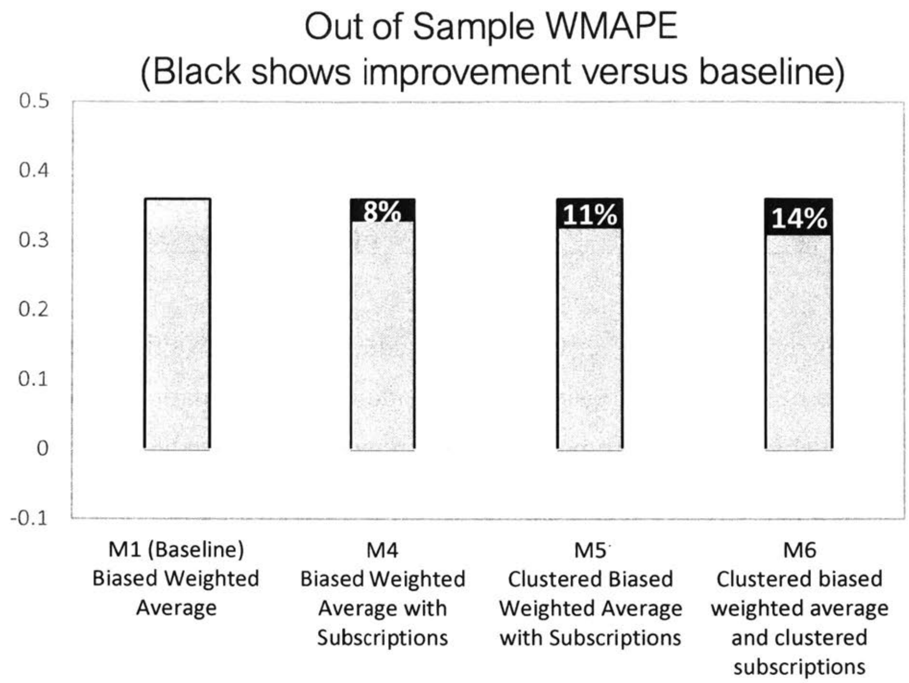

FIGURE 12: SUMMARY OF PERFORMANCE FOR M4, M5, AND M6 VERSUS M1. Models M4, M5, and M6 show significantly better performance versus M0, with M6 performing the best.

Models M4, M5, and M6 all showed significant improvement in out of sample WMAPE versus the baseline method for predicting demand in the presence of a stock-out event. Detailed regression results were compared for each model to determine how these models differ in structure and performance. These results are shown in Tables 9 and 10 below, which compare the regression results for the four models. The regression was performed using Training data for a single geographic region during one full selling season of articles.

Table 9 shows that by including subscriptions significantly changes the regression results. The coefficient value for observed sales decreases by approximately 15%, with a subscriptions coefficient of 0.054. These results suggest that there is not a 1:1 tradeoff between each subscription and each sale, but rather that each subscription is worth on average 7.5% equivalent of a previously observed sale.

TABLE 9 REGRESSION ANALYSIS FOR MODELS M1 AND M4. Regressions were performed using the filtered dataset including only article-dates with active subscriptions.

M1: Baseline					M4: Simple Regression with S				
Var.	β	Std. Error	t-value	Pr(>|t|)	Var.	β	Std. Error	t-value	Pr(>|t|)
W	0.8386	0.004726	177.5	<2e-16	W	0.70905	0.0045616	155.44	<2e-16
					S	0.05378	0.0008882	60.55	<2e-16

Comparing regression results between models M5 and M6 shows that clustering subscriptions based on observed previously sales adds additional information to the model. These results are shown in Table 10 below. In both regression models, all variables are highly significant. Model M6 shows that, like clustering on average sales (W), clustered subscriptions (S) results in coefficients that are both variable in magnitude and highly significant for each cluster. These results suggest that the improvement in WMAPE is due to new and relevant information being gleaned from each cluster.

TABLE 10: REGRESSION ANALYSIS FOR MODELS M5 AND M6. Regressions were performed using the filtered dataset including only article-dates with active subscriptions.

M5: Clustered W Regression with S					M6: Clustered W & S Regression				
Var.	β	Std. Error	t-value	Pr(>|t|)	Var.	β	Std. Error	t-value	Pr(>|t|)
W_clusterA	1.1439	0.213454	5.359	<2e-16	W_clusterA	0.98148	0.22924	4.281	1.9E-05
W_clusterB	1.0283	0.033663	30.548	<2e-16	W_clusterB	0.83147	0.03443	24.147	<2e-16
W_clusterC	0.8304	0.011007	75.454	<2e-16	W_clusterC	0.84522	0.01165	72.571	<2e-16
W_clusterD	0.7716	0.009033	85.42	<2e-16	W_clusterD	0.77191	0.00969	79.59	<2e-16
W_clusterE	0.6611	0.005285	125.09	<2e-16	W_clusterE	0.68796	0.00568	120.664	<2e-16
S	0.0525	0.052449	59.939	<2e-16	S_clusterA	0.07944	0.01568	5.074	4E-07
					S_clusterB	0.09944	0.00253	39.386	<2e-16
					S_clusterC	0.04728	0.00179	26.708	<2e-16
					S_clusterD	0.05232	0.00169	31.046	<2e-16
					S_clusterE	0.04152	0.00137	30.223	<2e-16

The detailed regression analysis is shown in Table 11 below for models M1, M4, M5, and M6. These results were determined using the same set of training data for each model. These results show that R^2, Adjusted R^2, and RMSE all favor M6, in line with the observed out of sample WMAPE results.

TABLE 11: MODEL FIT RESULTS FOR MODELS M1, M4, M5, AND M6

	M1	M4	M5	M6
R^2	0.636831	0.731704	0.731965	0.740949
Adj R^2	0.7635	0.8281	0.8343	0.8411
RMSE	14.78257	12.62834	12.40491	12.22474
OOS WMAPE	0.36	0.33	0.32	0.31

Predicting a decline in sales after fulfilling pent-up demand due to subscriptions

M6 was used as the foundation to continue model development. It best addressed the first goal of incorporating subscriptions into the demand forecast: M6 improved the prediction of an increase in demand immediately following a stock-out. The next phase of model development aimed to address the second goal, which was to predict a decline in sales after fulfilling pent-up demand due to subscriptions.

After investigating the available data, two methods were identified and iterated to address the goal of predicting a decrease in sales immediately after a short-term peak from fulfilled subscriptions. These methods are enumerated in M7 & M8. The approach of M7 was to modify the Days Available for Sale (DPV) such that the pent-up demand peak was spread more evenly over stock-out days with possible subscriptions. The approach of M8 was to discount observed sales if the Sale User ID matched a Subscription User ID for the relevant SKU & forecast period.

- **M7** is a linear regression with two features: "Adjusted" Weighted Average Sales and Active Subscriptions, with intercept of zero. It is similar to M6 but $\widehat{W}$ is calculated as shown, where $\widehat{DPV}_{MCC}$ includes the stock-out time an SKU was available for subscriptions.

$$\widehat{DPV}_{MCC} = \sum_{sizes}\sum_{1:i} n + \sum_{sizes}\sum_{1:i} n_{stockout}$$

$$\widehat{W}_t^{OB} = \frac{\sum v_t^{OB}}{\sum \widehat{DPV}_t}$$

$$D^R_{MCC} = \beta_{1,Vm}(\omega_1 \widehat{\mathrm{W}}^{\mathrm{OB}}_{\mathrm{SKU},1} + \omega_2 \widehat{\mathrm{W}}^{\mathrm{OB}}_{\mathrm{SKU},2} + \omega_3 \widehat{\mathrm{W}}^{\mathrm{OB}}_{\mathrm{SKU},3}) + \beta_{2,Vm} S_{MCC} \qquad \textbf{(M7)}$$

- **M8** is a linear regression with two features: "Adjusted" Weighted Average Sales and Active Subscriptions, with intercept equal to zero. It is similar to M6 but the Weighted Average Sales are calculated with a discounted $\widehat{W}$: only sales to *new users* are included in the regression. Sales that are associated with a User ID that also had a subscription to the SKU being forecasted ($v^{OB}_{t,subscribed}$) are subtracted from the total sales v^{OB}_t resulting in a new, lower $\hat{v}^{OB}_t$.

$$\hat{v}^{OB}_t = v^{OB}_t - v^{OB}_{t,subscribed}$$

$$\widehat{\mathrm{W}}^{\text{no Subs}}_{\mathrm{t}} = \frac{\sum \hat{v}^{OB}_t}{\sum DPV_t}$$

$$D^R_{MCC} = \beta_{1,Vm}(\omega_1 \widehat{\mathrm{W}}^{\text{no Subs}}_{\mathrm{SKU},1} + \omega_2 \widehat{\mathrm{W}}^{\text{no Subs}}_{\mathrm{SKU},2} + \omega_3 \widehat{\mathrm{W}}^{\text{no Subs}}_{\mathrm{SKU},3}) + \beta_{2,Vm} S_{MCC} \qquad \textbf{(M8)}$$

Figure 13 below shows that both M7 and M8 outperformed the Baseline forecast method. However, M7 does not perform better than M6, while M8 performs significantly better than M6. Therefore, M8 was selected as the best-performing model for final analysis and discussion.

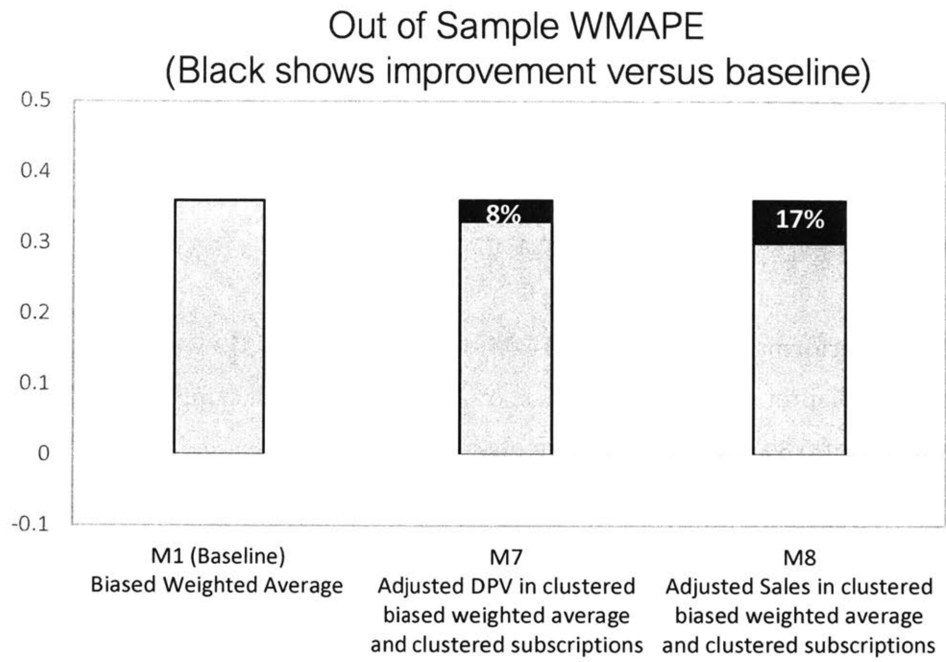

FIGURE 13: SUMMARY OF RESULTS FOR MODELS M7 AND M8 VERSUS M1. M7 shows moderate improvement, while M8 shows the most significant improvement with 16% lower error versus M1.

TABLE 12: REGRESSION ANALYSIS FOR MODEL 8. Regressions were performed using the filtered dataset including only article-dates with active subscriptions.

Feature	β	Std. Error	t-value	Pr(>|t|)
W_clusterA	1.241895	0.200567	6.192	6.18E-10
W_clusterB	0.869783	0.031174	27.901	<2e-16
W_clusterC	0.887257	0.01053	84.257	<2e-16
W_clusterD	0.851446	0.008926	95.388	<2e-16
W_clusterE	0.833282	0.006268	132.951	<2e-16
S_clusterA	0.074598	0.014218	5.247	1.58E-07
S_clusterB	0.100202	0.002314	43.308	<2e-16
S_clusterC	0.049901	0.0016	31.182	<2e-16
S_clusterD	0.052697	0.0015	35.081	<2e-16
S_clusterE	0.045531	0.001214	37.496	<2e-16

TABLE 13: MODEL FIT RESULTS FOR MODEL M8

	M8
R^2	0.7768187
Adj R^2	0.8636
RMSE	11.29964

Based on the relative performance of models M1-M8, models M6 and M8 were selected for further analysis (discussed in Chapter 6). These models show the lowest out of sample WMAPE and very high model fit, with R^2 of 0.84 and 0.86 respectively.

Models 9 – 16: Incorporating New Features

To determine which additional features added the most value to the demand prediction, each feature was tested iteratively versus the model including all features. To achieve the results shown, the data were filtered to only include SKU-dates with *no active subscriptions*. This cleaned the dataset to eliminate demand fluctuations due to stock outs and known pent-up demand. Then, a linear regression was performed using all available training data for all features, as shown in M9. Each subsequent model (M10 – 16) iteratively removes one feature at a time. The error rates and R^2 fit was tested for each case. A feature was considered meaningful if the error rate significantly worsened after the feature was removed from the regression.

As before, Weighted Sales History shall be referred to as $W_{t,MCC}$, where:

$$W_{t,MCC} = \omega_1 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},1} + \omega_2 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},2} + \omega_3 \mathrm{W}^{\mathrm{OB}}_{\mathrm{SKU},3}$$

- **M9** is a linear regression with all features found to be of potential importance, with intercept equal to zero: Weighted Average Sales of the SKU being forecasted, Weighted Average Sales of the Garment Cluster relevant to the SKU being forecasted, Position on the Website on the forecast date, Binary New In status, Article Age, Weighted Average Sales of the SKU across physical stores in the relevant geography to the Zara.com store being forecasted, and the One Day Change in net sales of the SKU and Zara.com store being forecasted.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_4 NewIn + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7\, OneDayChange \quad \textbf{(M9)}$$

- M10 is a linear regression with intercept equal to zero, identical to M9 but with Weighted Average Sales removed.

$$D_{MCC}^{R} = \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_4 NewIn + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7\, OneDayChange \quad \textbf{(M10)}$$

- M11 is a linear regression with intercept equal to zero, identical to M9 but with Garment Cluster Weighted Average Sales removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_3 Position + \beta_4 NewIn + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7\, OneDayChange \quad \textbf{(M11)}$$

- M12 is a linear regression with intercept equal to zero, identical to M9 but with Website Position removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_4 NewIn + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7\, OneDayChange \quad \textbf{(M12)}$$

- M13 is a linear regression with intercept equal to zero, identical to M9 but with the New In Binary indicator removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7\, OneDayChange \quad \textbf{(M13)}$$

- M14 is a linear regression with intercept equal to zero, identical to M9 but with Age removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_4 NewIn + \beta_6 \overline{W}_{t,MCC,physical} + \beta_7 OneDayChange \qquad \textbf{(M14)}$$

- M15 is a linear regression with intercept equal to zero, identical to M9 but with the Weighted Average Sales in Physical Stores removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_4 NewIn + \beta_5 Age + \beta_7 OneDayChange \qquad \textbf{(M15)}$$

- M16 is a linear regression with intercept equal to zero, identical to M9 but with One Day Change removed.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 \overline{W}_{t,Subfamily} + \beta_3 Position + \beta_4 NewIn + \beta_5 Age + \beta_6 \overline{W}_{t,MCC,physical} \qquad \textbf{M16}$$

Table 14 below shows the relative change in two different error metrics for each model M10 – M16 versus M9 where all features were included. Removing Weighted Average Sales (M10), provided a reference value to compare the relative importance of each feature (since it was proved that this was the most valuable feature in predicting demand). The shading indicates the impact each feature had on forecasting accuracy, where dark blue is the highest impact, red is no impact, and yellow is moderate impact.

TABLE 14: FEATURE IMPACT ON FORECAST ACCURACY: Each website feature was tested iteratively to determine which had highest predictive power for estimating future demand (blue = most impactful, yellow = some impact, red = zero impact).

Model	Feature Tested	Percent Change RMSE	Percent Change OOS WMAPE
M10	$W_{t,MCC}$	86%	165%
M11	$\bar{W}_{t,Subfamily}$	0%	0%
M12	$Position$	0%	0%
M13	$NewIn$	0%	0%
M14	Age	0.5%	0.7%
M15	$\bar{W}_{t,MCC,physical}$	0.4%	0.7%
M16	$OneDayChange$	2.5%	6%

Models 17 – 20: Age Feature Transformation

Previous research has demonstrated that logarithmic or declining exponential curves may more accurately describe retail demand patterns versus a linear regression [24]. The poor prediction accuracy observed with model M2 (logarithmic transformation of rolling Average Weighted Sales) suggests that a different feature should be used to model the expected decreasing exponential demand behavior.

After investigating the features available in the datasets, the time variable *Article Age* was selected to further explore exponential transformation models. To achieve the results shown, the data were filtered to only include SKU-dates with *no active subscriptions*.

As previously, rolling Weighted Sales History shall be referred to as $W_{t,MCC}$, where:

$$W_{t,MCC} = \omega_1 \mathrm{W}_{\mathrm{SKU},1}^{\mathrm{OB}} + \omega_2 \mathrm{W}_{\mathrm{SKU},2}^{\mathrm{OB}} + \omega_3 \mathrm{W}_{\mathrm{SKU},3}^{\mathrm{OB}}$$

- M17 is a linear regression with intercept equal to zero and two features: Weighted Average Sales ($W_{t,MCC}$), and the exponential of negative Article Age, in days.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 e^{-Age,days} \qquad \textbf{(M17)}$$

- M18 is a linear regression with intercept equal to zero and two features: Weighted Average Sales ($W_{t,MCC}$), and the exponential of negative Article Age, in weeks.

$$D_{MCC}^{R} = \beta_1 W_{t,MCC} + \beta_2 e^{-Age,weeks} \qquad \textbf{M18}$$

- M19 is a linear regression with intercept equal to zero and two features: logarithmic transformation of Weighted Average Sales ($W_{t,MCC}{}'$), and the exponential of negative Article Age, in weeks.

$$D_{MCC}^{R}{}' = D_{MCC}^{R} + 1$$

$$W_{MCC}^{OB}{}' = W_{SKU}^{OB} + 1$$

$$\ln(D_{MCC}^{R}{}') = \beta_1 \ln W_{t,MCC}{}' + \beta_2 e^{-Age,weeks} \qquad \textbf{M19}$$

- M20 is a linear regression with intercept equal to zero and one features: clustered Weighted Average Sales (like M2). It varies from M2 in that $\beta_{1,Age}$ is a vector of coefficients depending on Article Age, in weeks.

$$D_{MCC}^{R} = \beta_{1,Age,weeks} W_{t,MCC} \qquad \textbf{M20}$$

Table 15 below shows the regression results for each of models M1 and M18-M20. These results indicate that in model M17, the feature for age in terms of days does not explain the variation in the predicted demand. The regression for model M18 shows that the feature for age in terms of weeks is significant – these findings suggest that demand is not predictable on a day-by-day basis, but rather it is more predictable when aggregated in a week-by-week term. It is worth noting that M19 is a logarithmic transformation, so the coefficient and error values are not directly comparable to the other linear models.

TABLE 15: REGRESSION RESULTS FOR MODELS INCORPORATING AGE AS A PREDICTION FEATURE FOR DEMAND. Analyses performed using only articles that never experienced a stock out or subscription event. Models M1 and M17-20

Model	Coefficient	β	Std. Error	t-value	Pr(>|t|)	
M1	W	0.84437	0.00104	812.1	<2e-16	
M17	W	0.844632	0.001063	794.553	<2e-16	
	e^(-days)	-27.368901	23.024599	-1.189	0.235	*not significant*
M18	W	0.84912	0.00115	738.284	<2e-16	
	e^(-weeks)	-0.44612	0.0429	-9.637	<2e-16	
M19	ln_W	0.9425918	0.007015	1343.61	<2e-16	
	e^(-weeks)	-0.1404772	0.0061679	-22.77	<2e-16	
M20	W_ageClusterA	0.75595	0.001962	406	<2e-16	
	W_ageClusterB	0.854733	0.001681	508.4	<2e-16	
	W_ageClusterC	0.920893	0.001866	493.5	<2e-16	
	W_ageClusterD	0.8397	0.005847	143.6	<2e-16	

Table 16 shows the regression analyses of error using the same training data for each of the four models shown in Table 15 above. The WMAPE results were calculated using prediction error on out-of-sample test data, to determine how each model would perform on un-seen data. These results show that each of the four models performs very similarly. Models M19 and M20 demonstrate approximately 2% improvement versus the baseline. This improvement would be of significant importance to Zara, who forecasts demand for thousands of items each week. Given that model M20 follows the existing structure of typical demand forecasts in use at Zara, it was selected as preferable for implementation versus model M19.

In comparing the performance of the four models shown below in Table 16, it is worth noting that model M19 cannot be directly compared versus the other four linear models. The R^2, Adjusted R^2, and RMSE are all calculated on the training data against the transformed prediction variables. The OOS WMAPE however is translatable, as this metric is calculated versus the true Demand (rather than logarithmic transformation of demand). These results are depicted graphically in Figure 14 below.

TABLE 16: MODEL FIT RESULTS FOR MODELS M1, M17, M18, M19, AND M20. Results are shown for linear regression using training data which had no active subscriptions in the location-dates studied.

	M1	M17	M18	M19	M20
R^2	0.7679228	0.7679295	0.7688029	0.8048461	0.7737179
Adj R^2	0.8575	0.8575	0.8576	0.9549	0.8625
RMSE	2.559233	2.559308	2.556733	0.3297249	2.513514
OOS WMAPE	30.099%	30.099%	30.068%	29.642%	29.381%

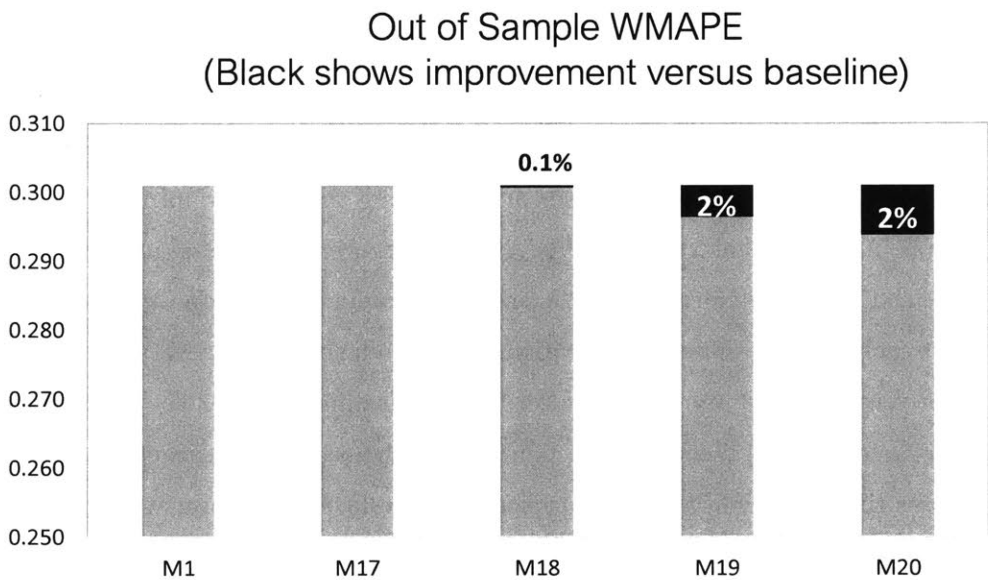

FIGURE 14: SUMMARY OF RESULTS FOR MODELS M17-M20 VERSUS M1. Models M17 and M18 show little to no improvement versus the baseline method, while models M19 and M20 show 2% improvement in WMAPE.

Methodologies without prediction improvement

Exploratory modelling lead to several failed attempts to predict e-commerce demand. The following models offer examples into strategies investigated which did not improve the forecast versus the baseline method. This list is not comprehensive, but instead is provided to illustrate the types of techniques studied.

- **Binary prediction: will sales increase versus last week?** Binary prediction methods using logistic regressions were studied with the goal of improving the ability to predict if future sales were likely to increase or decrease relative to the observed sales history. However, this method was unable to reliably distinguish between demand swings, and ultimately the binary indicator (increase vs. decrease) did not yield prediction improvements.

- **Probabilistic prediction: what is the probability that sales will increase versus last week?** Like the binary prediction method above, probabilistic methods using logistic regression were studied with the goal of improving the ability to predict if future sales were likely to increase or decrease relative to the observed sales history. Once again, this method was unreliable in distinguishing between demand swings and did not yield prediction improvements.
- **Binary prediction: will article be featured in New In category next week?** Binary logistic predictions were evaluated to determine if an article would be featured in the New In category. This method was evaluated under the hypothesis that articles in New In correlated to higher e-commerce demand. However, this thesis showed that New In is not a highly important input for forecasting e-commerce demand. Therefore, by the transitive property, it was determined that a binary indicator for whether an article would remain in New In would not improve the demand prediction.
- **Probabilistic prediction: what is the probability of a stock out next week?** Predicting a stock out is the result of two inputs: the demand prediction and the stock level. Therefore, using e-commerce data to predict the probability of a stock out is more complex than the primary goal of this thesis: to improve the accuracy of the demand forecast itself.
- **Linear regression: given probability of {stock out, New In, Demand Increase} what is estimated demand next week?** The methods discussed above were investigated and incorporated into linear models for e-commerce demand prediction. These results did not yield prediction improvements, which can be explained by the discussions above.

The unsuccessful modelling techniques are examples of creative uses of available data. Given the confines of this research to leverage statistical modelling tools, these methods yielded poor prediction accuracy. Further research should investigate if these methods can be incorporated into a machine learning model which addresses these additional features.

5.3 Final Results

Preliminary results of the 16 models described in Chapters 5.1 and 5.2 led the investigation to focus on two new methods: first, to evaluate models M6 and M8 for incorporating subscriptions information; second, to evaluate Age Clustering for articles without subscriptions.

In addition to WMAPE and R^2 across the entire out-of-sample dataset for each model, the clustered error and regression statistics were compared for models 6 and 8 versus the baseline. These results are shown in table 17 below. These results demonstrate that the t-value and P-value show high significance for the observed sales feature (W) and the subscriptions feature (S) for each cluster.

To meet the implementation requirements, the new models needed to demonstrate the business requirement of *robustness.* This requirement was met through two statistical methods in the context of this project: First, the new model(s) had to show improved performance accuracy for all items, and not just popular items with high sales. Therefore, the new model(s) were tested versus baseline in all clusters (A, B, C, D, E) by average sales. Second, the new model(s) were required to show improved performance in near-outlier cases. Therefore, the 90^{th} percentile error metric was tested for model performance versus baseline.

Figures 15 and 16 below shows the average error (Figure 15) and 90^{th} percentile error (Figure 16) of M6 and M8 in each cluster versus the baseline. From this analysis, it was observed that only the high W clusters demonstrated robust improvement with M8. Conversely, the low W clusters showed robust improvement with M6 versus the baseline. Intuitively, this outcome was in line with the two hypotheses regarding demand behavior after a stock-out. The first hypothesis stated that during a stock-out consumer purchases were censored, resulting in pent-up demand through subscriptions and the eventual rapid increase in purchases immediately after replenishment. This behavior was well predicted by adding Active Subscriptions as a feature in linear regression and applied to all clusters. The second hypothesis stated that after replenishment, the baseline prediction over-estimated true demand (due to the decline in sales after subscriptions were fulfilled through conversion). The demand decline was well predicted by removing Converted Subscriptions from the average sales feature (W). This model (M8) improved performance for clusters of articles with high average sales (high W) but did not improve performance for clusters of articles with low average sales.

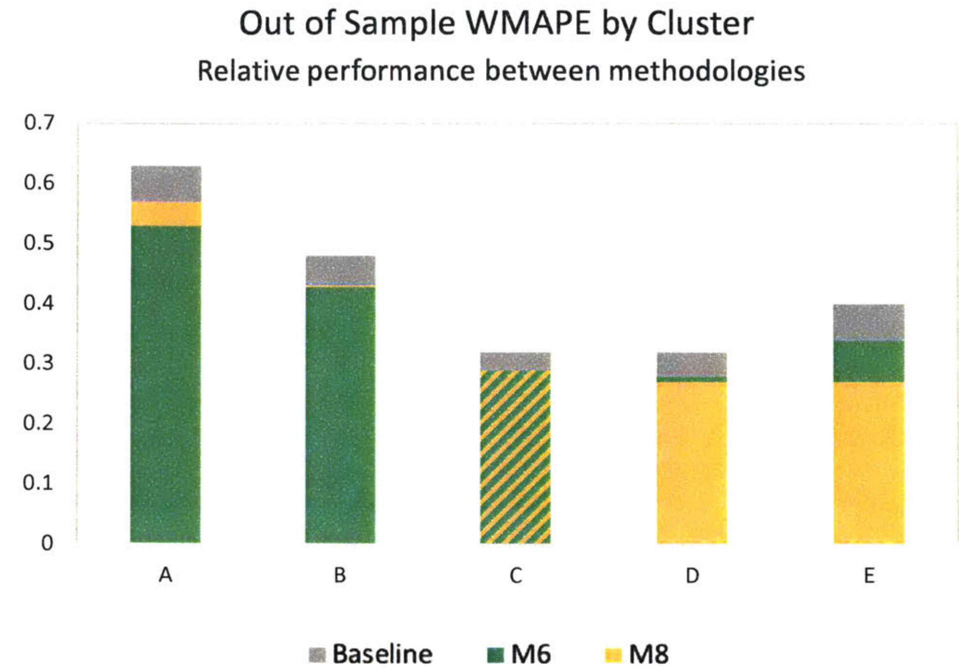

FIGURE 15: M6 AND M8 AVERAGE CLUSTERED ERROR (VERSUS M1): Model with lowest error is shown in green (M6) or yellow (M8). Marginal error from the other methodology and/or Baseline method is stacked.

Table 17 below shows the relative improvement percentage of each method M6 and M8 versus the baseline WMAPE for each cluster. These results suggest that the optimal methodology is to use decision tree logic to use model M6 for clusters A and B, and model M8 for clusters D and E. Cluster C requires further analysis.

TABLE 17: PERCENT IMPROVEMENT AVERAGE WMAPE BY CLUSTER WITH SUBSCRIPTIONS

Cluster	M6	M8
A	16%	10%
B	11%	10%
C	9%	9%
D	13%	16%
E	15%	33%

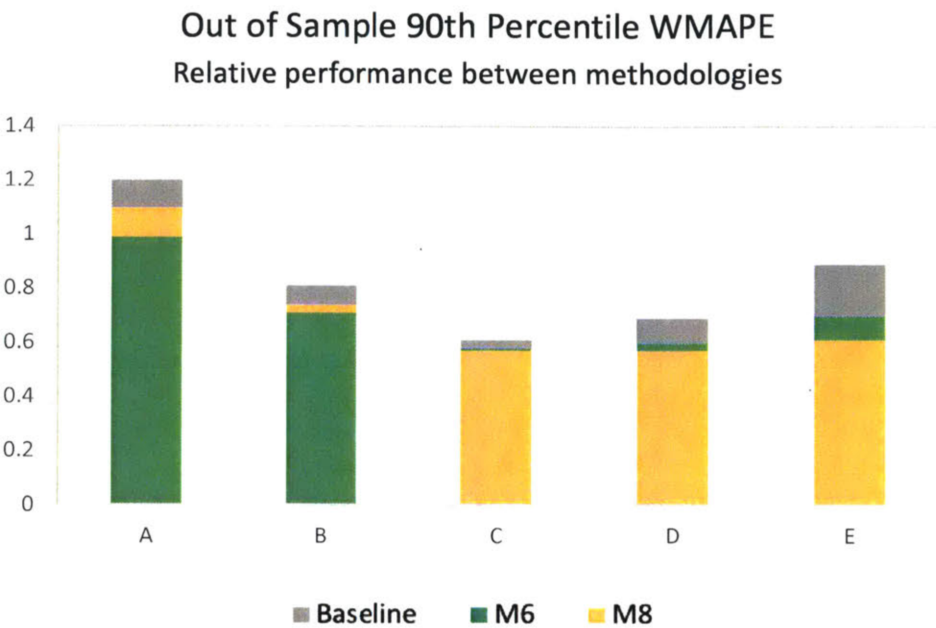

FIGURE 16: M6 AND M8 90TH PERCENTILE CLUSTERED ERROR (VERSUS M1): Model with lowest 90th percentile error is shown in green (M6) or yellow (M8). Marginal error from the other methodology and/or Baseline method is stacked.

Table 18 below shows the relative improvement percentage of the 90th percentile (P90) WMAPE for each method M6 and M8 versus the baseline P90 WMAPE for each cluster. These results are in line with the average WMAPE results shown above. Clusters A and B are best predicted by M6 and clusters C and D are best predicted by M8. Furthermore, these results indicate that cluster C, which showed equivalent *average* WMAPE improvement demonstrates better P90 WMAPE when predicted by M8.

TABLE 18: PERCENTAGE IMPROVEMENT 90TH PERCENTILE WMAPE BY CLUSTER WITH SUBSCRIPTIONS

Cluster	M6	M8
A	18%	8%
B	12%	9%
C	5%	7%
D	13%	17%
E	21%	31%

To summarize these findings, Table 19 presents the final methodology and model selection forecasting future demand for each cluster for articles that experience censored demand during a stock out.

TABLE 19: MODEL TYPE BY CLUSTER: Based on the clustered error analysis (shown above), the following implementation plan was recommended to update the Zara.com forecast models. Model M6 was recommended for clusters A & B, and Model M8 was recommended for clusters C, D, and E.

Cluster	Model	Methodology
A	M6	Add Subscriptions feature
B	M6	Add Subscriptions feature
C	M8	Add Subscriptions feature & adjust W for conversions
D	M8	Add Subscriptions feature & adjust W for conversions
E	M8	Add Subscriptions feature & adjust W for conversions

The robustness analysis for the new alternative website model features (age, physical stores, and one day change) demonstrated inadequate results for immediate implementation. While physical store sales and one day change initially showed high correlation to future sales, these features also showed high variability across articles. No distinct article characteristics were identified to indicate specific use cases that would make these features relevant on a case-by-case basis. Age showed more promising results.

Tables 20 and 21 below show the clustered percentage improvement versus the baseline prediction method. These results contrast two forecasting methods versus the baseline biased weighted sales average (M1). Given the improvements observed with M3, above, "W Cluster" is used as a benchmarking method to see if clustering on Age yields any improvement versus clustering on observed average sales history.

The results tables below show clustered percentage improvement for Average WMAPE by cluster (Table 20) and 90th Percentile WMAPE percentage improvement (Table 21). In both cases, the Age Clustering method yields significantly better results. In the case of the 90th percentile WMAPE, method M3 actually performs worse than the baseline method. It is worth noting that the majority improvement is observed in Cluster A, which yields 6.5% average improvement in WMAPE and 7% improvement in 90th percentile WMAPE. Based on these results, it is recommended that Zara investigate incorporating the age feature for cluster A (articles that have been available for a short period of time).

TABLE 20: PERCENTAGE IMPROVEMENT IN AVERAGE WMAPE BY CLUSTER WITH AGE. Results shown for methods M3 and M20 versus the Baseline method

Cluster	W Cluster (M3)	Age Cluster (M20)
A	3.2%	6.5%
B	0.0%	0.0%
C	0.0%	1.3%
D	0.0%	0.0%

TABLE 21: PERCENTAGE IMPROVEMENT IN P90 WMAPE BY CLUSTER WITH AGE. Results shown for methods M3 and M20 versus the Baseline method.

Cluster	W Cluster (M3)	Age Cluster (M20)
A	0%	7%
B	2%	2%
C	3%	3%
D	-1%	0%

5.4 Discussion

Two theories exist for explaining the improved performance versus baseline using models M6 and M8 for predicting demand after a stock-out. The first theory one could plausibly believe is that the new models are accurately incorporating new information to better forecast future demand. This is the rationale that is explained in the discussion above.

The second theory that could explain the improved performance is that these models demonstrate regression toward the mean – a well-documented outcome of statistical variation in population[25]. When this theory is applied to this study, it suggests that that articles in Clusters A & B have such low average sales (W) that they below the total population mean and are more likely to increase in future demand. This is simply due to statistical probability that they will incline toward the mean average sales. Model M6 accurately predicts this behavior by including Active Subscriptions, thereby increasing the prediction for future sales. Similarly, popular items with high average sales are above the population mean (e.g. Clusters D & E) and are likely to decrease in future demand, since statistically they will decline toward the mean average sales. Model M8 accurately predicts this behavior because items with high average sales (W) are predicted to decline in the future. Combining the two models, such that M6 predicts low-sales clusters and M8 predicts high-sales clusters, effectively creates a strategy where future predictions are normalized to converge toward the population mean average sales.

Regardless of the underlying theory to the new models, the results are evident. The new combined M6-M8 model approach with article clustering based on average sales (W) demonstrates 16% performance improvement in out-of-sample average error, for the geographies and seasons studied. Based on these findings, it was recommended that Zara implement this strategy in a test geography to further evaluate the possible improvement in demand forecasting.

This page intentionally left blank.

Chapter 6

Conclusion and Future Work

This research thesis investigates linear and logistic regression models for fashion retailers with the goal of improving e-commerce demand forecast using novel web features. It presents article-and-location specific forecasting methods that are implementable, communicable, scalable, and robust. By incorporating web data such as user subscriptions and article age, this research demonstrates that the models identified achieve as much as 16% WMAPE improvement versus baseline forecasting methods. Additional web features were identified and researched as possible inputs to the e-commerce demand forecast models.

6.1 Conclusion

The research presented discusses techniques to improve the demand forecast by leveraging the novel e-commerce data that has become available since the launch of Zara.com. This dataset was used as a case study to develop methods that may be used in other retail settings. The four objectives of this research were demonstrated using the Zara dataset:

1. Statistical models were used to develop simple and robust forecasting techniques that rely only on readily available e-commerce data.
2. Opt-in user tracking data (subscriptions) were incorporated to forecast censored demand of articles during a stock out.
3. Article Age was investigated as a method to predict article sales over time and was incorporated to forecast e-commerce demand.
4. Correlation and prediction error studies were performed to evaluate the relative importance of novel web features. This research narrowed the scope of web data that should be included in the improved statistical models for e-commerce demand forecasting.

The subscription and censored demand research yielded the highest versus the baseline forecast. This can be explained by two trends which are embedded in the corresponding subscription and demand data.

- Information about pent-up demand during a stock out (number of subscriptions) can be used to forecast an upcoming demand spike immediately after replenishment.

- Information about pent-up demand that later converts to fulfilled sales after replenishment can be used to forecast an upcoming decline in sales.

In addition to improving demand forecast accuracy for articles experiencing a stock out, this thesis identified a new method to improve the prediction accuracy for all articles. This method incorporates the Article Age as a novel feature for e-commerce demand prediction. Article Age is unique to online sales, where an article is introduced on the same date for an entire market. This contrasts physical store sales, where a new article may be introduced on different dates depending on each store location. The Article Age information yields a small improvement over a large number of articles – a finding that may be worth pursuing for a large fashion retailer with thousands of articles.

6.2 Future Work

Future investigations into improvements for user-friendly and implementable online sales forecasting may consider advancing to higher-order technology. This thesis compared several mathematical tools available to the sales forecaster: Traditional, Artificial Intelligence, and Hybrid approaches. As retail technology advances, logistics and distribution technology may be able to leverage the wide availability of data and computing power. These methods may enable faster response to changes in demand, lower prediction error, and less-intensive model maintenance over time.

The first benefit of introducing higher order technology is that the retail fashion demand forecaster may be able to incorporate a wider body of available data. Given the limitations of training statistical models, this research narrowed the scope of features to incorporate into the demand models. We compared the performance of each feature and selected only those features which demonstrated the most improvement in demand forecasting. However, with advanced tools, the computational power would be sufficient to include the entire dataset. This reduces the risk of leaving out highly important features.

The second benefit of introducing advanced modelling methods is that demand models do not require human updating. The baseline methods require continual monitoring and evaluation over time, since they include fixed constants and coefficients. Hybrid and AI approaches may update these coefficients automatically, reducing the maintenance workload for the forecasting team.

Future research into demand forecasting with censored demand may further evaluate the sensitivity of logistic models for predicting demand swings. This research investigated the logistic prediction across all items with censored demand. However, it is predicted that first clustering by similar articles may improve the overall prediction for relative increases and decreases in future demand.

This page intentionally left blank.

Sources Cited

[1] "How We Do Business," *Inditex.* [Online]. Available: www.inditex.com.
[2] "Annual Report 2010," Inditex, A Coruña, Spain, Annual Shareholder Report, Jun. 2011.
[3] I. Amed, A. Berg, L. Brantberg, and S. Hedrich, "The state of fashion | McKinsey & Company." [Online]. Available: https://www.mckinsey.com/industries/retail/our-insights/the-state-of-fashion. [Accessed: 27-Feb-2018].
[4] Gérard P. Cachon and Robert Swinney, "The Value of Fast Fashion: Quick Response, Enhanced Design, and Strategic Consumer Behavior," *Manag. Sci.*, no. 4, p. 778, 2011.
[5] "Annual Report 2016," A Coruña, Spain, Annual Shareholder Report, Jun. 2017.
[6] "Women's New In Clothes | New Collection Online | ZARA United States," *Zara.com.* [Online]. Available: https://www.zara.com/us/en/woman-new-in-l1180.html?v1=805003. [Accessed: 12-Apr-2018].
[7] J. Tennent, *Guide to Business Modelling. John Tennent and Graham Friend*, Main edition. London: Economist Books, 2011.
[8] G. B. Sproles, "Analyzing Fashion Life Cycles—Principles and Perspectives," *J. Mark.*, vol. 45, no. 4, pp. 116–124, Fall 1981.
[9] "Our story - inditex.com." [Online]. Available: https://www.inditex.com/en/about-us/our-story. [Accessed: 10-Feb-2018].
[10] F. Caro and V. Martínez-de-Albéniz, "Fast Fashion: Business Model Overview and Research Opportunities," in *Retail Supply Chain Management*, Springer, Boston, MA, 2015, pp. 237–264.
[11] "UNIQLO Business Model | FAST RETAILING CO., LTD." [Online]. Available: http://www.fastretailing.com/eng/group/strategy/uniqlobusiness.html. [Accessed: 16-Apr-2018].
[12] M. L. Fisher, J. H. Hammond, W. R. Obermeyer, and A. Raman, "Coping with Demand Uncertainty at Sport Obermeyer," *Harv. Bus. Rev.*, vol. 72, no. 3, pp. 90–90, Jun. 1994.
[13] B. D. Williams and M. A. Waller, "Top-Down Versus Bottom-Up Demand Forecasts: The Value of Shared Point-of-Sale Data in the Retail Supply Chain," *J. Bus. Logist.*, vol. 32, no. 1, pp. 17–26, Mar. 2011.
[14] Na Liu, Shuyun Ren, Tsan-Ming Choi, Chi-Leung Hui, and Sau-Fun Ng, "Sales Forecasting for Fashion Retailing Service Industry: A Review," *Math. Probl. Eng.*, p. 738675, Jan. 2013.
[15] S. Mou s. mou@auckland. ac. n., D. J. . Robb d. robb@auckland. ac. n., and N. DeHoratius nicole. dehoratius@ChicagoBooth. ed., "Retail store operations: Literature review and research directions," *Eur. J. Oper. Res.*, vol. 265, no. 2, pp. 399–422, Mar. 2018.
[16] J. Coles, "Cross-Channel Predictive Analytics for Retail Distribution Decisions," Masters, Massachusetts Institute of Technology, Cambridge, MA, 2017.
[17] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 6 edition. Boston: Pearson, 2015.
[18] T. Levitt, "Exploit the Product Life Cycle," *Harvard Business Review*, 01-Nov-1965. [Online]. Available: https://hbr.org/1965/11/exploit-the-product-life-cycle. [Accessed: 12-Apr-2018].
[19] D. Ivanov, A. Tsipoulanidis, and J. Schönberger, "Inventory Management," in *Global Supply Chain and Operations Management*, Springer, Cham, 2017, pp. 345–388.
[20] D. Ivanov, A. Tsipoulanidis, and J. Schönberger, "Demand Forecasting," in *Global Supply Chain and Operations Management*, Springer, Cham, 2017, pp. 301–315.
[21] D. Ivanov, A. Tsipoulanidis, and J. Schönberger, "Distribution and Transportation Network Design," in *Global Supply Chain and Operations Management*, Springer, Cham, 2017, pp. 189–232.
[22] J. C. F. (John C. F. Tengberg, "Inter-organizational information sharing of customer data in retail," Thesis, Massachusetts Institute of Technology, 2013.
[23] U. Ramanathan and L. Muyldermans, "Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK," *Int. J. Prod. Econ.*, vol. 128, pp. 538–545, Jan. 2010.
[24] A. Garro, "New Product Demand Forecasting and Distribution Optimization: A Case Study at Zara," Masters, Massachusetts Institute of Technology, Cambridge, MA, 2011.
[25] B. S. Everitt, *The Cambridge Dictionary of Statistics by B. S. Everitt.* Cambridge University Press, 1886.