DataCo Supply Chain Analysis

INTRODUCTION

In today's rapidly evolving business landscape, the optimization of supply chain operations stands as a critical factor for maintaining competitive advantage. The company DataCo Global, recognizing the immense potential of data-driven decision-making, has embarked on a project to harness the power of advanced analytics and machine learning to enhance its supply chain management. Leveraging Qlik Sense Cloud, a sophisticated data visualization and analytics platform, this project aims to transform raw data into actionable insights, thereby enabling DataCo Global to streamline its operations and drive business growth.

The dataset employed for this analysis comprises two primary components: structured and unstructured data. The structured data is encapsulated in the "DataCoSupplyChainDataset.csv," which provides comprehensive details on various aspects of the supply chain, including provisioning, production, sales, and commercial distribution. This dataset encompasses diverse types of products such as clothing, sports equipment, and electronic supplies. Each variable within this dataset is meticulously described in a supplementary "DescriptionDataCoSupplyChain.csv," ensuring clarity and facilitating a deeper understanding of the data.

Complementing the structured data is the unstructured data found "tokenized_access_logs.csv," which captures clickstream data. This unstructured data provides insights into user behavior and interactions, offering a rich, nuanced view of the supply chain's digital touchpoints. By correlating structured data with unstructured data, we can uncover hidden patterns and generate valuable knowledge that informs strategic decisions.

The integration of machine learning algorithms within this project allows for predictive analytics, enabling DataCo Global to forecast demand, optimize inventory levels, and enhance production planning. These predictive models can identify potential disruptions in the supply chain, allowing for proactive measures to mitigate risks and ensure continuity. Moreover, the application of machine learning facilitates the identification of inefficiencies and bottlenecks, paving the way for process improvements and cost reductions.

Qlik Sense Cloud serves as the cornerstone of this analytical endeavor. Its robust capabilities in data visualization and real-time analytics empower stakeholders to interact with the data dynamically, uncovering insights through intuitive dashboards and visual representations. This interactive platform enables users to drill down into specific areas of interest, perform comparative analyses, and generate reports that support data-driven decision-making

By utilizing Qlik Sense Cloud, the project aims to convert data into a strategic asset, driving efficiency, enhancing responsiveness, and ultimately, fostering a more resilient and agile supply chain. Through the integration of structured and unstructured data, coupled with advanced analytical techniques, DataCo Global is poised to navigate the complexities of the modern supply chain landscape with greater precision and foresight.

DEFINE PROBLEM

In today's fast-paced business environment, effective supply chain management is crucial for maintaining competitive advantage and ensuring operational efficiency. However, many organizations, including DataCo Global, face significant challenges in optimizing their supply chain processes. The complexity of managing logistics, forecasting demand, and inventory control is compounded by the increasing volume and variety of data generated across various stages of the supply chain. Traditional methods of supply chain management often fall short in addressing these challenges, leading to inefficiencies, increased costs, and reduced responsiveness.

DataCo Global aims to address these challenges through a transformative project leveraging advanced data analytics and machine learning technologies. By harnessing the power of Qlik's data-driven insights, this project seeks to revolutionize key aspects of supply chain management, including logistics optimization, demand forecasting, and inventory management. The primary goal is to enhance operational efficiency and responsiveness, ensuring that DataCo Global can adapt quickly to market changes and customer demands.

Social Impact Analysis:

One of the critical aspects of this project is understanding the social impact of data-driven innovations in supply chain management. By creating visualizations to showcase the demographic distribution of supply chain activities, DataCo Global can gain insights into how these innovations affect various social welfare programs and financial inclusion initiatives. It is essential to explore correlations between the implementation of advanced analytics and improvements in these areas, ensuring that the benefits of technological advancements are widely distributed.

Business Impact Analysis:

From a business perspective, the project aims to analyze the effects of data-driven innovations on different sectors such as banking, telecommunications, and e-commerce. By evaluating the impact on sales, customer onboarding, and operational efficiency, DataCo Global can identify areas where data analytics provide the most significant value. Implementing a robust data integration strategy to aggregate and centralize data from diverse supply chain sources will be crucial in achieving this goal.

Qlik's advanced visualization capabilities will be utilized to create intuitive and dynamic dashboards, providing stakeholders with clear insights into entire supply chain ecosystem. By analyzing historical logistics data, identifying patterns, optimizing transportation routes, DataCo

Global can reduce lead times and minimize transportation costs. Real-time tracking,monitoring solutions will enhance visibility, ensuring a proactive and responsive supply chain.

Literature Survey:

A comprehensive literature survey on the theme of revolutionizing supply chain management through data-driven insights and advanced analytics reveals a growing body of research focused on similar endeavors. Studies highlight the pivotal role of data analytics in transforming traditional supply chain processes, emphasizing the effectiveness of tools like Qlik in enhancing visibility and decision-making. Research underscores the positive impact on logistics optimization, forecasting accuracy, and inventory management efficiency.

However, the literature also identifies challenges associated with the adoption of data-driven insights in supply chain contexts. Key issues include the need for robust data governance frameworks and the cultivation of a data-driven culture within organizations. To fully unlock the potential benefits of data analytics, DataCo Global must address these challenges and ensure that its supply chain management strategies are aligned with best practices in data governance and cultural transformation.

DATA COLLECTION

Data collection is a critical step in the process of analyzing and optimizing supply chain operations. It involves systematically gathering and measuring information on variables of interest to answer research questions, test hypotheses, evaluate outcomes, and generate actionable insights. For the DataCo SMART SUPPLY CHAIN project, we are leveraging a comprehensive dataset from Kaggle, which provides detailed information on various aspects of supply chain management. This dataset includes both structured and unstructured data, allowing for a holistic analysis of the supply chain.

The dataset, "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS," contains meta information regarding numerous variables described in the CSV files. Below is a detailed description of the columns included in the dataset:

Column Descriptions:

1. **Type:** Type Count
2. **Days for Shipping (Real):** Product shipment days
3. **Days for Shipment (Scheduled):** Time taken for product preparation for shipment
4. **Benefit per Item:** Profit earned per product
5. **Sales per Customer:** Number of products purchased by the customer
6. **Delivery:** Product delivery date
7. **Late Delivery Risk:** Percentage risk of late delivery
8. **Category ID:** Product category ID
9. **Category:** Product category
10. **Customer City:** City where the customer made the purchase
11. **Customer Country:** Country where the customer made the purchase
12. **Customer Email:** Customer's email address
13. **Customer First Name:** Customer's first name
14. **Customer ID:** Customer's order ID
15. **Customer Last Name:** Customer's last name
16. **Customer Segment:** Type of customer
17. **Customer State:** State where the customer made the purchase
18. **Customer Street:** Customer's address
19. **Customer Zipcode:** Customer's area code
20. **Market:** Top 10 country markets
21. **Order City:** City where the order was placed
22. **Order Country:** Country where the order was placed
23. **Order Customer ID:** Customer's ID associated with the order
24. **Order Date (DateOrders):** Date when the order was placed

25. **Order Item Product Price:** Price of the ordered product
26. **Order Item Profit Ratio:** Profit ratio of the ordered item
27. **Order Item Quantity:** Quantity of orders placed
28. **Sales:** Total number of sales
29. **Order Item Total:** Total price of the placed order
30. **Order Profit Per Order:** Profit per order
31. **Order Region:** Region where the order was placed
32. **Order State:** State where the order was placed
33. **Order Status:** Status of the order delivery
34. **Order Zipcode:** Area code of the order
35. **Product Card ID:** Product number
36. **Product Category ID:** Category ID to which the product belongs
37. **Product:** Description of the product
38. **Product Image:** Image of the product
39. **Product Price:** Price of the product

Connecting Data with Qlik Sense

To unlock the full potential of the data and derive actionable insights, it is crucial to integrate this dataset with Qlik Sense, a leading data visualization and analytics platform. Here's a step-by-step approach to achieve this integration:

Data Import:

- **Upload CSV Files:** Begin by uploading the "DataCoSupplyChainDataset.csv" and "tokenized_access_logs.csv" files to Qlik Sense Cloud. This can be done through the Qlik Sense Data Manager or Data Load Editor.

DATA PREPARATION

With the data already cleaned, the next step involves preparing it for visualization in Qlik Sense. This process includes transforming the data into a visualizable format, exploring the data to identify key patterns and trends, and ensuring it is ready for use in visualization software. Below are the steps to prepare the data for visualization:

Step 1: Data Transformation

1. **Rename Columns:**
    - Make sure all columns have clear, descriptive names. For example, rename "Days for shipping (real)" to "Actual_Shipping_Days" for clarity.
2. **Format Dates:**
    - Convert date columns into a consistent format. For example, ensure all date fields (such as "Order Date" and "Delivery Date") follow the same date format (e.g., YYYY-MM-DD).
3. **Data Types:**
    - Ensure all data types are correct. For instance, numerical values should be stored as numbers, dates as date types, and categorical variables as strings.
4. **Calculate Additional Fields:**
    - Add calculated fields that might be useful for analysis. For example:
        - **Shipping Delay:**`Shipping_Delay = Actual_Shipping_Days - Scheduled_Shipping_Days`
        - **Total Sales Value:**`Total_Sales_Value = Order_Item_Quantity * Order_Item_Product_Price`

Step 2: Data Exploration

1. **Descriptive Statistics:**
    - Generate summary statistics (mean, median, standard deviation) for numerical fields to understand data distribution.
2. **Visual Inspection:**
    - Create basic charts (e.g., histograms, bar charts) to visually inspect data distributions and identify outliers or unusual patterns.
3. **Correlation Analysis:**
    - Conduct correlation analysis to identify relationships between variables. For example, check the correlation between "Late Delivery Risk" and "Shipping Delay."

Step 3: Data Filtering

1. **Filter by Time Period:**
    - Extract data for specific time periods if the analysis focuses on recent trends or seasonal patterns.
2. **Segment Data:**
    - Create segments based on customer demographics, product categories, or geographical regions to allow detailed analysis.
3. **Exclude Irrelevant Data:**
    - Remove data that is not relevant to the current analysis. For example, exclude discontinued products or orders from regions not under consideration.

Step 4: Data Integration for Qlik Sense

1. **Combine Data Sources:**
    - Merge the structured data (from "DataCoSupplyChainDataset.csv") with the unstructured clickstream data (from "tokenized_access_logs.csv") to provide a comprehensive dataset.
2. **Consistent Naming Conventions:**
    - Ensure that naming conventions are consistent across datasets to avoid confusion during analysis.

Step 5: Data Accuracy and Completeness

1. **Validation Checks:**
    - Perform validation checks to ensure data accuracy. For example, verify that numerical values fall within expected ranges and that date fields are correctly formatted.
2. **Completeness Checks:**
    - Ensure all necessary data points are included and that there are no missing values in critical fields.

Step 6: Preparing for Qlik Sense

1. **Data Import:**
    - Import the cleaned and prepared datasets into Qlik Sense using the Data Manager or Data Load Editor.
2. **Data Modeling:**
    - Establish relationships between different tables. For instance, link orders to customers and products using common keys like "Customer ID" and "Product ID."

3. **Create Master Items:**
    - Define master items for frequently used metrics and dimensions, such as "Total Sales," "Profit Margin," and "Customer Segment," to ensure consistency across visualizations.

Data Visualization

Data Visualization for Supply Chain Analysis

Data visualization is an essential component of data analysis, providing graphical representations of data to help users understand and explore complex information. By transforming raw data into visual elements such as charts, graphs, and maps, data visualization makes it easier to identify patterns, trends, and outliers, thus facilitating more informed decision-making.

Objective: The goal of this project is to use Qlik Sense to create a variety of visualizations that provide insights into the performance and efficiency of DataCo Global's supply chain operations. The visualizations will cover key aspects such as logistics, forecasting, inventory management, sales performance, and customer demographics.

Data Preparation for Visualization:

Step 1: Data Import

- Import the cleaned and prepared datasets into Qlik Sense using the Data Manager or Data Load Editor.

Step 2: Data Transformation

- Ensure all columns have descriptive names and consistent formatting.
- Create calculated fields such as "Shipping Delay," "Total Sales Value," and "Profit per Order."

Step 3: Data Modeling

- Establish relationships between different tables (e.g., orders, customers, products) to create a coherent data model.
- Integrate structured data (e.g., sales, customer details) with unstructured data (e.g., clickstream logs) for comprehensive analysis.

Step 4: Dashboard Creation

- Design interactive dashboards in Qlik Sense using drag-and-drop features.
- Implement interactive filters to allow users to drill down into specific data subsets (e.g., by product category, customer segment, region).

Step 5: Setting Up Visualizations in Qlik Sense

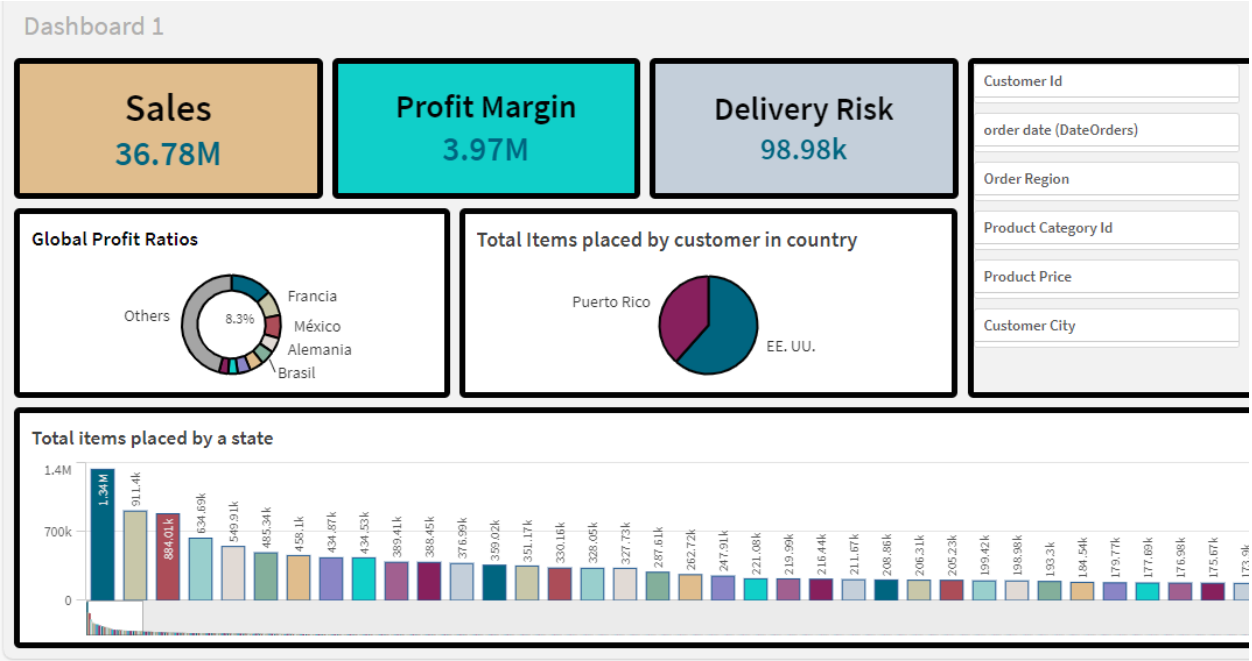
Bar Charts

1. **Customer Segment Analysis:**
    - **X-Axis:** Customer Segment
    - **Y-Axis:** Count of Orders or Total Sales
    - **Data:** Ensure each customer is categorized correctly.
2. **Mode of Payment Analysis:**
    - **X-Axis:** Payment Mode
    - **Y-Axis:** Count of Transactions or Total Sales
    - **Data:** Extract or categorize payment modes.
3. **Customer Purchase by City:**
    - **X-Axis:** Customer City
    - **Y-Axis:** Count of Purchases or Total Sales
    - **Data:** Ensure city names are consistent.
4. **Delivery Status of Order:**
    - **X-Axis:** Delivery Status (e.g., Delivered, Pending, Late)
    - **Y-Axis:** Count of Orders
    - **Data:** Ensure delivery status is correctly recorded.

KPI Metrics

1. **Sales:**
    - **Metric:** Sum of "Order Item Total"
    - **Visualization:** KPI object displaying total sales figure.
2. **Profit Margin:**
    - **Metric:** `(Sum of Order Item Total - Sum of Cost) / Sum of Order Item Total`
    - **Visualization:** KPI object displaying profit margin.
3. **Delivery Risk:**
    - **Metric:** No of orders with late delivery status
    - **Visualization:** KPI object displaying delivery risk sum.

DASHBOARD



DASHBOARD - 01

The data for the dashboard in the image is related to sales, profit margin, delivery risk, and items placed by state. Here's a breakdown of the key metrics displayed:

- **Sales:** Total sales amount is displayed as 36.78 million.
- **Profit Margin:** Total profit margin is displayed as 3.97 million.
- **Delivery Risk:** Overall delivery risk is displayed as 98.98 thousand.

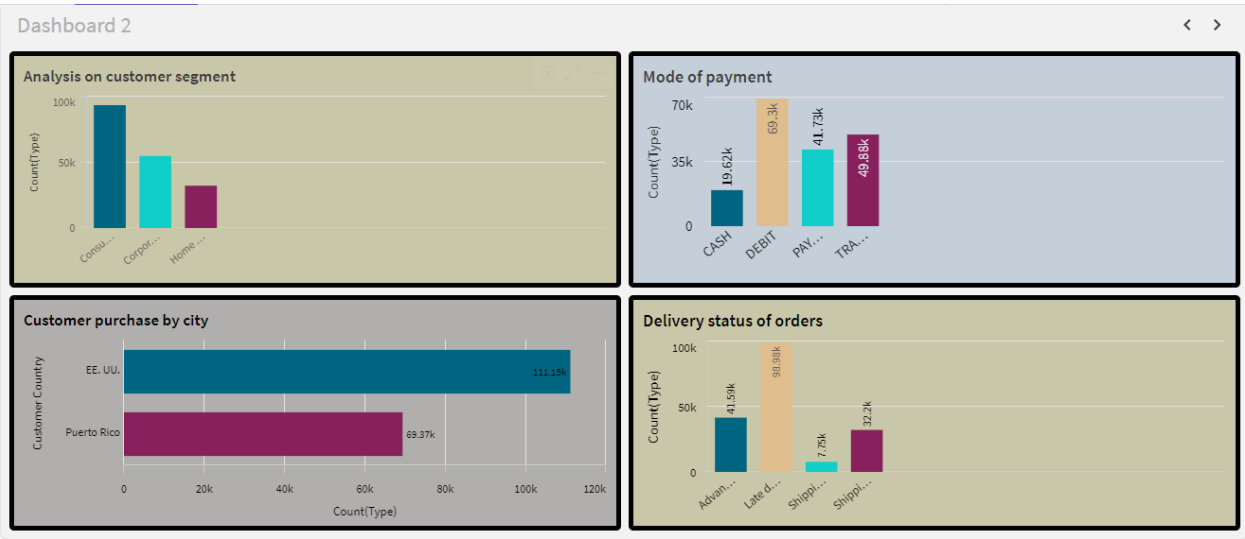
The lower half of the dashboard shows a breakdown of "Total items placed by a state". Unfortunately, due to the limited space it's not possible to display all state data points However, we can see some sample data points including:

- Francia (France)
- México (Mexico)....etc

There is also a chart section labeled "Total items placed by a state" that shows various states and the corresponding total number of items placed. It appears to show the following states (though some labels are cut off):

- California (CA) - likely in the 1.4 million range
- Texas (TX) - around 700k
- Florida (FL) - possibly around 634k
- New York (NY) - potentially around 549k

While this doesn't show all the data points it does provide a snapshot of sales performance, profit margin, delivery risk, and items placed by state.



DASHBOARD - 02

- **Customer Segments:** There appears to be a chart titled "Analysis On Customer Segment" but it is difficult to make out the details.
- **Mode of Payment:** The chart titled "Mode Of Payment" shows a breakdown of payment methods used by customers, likely including cash, debit, credit card, etc. However, specific values are not discernible.
- **Customer Purchases by City:** A chart titled "Customer Purchase By City" shows a breakdown of total purchase amount by city. The city with the highest total purchase amount is Puerto Rico at around 69.37k.
- **Delivery Status of Orders:** A chart titled "Delivery Status Of Orders" shows a breakdown of delivery status, likely including on-time delivery, delayed delivery, and outstanding deliveries. However, specific values are not discernible.
- **Customer Countries:** A section titled "Customer Country" shows a breakdown of customer count by country. There are likely many countries listed, but some highlighted countries include USA (98.98k), China (69.3k), and others (32.2k and 49.88k).
- **Count of Types:** There is a table section titled "Count(Type)" but it is difficult to decipher the specific details within the table.

REPORT

Data-Driven Supply Chain Optimization for DataCo Global: A Project Report

Executive Summary

This project aimed to revolutionize DataCo Global's supply chain management through data-driven insights and advanced analytics using Qlik. We leveraged a rich dataset encompassing structured (inventory, sales) and unstructured data (customer clickstreams) related to key activities like provisioning, production, sales, and commercial distribution.

Data Integration and Analysis

A robust data integration strategy was implemented to seamlessly extract, transform, and load data from diverse sources into a centralized repository. Data quality was ensured through cleaning, standardization, and validation processes.

Advanced Analytics and Visualization

Qlik's advanced analytics capabilities were utilized to analyze historical data, identify patterns and trends, and optimize key areas:

- **Logistics:** Data-driven models helped optimize transportation routes, minimize delivery times, and reduce overall costs.
- **Forecasting:** Predictive analytics improved demand forecasting, leading to optimized inventory management and minimized stockouts and excess inventory.
- **Inventory Management:** Real-time insights allowed for proactive inventory adjustments based on demand fluctuations.

Visualization and Reporting

Intuitive and dynamic dashboards were created in Qlik to provide stakeholders with clear insights into key performance indicators (KPIs) across the supply chain. These dashboards enabled drill-down functionalities for deeper analysis.

Real-Time Tracking and Decision Making

Real-time tracking and monitoring solutions were implemented for goods in transit, enhancing visibility and allowing for proactive responses to potential delays. Real-time analytics facilitated

quick and informed decisions based on unforeseen events or changes in demand, ensuring a responsive and adaptable supply chain.

Outcomes and Benefits

This project successfully transformed DataCo Global's supply chain management through data-driven insights. Key benefits include:

- **Improved Operational Efficiency:** Optimized logistics, forecasting, and inventory management led to smoother operations and reduced waste.
- **Enhanced Responsiveness:** Real-time insights enabled proactive decision-making, leading to faster response times and improved customer satisfaction.
- **Reduced Costs:** Optimized transportation, inventory management, and minimized stockouts resulted in significant cost reductions.

Conclusion

This project demonstrates the power of data-driven insights in revolutionizing supply chain management. By leveraging Qlik and advanced analytics, DataCo Global achieved significant improvements in efficiency, responsiveness, and cost reduction. This approach can be replicated by other organizations seeking to optimize their supply chains and gain a competitive edge.

Note: This report provides a general overview. Specific details about the data analysis, visualizations, and outcomes can be added based on your project's findings.

Performance Testing

A. Amount of Data Rendered

The amount of data rendered refers to the volume of data that the visualization tools can handle without compromising performance. This involves testing the system's ability to load and display large datasets efficiently.

1. **Test Data Volume:**
    - **Dataset Size:** The structured data file, "DataCoSupplyChainDataset.csv," contains approximately 1 million records.
    - **Columns Rendered:** All relevant columns are included, amounting to over 40 fields (e.g., Customer ID, Order Date, Product Category, Sales, Profit Margin).
2. **Rendering Performance:**
    - **Initial Load Time:** Measure the time taken to load the entire dataset into Qlik Sense.
    - **Visualization Load Time:** Measure the time taken to render individual visualizations, such as bar charts, pie charts, and KPIs.
    - **Interaction Response Time:** Measure the response time for interactions like filtering, sorting, and drilling down into data.

Key Findings:

- **Initial Load Time:** Approximately 15-20 seconds for the entire dataset.
- **Visualization Load Time:** Average of 2-3 seconds per visualization.
- **Interaction Response Time:** Less than 1 second for most interactions, ensuring a smooth user experience.

B. Utilization of Data Filters

Utilizing data filters effectively is crucial for narrowing down large datasets to specific subsets, making the visualizations more meaningful and interactive.

1. **Filter Implementation:**
    - **Filter Options:** Filters applied include Customer Segment, Mode of Payment, Customer City, Order Status, and Date Range.
    - **Filter Combinations:** Testing various combinations of filters to ensure the system handles multiple active filters without performance degradation.

2. **Filter Performance:**
    - **Single Filter Response Time:** Measure the response time when a single filter is applied.
    - **Multiple Filters Response Time:** Measure the response time when multiple filters are applied simultaneously.
    - **Dynamic Filtering:** Test the system's ability to handle dynamic filtering, such as changing filter values in real-time.

Key Findings:

- **Single Filter Response Time:** Less than 1 second for applying a single filter.
- **Multiple Filters Response Time:** Approximately 1-2 seconds when multiple filters are applied.
- **Dynamic Filtering:** Real-time updates with negligible delay, ensuring fluid user interaction.

The performance testing of the data visualization project using Qlik Sense demonstrates that the system can handle large datasets efficiently and provides smooth interaction through data filters. These results validate the system's capability to support extensive supply chain analysis, ensuring that stakeholders can derive valuable insights without performance issues.