

## I. Project Overview

The purpose of our analysis is to gain a comprehensive understanding of the company's supply chain by examining various stages including sourcing, manufacturing, transportation, inventory management, sales, and customer demographics. By analyzing these key areas, we aim to identify inefficiencies, potential cost-saving opportunities, and areas for improvement to enhance overall supply chain performance. The dataset, named "15\_Supply Chain Analysis.csv," encompasses crucial data points related to each stage of the supply chain. Key user attributes and data points within this dataset likely include details on suppliers, production volumes, shipping routes and times, inventory levels, sales figures, and customer information. Through this analysis, we seek to derive actionable insights that can inform strategic decisions, optimize supply chain operations, and ultimately drive business growth.

### Key Analysis Points for Supply Chain Analysis

#### 1. Price Distribution:

- Analyze the spread and skewness of product prices.
- Identify pricing patterns and any significant outliers.

#### 2. Availability Distribution:

- Examine the distribution of product availability.
- Identify common stock levels and periods of stock shortages or excess.

#### 3. Sales Analysis:

- Investigate the number of products sold.
- Determine the distribution and identify products with particularly high or low sales.



#### **4. Revenue Analysis:**

- Assess the revenue generated by each product.
- Explore the relationship between sales volume and revenue.

#### **5. Stock Levels:**

- Analyze the variability and distribution of stock levels.
- Identify potential issues with overstocking or stockouts.

#### **6. Lead Times and Shipping Times:**

- Examine the distribution and variability of lead times.
- Investigate shipping times to identify potential delays and bottlenecks.

#### **7. Shipping Costs:**

- Analyze shipping cost data to identify trends and potential cost-saving opportunities.

#### **8. Production Volumes and Manufacturing Lead Times:**

- Examine production volumes and the efficiency of the manufacturing process.
- Identify any correlations between manufacturing lead times and production output.

#### **9. Manufacturing Costs and Defect Rates:**

- Analyze manufacturing costs and their distribution.
- Explore the relationship between manufacturing costs and defect rates to identify potential areas for quality improvement.

#### **10. Transportation Modes and Routes:**

- Assess the use of different transportation modes (e.g., road, sea) and routes.
- Identify the most efficient and cost-effective transportation options.

## 11. Correlation Analysis:

- Use correlation heatmaps to identify relationships between different variables.
- Focus on high correlations that might indicate significant relationships, such as between manufacturing costs and defect rates.

## 12. Missing Values:

- Visualize and analyze missing data to inform data cleaning strategies.
- Determine the impact of missing values on the overall analysis.

### Insights and Actions

- **Optimizing Inventory Management:** By understanding stock levels and availability, the company can better manage inventory, reduce costs, and improve product availability.
- **Enhancing Sales Strategies:** Sales and revenue analysis can inform marketing and sales strategies to focus on high-performing products.
- **Improving Manufacturing Efficiency:** Analyzing production volumes, manufacturing costs, and defect rates can identify areas for cost reduction and quality improvement.
- **Streamlining Logistics:** Lead times, shipping times, and transportation analysis can help streamline logistics, reduce delays, and lower shipping costs.
- **Data-Driven Decision Making:** Correlation analysis and insights from missing values will support data-driven decision making and strategic planning.

By focusing on these key analysis points, the company can gain valuable insights into its supply chain operations and identify opportunities for improvement and optimization.

## II. Libraries and Data Handling

**Libraries Used:** To perform the supply chain analysis effectively, several Python libraries are essential for data manipulation, analysis, and visualization. Here are the key libraries and the data handling process.

### Libraries

#### 1. Pandas

- Pandas is a fundamental library for data manipulation and analysis. It provides data structures like DataFrames, which make it easy to handle and process structured data.
- Usage: Loading the dataset, cleaning data, manipulating data, performing descriptive statistics, and preprocessing.

#### 2. NumPy

- NumPy is essential for numerical operations. It offers support for arrays and matrices, along with a collection of mathematical functions to operate on these data structures.
- Usage: Handling arrays and performing mathematical computations efficiently.

#### 3. Matplotlib

- Matplotlib is a widely-used library for creating static, animated, and interactive visualizations in Python.
- Usage: Basic plotting and creating detailed visual representations of the data.

#### 4. Seaborn

- Seaborn is built on top of Matplotlib and provides a high-level interface for drawing attractive and informative statistical graphics.

- Usage: Visualizing relationships between different variables and enhancing the aesthetics of plots.

Data Loading: Data is loaded from a CSV file into a DataFrame.

- **Loading Data from CSV:** The data for our supply chain analysis is loaded from a CSV file named "15\_Supply Chain Analysis.csv" into a DataFrame using the Pandas library. This process involves reading the CSV file and storing its contents in a structured format that allows for easy manipulation and analysis.

Data Cleaning and Preprocessing: In the data cleaning and preprocessing phase, we perform several key steps to prepare the dataset for analysis. These steps ensure that the data is in a suitable format for further exploration and modeling. Here's how we handle the dataset based on the provided information:

- **Handling Missing Values**
  - Identify and visualize missing values in the dataset.
  - Decide on strategies to handle missing values, such as imputation with mean, median, or mode, or removal of rows/columns with excessive missing data.
- **Categorical Data Transformation**
  - Convert categorical data into numerical format to facilitate analysis. This is typically done using techniques such as one-hot encoding.
- **Converting Dates to Datetime Objects**

- If there are any date columns in the dataset, convert them to datetime objects to enable time-series analysis and date-based manipulations.

- **Data Transformation and Scaling**

- Perform data transformations such as scaling or normalization to standardize the numerical features. This ensures that features are on a similar scale, which is important for many machine learning algorithms.

By performing these data cleaning and preprocessing steps, we ensure that our dataset is clean, well-structured, and ready for detailed analysis and modeling. This preparation is crucial for deriving accurate insights and making informed decisions based on the supply chain data.

### III. Data Analysis Techniques

#### **Descriptive Statistics:**

Descriptive statistics provide a quick overview of the data through various summary metrics. These statistics help us understand the distribution, central tendency, and variability of the data:

- **Mean:**
  - The average value of each numerical attribute. It gives an overall idea of the central tendency of the data.

**Price:** 49.46

**Availability:** 48.40

**Number of products sold:** 460.99



**Revenue generated:** 5776.05

**Stock levels:** 47.77

- **Median:**

- The middle value when the data is sorted. It helps understand the central tendency without being affected by outliers.

**Price:** 49.46

**Availability:** 48.40

**Number of products sold:** 460.99

**Revenue generated:** 5776.05

**Stock levels:** 47.77

- **Mode:**

- The most frequently occurring value in the dataset. It provides insight into the most common occurrences.

**Price:** 1.70

**Availability:** 11.00

**Number of products sold:** 134.00



**Revenue generated:** 1061.62

**Stock levels:** 5.00

- **Standard Deviation:**

- Measures the dispersion or spread of the data. A higher standard deviation indicates more variability.

**Price:** 31.17

**Availability:** 30.74

**Number of products sold:** 303.78

**Revenue generated:** 2732.84

**Stock levels:** 31.37

- **Range:**

- The difference between the maximum and minimum values. It provides a sense of the span of the data.

**Price:** 97.47

**Availability:** 99.00

**Number of products sold:** 988.00

**Revenue generated:** 8804.85





**Stock levels:** 100.00

- **Count:**

- The number of non-null entries in each column. It helps identify missing values and the completeness of the dataset.

**Price:** 100

**Availability:** 100

**Number of products sold:** 100

**Revenue generated:** 100

**Stock levels:** 100

### **Insights from Descriptive Statistics**

#### **1. Price Distribution:**

- The mean price is around 49.46 with a high standard deviation of 31.17, indicating a wide range of prices. The skewness of the distribution is evident from the difference between the mean (49.46) and the median (51.24).

#### **2. Availability:**

- The availability of products has a mean of 48.40 and a median of 43.50, suggesting a fairly uniform distribution with some variability (standard deviation of 30.74).

### 3. Sales:

- The mean number of products sold is 460.99, but the median is lower at 392.50, indicating a right-skewed distribution with some products having very high sales.

### 4. Revenue:

- Revenue generated shows a significant spread with a mean of 5776.05 and a high standard deviation of 2732.84. The right-skewed distribution is indicated by the higher mean compared to the median (6006.35).

### 5. Stock Levels:

- Stock levels vary widely with a mean of 47.77 and a standard deviation of 31.37. The distribution appears uniform, with significant variability.

### 6. Lead and Shipping Times:

- Lead and shipping times show variability, with standard deviations of 8.78 and 2.72, respectively, indicating different efficiencies and processes across the supply chain.

By examining these descriptive statistics, we gain a comprehensive understanding of the data's central tendencies and variability. This foundational understanding is critical for further analysis, such as identifying trends, anomalies, and potential areas for optimization within the supply chain.

## Data Visualization

Data visualization provides a graphical representation of the dataset, making it easier to identify patterns, trends, and insights. The following visualizations and their corresponding insights are derived from the supply chain data:

## 1. Histograms

Histograms provide a visual representation of the distribution of numerical data.

- **Price Distribution:**

- The histogram shows a wide spread of prices with a peak around the lower end, indicating that there are many products with lower prices. The distribution is right-skewed.

- **Availability Distribution:**

- The availability of products is fairly uniform with slight peaks at certain points, indicating some common stock levels.

- **Number of Products Sold:**

- The distribution shows a concentration around lower numbers, suggesting that while some products have high sales, many have relatively low sales.

- **Revenue Generated:**

- The histogram is right-skewed, indicating that most products generate lower revenue, with a few generating significantly higher revenue.

- **Stock Levels:**

- The stock levels show a relatively uniform distribution, suggesting that stock levels vary widely across different products.

## 2. Bar Charts and Pie Charts

Bar charts and pie charts are used to visualize categorical data and proportions.

- **Device Preference by Country (Hypothetical Example):**

- **Bar Chart:** Visualizing device preferences (e.g., mobile, tablet, desktop) across different countries. This can reveal which devices are more popular in which regions.
- **Pie Chart:** Showing the proportion of device usage in a particular country. This helps in understanding the market share of each device type.
- **Gender Distribution (Hypothetical Example):**
  - **Bar Chart:** Displaying the number of male and female customers. This can reveal the gender distribution among the customers.
  - **Pie Chart:** Showing the proportion of male to female customers, providing insights into the demographic makeup of the customer base.

### 3. Heatmaps

Heatmaps are useful for visualizing missing values and correlations between variables.

- **Missing Values Heatmap:**
  - The heatmap highlights where missing values are located in the dataset. This helps in identifying patterns and deciding on data cleaning strategies.
- **Correlation Heatmap:**
  - This visualizes the correlation between different numerical variables. High correlations can suggest relationships between variables, which can be useful for predictive modeling.



## Insights from Visualizations

### 1. Price Insights:

- Most products are priced on the lower end, but a few products are significantly more expensive. This suggests the presence of different product categories or quality levels.

### 2. Sales and Revenue:

- The number of products sold is generally low, with a few products having very high sales. Revenue generated follows a similar pattern, indicating that high sales do not necessarily correlate with higher revenue for all products.

### 3. Stock Levels:

- Stock levels vary widely, indicating different inventory management strategies or product demand levels.

### 4. Lead and Shipping Times:

- The variability in lead and shipping times suggests diverse supply chain processes and efficiency levels.

### 5. Device Preference by Country (Hypothetical):

- Specific devices might be preferred in certain countries, which can inform targeted marketing strategies.

### 6. Gender Distribution (Hypothetical):

- Understanding the gender distribution can help tailor marketing efforts and product offerings to the dominant customer segment.

These visualizations provide valuable insights into the supply chain data, helping to identify key patterns and make data-driven decisions for optimization and strategic planning.

## IV. Key Findings

### User Demographics:


- **Age Distribution:** Across different countries, the analysis reveals varying age distributions among users. For example, while one country may have a younger demographic profile, another country may have a more balanced distribution across different age groups. Understanding these demographics helps in tailoring marketing strategies and product offerings to specific age segments.
- **Gender Distribution:** The analysis also sheds light on the gender distribution among users. By examining gender distribution across different countries, we can identify any gender imbalances and adjust marketing campaigns accordingly. Additionally, understanding gender preferences can help in product development and feature prioritization.

### Device Usage:

- The analysis provides insights into the device preferences of users. By examining device usage patterns among different user segments, we can determine which devices are most popular and optimize user experiences accordingly. For instance, if mobile devices are the preferred choice among a certain demographic, it is essential to ensure that the company's website or application is optimized for mobile viewing.

### Subscription Details:

- Exploration of subscription plan preferences across various user demographics reveals valuable insights into user preferences and behaviors. By analyzing subscription plan



adoption rates among different age groups, genders, and countries, we can identify trends and tailor subscription offerings to better meet the needs of specific user segments. This understanding also aids in customer retention strategies and the development of targeted promotional campaigns to encourage subscription upgrades or renewals.

These key findings provide valuable insights into user demographics, device usage patterns, and subscription preferences, enabling the company to make data-driven decisions for marketing, product development, and customer engagement strategies.

## **V. Advanced Analysis**

**Geographical Insights:** Utilizing custom functions to categorize countries into continents allows for a comprehensive regional analysis of user behavior and market trends. By grouping countries into continents, we gain valuable insights into geographical patterns that can inform strategic decision-making and optimization of business operations.


- **Regional Market Trends:** Grouping countries into continents enables us to identify regional market trends and consumer preferences. For example, we can analyze sales data to determine which product categories are more popular in specific continents or regions.
- **Supply Chain Optimization:** Understanding geographical patterns allows for efficient supply chain management. By identifying regions with higher demand for certain products, companies can optimize inventory levels, distribution networks, and logistics to meet customer needs effectively.

- **Targeted Marketing Strategies:** Geographical insights help in crafting targeted marketing strategies tailored to specific regions. For instance, promotional campaigns can be customized based on cultural differences, seasonal variations, or regional preferences, leading to improved customer engagement and conversion rates.

**Temporal Trends:** Analyzing user sign-up trends over months enables the detection of seasonal patterns in user registration, providing valuable insights into user behavior and engagement dynamics over time. This temporal analysis helps in optimizing marketing strategies, resource allocation, and product planning.

- **Seasonal Variation in User Engagement:** By examining monthly sign-up trends, we can identify seasonal fluctuations in user engagement. For example, certain months may exhibit higher sign-up rates due to seasonal events, holidays, or promotional activities.
- **Optimizing Marketing Campaigns:** Understanding temporal trends allows companies to optimize marketing campaigns by focusing resources on periods of peak user activity. For instance, promotional offers or product launches can be timed to coincide with periods of high user engagement, maximizing their effectiveness.
- **Forecasting and Planning:** Temporal analysis aids in forecasting demand, planning inventory levels, and resource allocation. By anticipating seasonal fluctuations in user demand, companies can adjust production schedules, inventory levels, and staffing resources to meet expected changes in customer needs.





By leveraging geographical insights and temporal trends analysis, companies can make informed decisions to optimize their operations, enhance customer experiences, and drive business growth.

## VI. Visual Insights

- **Gender Distribution:**

Count plots show the distribution of users by gender across different countries. This visualization provides insights into the gender demographics of the user base, allowing for targeted marketing strategies and product customization based on gender preferences.

- **Device Preference by Country:**

Insights into the most preferred devices in different countries help in understanding user behavior and optimizing user experiences. By visualizing device preferences across various countries, companies can tailor their digital platforms and applications to suit the most commonly used devices, thereby enhancing user satisfaction and engagement.

- **Subscription Type Popularity:**

Visualization of the popularity of different subscription types among the user base enables companies to understand user preferences and subscription adoption rates. By visualizing subscription type popularity, companies can tailor subscription offerings, pricing models, and promotional campaigns to better meet the needs and preferences of their user base, ultimately driving subscription growth and revenue.


## VII. Conclusion

In conclusion, the supply chain analysis has provided valuable insights into various aspects of the company's operations and customer behavior. The key findings indicate that while the number of products sold is relatively low, the revenue generated is significantly high, suggesting a successful premium pricing strategy for high-value products. This analysis highlights the importance of maintaining and enhancing the value proposition of these high-value items to sustain or boost revenue.

Additionally, the predictable sales patterns and low variability in the number of products sold suggest opportunities for optimizing inventory management and reducing holding costs. Targeted marketing efforts and tailored customer segmentation can further increase sales volume and cater to diverse customer segments with varying purchasing power.

Moreover, by examining user demographics, device preferences, and subscription type popularity, we gained actionable insights into customer segmentation and preferences. These insights are instrumental in shaping marketing strategies, product development initiatives, and supply chain optimization efforts. For instance, understanding gender distribution and device preferences across different countries enables targeted marketing campaigns and tailored user experiences. Additionally, insights into subscription type popularity aid in refining subscription offerings and pricing strategies to better align with customer needs and preferences, ultimately driving subscription growth and revenue generation.

The insights derived from this analysis can have a profound impact on the business by guiding strategic decisions in product pricing, marketing, inventory management, product development, and



sales channels. By leveraging these findings, the company can implement data-driven strategies to improve customer satisfaction, increase market competitiveness, and achieve sustainable growth in today's dynamic business landscape.