**ANALYSIS OF DATACO SUPPLY CHAIN**

A project report submitted for the degree of

**MASTERS**

In

**DATA SCIENCE**

by

**IBRAHIM KHALEELULLA MOHAMMAD (323002312)**

&

**KARTHIKEYA PONNEKANTI (321233913)**

Under the guidance of

**Dr. WASSNAA AL-MAWEE**

professor

**A picture containing text, sign

Description automatically generated**

**Department of Computer Science**

**Western Michigan University**

**Kalamazoo MI 49008-5343 USA**

**(269) 276-3420**

**MOTIVATION**

The DataCo Supply Chain Project aims to address the challenges and complexities of modern supply chain management using advanced data-driven technologies. Supply chains today are often global and highly intricate, involving numerous stakeholders, processes, and transactions. Managing these complex supply chains efficiently and effectively is crucial for businesses to maintain their competitive edge, optimize costs, improve customer satisfaction, and meet sustainability goals. However, traditional supply chain management approaches may fall short in the face of rapidly changing market dynamics, increasing customer demands, and growing environmental concerns.

The DataCo Supply Chain Project seeks to leverage data and advanced technologies to transform supply chain management, making it more agile, transparent, and sustainable. By harnessing the power of data analytics, machine learning, and blockchain, the project aims to provide real-time visibility into the supply chain, enable predictive analytics for demand forecasting, optimize inventory management, streamline logistics and transportation, enhance supplier collaboration, and promote sustainability through traceability and accountability.

**OVERVIEW**

The DataCo Supply Chain Project is a comprehensive initiative that encompasses various aspects of supply chain management. It involves the following key components:

1. Data Analytics: The project utilizes advanced data analytics techniques to collect, analyze, and process large volumes of data generated across the supply chain, including data from suppliers, manufacturers, logistics providers, retailers, and customers. By harnessing the power of data, the project aims to gain insights into supply chain operations, identify patterns, trends, and anomalies, and make data-driven decisions to optimize supply chain performance.
2. Machine Learning: The project incorporates machine learning algorithms to build predictive models for demand forecasting, inventory optimization, and transportation optimization. These models leverage historical data, real-time data, and external factors such as market conditions, weather patterns, and geopolitical events to generate accurate forecasts and recommendations for supply chain planning and execution.
3. Blockchain: The project employs blockchain technology to create a transparent and secure supply chain ecosystem. Blockchain enables immutable and transparent recording of transactions, ensuring trust and accountability among supply chain participants. The project leverages blockchain to enable end-to-end traceability, track and verify product provenance, enhance supplier collaboration, and ensure compliance with sustainability and regulatory requirements.
4. Sustainability: The project emphasizes sustainability as a core principle of modern supply chain management. It aims to promote environmentally responsible practices, such as reducing carbon emissions, optimizing transportation routes, minimizing waste, and promoting ethical sourcing. The project uses data and technology to measure, monitor, and report on sustainability performance across the supply chain, enabling businesses to make informed decisions and demonstrate their commitment to sustainability to customers, investors, and other stakeholders.

**REFERENCE:**

1. <https://www.kaggle.com/code/cheukhangtse/dataco-supply-chain-analysis>
2. <https://github.com/ktlszn/dataco-supply_chain>

**INITIAL QUESTIONS:**

1. Finding the null values and then deleting it.
2. Creating Plot on Order item discount distribution.
3. Count the type of transaction with the previous data then summarize with a plot.
4. Calculating number of average number of days to deliver and Create a plot.
5. Creating a bar graph on Order status.
6. Creating a plot on number of orders from different product catogeries between shipping modes.
7. Calculating the total number of orders from different product category.
8. Finding the summary of the project.

**DATA:**

A typical data set for a supply chain project like DataCo may include various types of data, such as:

1. Sales and demand data: This could include historical sales data, customer orders, and demand forecasts. It may also include data on product demand variability, seasonality, and trends, which can be used for demand forecasting and inventory optimization.
2. Supplier data: This could include information about suppliers, such as their contact details, location, capabilities, capacity, and performance metrics. It may also include data on supplier reliability, lead times, and quality of goods or services provided.
3. Transportation and logistics data: This could include data on transportation routes, carriers, shipping times, transportation costs, and logistics performance metrics. It may also include data on transportation modes, shipment tracking, and delivery status, which can be used for optimizing transportation and logistics operations.
4. Financial data: This could include data on costs, pricing, margins, and profitability. It may also include data on payment terms, invoice processing times, and financial performance metrics, which can be used for financial analysis and cost optimization.
5. Other relevant data: Depending on the specific requirements of the DataCo Supply Chain Project, there may be other types of data included in the dataset, such as customer data, product data, quality data, regulatory data, and market data, among others.

**Exploratory Data Analysis:**

* In this project we have used ggplot2, dplyr, tidyr, RColorBrewer, plotrix, lattice packages to generate plot/ run the code. We have deleted NA values in the dataset. We have also used Data Wrangling to segregrate columns to generate the plots. Generated shinyapp and uploaded into the Git hub repository.

**Narrative and Summary:**

* We found out the time taken for delivery is 2-3days from the exploratory analysis can be observed in the Average Time Delivery Plot.
* From the Count of Mode of Transaction, Debit card transactions are more.
* We have observed Order Item Discount per count