Sidharth Santhosh

AJ23SYD011

**DataCo Supply Chain Analysis**

**Problem Statement:**

A DataSet of Supply Chains used by the company DataCo Global was used for the analysis.

Order analysis to determine large volumes of sales: The goal of this study is to measure the sales volume for various markets and how delivery status influences sales. It also determines how long it takes for orders to arrive at their destination, which aids in determining the optimal delivery alternatives for markets.

Customer analysis based on total sales: This analysis determines total sales based on order area and customer segment. It aids in determining the benefits and total profit made by various consumer categories as well as their purchasing habits.

Top product category sales analysis: From 2015 to 2018, this analysis identifies the top product categories that contributed the most overall sales.

Market analysis for sales: Based on the market, this analysis determines the year with the biggest sales volume.

Fraud orders analysis: This data analysis shows the top five counties with the highest number of suspected fraud orders from 2015 to 2018.

Analysis of shipping options: This analysis examines several shipping options and determines the most often used shipping modes.

Late delivery analysis: This analysis determines how delayed shipments effect firm profitability across markets and consumer categories. Based on market data, calculates the year with the latest deliveries. identifies the various departments where delayed shipments occur and assists in the development of solutions to decrease delays.

Overall, the goal of these analyses is to identify the business's strengths and shortcomings to enhance its overall performance and profitability. They aid in the understanding of the market, consumer behaviour, and the impact of various elements such as shipping, delivery, and fraud on the firm.

**Dataset:**

Dataset Link:- <https://www.kaggle.com/code/cheukhangtse/dataco-supply-chain-analysis>

This dataset contains 53 columns and 180519 entries. The column contains :

'Type', 'Days for shipping (real)', 'Days for shipment (scheduled)', 'Benefit per order', 'Sales per customer', 'Delivery Status', 'Late\_delivery\_risk', 'Category Id', 'Category Name', 'Customer City', 'Customer Country', 'Customer Segment', 'Customer State', 'Department Name', 'Market', 'Order City', 'Order Country', 'Order Customer Id', 'order date (DateOrders)', 'Order Id', 'Order Item Discount', 'Order Item Discount Rate', 'Order Item Product Price', 'Order Item Profit Ratio', 'Order Item Quantity', 'Sales', 'Order Item Total', 'Order Profit Per Order', 'Order Region', 'Order State', 'Order Status', 'Product Name', 'Product Price', 'Product Status', 'shipping date (DateOrders)', 'Shipping Mode'

**Data Cleaning:**

Following extensive data exploration and feature analysis, the columns were cleaned/removed:

"Customer Email", "Customer Id", "Customer Fname", "Customer Lname", "Customer Password", "Customer Zipcode", "Latitude", "Longitude", "Order Item Cardprod Id", "Order Item Id", "Order Zipcode", "Product Card Id", "Product Image", "Product Category Id", "Order Item Id", "Department Id", "Customer Street","Product Description"

As a result of the aforementioned columns either having null values or being empty. These columns are not necessary according to the problem statement.

The columns that include dates were all brought together into a single format of month/day/year for all the columns because there is more data and years in this dataset.

**Exploratory Data Analysis:**

The distribution of numerical values in a dataset can be represented using a histogram graph, a kind of data visualisation. It operates by creating a series of bins from the data range and counting the number of data points that fall into each bin. After that, a bar graph is created using the obtained counts, with each bar representing a bin.

Chart

Description automatically generated

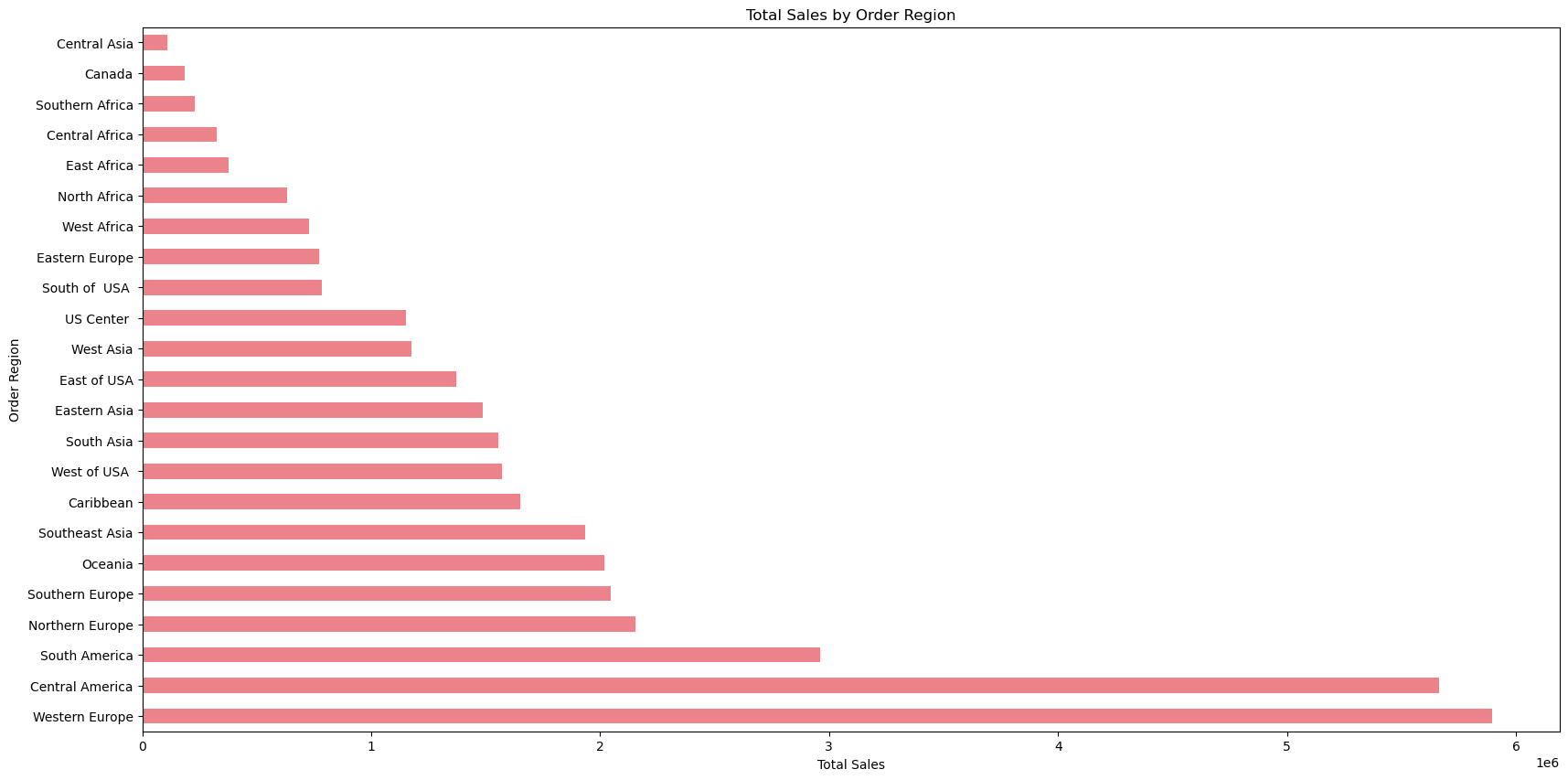
**Correlation Matrix:**

A correlation matrix is a table that displays the correlation coefficients between a set of variables. A statistical metric that shows how closely two variables are connected to one another is the correlation coefficient. Its values fall between -1 and +1, with +1 being a perfect positive correlation, 0 denoting no connection, and -1 denoting a perfect negative correlation.

**Chart, treemap chart

Description automatically generated**

**Order analysis to determine large volumes of sales:**

****From 2015 to 2018, Central Asia had the fewest sales, according to the observation, while Western Europe had the most.

**Market-based delivery status:**

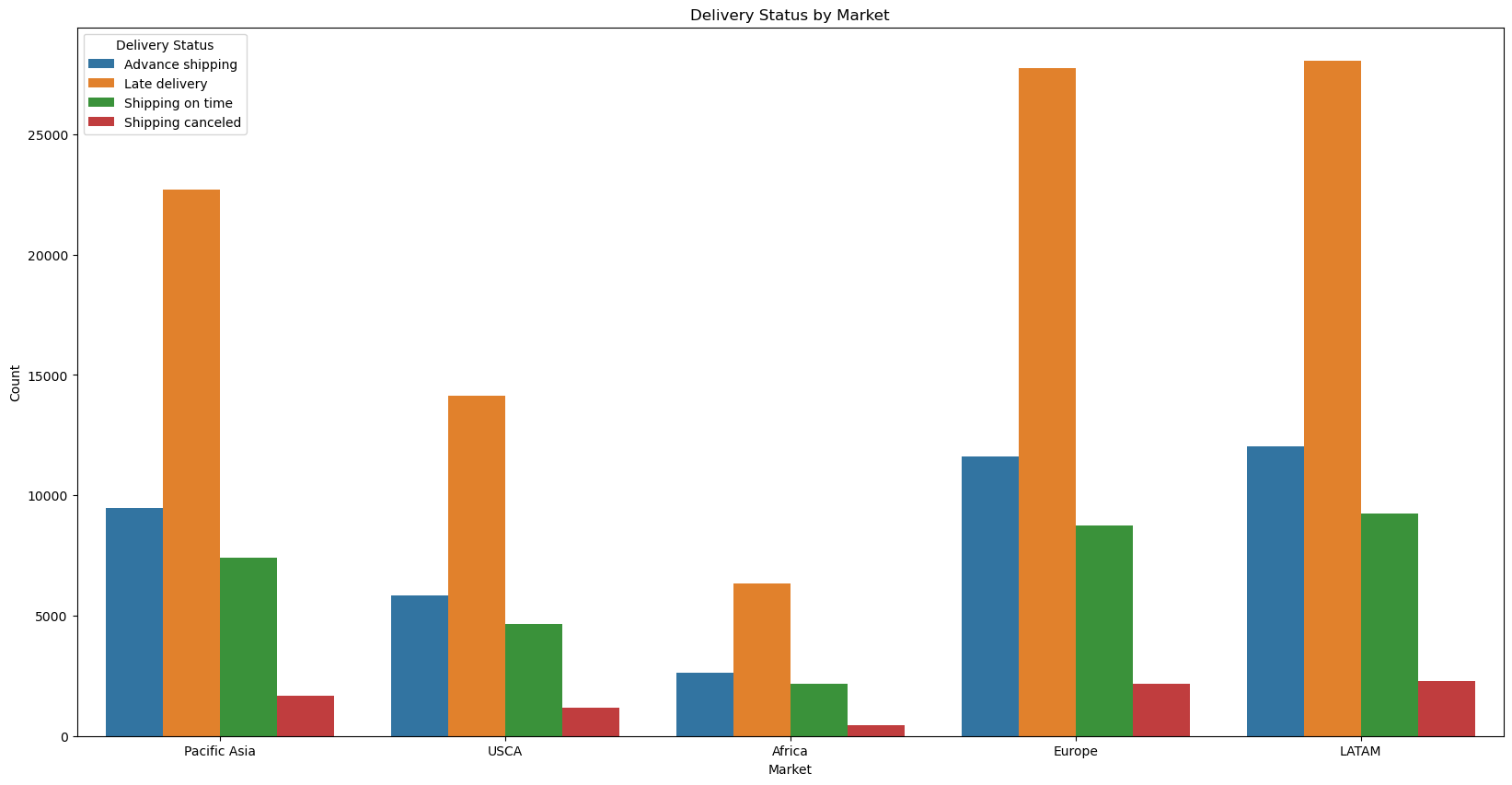
According to the data, there are four possible delivery statuses, which include:

1.Advance Shipping

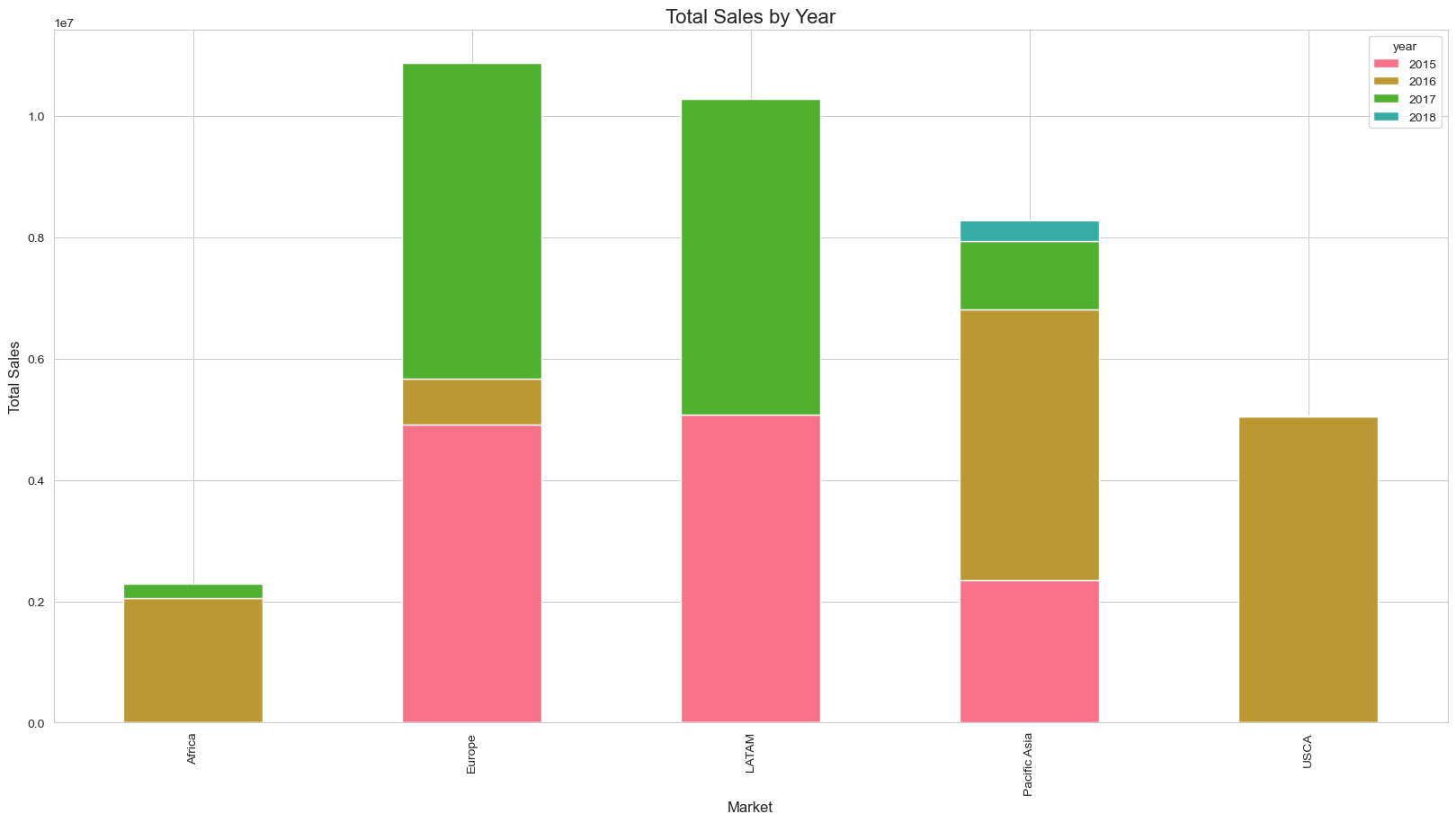
2.Late delivery

3.Shipping on time

4.Shipping cancelled

****

**Analysing which year had the highest amount of sales based on market:**

From 2015 through 2018, the year with the most sales was 2017. Europe and Latin America had higher sales than the rest of the Market.****

**Top ten most customers in terms of total sales:**

The most frequent item sold is fishing equipment, while computer equipment is the least popular.

Chart, bar chart

Description automatically generated

**Customer analysis based on total regional sales:**

**Chart, bar chart, box and whisker chart

Description automatically generated**

**Customer Analysis based on benefit and profit:**

**Chart, bar chart, box and whisker chart

Description automatically generated**

**Fraud orders analysis:**

The bulk of fraud orders have been reported from the Estados Unidos.Chart, bar chart

Description automatically generated

**Comparison on various Shipping options:**

As compared to other options, Standard class is used to deliver the majority of orders. Same-day shipments are not very common.

**Chart, pie chart

Description automatically generated**

**Market-based analysis of delayed shipments:**

African has the fewest amount of delayed shipments, while Europe and LATAM have the highest.

Chart, bar chart

Description automatically generated

**An analysis of late deliveries in multiple years:**

**Chart, bar chart

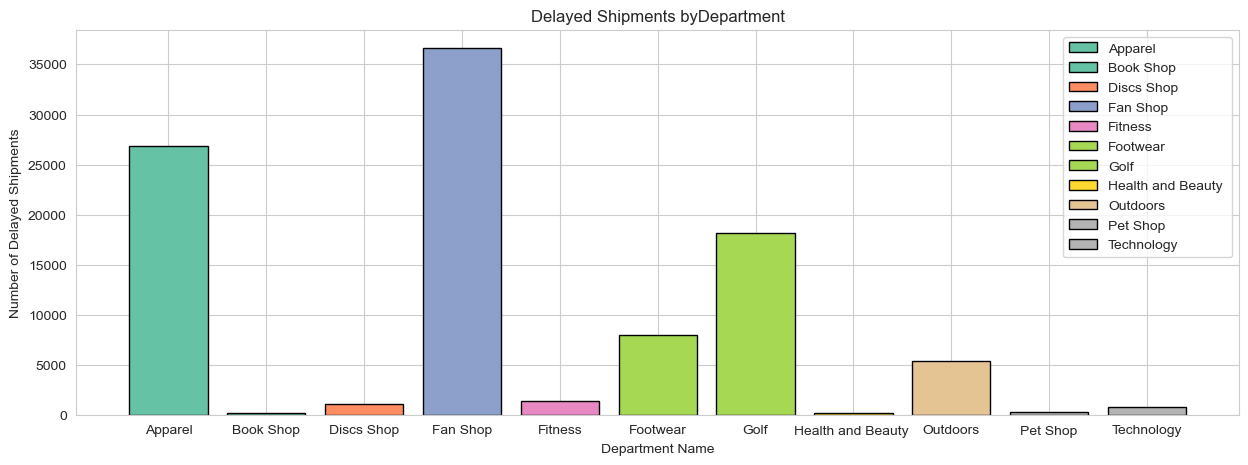
Description automatically generated**

**Customer Segment Analyses of Delayed Shipments:**

**Chart, bar chart

Description automatically generated**

**Analyses of Delays in Shipments Based on Department:**

****

**Based on the departments, different markets took longer to deliver the order:**

****