This report presents a holistic analysis of Amazon's supply chain operations, encompassing SQL-based data processing, Power BI visualizations, and Python-based machine learning insights.

Amazon Supply Chain Analysis

From Raw Data to Refined Insights

Prepared by: Praneeth Jajjara

SQL-Based Data Processing and Analysis of Amazon Orders

From Raw Data to SQL-Refined Insights

SUMMARY : This comprehensive set of data processing steps encompasses extraction, cleaning, transformation, feature engineering, and various analyses. These operations collectively contribute to a more refined and structured dataset, laying the groundwork for subsequent exploratory and analytical tasks.

1. Data Cleaning & Transformation

- **Checking for Null Values**
  To ensure data integrity, a thorough check for null values is performed on all columns.

- **Changing Date Format**
  In order to standardize the date representation, the format of the **Order Date** and **Delivery Date** columns is modified to **%d-%m-%Y**. This ensures a consistent and easily interpretable date format across the dataset.

2. Feature Engineering

- **Adding new parameter**
  A new feature, **fulfilment days**, is introduced to the dataset. This represents the duration between the order placement (**Order Date**) and the delivery date (**Delivery Date**). This additional information may prove valuable for further analysis.

- **Filtering Data**
  A filtering operation is conducted to count the number of orders where the quantity exceeds 1. This allows for the identification and analysis of orders with a higher quantity than the average.

- **Aggregate Function**
  An aggregate function is applied to calculate the total quantity of items supplied by each supplier. This facilitates a broader understanding of the distribution of supplied quantities among different suppliers.

- **Grouping Data**
  The dataset is grouped by location, and the average price of products in each location is calculated. This grouping operation provides insights into the average pricing trends across different geographical locations.

- **Subqueries**
  A subquery is employed to determine the count of high-value orders, considering those with prices exceeding 1000 Rupees. This analysis aims to identify and quantify high-value transactions within the dataset.

- **Conditional Logic**
  A new column "**volume"** is introduced to categorize orders based on quantity. This categorization includes 'HIGH' for orders with a quantity of 5, 'MEDIUM' for quantities greater than 2, and 'LOW' for all other cases. This additional classification adds a layer of granularity to the dataset, allowing for more nuanced analyses.

Power BI Report: Amazon Supply Chain Analysis

Tailored for Stakeholder Insightfulness

SUMMARY : Click on any element and witness the interconnected visualizations dynamically responding to provide a clear and comprehensive view of the data relationships and insights.

VISUALIZATION REPORT STRUCTURE :

I. **Location-wise Distribution of Suppliers (Table Visualization)**
- Overview of suppliers distributed across different locations.

II. **Quantity Sold by Location Overview (Table Visualization)**
- Breakdown of the quantity sold in each location.

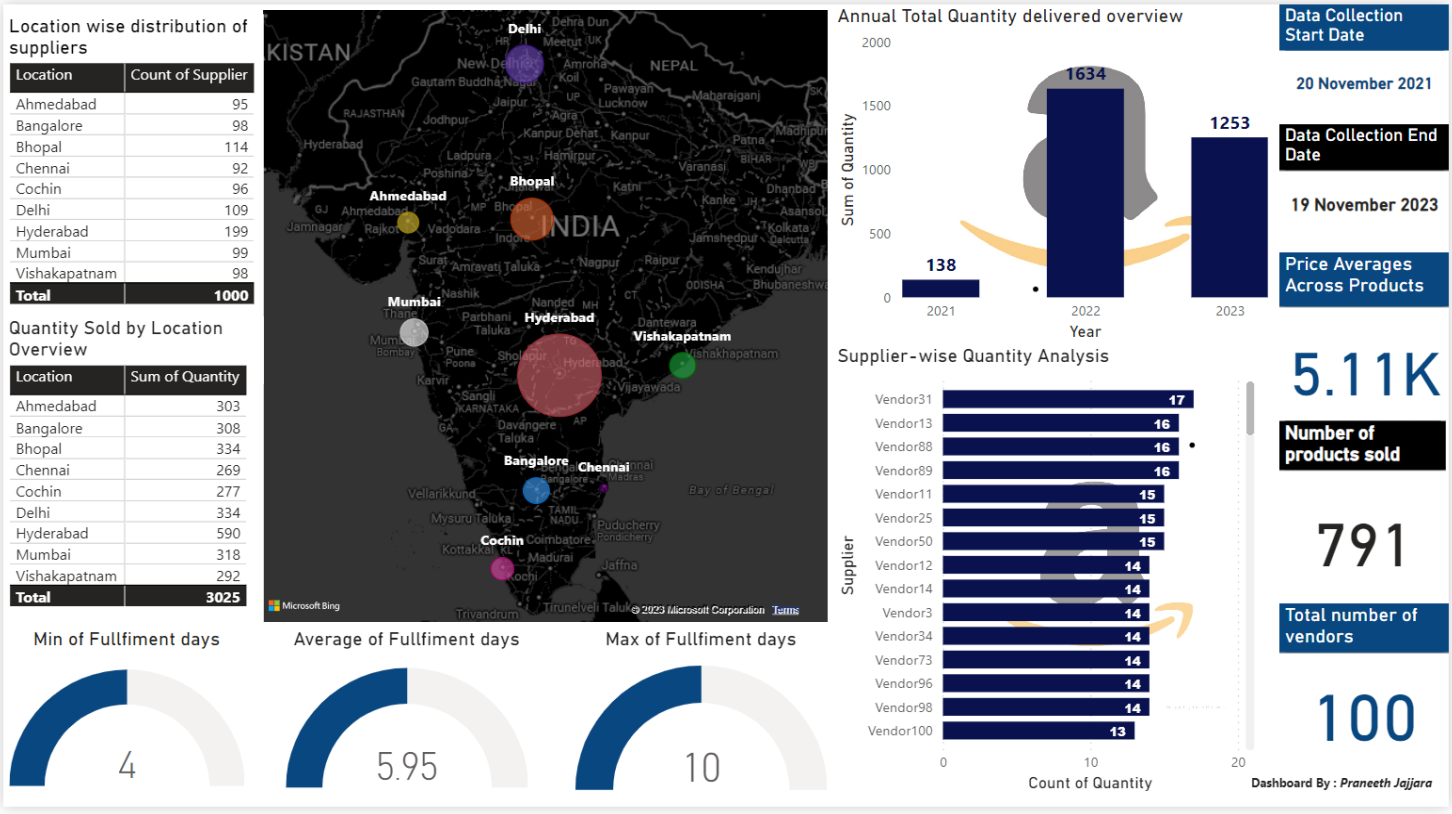
III. **Warehouse Locations and Product Sales (Map Chart)**
- Map chart displaying warehouse locations with bubbles indicating product sales count.

IV. **Fulfilment Days Metrics (Gauge Chart)**
- Metrics for fulfilment days: Min, Max, Avg using gauge chart.

V. **Annual Total Quantity Delivered (Stacked Column Chart)**
- Stacked column chart showcasing the annual total quantity delivered overview.

VI. **Supplier-wise Quantity Analysis (Stacked Bar Chart)**
- Stacked bar chart providing insights into the quantity sold by each supplier.

VII. **Additional Metrics (Card Visualization)**
- Start & End date of data collection.
- Total number of products sold.
- Average price across products.
- Total number of vendors.



PYTHON - Machine Learning Insights from Amazon Data

Harvesting Intelligence

Summary : This comprehensive approach combines data exploration, trend analysis, metric development, and stakeholder engagement to optimize Amazon's supply chain operations. The insights gained can serve as a foundation for continuous improvement and informed decision-making.

The procedural framework for the methodology is as follows:

1) Data Exploration and Understanding:

a) Dataset Overview:

- The dataset, sourced from 'Book1.csv', contains information related to Amazon's supply chain operations, including order details, fulfilment times, and product information.

b) Initial Exploration:

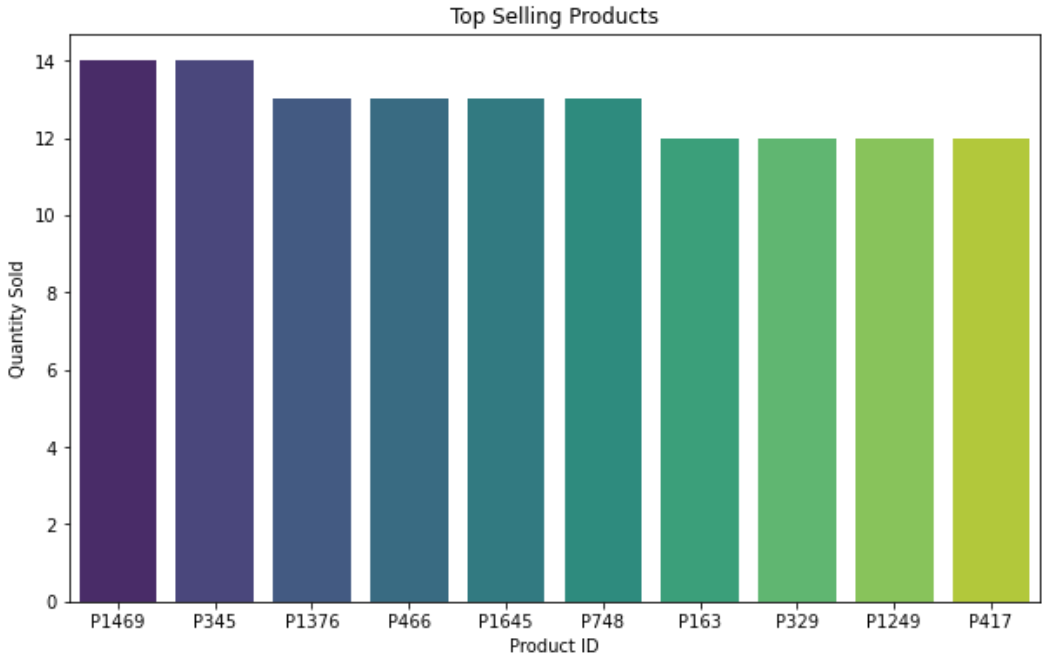
- Explored key variables such as 'ProductID,' 'Quantity,' 'Location,' 'Supplier,' 'Fulfilment days,' and 'Order Date.'
- Assumed 'Order Date' is in datetime format for time-series analysis.

	Sno	ProductID	OrderDate	DeliveryDate	Quantity	Price	Supplier	Location	Fullfiment days	Volume
0	1	P1171	01-01-2022	05-01-2022	3	7859	Vendor41	Ahmedabad	4	Medium
1	2	P1048	12-11-2023	16-11-2023	2	6674	Vendor95	Hyderabad	4	Low
2	3	P1052	02-03-2023	06-03-2023	4	2248	Vendor6	Hyderabad	4	Medium

2) Trend Analysis:

- **Objective**:
  Identify the best-performing products (Top Selling Products) for efficient inventory management.

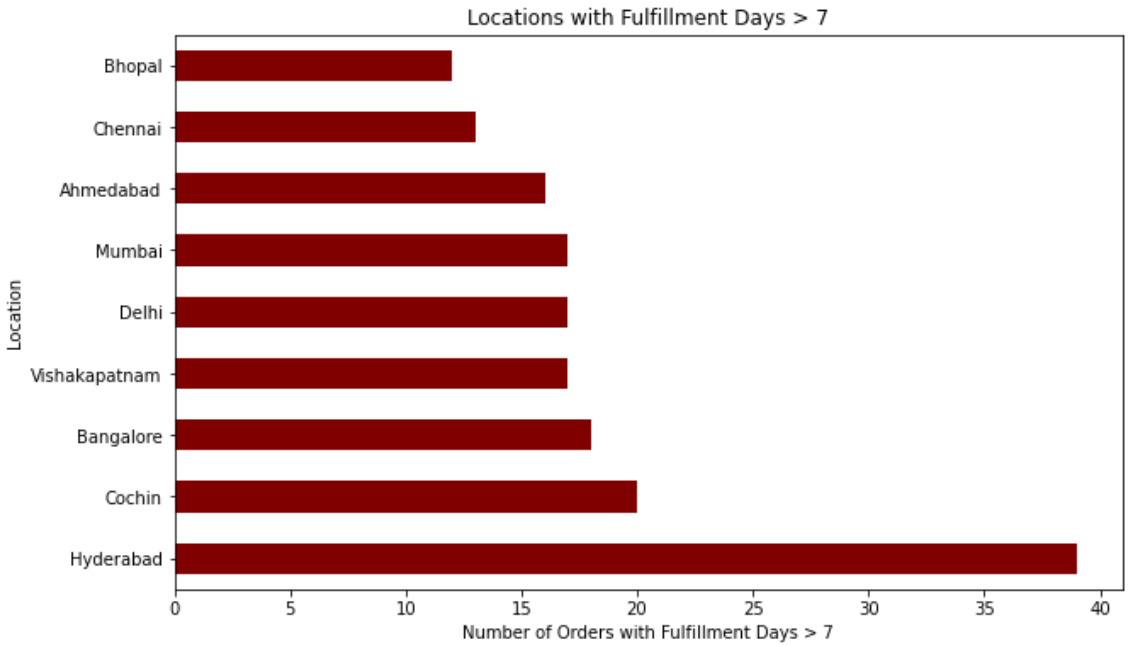
- **Business Findings**:
  - List of top 10 selling products based on historical sales data,recommending prioritization of these products for effective inventory management.
  - Explanation of how optimizing inventory for these products can enhance stock management.



3) Improving Operational Efficiency:

- **Objective:**
  Address issues related to fulfilment days exceeding 7 days.

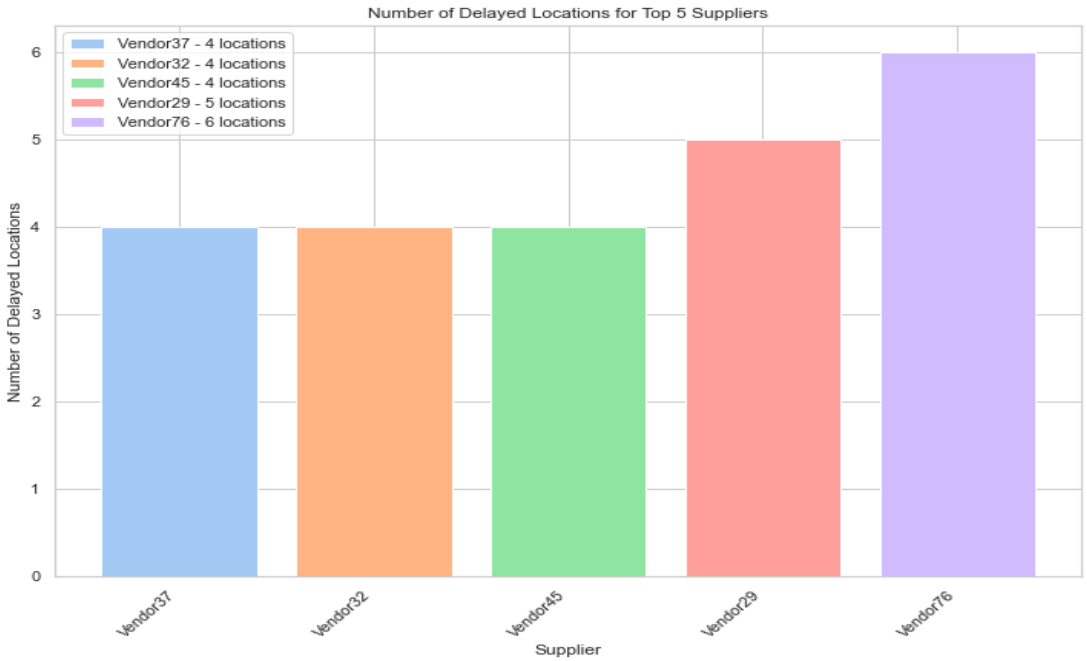
- **Business Findings:**
  - Extracted information on the number of orders for locations with fulfilment days exceeding 7 days.
  - Recommendations for quicker and more reliable deliveries contribute to enhanced customer experience.



4) Identify Areas of Improvement:

- **Objective:**
  Identify the top 5 suppliers causing delays.

- **Business Findings:**
  - Utilized analytical insights to identify areas for operational improvement.
  - Supplier Performance Analysis

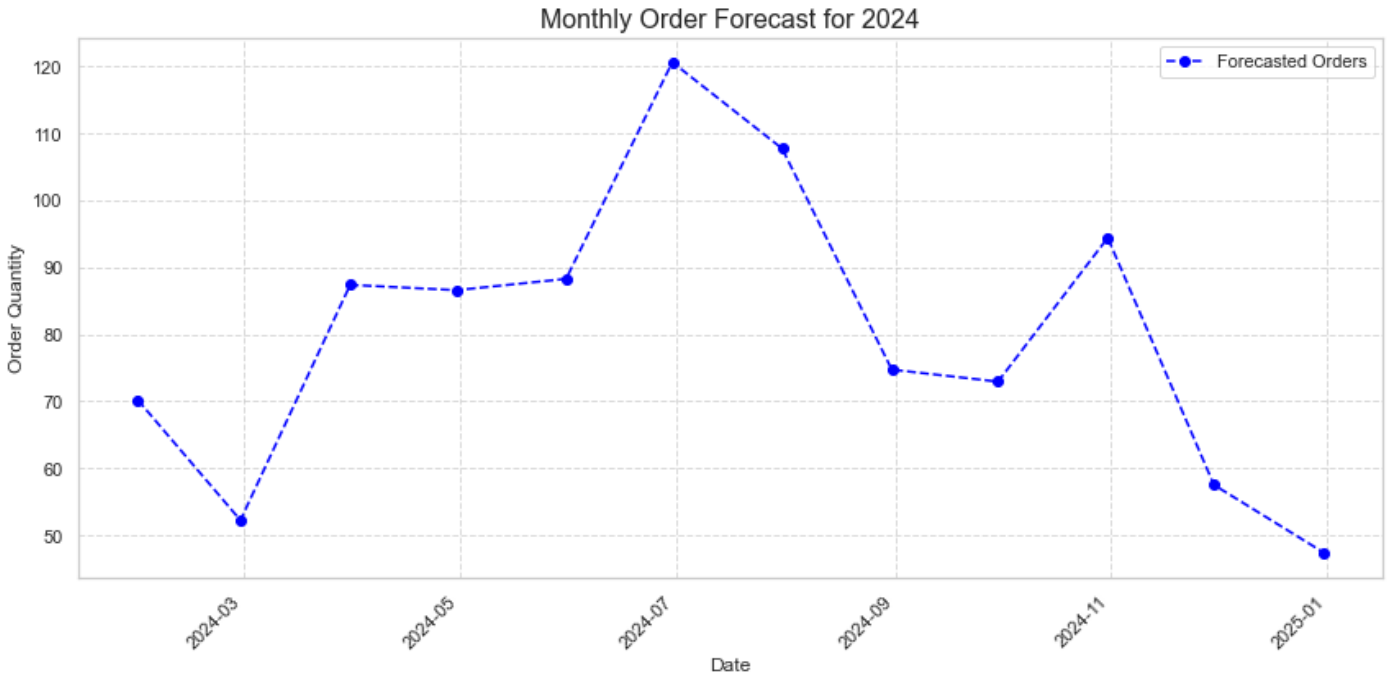


5) SALES FORECASTING (USING MACHINE LEARNING MODEL – SARIMA)

- **Objective:**
  Forecast future year sales using the SARIMA model.

- **Business Findings:**
  Sales Forecasting helps **in** improve inventory management, order fulfilment, and supplier management.

Month	Predicted Orders
Jan 2024	70
Feb 2024	52
Mar 2024	87
Apr 2024	87
May 2024	88
Jun 2024	121
Jul 2024	108
Aug 2024	75
Sep 2024	73
Oct 2024	94
Nov 2024	58
Dec 2024	47



Summary of Data Analysis Leveraging ML:

Utilized data analysis and machine learning techniques for comprehensive insights. Explored top-selling products, optimized inventory, and streamlined order fulfilment through data-based strategies. Identified supplier delays using data analysis, and employed SARIMA model for predictive sales analytics, culminating in actionable recommendations for enhanced business operations.

Limitations and Future Work:

Limitations: Acknowledge any limitations in the analysis.

Future Work: Suggest areas for further investigation or improvement.