**EDA Document**

**Optimization in Supply Chain Management**

**Date 29-8-2023**

# **Introduction: Optimization in Supply Chain Management**

Supply chain management is a complex and dynamic process that spans from raw material procurement to the final delivery of goods to customers. In today's competitive landscape, businesses recognize the need to optimize their supply chain operations to gain a competitive edge.

* **Efficient Inventory Management**: Maintaining the right level of inventory to meet customer demand while minimizing carrying costs and the risk of stockouts.
* **Effective Demand Forecasting**: Accurately predicting future demand patterns to ensure that products are available when needed, minimizing excess inventory and waste.
* **Transportation and Logistics Optimization**: Efficiently managing the movement of goods, choosing optimal routes, and selecting appropriate transportation modes to reduce costs and delivery times.
* **Supplier Collaboration**: Collaborating closely with suppliers to ensure a reliable supply of materials and components.
* **Data-Driven Decision-Making**: Leveraging data analytics to make informed decisions, identify bottlenecks, and continuously improve supply chain processes.

In this EDA project, we aim to uncover valuable insights from our supply chain data to support optimization efforts. By examining historical data, identifying patterns, and understanding correlations, we seek to make data-driven recommendations that can lead to more efficient supply chain management.

The following sections will delve into the details of our EDA, providing insights, visualizations, and actionable recommendations to enhance supply chain operations.

**Data Sources**:

The past is your treasure trove of knowledge. Historical records are your time capsules, preserving the legacy of your supply chain's evolution. They reveal trends, patterns, and lessons learned.

**Data Preparation**:

Certainly, data preparation is a crucial step in any exploratory data analysis (EDA) process. It involves cleaning and preprocessing the data to ensure that it's ready for analysis. Here are the steps you might take to clean and preprocess supply chain management data using SQL queries and Python:

1. Data Import:

SQL: In SQL, you would typically start by importing the data from your data source into a database or a temporary table.

Python: In Python, you can use libraries like Pandas to read data from various formats (e.g., CSV, Excel, databases) into a Data Frame.

2. Handling Missing Data:

SQL: Use SQL queries to identify missing values in your dataset. You can use queries like `SELECT COUNT (\*) FROM table WHERE column name IS NULL` to count missing values.

Python: In Python, Pandas provide functions like `insula ()` and `fillna()` to identify and handle missing data. You can use `dropna()` to remove rows with missing values or `fillna()` to fill missing values with appropriate values (e.g., mean, median).

3. Handling Outliers:

SQL: Identify outliers using SQL queries with aggregation functions or window functions. For example, you can use `SELECT column name FROM table WHERE column name > Q3 + 1.5 \* IQR` to find outliers.

Python: In Python, libraries like Pandas and NumPy are useful for detecting and handling outliers. You can use statistical methods or visualization techniques to identify outliers and decide whether to remove or transform them.

4. Data Transformation:

SQL: SQL queries can be used for various data transformations, including aggregations, filtering, and joining with other tables. For example, you might calculate aggregate statistics like sums, averages, or counts.

Python: Pandas provide powerful data transformation capabilities. You can use functions like `group by () `, `pivot () `, and `merge () ` to transform and combine data. Additionally, you can apply custom functions to manipulate data.

**Data Exploration:**

**That's great to hear that you've performed various data exploration techniques, including boxplots, histograms, bar plots, and Sweet viz. These visualizations and tools are essential for gaining insights into your supply chain management data. Let's briefly discuss the significance of each of these techniques in the context of data exploration:**

**1. Boxplot (Box-and-Whisker Plot):**

**Purpose: Boxplots are excellent for visualizing the distribution of numerical variables and identifying potential outliers.**

**Insights: They provide a summary of key statistics, such as the median, quartiles, and potential outliers, making it easier to understand the central tendency and spread of your data.**

**2. Histogram:**

**Purpose: Histograms are used to visualize the distribution of a single numerical variable.**

**Insights: They show the frequency or density of data points within different bins or intervals. Histograms help identify data skewness, central tendencies, and the presence of multiple modes.**

**3. Bar Plot (Bar Chart):**

**Purpose: Bar plots are ideal for visualizing categorical data, including comparisons between different categories or groups.**

**Insights: They help you understand the frequency, proportion, or performance of different categories within your supply chain, such as product categories, suppliers, or regions.**

4. Sweet viz:

**Purpose: Sweet viz is a powerful Python library for automated exploratory data analysis. It provides a comprehensive overview of data, including comparisons between two datasets (e.g., training vs. testing data).**

**Insights: Sweet viz generates detailed visualizations and statistics, making it easier to compare datasets, identify differences, and gain insights into data distributions, relationships, and missing values.**

**By using these data exploration techniques, you've likely gained valuable insights into your supply** chain management data, such as understanding the distribution of key variables, spotting potential outliers, and identifying trends and patterns. These insights serve as a foundation for making informed decisions and further optimizing your supply chain processes. It's excellent practice to document and present these visualizations in your EDA report to communicate your findings effectively.

I have created a dashboard using pallet manufacturing Company Sales and Inventory Analysis. In dashboard I have used the cards, Pie charts and a line graph.

SQL: SQL can be used to perform data normalization, but it's more commonly done in Python using libraries like Scikit-Learn. SQL can be used to create new tables with normalized data.

Python: Scikit-Learn provides tools for data preprocessing, including normalization and standardization. You can use `Minmax Caler` or `Standards Caler` to scale your data.

SQL: In SQL, you can use `CASE` statements or subqueries to handle categorical data by creating new columns with numeric values representing categories.

Python: Pandas provide functions like `get dummies () ` to convert categorical variables into dummy/indicator variables.

SQL: Perform data validation and integrity checks using SQL constraints like primary keys, unique constraints, and foreign keys.

Python: You can use Python code to validate data, checking for consistency and adherence to business rules.

- SQL: If you've performed data transformations in SQL, you can export the cleaned and preprocessed data back to a new table or file.

-Python: If you've used Python for data preprocessing, you can export the cleaned data to a new file using Pandas.

By following these steps and using both SQL and Python as appropriate, you can ensure that your supply chain management data is cleaned, transformed, and ready for exploratory data analysis.

I have created a dashboard using a Pallet Manufacturing Company using Sales and Inventory Analysis. I have used the pie chart and cards and the line graphs by region, state, city and Cust Name.