Adaptability of Manufacturing Operations through Digital Twins

by

Maria Fernanda Reyes Castillo

Bachelor of Science, Industrial Engineering, Universidad Tecnológica Centroamericana, Honduras, 2013

and

Sachin Kumar Garg

Master of Business Administration, Indian Institute of Management, Lucknow, India, 2019

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Signature of Author: _____

<div align="right">

Department of Supply Chain Management
May 14, 2021

</div>

Signature of Author: _____

<div align="right">

Department of Supply Chain Management
May 14, 2021

</div>

Certified by: _____

<div align="right">

Özden Tozanli Yilmaz
Postdoctoral Associate
Capstone Advisor

</div>

Accepted by: _____

<div align="right">

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

</div>

Adaptability of Manufacturing Operations through Digital Twins

by

Maria Fernanda Reyes Castillo

and

Sachin Kumar Garg

Submitted to the Program in Supply Chain Management
on May 14, 2021 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Manufacturing companies are under pressure to build faster and more efficient decision-making capabilities due to the rapidly changing customer demand and expectations. The conventional analytical models are no longer sufficient to capture the complexities of the supply chain. Companies are looking to embark in a digital transformation to address these challenges. One of the digital technologies that offer manufacturers a way to navigate this journey is digital twins, a virtual replica of an object, process or system. Our project focused on studying how digital twins can react to a complex and dynamic environment to create an adaptive mechanism and how can digital twins add value to increase operational efficiency. To answer these questions, we created a conceptual framework of digital twin, AI model and developed a learning feedback loop between simulation and artificial intelligence algorithm. We modeled the supply chain network by using data from a beverage industry and created what-if scenarios that involved varying customer demand and lead time through discrete-event simulation. The output of the simulation was fed into the AI algorithm. The AI prediction was simulated again and results were analyzed. Our research provides insights and discover value associated with adopting these technologies for better decision making. Our recommendation from this study will help supply chain managers understand that a digital twin and AI model framework can be developed, and can be utilized to foresee patterns in supply chain, and proactively take actions to resolve any bottlenecks and constraints.

Capstone Advisor: Özden Tozanli Yilmaz
Title: Postdoctoral Associate

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Manufacturing companies across the globe face challenges in reacting to rapidly changing customer demand and expectations while achieving efficiencies in their supply chains. Many companies today are extensively broadening their global network of suppliers, distribution centers, warehouses, and plants, which continuously interact with each other to satisfy customer needs in the most effective and efficient way. This ongoing transformation of global markets continually contributes to the expanding complexity in managerial and operational layers along supply chain networks.

However, conventional analytical models are no longer sufficient to capture all the interactions between different stakeholders of such complex supply chains (Long et al., 2011). To overcome this challenge, companies now are under pressure to build faster and more efficient decision-making capabilities in more customer-centric ecosystems. This problem is being faced by many industries including BeverageCo, the sponsor company. BeverageCo, is a multi-category beverage leader of the fast-moving consumer goods (FMCG) sector and is under pressure to build a more efficient and agile supply chain.

BeverageCo serves 290 million people across 10 countries in Latin America, through 52 plants, 263 production lines (see Figure 1) and 297 distribution centers. BeverageCo plants have a varying production capacity depending on their vicinity to customers and geography. The company has a very rich portfolio of products, ranging from low to no-sugar carbonated beverages, juices, purified water and energy drinks.

BeverageCo's strategic framework is to transform its operating models through the deployment of world class supply chain solutions across its operations, enabled by digital processes, technology, and innovation. For this reason, the company is creating a sustainable competitive advantage, supported by leading edge strategic capabilities across its value chain.

Figure 1:

BeverageCo - Plants and Production lines per Country
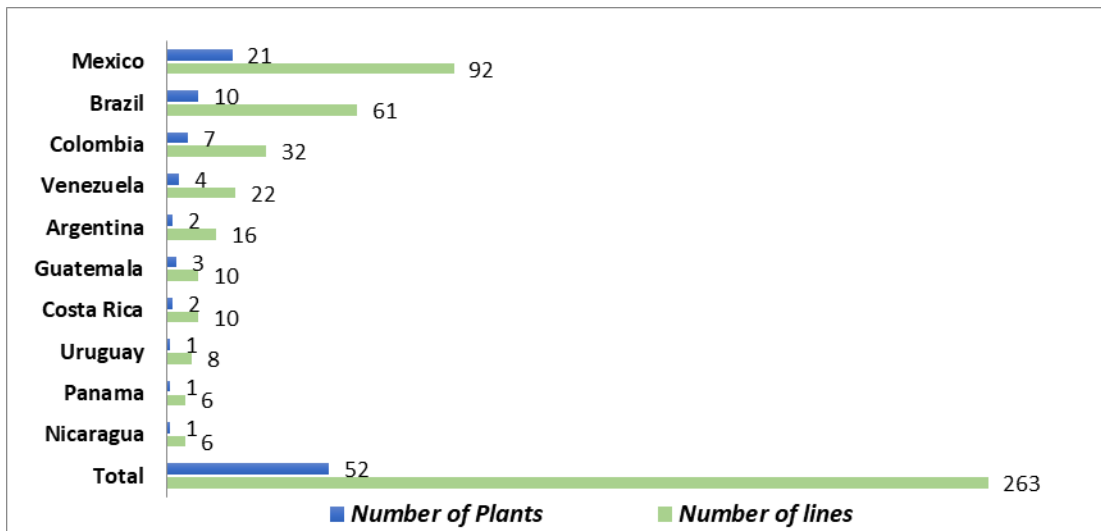
Although BeverageCo has started to transform its manufacturing and supply chain strategy, the company wants to explore digital technologies that can help them navigate through their digital transformation journey. Nevertheless, digital transformation is not just about introducing digital technologies such as big data, internet of things, artificial intelligence, and digital twins (DTs). It must consider evaluating the added value of these technologies to the operations of the business. One of the digital technologies that BeverageCo now is interested in exploring the added value are digital twins.

DTs are virtual replicas of physical objects, processes or systems (Purdy et al., 2020). DTs are considered to be living models of their physical representations (Liu et al., 2018). These models are updated with ubiquitous flow of data collected through a network of high-quality sensors such Internet-of-Things (IoT) devices. This capability helps DTs constantly adapt to the changes in the environment based on the interchange of information from the virtual model to the physical counterpart and vice-versa.

In terms of the supply chain (SC), a digital SC twin can be defined as "a model that can represent the network state for any given moment in time and allow for complete end-to-end SC visibility to improve

resilience and test contingency plans" (Ivanov et al., 2019). This model provides insights from the dynamics of the real-world systems that they emulate in a simulated environment.

This project evaluates a DT model for BeverageCo to examine the behavior of the company's manufacturing and supply chain operations and understand its value-added. We aim to explore patterns of adaptability of the supply chain network by studying the plant and distribution center interactions to improve visibility and value sharing. The plant and distribution center interactions were studied by varying demand and expected customer lead-times.

The research questions of this project are:

1. How can digital twins react to the complex environment and create a dynamic adaptive mechanism in manufacturing operations along SCs?
2. How can digital twins help increase operational efficiency and value sharing in a supply chain setting?

To answer these research questions, this study investigated a data-driven learning system. This learning system was achieved by developing a feedback loop between a supply chain network simulation and an artificial intelligence model to predict the patterns of adaptability under four different scenarios by varying demand and customer expected lead-time values.

Customer demands and expectations, on the other hand, are the sources of complexity in the supply chain. To better understand the complex and adaptive dynamics of the supply chain under the changing circumstances, we studied a complex adaptive system (CAS) framework and reviewed the relevant literature. Focusing on this, CAS has the property of adaptation, meaning that it has the "ability to consciously alter its system configuration and influence its current and future survival" (McCarthy, 2003). Given these characteristics of adaptiveness and found in environments with many interactions, supply chain networks can be considered CAS (Pathak et al., 2007).

Inspired by the concept of CAS, allowed to perform a data analysis of the line, plant and country level components of BeverageCo's manufacturing operations. The data analysis was made using a one-year scorecard data from BeverageCo's enterprise resource planning (ERP) system. This analysis helped in understanding and conceptualizing BeverageCo's supply chain digital twin.

The hypothesis of this project was that digital twins can be used to develop a feedback loop between simulation and AI resulting in a learning system to understand patterns for the adaptability of manufacturing operations and the whole supply chain. The simulation model was used to generate inventory levels per distribution center by varying demand and customer lead-time values. The inventory levels were used to train the AI model. The AI model learned the patterns of inventory levels at each distribution center and provided predictions of those inventory levels. This created a feedback loop between the simulation model and AI algorithm. The interactions between distribution centers to satisfy customer demand were observed providing useful information for better decision making.

Although the project makes recommendations for the value-added in implementing digital twins in this specific beverage company, the study provides a practical reference for future researchers and practitioners in examining the potential value of utilization of DTs in the FMCG industries by filling a gap in the literature. The conceptual framework and recommendations can be applied to similar interactions between individual components of a supply chain.

This document is organized in the following sections: An Introduction was provided in Section 1. Section 2 gives a comprehensive literature review on complex adaptive systems, digital twins, and artificial intelligence. Section 3 describes the project's methodology. In Section 4, the results of simulation and AI are explained. Based on these findings, insights and managerial implications are presented in Section 5. Finally, Section 6 provides concluding remarks and recommendations and points to future research directions.

2 LITERATURE REVIEW

This chapter covers the literature review on Complex Adaptive Systems (CAS) and Digital Twins (DTs).

2.1 Complex Adaptive Systems (CAS)

A system is an integrated whole that consists of interconnected components (Merali & Allen, 2011). Throughout many years and inputs from different disciplines, the study of systems has become known as complex system theory, or some call it, complex theory, or complexity science (McCarthy, 2003). Complex systems theory models and provides understanding of the behavior of systems that possess: 1) a configuration made up of large number of elements 2) significant interactions among these elements and 3) organization in the system (McCarthy, 2003).

A special type of complex system is identified as a complex adaptive system (CAS). Holland (2006) defines CAS as a system that has many components, referred to as agents, that adapt or learn as they interact with each other. CAS exhibits four key characteristics:

1. Parallelism: Continuous and simultaneous signal emissions between large numbers of agents (Holland, 2006). The dynamic system is constantly exchanging information and energy with its environment (Nilsson & Darley, 2006).

2. Conditional Action: Action of the agents in CAS depend on the signals they receive (Holland, 2006). Agents act in correlation and interdependence according to certain policies that influence their behavior (Nilsson & Darley, 2006).

3. Emergence: Interactions among agents produce a collective behavior or pattern that the agents cannot create individually (Nilsson & Darley, 2006). The aggregate behavior of CAS emerges from the interactions of its components (agents). When seeking to adapt, they develop rules that anticipate the consequences of certain responses (Holland, 1992).

4. Adaptation and coevolution: Agents change as time goes by due to adaptations that improve performance, rather than due to random alterations (Nilsson & Darley, 2006). The ability to adapt and learn is the fundamental property of CAS (Holland, 1992).

Many difficult contemporary problems from various fields center on CAS; these include encouraging innovation in dynamic economies, preserving ecosystems, predicting changes in global trade, controlling the internet, strengthening the immune system, and understanding markets (Holland, 2006).

2.1.1 Complex Adaptive Systems in Supply Chain Management

In 2001, Thomas Choi first introduced the idea of treating supply networks as CAS. A CAS is a network composed of various interconnected entities (or agents) that display adaptive characteristics in response to changes in the interactions of agents and in the environment (Choi et al., 2001). In 2005, Amit Surana also explored the idea of supply networks from a CAS perspective. Surana et al. (2005) states: "The concept of CAS allows one to understand how supply networks as living systems co-evolve with the rugged and dynamic environment in which they exist and identify patterns that arise in such an evolution."

Pathak et al. (2007) identified four key elements of CAS in supply chains: 1) organizational entities exhibiting adaptivity, 2) a topology with interconnectivity between multiple supply chains, 3) self-organizing and emergent system performance, and 4) an external environment that coevolves with the system.

A conceptual framework (Figure 2) built by Nair & Reed-Tsochas (2019) considers the CAS perspective in the operations and supply chain management research.

12

Figure 2:

Conceptual model of CAS in supply chain management (adapted from Nair & Reed-Tsochas, 2019)
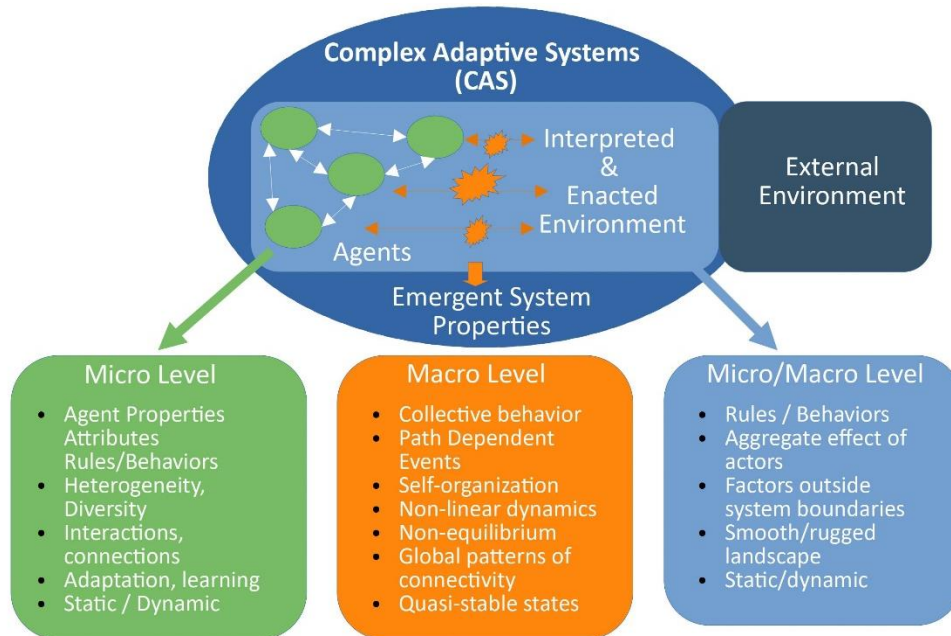
According to Figure 2, Nair and Reed-Tsochas (2019) established properties found in the micro and macro levels of the system, pointing out that agents are in the micro-level and emergent system properties that result from agent interactions with the environment are found in the macro level. However, they divided the environment into two levels: the "interpreted and enacted environment," which directly interacts with the agents, and the "external environment," which does not directly affect the agents. They describe this boundary between "interpreted and enacted environment" and "external environment" as not being static but dynamic since agents may interact with factors in the external environment at different moments in time. Therefore, the environment can possess properties from either the micro or macro level.

Companies have started to look for new ways of achieving competitive advantage, different than cost, quality and time (Giannoccaro, 2015). Adaptation has become a key element for companies to maintain competitive advantage (Reeves & Deimler, 2011). As mentioned previously, the CAS perspective is

relatively new in the supply chain management domain. Some researchers have begun to investigate and apply this perspective to solve complex supply chain problems.

The relationship between learning and adaptation in supply chains located within industrial districts was studied by Giannoccaro (2015), with the objective of identifying the best adaptive supply chain. The author followed a complexity science approach, in which firms accomplish learning processes developing new knowledge to changes in its environment. The tool used in Giannoccaro's (2015) study is NK modeling. In the NK fitness model, the firm consists in a set of N interaction decisions on how to perform specific actions. The configuration of the firm displays the specific set of options on the decisions made by the firm and is modeled by the vector $d = (d_1, d_2 \dots d_N)$ where the decisions are usually assumed to be binary. All the possible configurations the firm can adopt are then 2^N fitness value (payoff) is related with each configuration and the firm is optimized to find the best configuration (highest payoff) (Giannoccaro, 2015). The study addressed the design of adaptive supply chains by analyzing physical features (e.g., the number of firms) and organizational features (vertical integration) that lead to adaptation. In contrast, our research analyzes behaviors of plant and DC nodes by studying their reactions under different scenarios (e.g., demand variability and lead-time) to improve value sharing in the whole SC of a beverage industry.

A study by Zhao et al. (2019) focused on investigating a firm's adaptive strategies against disruptions in a supply chain network. They analyzed the effects of the supply chain network where in the face of a disruptions, the firms' interconnected parts of the network have the ability to adapt and restructure their connections. Their study identified the structural complex properties of the network (supplier-customer and competitor) and used agent-based simulation models in order to understand if the firm's adaptive behaviors can reduce the impact of disruptions. Several simulation runs were executed, and the model followed two steps: 1) An initial disruption by removing a supplier node, and 2) After the removal of a supplier node, each customer of this node tried to find alternative suppliers (Zhao et al., 2019). Zhao's

study examined the adaptive behaviors of the supplier and customer in a supply chain network by using agent-based modeling.

Inspired by the concept of CAS, this project models the BeverageCo's supply chain network. CAS theory helped us analyze the supply chain from a top-down and bottom-up perspective. It allowed for better understanding of the interactions happening at the line level, plant level and country level.

2.2 Concept of Digital Twins

In this section, an overview of Digital Transformation is presented with a focus on the studies in the field of Digital Twins. These studies concentrate on optimizing operations and achieving efficiencies through digital twins. The previously published work on various digital technologies is also explored in this section. Finally, a review on the conceptual architecture of digital twins is provided.

2.2.1   Digital Transformation

Digital transformation is the process of using digital technologies to transform existing and non-digital business processes or services, or creating new ones, to meet evolving market conditions and customer expectations. Digital transformation today is altering the way businesses operated traditionally and has become crucial for all companies from small firms to large-scale enterprises. Many companies have started a major transformational change by substantially integrating digital technologies into their businesses (Hess et al., 2016). Digital technologies make it possible to accelerate experimentation with business models, providing new opportunities (Massa & Tucci, 2013). Crittenden & Peterson (2019) define digital transformation as a new way of doing business, whereas, for Caro and Sadr(2019), it is the adoption of state-of-the-art concepts such as Industry 4.0, smart factory, or digital twins.

However, digital transformation is not about technology, it is about the value addition to the enterprise. This value addition can only be achieved by implementing technologies that can strategically fit in

company's vision. Hence, the first step in starting a digital transformation journey is to have a strategic framework in place, taking into account short and long-term goals of the company (Kane et al., 2015). This strategic framework should consider existing capabilities, target state, and the gap between current state and target state (Ustundag & Cevikcan, 2018).

A winning digital transformation strategy should strive to provide a coherent plan of action and a guiding policy to reach from current to target state. However, digital transformation is not a static process; it is an iterative discovery-driven process, as the approach needs to be revised based on tests and experiments with different prototypes and solutions (Blank, 2007).

As a starting point towards digital transformation journey, companies should build an initial hypothesis identifying technologies incorporated in this transformation. These technologies may not always provide the expected business value, and may require a replacement by another technology, or a change in implementation methodology. For instance, an advanced autonomous robot may not provide the expected business value to a manufacturing company.

Following the best practices for Digital Transformation strategies, specifically on digital twin and artificial intelligence, found in the literature, the goal of this project is to develop a conceptual model of digital twin and artificial intelligence, and evaluate if this model can add value to BeverageCo.

2.2.2   Digital Twins

In 2003, the University of Michigan Executive Course on Product Lifecycle Management (PLM) publicly introduced the concept of digital twins (DT) as a virtual, digital equivalent of physical object with data flow from physical space to virtual space, and Information and insights shared by virtual space to physical space (Grieves, 2015). National Aeronautics and Space Administration (NASA) came up with a more specific definition of DT (Glaessgen & Stargel, 2012) "an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates,

fleet history, etc., to mirror the life of its corresponding flying twin." The idea has since then, gain importance from both academicians and industry practitioners.

DTs are not just the mirrors of the static physical objects, processes, or systems but also dynamic simulations of these physical assets (Qi et al., 2018). This dynamic capability allows virtual simulation models to understand, predict, learn, and improve performance by guiding physical systems in responding to changes in their environment (Fei Tao et al., 2017).

Internet of Things (IoT) here is one of the major drivers behind connectivity and adaptability of DT. IoT devices such as high-quality sensors and microprocessors enable real-time data collection through devices connected to physical assets. DT is capable of collecting data through networked IoT sensors from a physical system to the digital twin, and continuously update its model in real-time. Since, the interaction is bi-directional, a physical system can use the knowledge acquired from one or more digital twins to improve its performance (Madni et al., 2019).

Performance and learning capacity of the digital twins can be enhanced by connecting them to the digital thread. According to D'Amico et al. (2019), the stream of data circulating underneath the digital world is called the digital thread, which keeps the real and virtual world synchronized. The digital thread allows a connected network of data flow and integrated view of products or systems throughout their lifecycle, creating a universal access to data. When connected to the digital thread, DTs can create their own network of feedback loop and become more powerful as they feed each other with a strand of their historical and real-time data. Such network of DTs delivers a number of value-adding outcomes such as real-time monitoring of systems, optimized asset maintenance, faults detection and risk mitigation, and reduced unplanned downtimes.

DT can use simulation models to explore physical system behavior under various what-if scenarios to optimize a physical system. In a DT model of a physical supply chain, simulations can be run to understand

the behavior of a logistics network over time, and any descriptive statistics of the working of supply chain can be collected. For example, which manufacturing plants can be used in case of a sudden increase in customer demand. Such rules can be combined and tested against any disruption events from machine shutdown to natural disasters to understand supply chain reaction in different scenarios. Adaptive DTs can learn from the predictions captured using technologies such as Artificial intelligence (AI) and machine learning (ML). AI and ML will be further discussed in the Section 2.2.4, "Digital Twin Technologies".

The main sources of data that a DT can use are historical data, sensor data, user inputs, data from a network of digital twins, and simulation data. AI model can be implemented to obtain patterns in this data and generate insights such as the impact of a sudden decrease or increase in demand on manufacturing plants or to understand the implications of a new environmental policy on production. The output can then be fed back into physical system, creating a feedback loop to obtain operational and supply chain efficiencies.

A DT reference model was presented by Bevilacqua et al. (2020) in order to define conceptual guidelines for the implementation of DT for risk prediction and prevention. For example, DT models can increase the accuracy of machine anomaly detection and increase machine life. As another example by Madini et al. (2019), DT can be used to combine data from physical systems and IoT devices to optimize manufacturing processes, for example, to improve logistics support.

2.2.3   Digital Twins - Interaction & Collaboration

According to Tao et al. (2019), digital twins must interact and collaborate with each other to tackle complex problems. Digital thread enables this interaction by providing the ability to access, integrate, analyze, and transform data from multiple DTs. Through this interaction and collaboration, multiple DTs can be optimized through the synchronization between themselves, and with their physical counterparts. Focusing on this, physical twins can be dynamically adjusted based on direct orders from the digital model.

At present, only a few authors have explored interaction and collaboration between DTs. Vachálek et al. (2017) argue that DTs allow physical systems to respond more quickly to unexpected events based on constant interaction between the physical and digital systems. According to Rosen et al. (2015) , a digital space can capture all available data such as system sensor data, data from IoT devices, among others, from a physical space over a digital thread.

2.2.4 Digital Twin Technologies

DT encompasses a wide array of digital technologies and work in conjunction with various such technologies.

Following sections present the use of Digital Twins with simulation, Industrial internet of Things (IIoT), artificial intelligence (AI), and machine learning (ML).

2.2.4.1 Digital Twin and Simulation

Simulations can be used to run what-if scenarios in the virtual environment and predict physical system behavior in the real world. Weyer et al. (2016) suggested that DT simulations could help monitor and optimize complex manufacturing systems. According to D'Amico et al. (2019), a virtual twin which represents attributes of a physical system can be combined with environmental attributes such as data from other assets or external information which does not exist in physical system. Environmental data is a major difference between physical and virtual systems. It is possible to enable simulation of what-if scenarios where output can be predicted by combining data from the physical system, environmental data, and user inputs.

2.2.4.2 Digital Twin and Industrial Internet of Things (IIoT)

Tozanlı et al., 2020 suggested that IIoT devices can be considered as one of the primary enablers of the digital twin technology. According to Boyes et al. (2018) , industrial IoT (IIoT) is a system of devices

connected by communication software to monitor, collect, exchange, and analyze information to change their behavior or of their manufacturing environment. IIoT devices provide real-time data from multiple sources to digital twins, increasing the use and applicability of DT. According to Fuller et al. (2020), the growth of digital twins in the manufacturing industry is possible because of a similar growth of IIoT devices. On the other hand, it should be also noted that DTs are not part of IIoT technology. Utilizing IIoT devices facilitates ubiquitous real-time data to DTs and enhances their capabilities.

2.2.4.3 Digital Twin and Artificial Intelligence (AI)

Artificial intelligence (AI) refers to the ability of a computer or machine to mimic the capabilities of human mind. Artificial Intelligence encompasses several techniques. These techniques have the property that they can predict or prescribe efficiently and intelligently without being explicitly instructed (Cagle, 2019). Machine learning is a technique from Artificial Intelligence practices. Machine Learning (ML) algorithms learn from past data and improve themselves recursively to better identify patterns and predict (Cagle, 2019).

2.2.5 Digital Twin Architecture

This section presents an architecture of digital twin showing the information flow between a physical system, simulation, and an AI model through IoT devices, and transferring the information back to physical system. Combining inputs from Grieves (2015), Tao et al. (2019), Bevilacqua et al. (2020), and Canedo (2016), Parrott and Warshaw (2017) constructed a digital twin architecture with the following feedback loop: Concept, Communicate, Analyze, and Improve. The digital twin architectural diagram depicted in Figure 3 is an adopted version constructed for this study to provide the readers with an understanding of information flow within a digital twin.

Figure 3:

Digital Twin Conceptual Architecture (adopted from Parrott & Warshaw (2017))
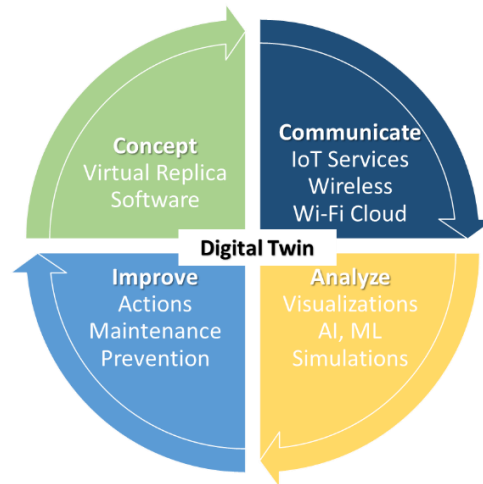
In this regard, the four keystones of the digital twin architecture demonstrated in Figure 3 are elaborated in the following sections.

2.2.5.1 Concept

Initial design is envisioned at this stage, and implemented by creating a replica of physical object, process, and system. The concept stage encompasses replicating physical processes and objects to digital systems with the use of enterprise resource planning systems, supply chain systems, CAD models, and IIoT devices. For example, several IIoT devices can be used to capture real-time data such as performance data, external data affecting operations, from physical objects; physical assets can be represented by 3D model using CAD.

2.2.5.2 Communicate

The communicate step provides the seamless, bi-directional, real-time connectivity between the digital twin and physical system. Several communication interfaces are necessary to transfer information from physical systems, for example, through IIoT devices to the digital twin. Finally, the data gathered is aggregated to ingest into a data repository, and then processed and prepared for analytics.

2.2.5.3 Analyze

In the analyze step, data is analyzed and visualized to obtain insights from data. Data analysis involves data collection, cleaning, transforming, modeling, and interpreting data to discover insights and guide decision-making. Data analysts can use advanced technologies such as artificial intelligence to develop models to provide recommendations to business.

2.2.5.5 Improve

In the improve step, actionable insights from the analyze step can be fed back to physical system. These insights can be used to recommend preventive measures, and to reduce maintenance costs in plants. This connection between digital twin and physical system completes the feedback loop between the two systems.

With a focus on the beverage industry, this project investigates the utilization of DTs to examine the adaptability of manufacturing and supply chain operations in complex and dynamic ecosystems. This study aims to fill a gap in the literature by analyzing the value-added of using DTs by creating a data-driven learning system with the utilization of simulation and artificial intelligence.

3 DATA & METHODOLOGY

This section outlines the methods and tools used to answer research questions discussed in the previous section.

This study examined a data-driven learning system to analyze complex and adaptive dynamics of the utilization of digital twins in manufacturing operations along the supply chain of a beverage manufacturer. To understand and run what-if scenarios, a feedback loop between simulation and AI to foresee the patterns for the adaptability of supply chain agents (plants and distribution centers) and within the whole supply chain (SC) settings was created. The project analyzed how different supply chain players such as manufacturing plants and distribution centers interact in a supply chain network, and respond to the dynamic environment both internally and externally to create an autonomy in decision-making processes to achieve operational efficiency and value sharing.

This project was conducted in six phases, as depicted in Figure 4.

Figure 4:

Methodology Overview
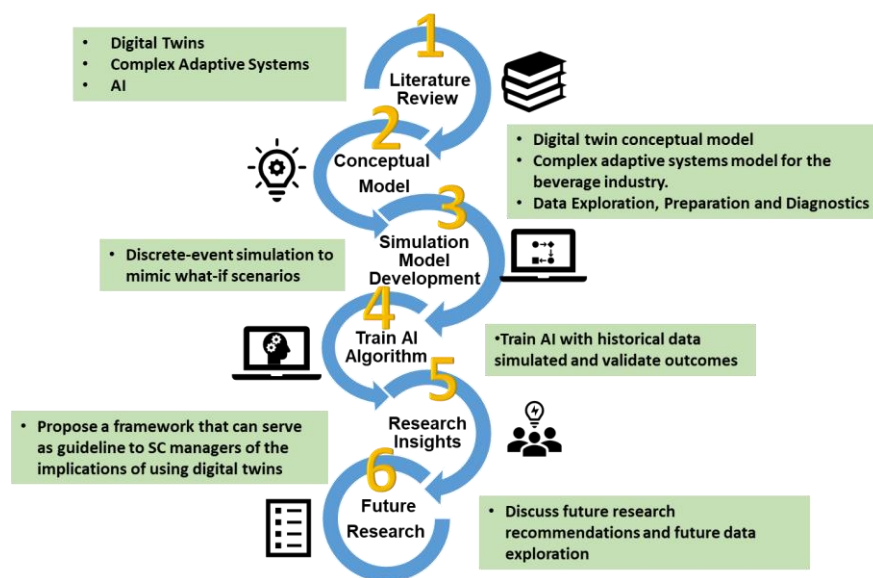
In this regard, Phase 1 was detailed in Section 2, 'Literature review'. Phases 2, 3, and 4 are explained in this section. Phase 5 and Phase 6 will be further elaborated in Section 5, 'Discussion and management insights' and Section 6, 'Conclusion', respectively.

First, a conceptual model (Phase 2) of a digital twin mechanism in the beverage industry was developed. This model represents a replica of the physical supply chain of BeverageCo. Then, a discrete-event simulation model was created and implemented to mimic various real-world situations (Phase 3). The simulation output was then used to train an AI model (Phase 4). Finally, a feedback loop was established between the simulation and AI model to share AI predictions with simulation model. These predictions can be reviewed and analyzed by BeverageCo supply chain users, and can be fed back to the physical supply chain completing the closed connection between physical supply chain, digital twin and AI model. The following sections discuss digital twin conceptual model and complex adaptive systems. Subsequently, our methodology for data preparation, simulation and AI model design, and for feedback loop creation is provided.

3.1 Digital Twin Conceptual Model

Our digital twin conceptual model for the beverage industry is adapted from Modoni et al. (2019) as shown in Figure 5. According to Figure 5, a continuous mirroring between the physical manufacturing plants and its digital twin is realized by means of two data flows. The first one (from left to right) denotes the real-time data flow and includes performance metrics from the real physical manufacturing plants to the digital twin. The second flow (from right to left) represents the actions performed based on the feedback from the digital twin to the physical manufacturing plants. These actions are the decisions that managers can make with the use of digital twins.

We enhanced this digital twin conceptual model developed by Modoni et al. (2019) by adding the concept of the data-driven learning system. Focusing on this, the learning system addresses the utilization of AI

and simulation to provide supply chain managers opportunities to understand the impact of AI predictions on supply chain behavior of BeverageCo. DTs can be made more intelligent and can be used to prescribe actions in the physical world using the data obtained from AI algorithms. AI-driven predictions can enrich the capabilities of DT by recommending optimized future actions to users.

Figure 5:

Digital Twin Conceptual Model (adopted from Modoni et al. (2019))
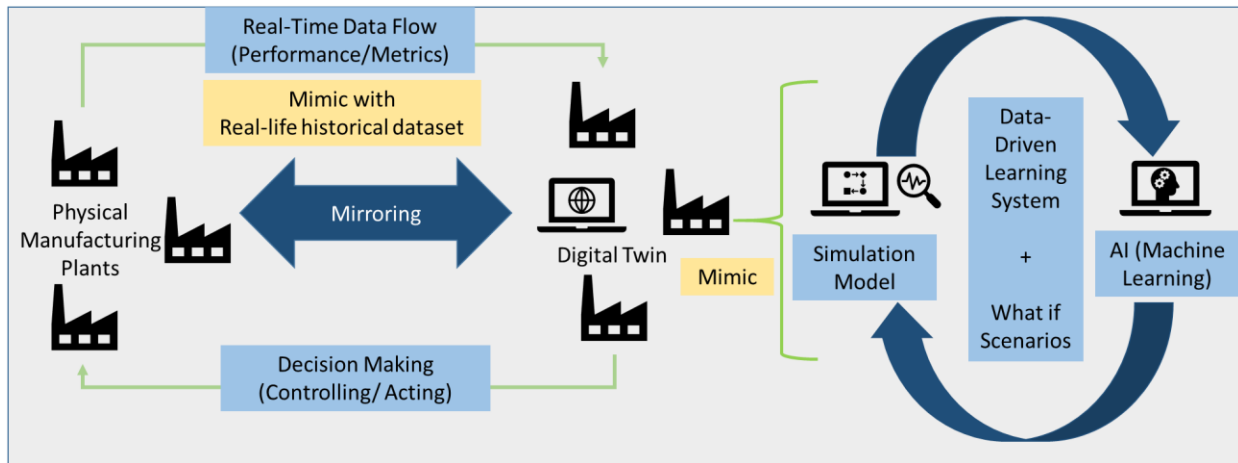
Although the conceptual model presented in Figure 5 is how it will look if digital twins are implemented in real life, this capstone does not focus on creating a real-time digital twin where data is transferred through IoT sensors. This research offers a conceptual view of digital twins based on a real-life historical dataset from BeverageCo. The digital twin created in this study consisted of historical data, simulation, an AI model, and a feedback loop between AI and simulation model.

3.2 A Complex Adaptive System (CAS) Conceptual Framework of the Beverage Industry

The conceptual framework was built using Nair & Reed-Tsochas (2019)'s consideration of the CAS perspective in operations and supply chain management research. Inspired by the CAS theory, in Figure 6 a conceptual framework is created to analyze the manufacturing and supply chain network of BeverageCo.

Figure 6:

CAS Conceptual Framework in Manufacturing Operations
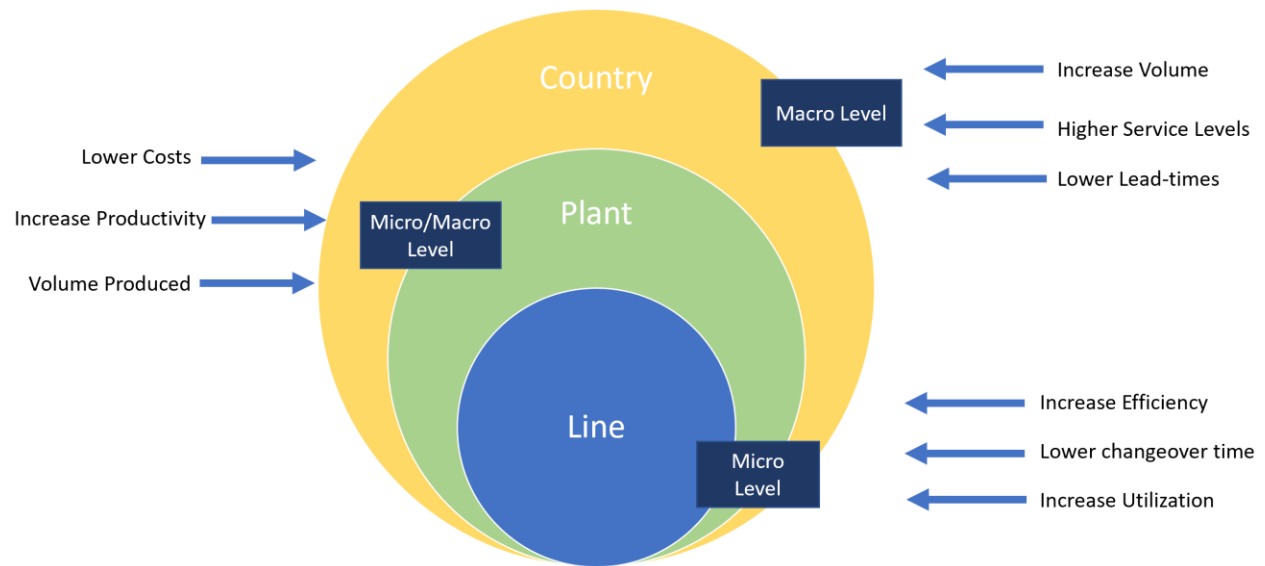
According to the figure, the line level is the macro level in which the operation aims to increase efficiency, lower changeover times and increase utilization. Each production line has its own configuration and performance in order to achieve the targets for efficiency, changeover times and utilization. The interaction between production lines creates the micro/macro level which is the plant level. At this level, the operation is looking to lower costs, increase productivity and secure production volumes. The interaction between different production plants creates the macro level which is the country level. The operation at this level is looking to increase volume, increase service levels and reduce lead-times to satisfy customer demands and expectations.

In this project the focus of the scenarios created during simulation was at the country level. The four scenarios were created by varying customer demand and expected customer lead-times values, and the interaction between manufacturing plants, distribution centers and customers were observed.

3.3 Research setting and Data

3.3.1   Process Flow

The following section explains the execution process flow. This section starts by providing a high-level process flow to the reader. Each of the steps and sub-steps are explained in following sections in detail and can be visualized in Figure 7.

Figure 7:

Process Flow Diagram
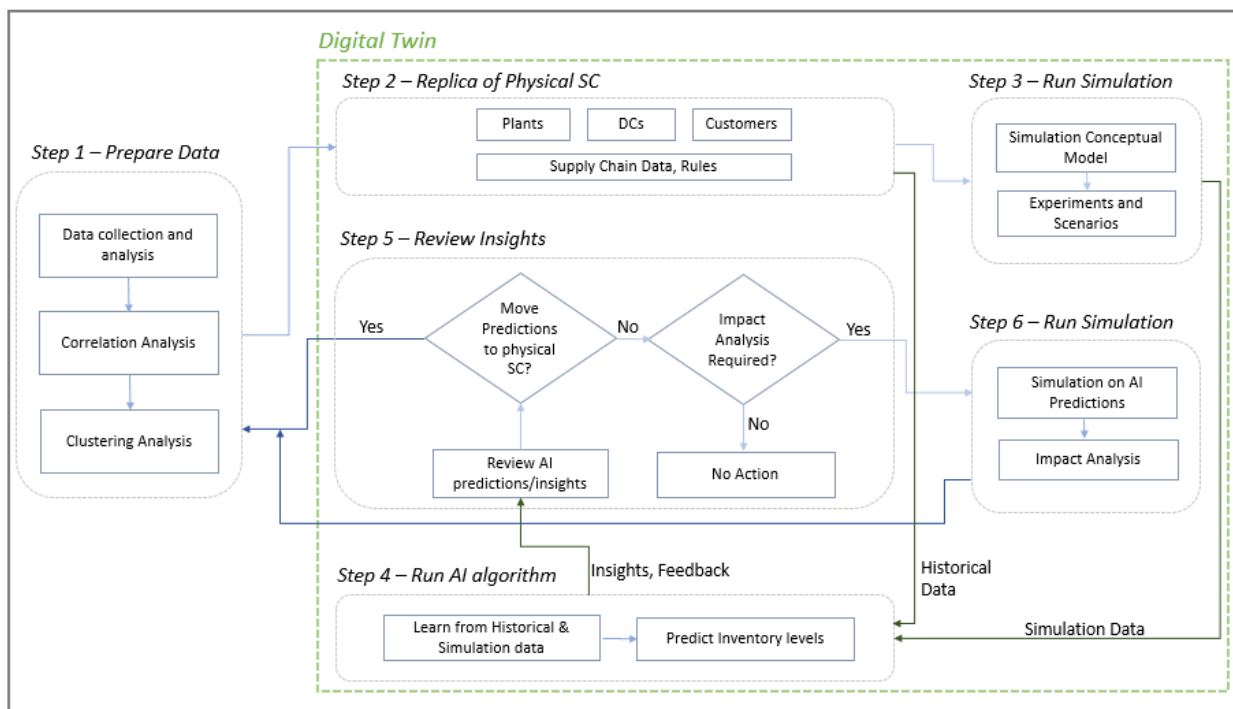
<u>Step 1</u>: Execution starts with data preparation including data collection and analysis, followed by correlation and clustering analysis. The output of Step 1 is the 'data' required for the creation of conceptual model of digital twin.

Step 2: In this step, we create a digital twin to replicate physical supply chain of BeverageCo. Digital twin consists of manufacturing plants, distribution centers, customers, and supply chain rules or parameters of physical supply chain of BeverageCo.

Step 3: Several experiments and scenarios are run in this step. We run multiple experiments by increasing demand, changing lead times to generate inventory levels for different scenarios. The data from simulation is shared with AI model.

Step 4: AI model accepts historical data directly from digital twin and simulation data from simulation model. AI algorithm learns from historical and simulation data, and share insights/predictions with digital twin. This transaction completes the feedback loop between digital twin and AI model.

Step 5: Supply chain managers can review the predictions shared by AI model in this step. Various data interpretation and visualizations techniques can be used to review insights shared by AI model. SC managers can decide to fed predictions to physical supply chain or perform an impact analysis to understand the impact of these predictions on BeverageCo supply chain.

Step 6: SC managers can perform an impact analysis in this step to understand the supply chains reaction to AI predictions. Various what-if scenarios can be run on AI predictions and historical data obtained from the physical supply chain. Depending on the results of these scenarios, SC managers can manipulate the physical supply chain with AI predictions. There is an opportunity to alter AI predictions and share refined numbers with physical supply chain as well. These modifications may be necessary to include market intelligence and real-world knowledge of SC managers.

3.3.2   Data Selection and Preparation

Data was obtained from BeverageCo's enterprise resource planning (ERP) system and was collected for all the countries where the company operates. Among these countries, data from Brazil was found to be

the most comprehensive and sufficient for this study. Brazil has second highest number of plants and distribution centers (DCs) providing enough data for the study and at the same time, reducing complexity which might incur if data from Mexico, country with highest number of plants is used. Further, Mexico data is incomplete, and data for all plants is not available in the ERP system. Hence, Brazil data was selected for the study and to build the digital twin model. Brazil operations consist of 10 manufacturing plants with 61 production lines, and 5 distribution centers.

Historical data consists of monthly data available from Jan 2018 to Mar 2019, including quantitative information of manufacturing plants. Based on data availability, relevant data variables were identified for the plant and distribution center selection. Data was examined at three different levels – Production Line, Plant, and Country level. Figure 8 depicts the significant variables identified for manufacturing at these three levels.

Since majority of the variables including costs and yield values were available at the 'Plant' level, the plant level data was selected for further data analysis.

Figure 8:

Key variables identified at different location levels
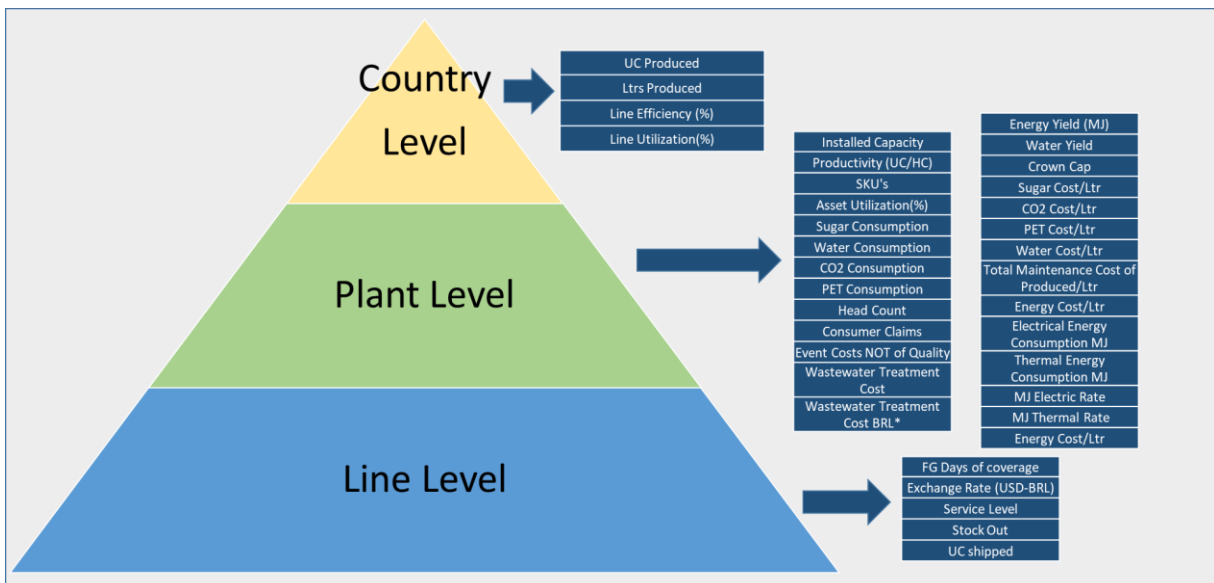
3.3.3 Data Analysis

For preliminary data analysis, we first analyzed the correlation between variables and performed a clustering analysis. Correlation results were shared as an input to identify right variables for clustering. Clustering findings were used to select plants for digital twin and AI model. To increase the testing efficiency, in the data selection process, the focus was to identify the plants and distribution centers most relevant for developing the digital twin model. To overcome this limitation of number of plants and DCs, we used the results from the clustering analysis to select manufacturing plants and DCs for the analysis.

3.3.3.1 Correlation Analysis

Correlation is a statistical measure that indicates the degree to which two variables fluctuate in relation to one another. In this study, the main objective of correlation analysis was to identify suitable variables for clustering. The output of correlation analysis was used to limit the variables required for the clustering analysis. For example, it was evident from the data that liters produced and the installed capacity have a very high correlation; hence, both these variables were not required for clustering analysis. A positive correlation indicates the extent to which two variables are associated with a positive relationship, i.e., the variables increase or decrease in parallel, while a negative correlation is when two variables are associated with a negative relationship, i.e., increase in one variable decrease other.

3.3.3.2 Clustering Analysis

Clustering is a technique of unsupervised learning. It is not necessary to provide a priori an example or label of the clusters, but the algorithm creates the groups based only on the set of data. It is a common method for statistical data analysis used to discover similarities between the data items and group them according to these common characteristics into several categories, known as clusters. It is also used to discover patterns (Milligan & Cooper, 1985).

In this study, clustering analysis was used to cluster 10 manufacturing plants of Brazil into smaller groups with similar characteristics. Plants within resultant clusters or group of plants are more similar to each other than to the rest of the plants based on certain identified parameters.

Clustering analysis involves two important identifications – identify appropriate numbers of clusters and identify the variables to discriminate between groups. The elbow method was used to find the number of possible groups with similar patterns and k-means was used to help identify the variables that allow discrimination between groups.

Elbow method

In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of the clusters, and picking the elbow of the number of clusters to use. The relationship is graphed between the number of clusters and within cluster sum of squares (WCSS) and then the number of clusters is selected where the change in WCSS begins to level off (Syakur et al., 2018).

K-means clustering

The k-means clustering is a partitional algorithm. The goal of using this algorithm in our analysis was to group the data items in a way that the items with similarities are in one group. K-means clustering is one of the most used algorithms in research. The algorithm consists of following steps:

a. Identify number of clusters using Elbow method explained in 'Elbow method' section. Within-Cluster-sum-of-squares is used as a measure to find the optimum number of clusters. Within-Cluster-sum-of-squares is defined as the sum of the squared distance between each member of the cluster and its centroid.

$$WCSS = \sum_{i=1}^{n}(x_i - c_i)^2$$

where,

x_i = data point

c_i = closest point to centroid

The WCSS is measured for each value of k. The value of k, that which has the least amount of WCSS is considered optimum.

b. Choose K random data items representing k clusters to initiate the algorithm. These k data points will be the cluster centers called centroids.

c. Assign each data item (x_i) to the closest cluster by minimizing following objective function:

$$J_{KM} = \sum_{j=1}^{k} \sum_{i=1}^{n} w_{ij} \left(\left\| x_i - c_j \right\| \right)^2$$

where,

k = number of clusters

n = the amount of data items in the data set

$w_{ij} = \begin{cases} 1 \ if \ c_j \ is \ the \ centroid \ closest \ to \ x_i \\ 0 \ otherwise \end{cases}$

$\left\| x_i - c_j \right\|$ = the Euclidean distance between each data item x_i and the centroid c_j

d. Identify new centroids by taking the average of the assigned points

e. Keep repeating step c and step d until convergence is achieved

The results of clustering analysis provided us with the final list of plants and DCs used in the digital twin model.

3.4 Simulation Modelling Approach

The steps in our simulation modelling are demonstrated in Figure 9. The modelling process is comprised of two stages: the development stage and the experimental stage.

Figure 9:

Simulation Modelling Process

Development Stage	Experimental Stage
Supply Chain Network Description	Scenarios and Experiments
Conceptual Model	Simulation and Evaluation Results
Model Implementation	Documentation
Verification & Validation	Analysis
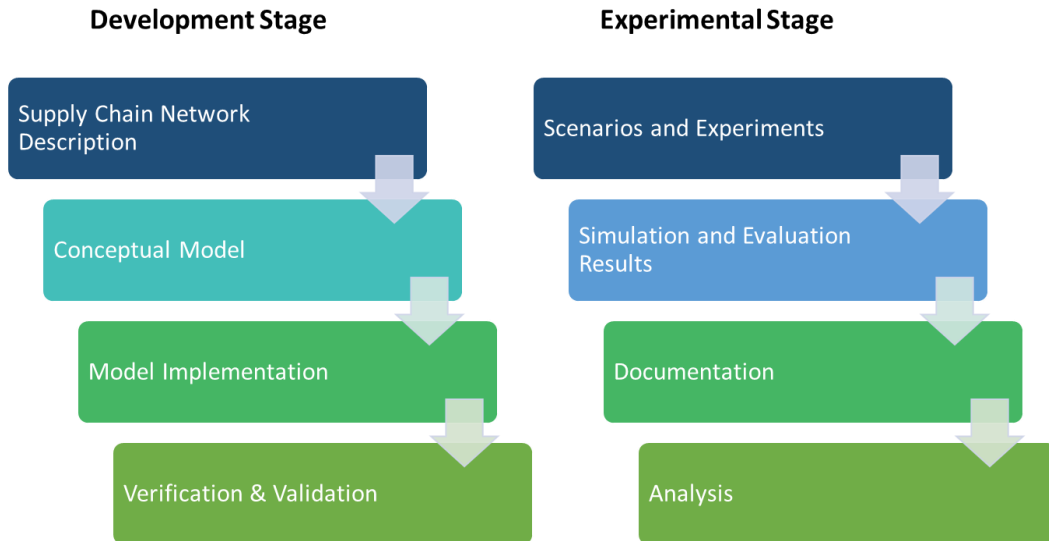
The following sections elaborates on the stages of the simulation modeling in further detail.

3.4.1 Supply Chain Network Description

The BeverageCo supply chain network used for the purpose of this study consisted of 10 manufacturing plants from Brazil operations. As explained in the Section 4, Results , five manufacturing plants from Cluster 1 (see Section 4.1.2, Clustering results, Table 8) were selected to be modeled in the simulation together with five distribution centers and ninety-nine customers. All customers designated to be fulfilled from the selected plants are included in the study.

In the existing supply chain network, each manufacturing plant serves to one distribution center, and a distribution center serves a group of customers based on the nearest distance proximity (see Figure 10). The manufacturing plants locate in Jundiaí, Maringa, Marilia, Itabirito, and Curitiba, which are identified as BAAD, BAAJ, BAAK, BAAG and BAAI in Figure 10, respectively. Similarly, the distribution centers are

placed in Assis Brasil, Marilia, Macro, Curitiba, and Maringa and referenced as DC_AB, DC_ML, DC_MC, DC_CB, and DC_MG, respectively. The customers are numbered from 1 to 99, and shown in blue circles.

Figure 10:

BeverageCo Supply Chain Network
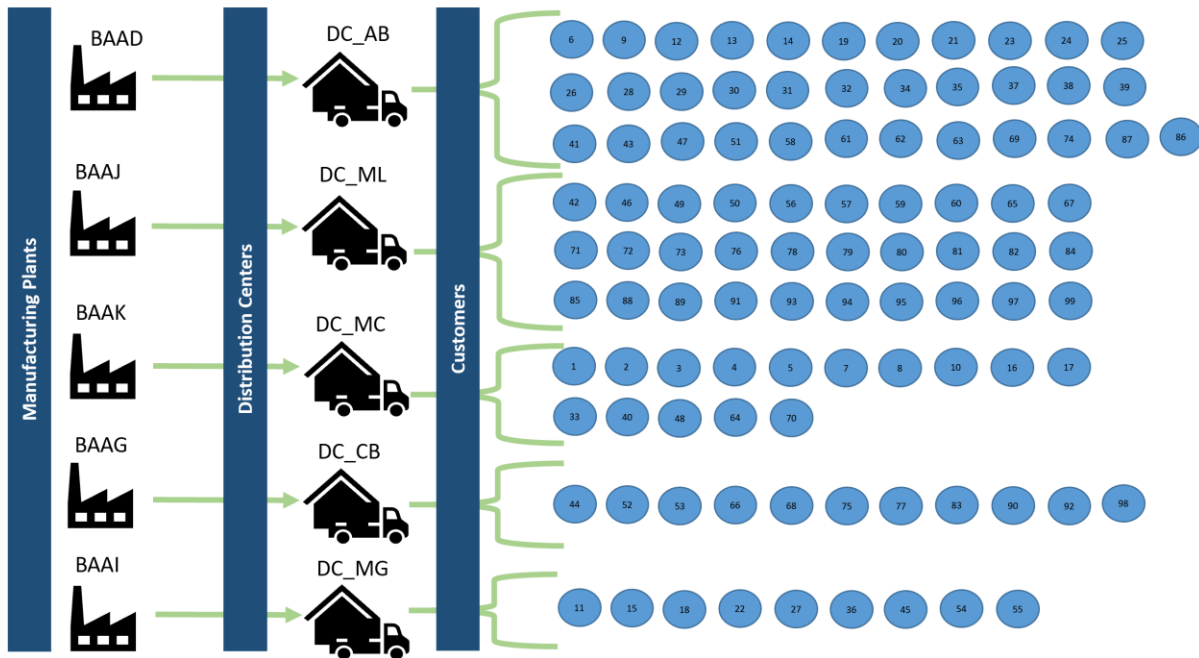
In the ERP system, all the data required to build the model was not present; hence, we structured certain assumptions on customer demand, lead time, and inventory policy to be considered in the simulation model. These assumptions are explained below.

<u>Customer Demand:</u>

Customer demand is at the daily level and can be fulfilled by any DC. Demand was computed using the one-year shipment quantity data available in the ERP system. Triangular distribution was used as only limited shipment data for few plants and distribution centers was available.

Triangular distribution is a continuous, bounded probability distribution, which serves as a good way to express uncertainty by providing the minimum, maximum and most likely (mode) values. We developed

triangular distribution using min shipment quantity, max shipment quantity, and mean shipment quantity

obtained from data. Triangular distribution was formulated as follows:

Triangular distribution (min, mode, max):

$$Min = Min\ (i)$$

$$Max = Max\ (i)$$

$$Mode = 3 * Mean\ (i) - Min\ (i) - Max\ (i)$$

where;

 i ∈ shipment data.

Min, Max and Mode values, in liters for each distribution center are summarized as shown in Table 1.

Table 1:

Triangular Distribution for Customer

Distribution Center Customers	Min	Max	Mode
DC_AB	99,430	164,363	102,306
DC_ML	27,320	46,072	28,893
DC_MC	35,619	82,270	41,787
DC_CB	64,083	93,102	69,171
DC_MG	131,919	251,284	202,948

DC Inventory Policy:

DC inventory policy used is assumed as a "Min-Max" policy. The "Min-Max" policy also referred as (s,S)

policy is an inventory policy where products are ordered when the inventory level falls below a fixed

inventory level 'Min' (s). The ordered quantity is set to such a value that the resulting inventory quantity

equals 'Max' (S). Same inventory policy was used for all the DCs and values are shown in Table 2. The units

are in liters and were estimated based on the average inventory days of sales found in BeverageCo's

historical dataset. The inventory holding cost was also provided as input to simulation model, which was

assumed to be $0.001 per day per liter.

Table 2:

Min-Max Policy Values per DC in Liters

Distribution Center	s	S
DC_AB	12,910,636	17,214,181
DC_ML	3,389,664	4,519,552
DC_MC	2,742,755	3,657,007
DC_CB	7,933,417	10,577,889
DC_MG	5,880,281	7,840,375

Customer expected lead time (ELT):

Customer expected lead time is the time an order should be fulfilled in. In BeverageCo's data, the distribution center to customers lead time varied from 1 to 4 days.

SKU Selection:

All the SKU's of beverage company have a similar product flow. The replenishment to DC always happens from the plant, and the DC distributes the supply to customers. However, in the ERP system, replenishment data is available for only few SKU's. As a result, we selected only 1 SKU – 'PET 2,0L' for the analysis as this SKU was produced at all plant locations and shipped to all DCs and customers.

3.4.2   Simulation Conceptual Model

The conceptual model used for our study is shown in Figure 11. The notation used is the following:

Notation

Indices for the supply chain entities:

- C = Customer

- D = Distribution Center

- P = Manufacturing plant

- I_{CD} = Information flow from entity C to entity D

- t = Time period in days

The simulation begins at period t = 1, and the information flows from customers to DCs. Customers generate demand orders to closest DC. If the closest DC does not have inventory to satisfy the order quantity, the customer will look for another closer DC to satisfy demand from. If there is inventory at DC to satisfy the order quantity, the DC receives the order and checks if inventory is below min quantity $s.$

If inventory is not below min quantity, DC generates shipment order, updated inventory and customer receives shipment order. If inventory at DC is below min quantity, it orders the quantity to reach target max quantity from manufacturing plant. Plants receive the demand order from DC, then generates the shipment order. The DC receives shipment order, then updates its inventory level.

The simulation starts from t = 1, and is repeated for each day till it ends when $t = 365.$

Figure 11:

Simulation Conceptual Model
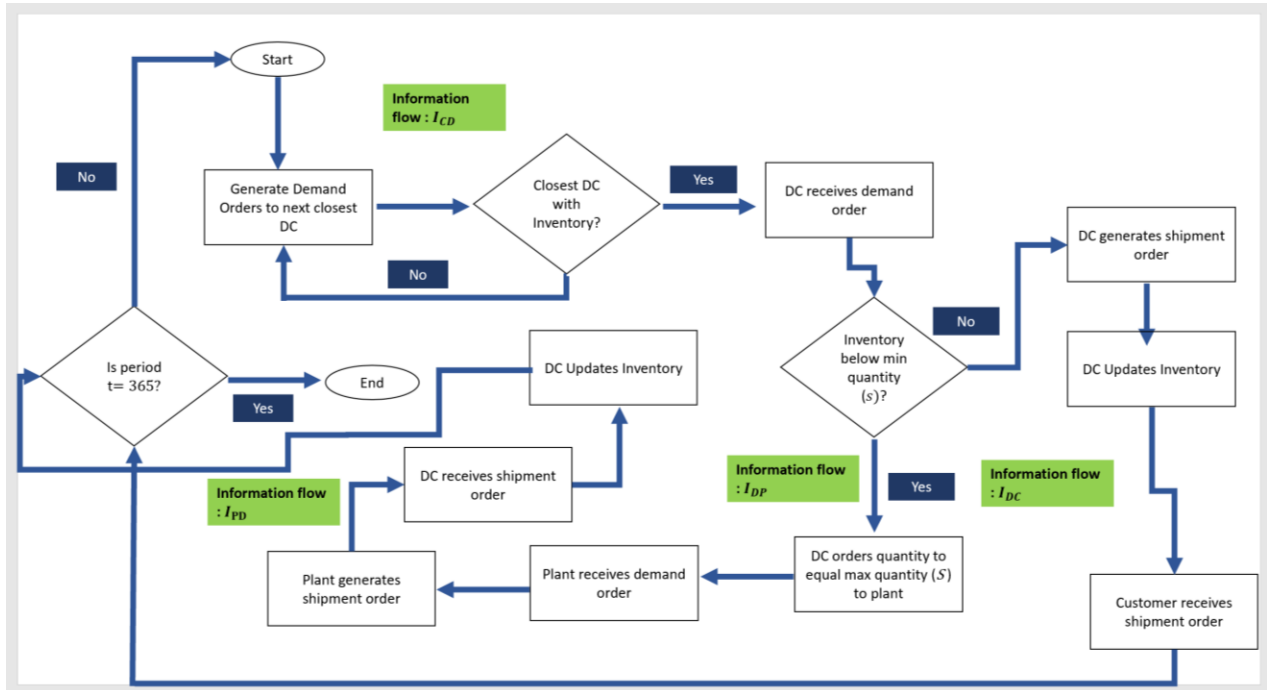
3.4.3   Scenario and Experimental Design

Using BeverageCo's historical dataset of production volumes as guidance, it was analyzed that the peak season is during the months of November and December. The scenarios that were simulated had the objective of studying the effects of increasing the demand in November and December and adjusting the customer expected lead-time (ELT). ELT is the time-period within which the ordered product is expected to be received by the customer. The results of the simulation model will enable managers take actions and decisions to better serve their customers.

Figure 12 shows the input and outputs of this scenario.

Figure 12:

Scenario Overview
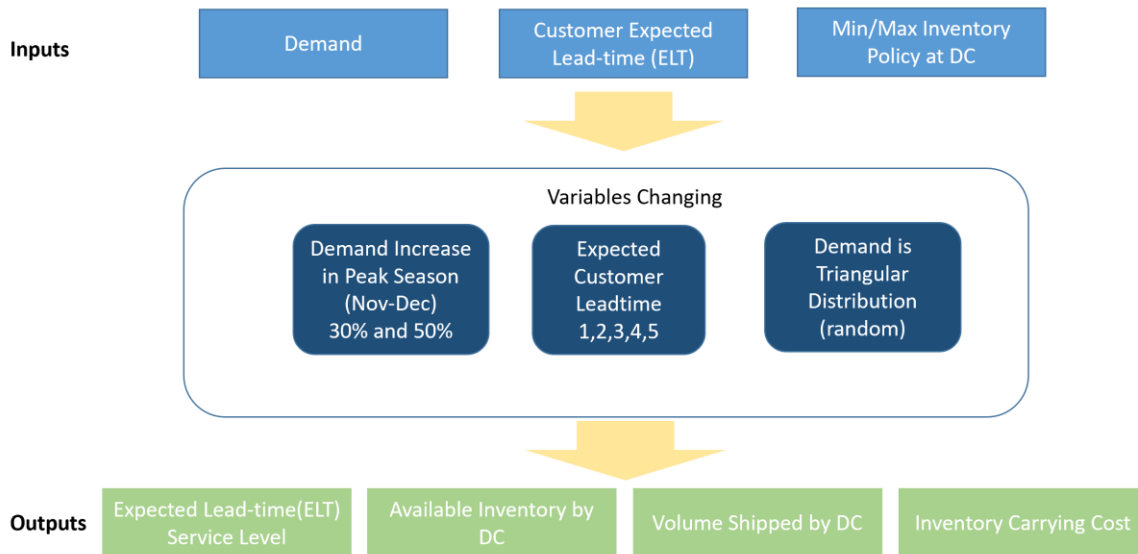
A 2^k factorial design was used to create four experiments, and to measure the effects of factors on the response with which interactions were measured. In experimental-design terminology, the input parameters and structural assumptions comprising a model are called factors, and output performance

measures are called responses (Law, 2013). Factors have two levels denoted with $(-, +)$ signs and the 2^k possible factor-level combinations, are called design points.

The 2^2 factorial design is shown in Table 3. The factors j_1 and j_2 are the expected customer lead-time (ELT) and the demand factor (Df). The notation R_i is the value of the response when running the simulation with the ith combination of factor levels. The response for our experiments was the average inventory level at DC. For example, R_3 is the response resulting from running the simulation with factor (ELT) at $(-)$ level and factor (DF) at its $(+)$ level.

Table 3:

Experimental Design Matrix for the 2^2 factorial design on ELT and Df

Factor Combination (design point)	$j_1 = ELT$	$j_2 = Df$	$ELT \ x \ Df$	R_i
1	-	-	+	R_1
2	+	-	-	R_2
3	-	+	-	R_3
4	+	+	+	R_4

The main effect of each factor, denoted by e_j, is the average change in the response due to moving factor j from its $(-)$ level to its $(+)$ level. This average was taken over all combinations of the other factor levels in the design.

$$e_1 = \frac{(R_2 - R_1) + (R_4 - R_3)}{2} \qquad e_2 = \frac{(-R_2 - R_1) + (R_4 + R_3)}{2}$$

The signs of column $ELT \ x \ Df$ in Table 3 were calculated by multiplying the ith sign in the factor (ELT) column by the ith sign in factor Df column (the product of like signs is a $(-)$ and the product of opposite signs is a $(+)$). This column gives the signs of the R_i used to form the interaction effect denoted by $e_{j_1 j_2}$. The measure of the interaction is the difference between the average effect of factor j_1 when factor j_2 is at its $(+)$ level (and all other factors other than j_1 and j_2 are held constant) and the average effect of factorj_1 when factor j_2 is at its $(-)$ level.

$$e_{j_1 j_2} = \frac{(R_1 - R_2) + (R_4 - R_3)}{2}$$

The matrix design of the four scenario and experiments conducted are presented in Table 4-7.

Table 4:

Scenario 1: Experimental Design

Factor	-	+
ELT	1	2
Df	1.3	1.5

Expected customer lead-time varies from 1 to 2 days and demand increase factor varies from 1.3 to 1.5.

Table 5:

Scenario 2: Experimental Design

Factor	-	+
ELT	2	3
Df	1.3	1.5

Expected customer lead-time varies from 2 to 3 days and demand increase factor varies 1.3 to 1.5.

Table 6:

Scenario 3: Experimental Design

Factor	-	+
ELT	3	4
Df	1.3	1.5

Expected customer lead-time varies from 3 to 4 days and demand increase factor varies from 1.3 to 1.5.

Table 7:

Scenario 4: Experimental Design

Factor	-	+
ELT	4	5
Df	1.3	1.5

Expected customer lead-time varies from 4 to 5 days and demand increase factor varies from 1.3 to 1.5.

3.5 Artificial Intelligence (AI) Model

An AI time-series algorithm was used to predict inventory levels at each DC from the output of the simulation experiments. There are many AI models that can be applied for time-series forecasting. A gradient boosting model was selected for this study. Gradient boosting is an ensemble artificial intelligence method that uses regression trees that are added sequentially. In each step, the new predictor is fit into the residual errors of the prior models. It can be used for both classification and regression predictive modeling.

Extreme Gradient Boosting (XGBoost) is an efficient, fast and effective open-source gradient boosting algorithm (Chen & Guestrin, 2016). The evolution and concepts of XGBoost from decision trees is shown in Figure 13. It uses parallel processing and tree-pruning. Because favorable prediction results have been obtained by using this model, since it avoids overfitting and bias, it is widely used in machine learning competitions. The XGBoost model performs excellently on the prediction of the sales volume, stock price, and traffic flow (Gumani et al., 2017; Wang & Guo, 2020; Lu et al., 2020).

Figure 13:

Evolution of XGBoost Algorithm from Decision Trees
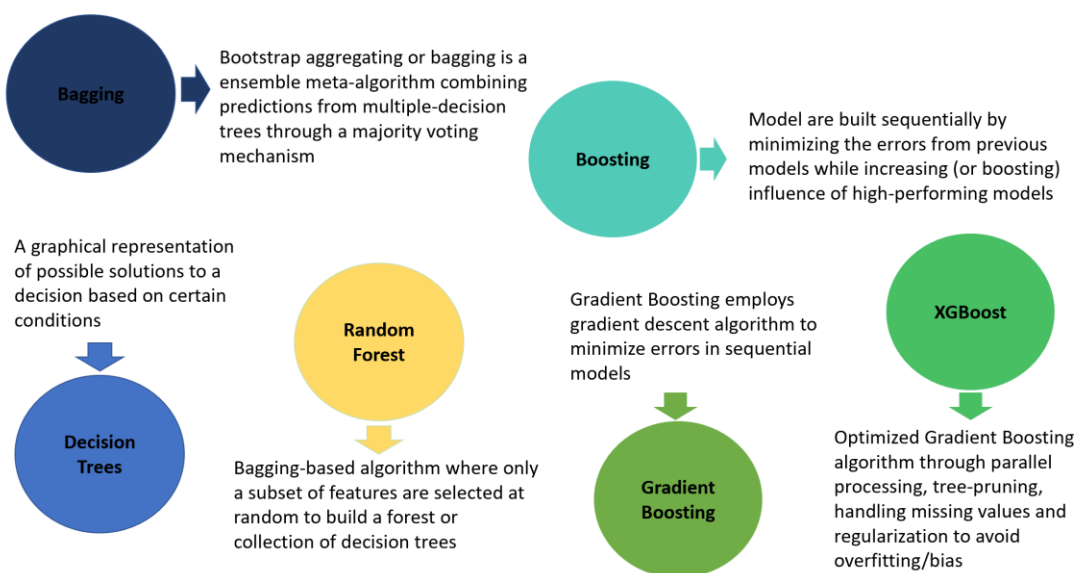
3.6 Feedback loop

In this step, a feedback loop was established between AI model and DT to improve decision-making. In real world, AI model share actionable insights with decision makers to predict behavior of physical system. However, in this study, the output of AI model was shared with digital twin. A simulation was then run to validate the output of AI model.

AI model was run to predict Inventory levels at each DC for a future period. Simulation model takes the predicted inventory levels as input, in addition to demand, lead time, inventory policy, which were other inputs to the model. Simulation model was run to validate if the predicted inventory levels are correct based on the future what-if scenario provided to model. For instance, simulation model can be run with demand coefficient (1.5), lead time (5), inventory policy (Min-Max), and inventory level predicted by AI model. In this scenario, if the AI model predictions are correct, DC's will fulfill customer demand from available inventory without replenishing from plant, and expected service levels are close to 100%. This interaction completes the feedback loop connection between the AI model and digital twin.

4 RESULTS

This chapter presents data analysis, results from simulation, artificial intelligence model, and the feedback loop.

4.1 Clustering Results

4.1.1 Correlation

The main objective of correlation analysis was to identify suitable variables for clustering. Correlation analysis was executed through Python. Figure 14 is the heatmap result of the correlation analysis.

Figure 14:

Correlation results for the Brazil plant
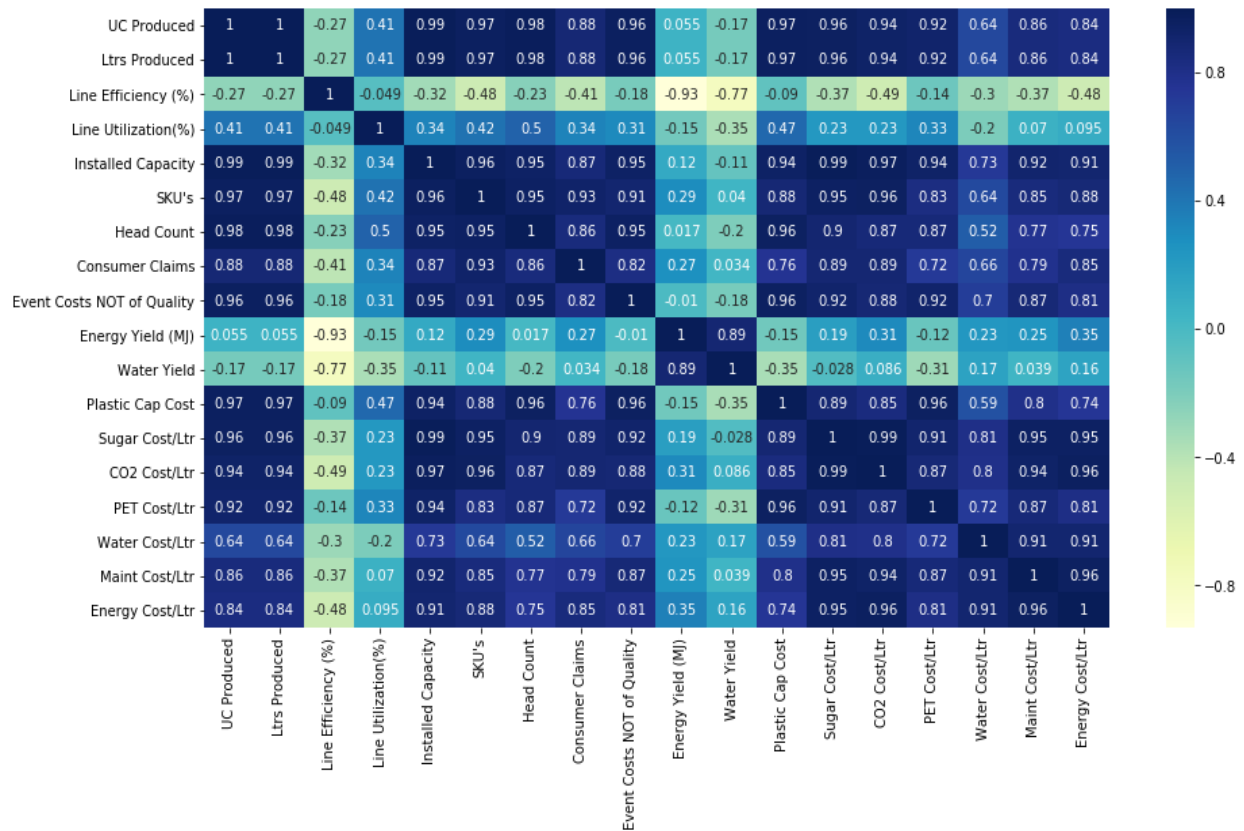
The results of Figure 14 are interpreted as follows. A perfect positive correlation has a 1 value and a perfect negative correlation has a -1 value. No correlation represents a 0 value. For both positive and negative correlation, values between 0.5 and 0.99 are described as having high correlation, and values between 0.3 and 0.49 as moderate correlation. The values below 0.29 are considered of low correlation.

Correlation results show a high correlation between UC produced, Ltrs produced, SKUs and all the cost variables such as Energy cost/Ltr and Sugar cost/Ltr. Hence, out of all these variables, only Ltrs produced is selected for clustering analysis. Similarly, as water yield and energy yield have high correlation, only energy yield was selected. Both line efficiency (%) and line utilization (%) were selected for clustering analysis.

4.1.2 Clustering

The output of correlation analysis was used to limit the variables required for the clustering analysis. Clustering analysis was performed to group the 10 manufacturing plants of Brazil into smaller subgroups with similar characteristics. As described in methodology section, clustering analysis consists of two important steps which are identifying the numbers of clusters and identifying the variables to differentiate between groups.

Elbow method

Elbow method was used to identify number of groups, and *k*-means helped in identifying the variables. To determine the optimal number of clusters, the value of k at the "elbow" i.e. the point after which the distortion/inertia start decreasing in a linear fashion, was selected. Thus, according to elbow method (Figure 15), three is an appropriate number of clusters for this study.

Figure 15:

Elbow method to identify number of clusters
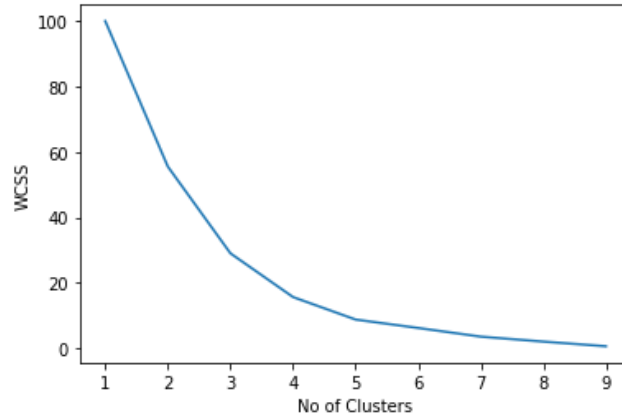
_k_-means clustering method

$k = 3$ identified from the elbow method is an input to *k*-means algorithm. Clusters were created using Python. Table 8 shows the output of *k*-means algorithm. Cluster 1 is largest cluster with five plants followed by Cluster 3 with three plants, and Cluster 2 with two plants.

Table 8:

Clustering Results

ID	PLANT	CLUSTER
Plant Bauru	B0AC-Bauru	Cluster 2
Plant Jundiaí	BAAD-Jundiaí	Cluster 1
Plant Mogi	BAAE-Mogi das Cruzes	Cluster 3
Plant Campo Grande	BAAF-Campo Grande	Cluster 3
Plant Itabirito	BAAG-Itabirito	Cluster 1
Plant Curitiba	BAAI-Curitiba	Cluster 1
Plant Maringá	BAAJ-Maringa	Cluster 1
Plant Marília	BAAK-Marilia	Cluster 1
Plant Antônio Carlos	BAAL-Antônio Carlos	Cluster 2
Plant Porto Alegre	BAAM-Porto Alegre	Cluster 3

All five DCs from Cluster 1 were selected to have a comprehensive view of beverage company supply chain in terms of fulfilling demand of all the regions. Customers are located across regions and fulfilled by DCs.

4.2 Simulation Results

4.2.1 Model Implementation, Verification and Validation

The conceptual model was implemented using discrete-event simulation. We verified and validated that

the conceptual model and simulation system behavior was in sync. We did this by comparing the results

of simulation model with historical data from BeverageCo. The what-if scenarios were run with the

historical data – demand, lead time and supply chain parameters of BeverageCo to understand supply

chain behavior – DC shipments and inventory levels for a historical period. To validate the results of

simulation model, the results from simulation were compared with actual data from BeverageCo.

4.2.2 Scenarios and Experimental Design Results per DC

As shown in Tables 4-7 of methodology section, we ran four simulation scenarios and our factors were

the expected customer lead-time from 1 to 5 days and demand increase factor that was varied from 1.3

to 1.5, which means a 30% and 50% respectively in the period of November and December. The response

measured was the average inventory level at each DC in time period 307, month of November. For each

scenario, 5 replications were made per lead-time.

In the following Tables 9 to 13, the results of the experimental design from each distribution center per

scenario are shown. It is demonstrated that the effect of variables *ELT* and *Df* on each distribution center

was different. Indicating that each DC was changing and adapting in order to satisfy the demand increase

and desired customer lead-time.

Table 9:

Scenarios and Experimental Design Results DC_AB in Liters

Distribution Center : DC_AB

Factor	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	-	+	-	+	-	+	-	+
ELT	1	2	2	3	3	4	4	5
Df	1.3	1.5	1.3	1.5	1.3	1.5	1.3	1.5

Design Point	Response: Inventory Levels			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
1	5157934	6331467	6610925	6438071
2	6331467	6610925	6438071	5205807
3	3222825	5530065	5009138	5167189
4	5530065	5009138	5167189	4152274

Effect	Scenario 1	Scenario 2	Scenario 3	Scenario 4
e_{ELT}	1740387	-120734	-7402	-1123589
e_{Df}	-1368256	-1201595	-1436335	-1162208
$e_{ELT \, x \, Df}$	566853	-400192	165452	108674

The results of DC_AB shown in Table 10 demonstrate that both increase in *ELT* (expected demand lead-time) and D_f demand factor affect the inventory levels. The effect e_{ELT} is negative in Scenarios 2, 3 and 4. This means that when the expected customer lead-time factor (*ELT*) is increased the inventory levels decrease. The results showed that the inventory levels of the DC as the *ELT* increase were stable and the DC was replenishing slower as the *ELT* increased.

The opposite will apply for Scenario 1. Therefore, changing *ELT* from 1 to 2 in this DC_AB does not lower inventory levels in the period studied. The inventory increased because the DC had to replenish faster and at higher inventory levels in order to satisfy customer demand.

The effect e_{Df} is negative in all scenarios. Since the *ELT* x *Df* interaction is positive in Scenarios 1,3 and 4 , there is further indication that lower average inventory levels are observed by setting *ELT* and *Df* at opposite levels in DC_AB. The opposite will apply for scenario 2.

Table 10:

Scenarios and Experimental Design Results DC_CB in Liters

Distribution Center: DC_CB

Factor	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	-	+	-	+	-	+	-	+
ELT	1	2	2	3	3	4	4	5
Df	1.3	1.5	1.3	1.5	1.3	1.5	1.3	1.5

Design Point	Scenario 1	Scenario 2	Scenario 3	Scenario 4
	Response: Inventory Levels			
1	5157934	1883669	2263459	2012574
2	1883669	2263459	2012574	3084707
3	3222825	424351	495415	950520
4	424351	495415	950520	853889

Effect	Scenario 1	Scenario 2	Scenario 3	Scenario 4
e_{ELT}	-3036370	225427	102110	487751
e_{Df}	-1697214	-1613681	-1415049	-1646435
$e_{ELT \, x \, Df}$	237896	-154363	352995	-584382

The results of DC_CB shown in Table 10 demonstrate that both increase in *ELT* (expected demand lead-time) and D_f demand factor affect the inventory levels. The effect e_{ELT} is positive in Scenarios 2 to 4.

This means that when the expected customer lead-time factor (*ELT*) is increased the inventory levels increase. The results showed that the inventory levels of the DC as the *ELT* increase were not stable and the DC was replenishing faster as the *ELT* increased. This particular DC serves customers with longer distances therefore, its replenishment cycle varies.

The opposite applies to Scenario 1. The effect e_{Df} is negative in all scenarios. This means the demand factor D_f on inventory levels has an adverse effect on inventory levels.

Since the *ELT* x *Df* interaction is positive in Scenarios 1 and 3, but negative in Scenarios 2 and 4 there is no further indication that lower average inventory levels are observed by setting *ELT* and *Df* at opposite levels in DC_CB.

Table 11:

Scenarios and Experimental Design Results DC_MC in Liters

Distribution Center: DC_MC

Factor	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	-	**+**	**-**	**+**	**-**	**+**	**-**	**+**
ELT	1	2	2	3	3	4	4	5
Df	1.3	1.5	1.3	1.5	1.3	1.5	1.3	1.5

Design Point	Scenario 1	Scenario 2	Scenario 3	Scenario 4
	Response: Inventory Levels			
1	1188255	1493986	1305220	1198458
2	1493986	1305220	1198458	1221288
3	1292952	1260886	1263703	647323
4	1260886	1263703	647323	1277250

Effect	Scenario 1	Scenario 2	Scenario 3	Scenario 4
e_{ELT}	136833	-92974	-361571	326378
e_{Df}	-64202	-137309	-296326	-247587
$e_{ELT \, x \, Df}$	-168899	95792	-254809	303549

The results of DC_MC shown in Table 11 demonstrate that both increase in *ELT* (expected demand lead-time) and D_f demand factor affect the inventory levels. The effect e_{ELT} is positive in Scenarios 1 and 4.

This means that when the expected customer lead-time factor (*ELT*) is increased the inventory levels increase. The opposite applies to Scenarios 2 and 3. The effect e_{Df} is negative in all scenarios. This DC carries higher inventory levels and helps other DC satisfy demand when needed. Therefore, its replenishment cycle varies.

Since the *ELT* x *Df* interaction is positive in Scenarios 2 and 4, but negative in Scenarios 1 and 3, there is no further indication that lower average inventory levels are observed by setting *ELT* and *Df* at opposite levels in DC_CB.

Table 12:

Scenarios and Experimental Design Results DC_ML in Liters

Distribution Center: DC_ML

Factor	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	-	+	-	+	-	+	-	+
ELT	1	2	2	3	3	4	4	5
Df	1.3	1.5	1.3	1.5	1.3	1.5	1.3	1.5

Design Point	Scenario 1	Scenario 2	Scenario 3	Scenario 4
	Response: Inventory Levels			
1	105443	210936	7568	17172
2	210936	7568	17172	14350
3	27383	12754	20122	500530
4	12754	20122	500530	16869

Effect	Scenario 1	Scenario 2	Scenario 3	Scenario 4
e_{ELT}	45432	-98000	245006	-243241
e_{Df}	-138120	-92814	247956	242938
$e_{ELT \, x \, Df}$	-60061	105368	235402	-240419

The results of DC_ML shown in Table 12 demonstrate that both increase in *ELT* (expected demand lead-time) and D_f demand factor affect the inventory levels. The effect e_{ELT} is positive in Scenarios 1 and 3.

This means that when the expected customer lead-time factor (*ELT*) is increased the inventory levels increase. The opposite applies to Scenarios 2 and 4. The effect e_{Df} is negative in Scenarios 1 and 2. This is explained similarly to DC_MG. This DC carries higher inventory levels and helps other DC satisfy demand when needed. Therefore, its replenishment cycle varies.

This means the demand factor D_f on inventory levels has a negative inversely effect on inventory levels. If D_f increases, the inventory levels decrease. The opposite applies for Scenarios 3 and 4.

Since the *ELT* x *Df* interaction is positive in Scenarios 2 and 3, but negative in Scenarios 1 and 4 there is no further indication that lower average inventory levels are observed by setting *ELT* and *Df* at opposite levels in DC_ML.

Table 13:

Scenarios and Experimental Design Results DC_MG in Liters

Distribution Center: DC_MG

Factor	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	-	+	-	+	-	+	-	+
ELT	1	2	2	3	3	4	4	5
Df	1.3	1.5	1.3	1.5	1.3	1.5	1.3	1.5

Design Point	Scenario 1	Scenario 2	Scenario 3	Scenario 4
	Response: Inventory Levels			
1	2345525	2161989	2392350	2204257
2	2161989	2392350	2204257	1440273
3	1367423	2490899	1986723	1422163
4	2490899	1986723	1422163	1768793

Effect	Scenario 1	Scenario 2	Scenario 3	Scenario 4
e_{ELT}	469970	-136908	-376327	-208677
e_{Df}	-324596	-38359	-593861	-226787
$e_{ELT \; x \; Df}$	653506	-367269	-188234	555308

The results of DC_MG shown in Table 13 demonstrate that both increase in *ELT* (expected demand lead-time) and D_f demand factor affect the inventory levels. The effect e_{ELT} is negative in Scenarios 2 to 4. This means that when the expected customer lead-time factor (*ELT*) is increased the inventory levels decrease. The results showed that the inventory levels of the DC as the *ELT* increase were stable and the DC was replenishing slower as the *ELT* increased.

The opposite applies to Scenario 1. The effect e_{Df} is negative in all scenarios. This means the demand factor D_f on inventory levels has a negative inversely effect on inventory levels. If D_f increases, the inventory levels decrease.

Since the *ELT* x *Df* interaction is positive in Scenarios 1 and 4, but negative in Scenarios 2 and 3, there is no further indication that lower average inventory levels are observed by setting *ELT* and *Df* at opposite levels in DC_CB.

4.2.3   Peak Season Demand Increase and Customer Expected Leadtime

The results for the scenarios show that the DCs that are most affected by the increase in customer demand (30% and 50%), in addition to the pressure to satisfy a targeted customer lead-time (ELT), are DC_CB and DC_ML. The supply chain network adapted to satisfy the increase in demand and lead-time, satisfying the additional supply from DC_MC and DC_ML. This occurred because both of these DC's kept more inventory levels than the other DC's making it possible to have additional inventory to satisfy the customers demand.

Consistently, in all scenarios both of these DCs experienced the highest demand. The results also showed that in each scenario the customers were served by more than one DC (see Table 14).

The DCs that serviced the same customers consistently were DC_MC and DC_ML, DC_MG and DC_ML, DC_CB and DC_AB.

Table 14:

Number of Customers served per DC

Source	Baseline	Replications									
		ELT = 1 day		*ELT = 2 days*		*ELT = 3 days*		*ELT = 4 days*		*ELT =5 days*	
		1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
DC_AB	33	33	36	36	36	33	36	33	46	33	33
DC_CB	11	37	47	47	86	42	57	47	64	37	54
DC_MC	15	19	19	19	31	19	35	19	40	19	35
DC_ML	31	50	53	53	57	49	63	50	53	50	53
DC_MG	9	50	48	49	50	51	64	40	58	50	50
Total	99	189	203	204	260	194	255	189	261	189	225

The inventory levels from each scenario are used as inputs to the AI model. The AI model will predict the inventory levels from the simulation and results are presented in Section 4.3. The inventory levels predicted by AI were then used to run scenarios again and results are presented in Section 4.4.

4.3 Gradient Boosting Results

Python language was used to run the AI algorithm. XGBoost was used for time series forecasting of inventory levels and the dataset was transformed into a supervised learning problem. Additionally, a special technique for evaluating the model called walk-forward validation was used.

4.3.1 Data Transformation

The inventory levels generated in the simulation model were used as input to the time series forecast model. The dataset was transformed as a supervised learning problem by applying the lag method. The value at the previous time step was used to predict the value at the next time-step.

4.3.2 Walk-forward validation

Walk-forward validation technique was used to split the data in training and test sets. A cut point of the last 14 days was selected. All data except the last 14 days was used for training, and the last 14 days was used for testing. The decision of using 14 days was based on the assumption that for a real-life application supply chain manager can use AI for their tactical decisions, for example predict the next two weeks of inventory levels in each distribution center.

4.3.3 XGBoost Results

The prediction results for each DC are explained in Figures 16-20 and Tables 15-19. The MAE (Mean Absolute error) and MAPE (Mean Absolute Percentage Error) metric were used to evaluate the predictive capability of XGBoost.

The overall algorithm performance yielded a 59,456 MAE and a 4.37% MAPE. This result indicates that XGBoost a has very good performance for predicting inventory levels as a time series forecast model. It is important to note that since the inventory policy defined in the simulation was a min-max policy, this produces stable inventory patterns that can be identified by the XGBoost algorithm. Therefore, the model will need to be tested on the real-life inventory patterns of BeverageCo.

The forecast results for distribution center DC_AB are shown in Figure 16 and Table 15. The XGBoost model delivered a low MAE of 79,159 and a MAPE of 4.38%.

Figure 16:

Line Graph of XGBoost Forecasting Results for DC_AB in Liters
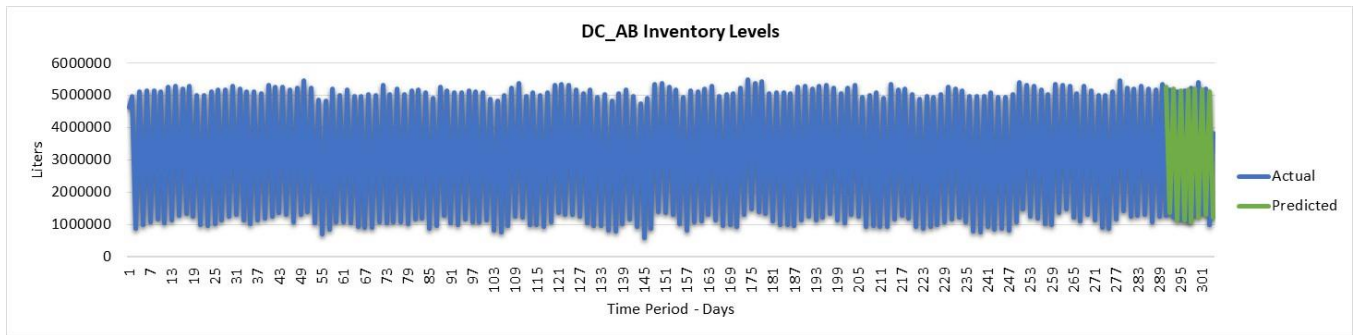
Table 15:

XGBoost Forecasting Metrics for DC_AB in Liters

Time	Actual	Predicted	Absolute Error	Absolute % Error
1	5,336,526	5,251,274	85,252	1.6%
2	1,264,934	1,354,303	89,369	7.1%
3	5,173,921	5,192,447	18,525	0.4%
4	1,206,163	1,131,860	74,303	6.2%
5	5,113,985	5,149,135	35,150	0.7%
6	1,149,013	1,127,796	21,217	1.8%
7	5,137,972	5,177,753	39,781	0.8%
8	1,096,644	1,079,558	17,086	1.6%
9	5,232,835	5,199,922	32,913	0.6%
10	1,315,996	1,209,194	106,801	8.1%
11	5,405,408	5,182,849	222,559	4.1%
12	1,348,214	1,304,180	44,034	3.3%
13	5,199,575	5,107,373	92,202	1.8%
14	981,094	1,210,130	229,036	23.3%

MAE	79,159
MAPE	4.38%

The forecast results for distribution center DC_CB are shown in Figure 17 and Table 16. The XGBoost model delivered a low MAE of 36,460 and a MAPE of 0.52%.

Figure 17:

Line Graph of XGBoost Forecasting Results for DC_CB in Liters

Table 16:

XGBoost Forecasting Metrics for DC_CB in Liters

Time	Actual	Predicted	Absolute Error	Absolute % Error
1	7,234,019	7,229,773	4,246	0.1%
2	6,410,265	6,394,379	15,886	0.2%
3	8,947,670	8,915,870	31,800	0.4%
4	8,113,781	8,095,985	17,797	0.2%
5	7,284,914	7,264,545	20,369	0.3%
6	6,452,374	6,458,597	6,223	0.1%
7	8,891,925	8,923,088	31,163	0.4%
8	8,051,629	8,080,858	29,229	0.4%
9	7,223,281	7,228,534	5,253	0.1%
10	6,408,414	6,418,126	9,712	0.2%
11	8,965,703	8,936,668	29,035	0.3%
12	8,109,697	8,086,689	23,008	0.3%
13	7,245,578	7,290,410	44,832	0.6%
14	6,182,144	6,424,035	241,891	3.9%

MAE	36,460
MAPE	0.52%

The forecast results for distribution center DC_MC are shown in Figure 18 and Table 17. The XGBoost

model delivered a low MAE of 48,190 and a MAPE of 6.20%.

Figure 18:

Line Graph of XGBoost Forecasting Results for DC_MC in Liters
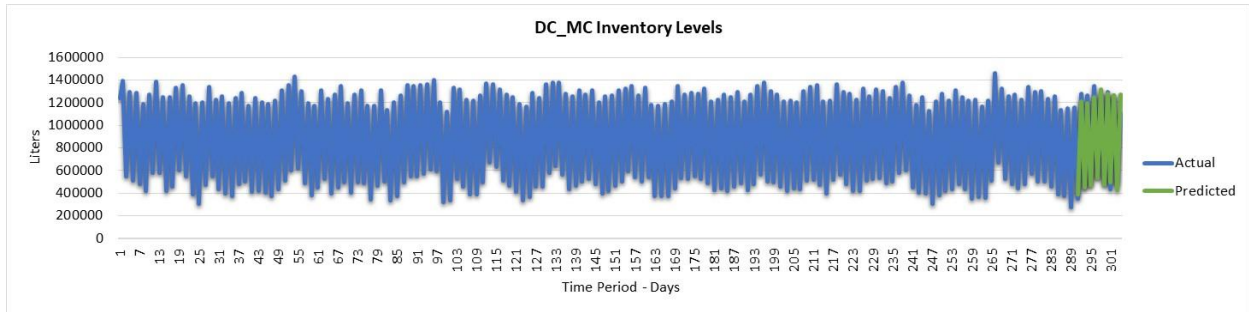
Table 17:

XGBoost Forecasting Metrics for DC_MC in Liters

Time	Actual	Predicted	Absolute Error	Absolute % Error
1	350,314	389,660	39,345	11%
2	1,277,402	1,213,468	63,934	5%
3	454,136	451,421	2,715	1%
4	1,266,604	1,192,843	73,761	6%
5	483,801	462,987	20,814	4%
6	1,348,034	1,249,348	98,686	7%
7	547,769	537,981	9,789	2%
8	1,278,762	1,314,706	35,945	3%
9	491,376	471,652	19,724	4%
10	1,290,401	1,278,893	11,508	1%
11	435,873	500,262	64,389	15%
12	1,264,833	1,266,166	1,332	0%
13	492,751	430,270	62,481	13%
14	1,101,258	1,271,490	170,232	15%

MAE	48,190
MAPE	6.20%

The forecast results for distribution center DC_MG are shown in Figure 19 and Table 18. The XGBoost model delivered a low MAE of 105,437 and a MAPE of 9.06%.

Figure 19:

Line Graph of XGBoost Forecasting Results for DC_MG in Liters
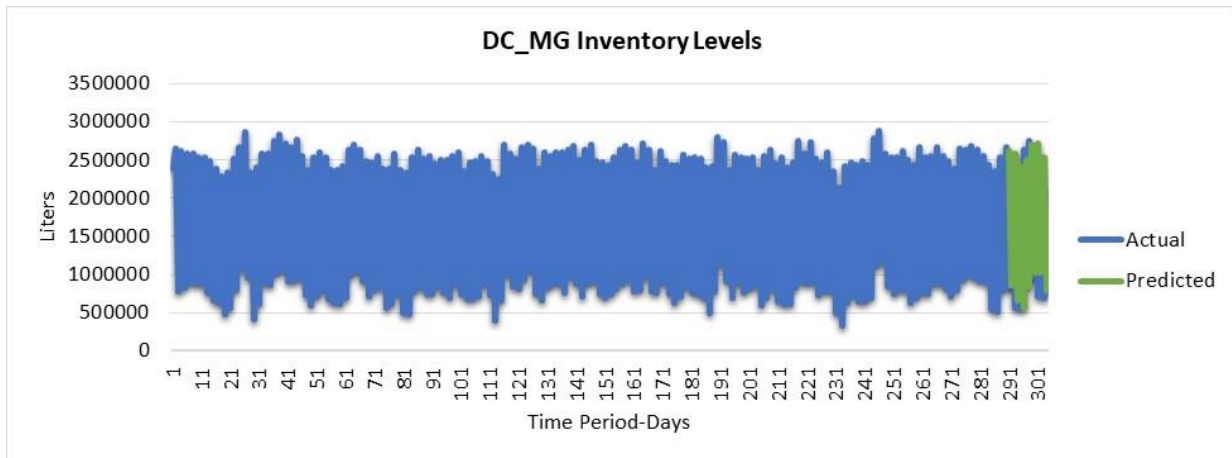
Table 18:

XGBoost Forecasting Metrics for DC_MG in Liters

Time	Actual	Predicted	Absolute Error	Absolute % Error
1	2,671,292	2,618,202	53,090	2.0%
2	850,934	860,753	9,819	1.2%
3	2,422,332	2,578,032	155,700	6.4%
4	555,672	664,443	108,771	19.6%
5	2,303,886	2,427,160	123,274	5.4%
6	558,978	553,751	5,227	0.9%
7	2,632,779	2,503,201	129,578	4.9%
8	918,837	835,465	83,373	9.1%
9	2,748,652	2,699,688	48,963	1.8%
10	973,399	1,019,728	46,328	4.8%
11	2,573,959	2,719,362	145,403	5.6%
12	706,310	1,022,408	316,098	44.8%
13	2,377,164	2,530,532	153,368	6.5%
14	691,629	788,755	97,127	14.0%

MAE	**105,437**
MAPE	**9.06%**

The forecast results for distribution center DC_MG are shown in Figure 20 and Table 19. The XGBoost model delivered a low MAE of 28,032 and a MAPE of 1.69%.

Figure 20:

Line Graph of XGBoost Forecasting Results for DC_ML in Liters

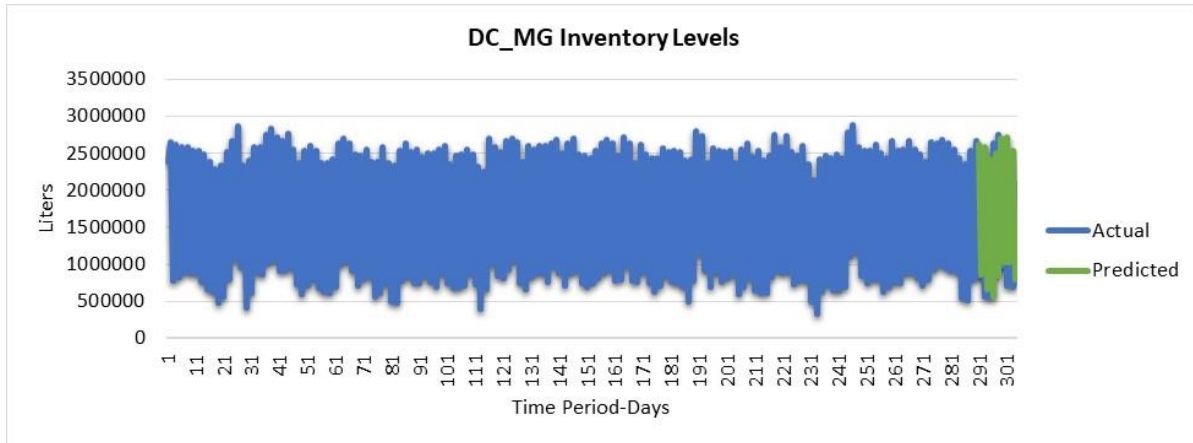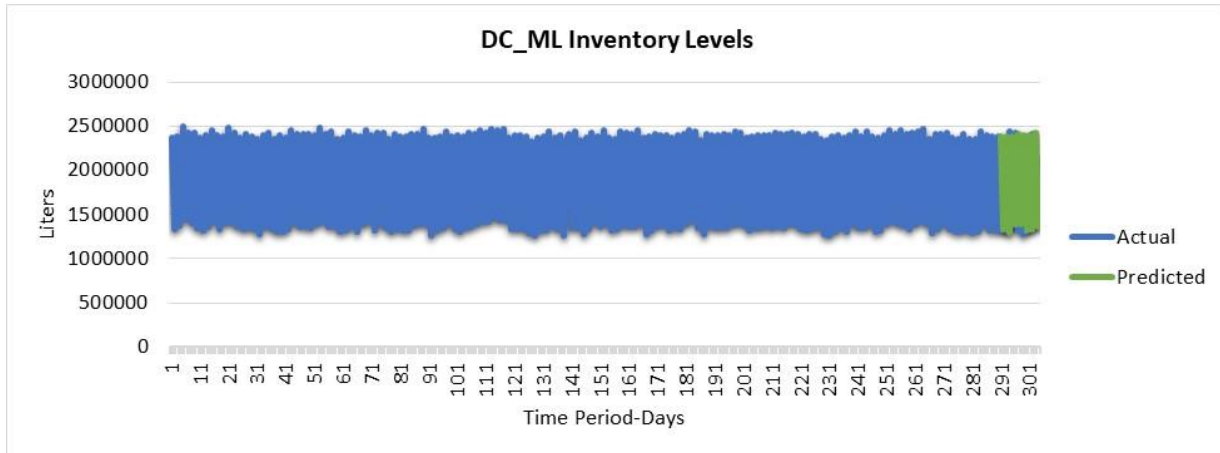**Table 19:**

XGBoost Forecasting Metrics for DC_ML in Liters

Time	Actual	Predicted	Absolute Error	Absolute % Error
1	2,392,780	2,381,231	11,550	0.5%
2	1,320,235	1,339,797	19,561	1.5%
3	2,354,549	2,384,045	29,496	1.3%
4	1,325,023	1,287,905	37,119	2.8%
5	2,444,245	2,385,437	58,808	2.4%
6	1,370,077	1,393,384	23,307	1.7%
7	2,426,445	2,416,700	9,746	0.4%
8	1,322,040	1,384,958	62,919	4.8%
9	2,361,365	2,405,242	43,877	1.9%
10	1,274,221	1,318,872	44,652	3.5%
11	2,388,092	2,395,324	7,232	0.3%
12	1,336,655	1,329,159	7,496	0.6%
13	2,412,371	2,429,285	16,914	0.7%
14	1,375,801	1,356,034	19,767	1.4%

MAE	**28,032**
MAPE	**1.69%**

4.4 Feedback Loop Results

A feedback loop was developed between AI model and digital twin for information sharing. AI model predicted on hand inventory at each DC and it was sent to the digital twin simulation model.

The following four scenarios were selected for the feedback modelling. As demand is peaked in November and December and varies month over month, to consider monthly peaks and variation in demand, we selected four data points from the months of November and December. We could verify the results for one-year demand; however, we concentrated on these four data points to increase the testing efficiency. ELT and Demand coefficient were selected randomly to validate results in multiple cases.

Scenario 1: Nov, Day 1, Demand Coefficient 1.3, ELT 2

Scenario 2: Nov, Day 15, Demand Coefficient 1.3, ELT 3

Scenario 3: Dec, Day 1, Demand Coefficient 1.5, ELT 4

Scenario 4: Dec, Day 15, Demand Coefficient 1.5, ELT 5

Above scenarios were selected to consider a mix of variability within different input parameters. The predicted inventory level for scenarios was shared as input with the simulation model. The simulation model was then run to validate if predicted inventory levels were sufficient to fulfill DC demand.

Results of simulation run shows that the service levels are 100% for all DC's. We identified two reasons for such service levels:

1. Predictions from AI algorithm – Inventory numbers predicted by AI algorithm ensure complete fulfillment of the DC demand from the on-hand inventory at the respective DC.
2. Min-max inventory policy – This inventory policy ensures that whenever, inventory levels falls below a certain threshold - Min, a replenishment is made to reach target - Max.

Figure 21 shows the monthly variation in demand for all the distribution centers. AI predicted inventory was sufficient to fulfill demand at all DC's, leading to a service level of 100%. In the simulation output, a shipment is present against each DC demand from the same DC.

Figure 21:

DC Demand (Shipment) for different scenarios in Liters
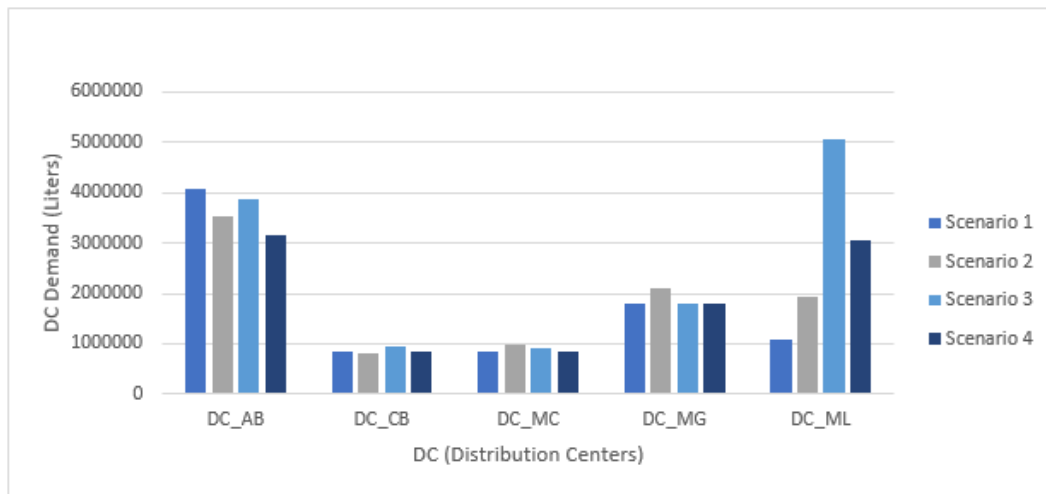
Y-axis shows DC demand, X-axis display 5 distribution centers, and as explained above, output of four identified scenarios was used for the simulation model.

Further, as shown in Figure 22, each DC is maintaining a day of supply between 1 to 2 days, with exception of DC_CB where for Scenario 1, days of supply is 2.1. Note that the simulation was run with daily demand, and model was expected to replenish DC inventory at a daily frequency. Since, the Min-Max inventory policy with Min and Max approximating 3 and 4 days of supply respectively is used, model was expected to have around 2 to 4 days of supply at each DC. Days of supply, right after a shipment is received at DC is around 4 days. Similarly, days of supply calculated right before shipment is received at DC is approximately 2 days. This is true for cases with low demand variability and a daily inventory review. However, demand variability will influence the days of supply at DC. In the results displayed in Figure 22, note that the days of supply was calculated after fulfilling the daily demand, or from ending on hand quantity.

Figure 22:

DC Days of Supply for different scenarios
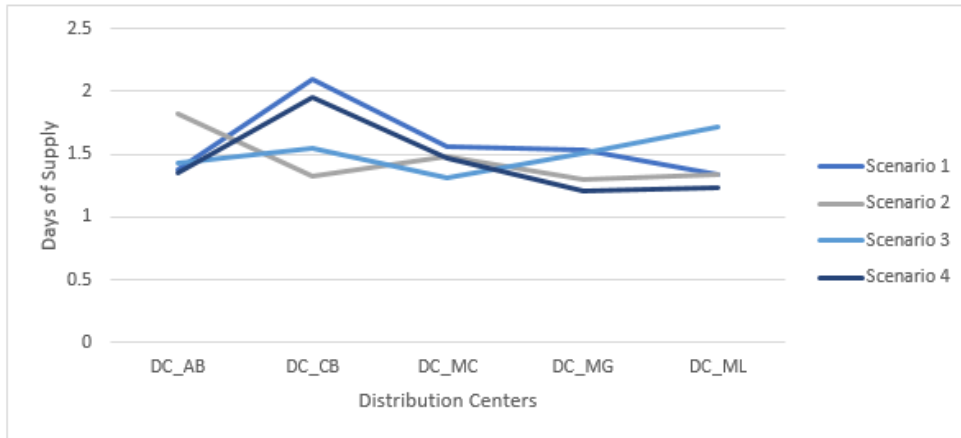
Y-axis shows DC days of supply, X-axis display 5 distribution centers, and as explained above, output of four identified scenarios was used for simulation model.

4.5 Hypothesis Analysis

The hypothesis of the capstone was that a learning system can be built through a feedback between simulation and AI model to understand patterns for the adaptability of manufacturing operations and the whole supply chain. This was demonstrated through the creation of what-if scenarios and experimental design results. The results showed that the distribution centers behaved differently under each change of increase in demand and expected customer lead-time. The feedback loop between AI and simulation allowed for further insights on how can supply chain managers better take decisions. The simulation provided inventory levels visibility and the feedback loop allowed analysis on how the days of inventory at each DC can change to allow 100% service level.

5 DISCUSSION AND MANAGERIAL RECOMMENDATIONS

This section discusses the results of our study conducted for BeverageCo and offers managerial recommendations that the company should consider for the creation of a digital twin.

This capstone project presented a conceptual model of a digital twin and explored opportunities of using digital twin to obtain operational efficiencies. In this section, we provide recommendations to BeverageCo to implement a digital twin. Based on the conceptual model of a digital twin discussed throughout this document, the following discussion and recommendations are categorized into three sub sections : Digital twin, Artificial intelligence model, and Feedback loop.

5.1 Digital Twins

The results of this research provide evidence that DTs can add value to BeverageCo since it allows to better understand the interactions between each component of the supply chain network and the behavior of the supply chain under uncertainties such as customer demands and expectations. As mentioned in Section 4.3, DTs provide BeverageCo with a feedback loop created between simulation and AI that result in a learning system providing the ability to run simulations under a number of what-if scenarios such as varying demand and lead times to understand the supply chain behavior under uncertain circumstances. As an example, in the scenario where demand was increased to 1.3 or 1.5 times the average demand, secondary DCs (not the closest DC) were used to fulfill demand, resulting in DC interactions across supply chain. Additionally, increasing ELT provided flexibility to supply chain by allowing multiple distribution centers (DC) to fulfill customer demand. In experiments where ELT is increased, we discovered that changing ELT from 2 to 3, from 3 to 4, and from 4 to 5 leads to a reduction in inventory levels at few DCs, while increasing inventory levels at other DCs. This can be explained by the fact that multiple DCs' are now able to fulfill customer demand by interacting with each other. From this interaction, they identify

the opportunities to optimize inventory levels across the supply chain. These optimized inventory levels lead to higher service levels across BeverageCo's supply chain.

In the BeverageCo's existing supply chain network, the closest DC fulfills the customers' demand. However, the creation of a digital twin provides an opportunity to BeverageCo's SC managers to explore supply chain behavior under additional network configurations.

Focusing on this, BeverageCo should consider the creation of a digital twin to mimic the supply chain, making possible to understand supply chain behavior and take strategic decisions to better prepare upstream and downstream systems, processes, and human resources for different scenarios. For instance, in the above scenario, SC managers have to make arrangements with different supply chain players - with customers to increase lead time with manufacturing plants to adjust production and with distribution centers for storage and logistic needs.

Furthermore, DTs allows SC managers to understand supply chain behavior in disruption scenarios such as plant shutdowns or region lockdowns leading to supply chain network disconnect. In addition, another usage of DTs not discussed in this study, is real-time simulations. DTs can help answer operational questions such as, what is the best way to fulfill customer demand in case a truck, on the way to a customer location breaks or which manufacturing plants should increase production in case of a strike at a distribution center. This use of digital twin will equip BeverageCo SC managers with a powerful tool, allowing them to make insight driven informed decisions. Therefore, the creation of digital twin is a key element to achieve supply chain efficiencies and agility.

5.2 Artificial intelligence Model

An artificial intelligence (AI) model was created to provide supply chain managers an ability to foresee patterns in supply chain to improve decision-making, and make supply chains resilient. As mentioned in Section 4.4 and explained in Figures 16-20, AI model was used to predict DC inventory levels from

historical and simulation data obtained from DT. The algorithm performance yielded an overall MAPE of 4.37%, which indicates that XGBoost, an AI algorithm used for time series forecasting, has very good performance for predicting inventory levels as a time series forecast model. From these results, it is clear that AI model is able to predict inventory levels only based on the data shared by DT. Based on this finding, we recommend that BeverageCo use this AI model, together with DT model to predict future DC inventory levels.

Supply chain variables such as customer demand, production quantity, as well as, relatively fixed variables such as lead times, inventory policies, and production capacity can be used to predict inventory, demand, DC shipments, or logistic needs of future. For example, based on the inventory prediction model created in the study, BeverageCo SC users can create a similar AI model to predict monthly production quantities, warehouse space requirements or distribution workforce required. These predictions will, of course, change and improve by the continuous data feed from digital twin, and through self-learning by AI model.

5.3 Feedback Loop

A feedback loop was created between discrete simulation model and the AI model to create a data-driven learning system. Data from AI model was shared with digital twin to analyze the impact of AI predictions and provide SC managers opportunities to run what-if scenarios to understand supply chain response to the AI predictions. According to Section 4.5, results from the feedback loop demonstrate that the AI predicted inventory levels validates our hypothesis about service levels and days of supply at DCs. Using predicted inventory level from AI model, all the DC's were able to fulfill complete demand, and maintain more than 1 day of supply. BeverageCo SC managers can analyze AI predictions, and use DT again to run simulations and take corrective actions based on AI predicted inventory levels.

In addition, a feedback loop provides flexibility to BeverageCo SC managers to maintain predicted inventory levels at the DCs, or wait to come closer to the predicted time-period, and execute feedback loop again. If the decision is to wait, simulations will run with latest available data, and AI model will

regenerate predictions. New predictions are inclined to have better accuracy as they consider latest data from system. There is a tradeoff between taking action early and waiting to get better predictions, which should be considered before taking any ordering or shipping decision. Although, the lead time and safety stock information at manufacturing locations is not present to make precise recommendations, we suggest BeverageCo to run AI model on a weekly basis and share the output with digital twin. Based on our study of plant to DC and DC to customer lead times, one week is enough to ship inventory to any customer location.

One possible opportunity that would enrich the analysis is to include supply chain planning data such as data on safety stock and order interval is not available in ERP system. This data was not considered during analysis. In case the data becomes available, it can be included in the simulation model. This would allow the model predictions to mimic more closely the real-world supply chain complexities of BeverageCo. Moreover, the beverage company's new additions of plants and DC's can be considered if data is available. In the methodology section, we mentioned that some plants and DCs did not have information available in ERP system. In case this data is available in future, our study can be readily applied to them and these new locations can be added to model. Based on this, BeverageCo could consider providing managerial recommendations to new plants as well.

6 CONCLUSION

The research project began with an objective to understand how digital twins can add value to the supply chain network of a beverage industry. The study started with an initial hypothesis that digital twins can be used to develop a simulation model to understand patterns for the adaptability of manufacturing operations and the whole supply chain. To prove the hypothesis, we created a conceptual framework of digital twin and AI model, and developed a learning feedback loop between artificial intelligence algorithm and digital twin model.

First step towards building the conceptual framework was to create digital twin of BeverageCo supply chain. To create the digital twin, data from BeverageCo's ERP system was used. The data was analyzed to identify the variables that has similar characteristics in all manufacturing plants. Following this, a correlation analysis was performed to reduce the variables for clustering analysis. Clustering analysis was conducted to identify the manufacturing plants for digital twin.

We developed a Digital twin to replicate BeverageCo supply chain, and provided supply chain managers an ability to run simulations with varying demand and lead times. Simulation was run with different demand coefficients and different customer expected lead time (ELT) to explore the impact of changing demand expectations, and the impact of increasing ELT on supply chain. The results from simulation analysis demonstrate that by increasing ELT, BeverageCo supply chain becomes more flexible, allowing multiple distribution centers (DC) interact across supply chain to fulfill customer demand. This flexibility and interaction lead to higher collaboration between distribution centers, leading to higher customer service levels. This is an important conclusion for BeverageCo, as the digital twin provides insights to supply chain managers to better understand supply chain behavior under various conditions and future possibilities, and proactively plan and prepare for such situations.

In second step, we created an artificial intelligence (AI) model to provide supply chain managers predictions based on data from digital twin. The AI model was used to predict DC inventory levels from historical and simulation data obtained from DT. The AI algorithm results had a MAPE of 4.37% showing that the AI model developed has good performance for predicting inventory levels as a time series forecast model. An important conclusion that can be drawn from the results of AI model is showing the applicability of using an AI model within a digital twin in predicting supply chain variables.

Finally, a feedback loop was established between the AI model and the digital twin model to demonstrate that the predictions from the AI model can be fed to the digital twin model to create a data driven learning system. The output of AI model, DC inventory levels was shared with digital twin to understand if the predicted inventory is sufficient to fulfill DC demand. The feedback loop results validate another hypothesis that the AI predicted DC inventory levels are sufficient to fulfill customer demand from the selected DC. Feedback loop enables BeverageCo supply chain users take informed decisions by providing an integrated view of historical data, simulation data and the predictions from AI.

The main conclusion that can be drawn from this study is that a conceptual framework of digital twin and AI model can be developed, and can be utilized to foresee patterns in supply chain, and proactively take actions to resolve any bottlenecks and constraints. The conceptual framework created in the study allows BeverageCo SC managers analyze AI predictions, use digital twin to run what-if scenarios, and take corrective actions to bring supply chain efficiencies.

In this study, we presented a conceptual digital twin model, which can learn and adapt, for a beverage company. The developed methodology represents our efforts in systematically using latest digital technologies to bring efficiencies in supply chain.

From a managerial perspective, our study provides new insights and important implications of using digital twin in BeverageCo supply chain. Digital twin is a virtual image of BeverageCo's supply chain providing

real-time visibility of its supply chain by establishing a live connection with physical supply chain, allowing supply chain managers remote monitoring and control of the supply chain.

DT allows to run simulations on live data, making the simulation results mimic real-world conditions. In this study, we ran simulation with different demand and lead-time scenarios to understand the impact on inventory levels at distribution centers. Our recommendation to BeverageCo SC managers is to utilize DT to get end-to-end visibility of supply chain, which will provide them capability to identify potential bottlenecks and supply risks in their supply chain. For example, DT will provide visibility on late deliveries, about use of alternate mode of transports or use of secondary suppliers.

From the findings discussed in Section 4, we recommend SC managers to use DT to run simulations to understand supply chain behavior in different scenarios. DT can simulate multiple scenarios to derive both strategic and execution level decisions such as expanding production at a plant, or ship more to a particular distribution center. As discussed in Section 3, our conceptual model of digital twin consisted of an AI algorithm, producing output that is more intelligent and valuable. AI algorithm combines intelligence obtained from historical data, for example, about all the disruptions happened in past 2 years, with the simulation results, and provide SC managers better predictions.

In addition, DT provides SC managers capability to test, validate and refine assumptions. For instance, in this study, to test the impact of lead-time, we created a scenario by changing the customer expected lead-time from 4 to 5 days. The result was increased flexibility within supply chain, allowing multiple distribution centers interact with each other, and fulfill customer demand more optimally. This increased the overall customer service levels in supply chain. Hence, the result also lead us to a conclusion that use of DT increased the efficiency in supply chain, and enabled BeverageCo to make the right inventory decisions.

Further, another benefit of DT is in collaboration between different teams. By providing complete visibility of supply chain, DT helps in documenting and communicating supply chain processes to all the

stakeholders within the company. This leads to faster and informed decision-making within the organization.

The methodology we used in the study provides a systematic procedure to create this learning system.

For the overall strategy, we propose that BeverageCo work in three main areas to implement a digital twin learning model:

- Create a digital twin to replicate physical supply chain of BeverageCo. BeverageCo can decide to implement DT in steps by first implementing the virtual replica of physical supply chain. This will provide them end-to-end visibility. SC managers will need some time to understand new technology, change in processes and to adapt new solution. After realizing benefits from this version of DT, SC managers can move to create an artificial intelligence algorithm, which is the next step towards empowering DT.

- Once SC users adopt the first version of DT, develop an artificial intelligence model to provide supply chain managers an ability to foresee patterns in supply chain to improve decision-making. AI model will provide BeverageCo users prescriptive solutions, which can still be validated using digital twin before making changes in physical supply chain.

- Finally, to harness real power of digital twin, establish a feedback loop between the AI model and the simulation model to share predictions from the AI model with the simulation model to create a data driven learning system.

Providing a framework to create a digital twin learning system is the main conclusion of our research.

6.1 Future Research

Although in this study, we implemented multiple scenarios related to demand and lead-time variability, the same model can be used for other scenarios such as supply chain disruptions. Implementing the scenarios used in this study can be a good starting point towards the transformation journey of using digital twin. However, SC managers should equip the simulation model with all the other possibilities,

69

which may arise in supply chain. This will allow the AI model to learn different situations and provide predictions based on real-world situations.

In the conceptual model presented in this study, predictions and insights from AI model were shared with BeverageCo. SC managers review these predictions, run what-if analysis to interpret, and analysis impact of these predictions. From this analysis, they decide on feeding physical supply chain with AI predictions. As an extension to this model, we propose that future researchers and practitioners examine the possibilities of empowering digital twin by automating this entire process, and provide prescriptions after doing impact analysis. We believe digital twin can execute multiple what-if scenarios on its own, without any input from SC users.

Though, this study provides a useful reference for future researchers in examining the potential value of utilization of DTs in the FMCG industries, the conceptual framework and recommendations can be applied to similar interactions between individual components of other supply chains. We propose that the model we prepared can be used in other manufacturing industries as well. However, further research to establish the applicability of this model in other industries is required.

REFERENCES

Bevilacqua, M., Bottani, E., Ciarapica, F. E., Costantino, F., Di Donato, L., Ferraro, A., Mazzuto, G., Monteriù, A., Nardini, G., Ortenzi, M., Paroncini, M., Pirozzi, M., Prist, M., Quatrini, E., Tronci, M., & Vignali, G. (2020). Digital Twin Reference Model Development to Prevent Operators' Risk in Process Plants. *Sustainability*, *12*(3), 1088. https://doi.org/10.3390/su12031088

Bevilacqua, M., Bottani, E., Ciarapica, F. E., Costantino, F., Donato, L. D., Ferraro, A., Mazzuto, G., Monteriù, A., Nardini, G., Ortenzi, M., Paroncini, M., Pirozzi, M., Prist, M., Quatrini, E., Tronci, M., & Vignali, G. (2020). *Digital Twin Reference Model Development to Prevent Operators' Risk in Process Plants*. 17.

Blank, S. G. (2007). *The four steps to the epiphany: Successful strategies for products that win* (3rd ed). S. G. Blank.

Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in Industry*, *101*, 1–12. https://doi.org/10.1016/j.compind.2018.04.015

Cagle, K. (2019). *What Is Artificial Intelligence?* Forbes. https://www.forbes.com/sites/cognitiveworld/2019/08/20/what-is-artificial-intelligence/

Canedo, A. (2016). Industrial IoT lifecycle via digital twins. *Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, 1. https://doi.org/10.1145/2968456.2974007

Caro, F., & Sadr, R. (2019). The Internet of Things (IoT) in retail: Bridging supply and demand. *Business Horizons*, *62*(1), 47–54. https://doi.org/10.1016/j.bushor.2018.08.002

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Choi, T. Y., Dooley, K. J., & Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: Control versus emergence. *Journal of Operations Management*, *19*(3), 351–366. https://doi.org/10.1016/S0272-6963(00)00068-1

Crittenden, V., & Peterson, R. A. (2019). Digital Disruption: The Transdisciplinary Future of Marketing Education. *Journal of Marketing Education*, *41*(1), 3–4. https://doi.org/10.1177/0273475319825534

D'Amico, D., Ekoyuncu, J., Addepalli, S., Smith, C., Keedwell, E., Sibson, J., & Penver, S. (2019). Conceptual framework of a digital twin to evaluate the degradation status of complex engineering systems. *Procedia CIRP*, *86*, 61–67. https://doi.org/10.1016/j.procir.2020.01.043

Fuller, A., Fan, Z., & Day, C. (2020). *Digital Twin: Enabling Technologies, Challenges and Open Research*.

Giannoccaro, I. (2015). Adaptive supply chains in industrial districts: A complexity science approach focused on learning. *International Journal of Production Economics*, *170*, 576–589. https://doi.org/10.1016/j.ijpe.2015.01.004

Glaessgen, E., & Stargel, D. (2012, April 23). The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference
20th AIAA/ASME/AHS Adaptive Structures Conference
14th AIAA*. 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference
20th AIAA/ASME/AHS Adaptive Structures Conference
14th AIAA, Honolulu, Hawaii. https://doi.org/10.2514/6.2012-1818

Grieves, M. (2015). *Digital Twin: Manufacturing Excellence through Virtual Factory Replication*.

Gumani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., & Bhirud, S. (2017). *Forecasting of sales by using fusion of machine learning techniques*. 93–101. https://doi.org/10.1109/ICDMAI.2017.8073492

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for Formulating a Digital Transformation Strategy. *MIS Quarterly Executive*, *15*(2), 123–139.

Holland, J. H. (1992). Complex Adaptive Systems. *Daedalus*, *121*(1,), 17–30.

Holland, J. H. (2006). Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity*, *19*(1), 1–8. https://doi.org/10.1007/s11424-006-0001-z

Ivanov, D., Dolgui, A., & Sokolov, B. (Eds.). (2019). *Handbook of Ripple Effects in the Supply Chain* (Vol. 276). Springer International Publishing. https://doi.org/10.1007/978-3-030-14302-2

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). *Strategy, not Technology, Drives Digital Transformation*. 29.

Law, A. M. (2013). *Simulation modeling and analysis* (Fifth edition). McGraw-Hill Education.

Liu, Z., Meyendorf, N., & Mrad, N. (2018). *The role of data fusion in predictive maintenance using digital twin*. 020023. https://doi.org/10.1063/1.5031520

Long, Q., Lin, J., & Sun, Z. (2011). Modeling and distributed simulation of supply chain with a multi-agent platform. *International Journal of Advanced Manufacturing Technology*, *55*(9–12), 1241–1252. https://doi.org/10.1007/s00170-010-3148-7

Lu, W., Rui, Y., Yi, Z., Ran, B., & Gu, Y. (2020). A Hybrid Model for Lane-Level Traffic Flow Forecasting Based on Complete Ensemble Empirical Mode Decomposition and Extreme Gradient Boosting. *IEEE Access*, *8*, 42042–42054. https://doi.org/10.1109/ACCESS.2020.2977219

Madni, A., Madni, C., & Lucero, S. (2019). Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems*, *7*(1), 7. https://doi.org/10.3390/systems7010007

Massa, L., & Tucci, C. (2013). *Business Model Innovation* (BOOK_CHAP). Oxford Handbook of Innovation Management; Oxford University Press. http://infoscience.epfl.ch/record/201604

McCarthy, I. P. (2003). Technology management a complex adaptive systems approach. *International Journal of Technology Management*, *25*(8), 728. https://doi.org/10.1504/IJTM.2003.003134

Merali, Y., & Allen, P. (2011). *Complexity and systems thinking* (pp. 31–52). https://doi.org/10.4135/9781446201084.n1

Milligan, G. W., & Cooper, M. C. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometriky*, *50*(2), 159–179.

Modoni, G. E., Caldarola, E. G., Sacco, M., & Terkaj, W. (2019). Synchronizing physical and digital factory: Benefits and technical challenges. *Procedia CIRP*, *79*, 472–477. https://doi.org/10.1016/j.procir.2019.02.125

Nair, A., & Reed-Tsochas, F. (2019). Revisiting the complex adaptive systems paradigm: Leading perspectives for researching operations and supply chain management issues. *Journal of Operations Management*, *65*, 80–92. https://doi.org/10.1002/joom.1022

Nilsson, F., & Darley, V. (2006). On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings: Experiences from a packaging …. *International Journal of Operations & Production Management*, *26*. https://doi.org/10.1108/01443570610710588

Parrott, A., & Warshaw, L. (2017). *Industry 4.0 and the digital twin technology | Deloitte Insights*. https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html

Pathak, S. D., Day, J. M., Nair, A., Sawaya, W. J., & Kristal, M. M. (2007). Complexity and Adaptivity in Supply Networks: Building Supply Network Theory Using a Complex Adaptive Systems Perspective*: Complexity and Adaptivity in Supply Networks. *Decision Sciences*, *38*(4), 547–580. https://doi.org/10.1111/j.1540-5915.2007.00170.x

Purdy, M., Eitel-Porter, R., Krüger, R., & Deblaere, T. (2020). *How Digital Twins Are Reinventing Innovation*. 9.

Qi, Q., Tao, F., Zuo, Y., & Zhao, D. (2018). Digital Twin Service towards Smart Manufacturing. *Procedia CIRP*, *72*, 237–242. https://doi.org/10.1016/j.procir.2018.03.103

Reeves, M., & Deimler, M. (2011, July 1). Adaptability: The New Competitive Advantage. *Harvard Business Review*, *July–August 2011*. https://hbr.org/2011/07/adaptability-the-new-competitive-advantage

Rosen, R., von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About The Importance of Autonomy and Digital Twins for the Future of Manufacturing. *IFAC-PapersOnLine*, *48*(3), 567–572. https://doi.org/10.1016/j.ifacol.2015.06.141

Surana, A., Kumara *, S., Greaves, M., & Raghavan, U. N. (2005). Supply-chain networks: A complex adaptive systems perspective. *International Journal of Production Research*, *43*(20), 4235–4265. https://doi.org/10.1080/00207540500142274

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2415. https://doi.org/10.1109/TII.2018.2873186

Tao, Fei, Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2017). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, *94*(9–12), 3563–3576. https://doi.org/10.1007/s00170-017-0233-1

Tozanlı, Ö., Kongar, E., & Gupta, S. M. (2020). Trade-in-to-upgrade as a marketing strategy in disassembly-to-order systems at the edge of blockchain technology. *International Journal of Production Research*, *58*(23), 7183–7200. https://doi.org/10.1080/00207543.2020.1712489

Ustundag, A., & Cevikcan, E. (2018). *Industry 4.0: Managing The Digital Transformation*. Springer International Publishing. https://doi.org/10.1007/978-3-319-57870-5

Vachálek, J., Bartalsky, L., Rovný, O., Sismisova, D., Morhac, M., & Loksik, M. (2017). *The digital twin of an industrial production line within the industry 4.0 concept*. 258–262. https://doi.org/10.1109/PC.2017.7976223

Wang, Y., & Guo, Y. (2020). Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications*, *17*(3), 205–221. https://doi.org/10.23919/JCC.2020.03.017

Weyer, S., Meyer, T., Ohmer, M., Gorecky, D., & Zühlke, D. (2016). Future Modeling and Simulation of CPS-based Factories: An Example from the Automotive Industry. *IFAC-PapersOnLine*, *49*(31), 97–102. https://doi.org/10.1016/j.ifacol.2016.12.168

Zhao, K., Zuo, Z., & Blackhurst, J. (2019). Modelling supply chain adaptation for disruptions: An empirically grounded complex adaptive systems approach. *Journal of Operations Management*, *65*, 190–212. https://doi.org/10.1002/joom.1009