**INDUSTRY OVERVIEW**

The **supply chain** industry manages how goods move from where they are made to where people use them. It includes planning, storing, and transporting products. Companies like **FedEx** form an essential link in this chain. They collect packages, sort them in hubs, and deliver them to customers across many regions. The industry’s main goal is to move items quickly, safely, and at the lowest cost possible. To do this, companies depend on accurate data about distance, time, and demand. Small delays or errors in one stage can affect the entire system. As online shopping has grown, the need for efficient delivery systems has also increased. Data analysis now plays a key role in improving delivery routes, reducing idle time, and managing shipment delays.

The **stock market** connects closely to how well companies in the supply chain operate. When a logistics company manages deliveries efficiently, it often reduces costs and increases profit, which can raise its stock value. Investors watch delivery performance, customer satisfaction, and on-time rates to judge a company’s stability. For a company like FedEx, consistent delivery service builds customer trust and influences long-term growth. By studying FedEx’s delivery data, we can learn how operational factors affect both day-to-day efficiency and the company’s broader financial position.

**PROBLEM STATEMENT**

FedEx operates a large and complex delivery network that moves thousands of shipments every day. Timely delivery is a key measure of its performance. When shipments are delayed, customer trust and operational efficiency can decline. Understanding why delays happen helps improve planning and service quality.

This project focuses on studying shipment data to find patterns that cause delivery delays. It examines how factors like distance, weight, and shipment date affect the delivery time. It also studies demand changes over different days, weeks, and months to identify peak periods that may strain resources.

Furthermore, the project explores the link between delivery performance and FedEx’s stock market data. By comparing operational results with stock movements, it seeks to understand whether logistics efficiency reflects the company’s financial stability. The goal is to provide clear insights that can help improve delivery reliability and support better decision-making.

**COMPANY OVERVIEW**

**FedEx Corporation** is a global leader in logistics and transportation. It provides services that include express delivery, freight, e-commerce, and supply chain management. Founded in 1971 and headquartered in Memphis, Tennessee, FedEx connects businesses and customers across more than 200 countries and territories. Its core strength lies in its ability to move packages quickly and reliably through an extensive network of air and ground transportation.

The company operates several business segments, such as FedEx Express, FedEx Ground, and FedEx Freight. Each segment serves different customer needs but shares the same goal — to deliver shipments safely and on time. FedEx Express focuses on fast air delivery, while FedEx Ground manages cost-effective shipping within regions. FedEx Freight handles large and heavy goods for commercial customers.

Technology plays a major role in FedEx operations. The company uses tracking systems, data analytics, and automation to monitor shipments and improve performance. These tools help manage millions of packages each day while reducing errors and costs. FedEx continues to invest in innovation to strengthen its logistics network and maintain its position in a competitive global market.

**DATA DESCRIPTION**

The dataset records FedEx shipment operations in a structured format. It contains **15 columns** with both numeric and categorical data. Each record shows when and where a shipment was sent, how far it traveled, and how long it took to reach its destination. The data includes planned and actual delivery times, which helps measure how well the company meets its delivery goals. It also captures delay information and distance values that support detailed performance analysis. This dataset forms the basis for studying delivery efficiency, shipment delays, and how FedEx’s operational performance connects to its broader supply chain activities.

Below is the summary of each column and its purpose:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Year** | The year when the shipment was made. This helps in grouping and trend analysis across different years. |
| **Month** | The month of the shipment. Used to identify seasonal patterns or monthly delivery trends. |
| **DayofMonth** | The specific day in the month when the shipment occurred. Supports fine-grained daily analysis. |
| **DayOfWeek** | Numeric representation of the day (1–7). Helps track performance differences between weekdays and weekends. |
| **Actual\_Shipment\_Time** | The actual time taken for a shipment to reach its destination, recorded in hours or minutes. Used to evaluate performance accuracy. |
| **Planned\_Shipment\_Time** | The expected time required for the shipment based on schedule or service commitments. |
| **Planned\_Delivery\_Time** | The target delivery time promised by FedEx for the shipment. Used to assess on-time performance. |
| **Carrier\_Name** | The name of the carrier or transport service handling the shipment. It helps compare performance between carriers. |
| **Carrier\_Num** | A numeric identifier assigned to each carrier, useful for grouping and aggregation. |
| **Planned\_TimeofTravel** | The planned duration of travel from source to destination, derived from scheduling data. |
| **Shipment\_Delay** | The delay time in delivery, calculated as the difference between actual and planned times. A positive value indicates late delivery, while a negative value indicates early delivery. |
| **Source** | The city or location where the shipment originated. Used for route-level analysis and network mapping. |
| **Destination** | The delivery city or endpoint of the shipment. Enables comparison of delivery times and delays across destinations. |
| **Distance** | The physical distance between the source and destination (in kilometers or miles). Helps study how distance impacts travel time and delivery reliability. |
| **Delivery\_Status** | A numeric or coded value showing whether a shipment was delivered on time, delayed, or still in transit. Used as a performance indicator. |

**DATA GAPS**

Total no. of entries: **3604175**

Based on the analysis of missing values, the following columns exhibit data gaps:

\*   **Actual\_Shipment\_Time**: 81602 missing values

\*   **Planned\_TimeofTravel**: 547 missing values

\*   **Shipment\_Delay**: 81602 missing values

\*   **Delivery\_Status**: 81602 missing values

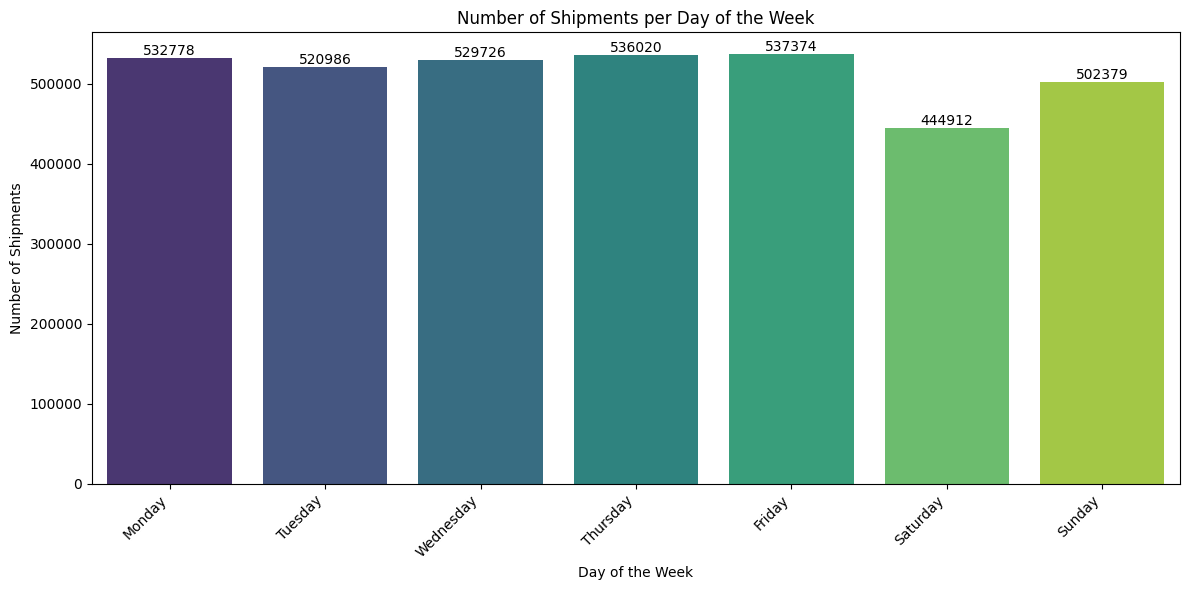
A notable observation is the identical count of missing values (81602) across `Actual\_Shipment\_Time`, `Shipment\_Delay`, and `Delivery\_Status`. This strong correlation in missingness suggests a potential dependency between these variables. It is possible that if the `Actual\_Shipment\_Time` is not recorded, it consequently impacts the calculation of `Shipment\_Delay` and the determination of `Delivery\_Status`. This could indicate scenarios where shipments were not initiated or tracked properly, leading to these data gaps.

The `Planned\_TimeofTravel` column has a smaller number of missing values (547). This could be attributed to different factors, such as instances where a planned travel time was not applicable or assigned, possibly for shipments that were not dispatched as initially planned.

**EXPLORATORY DATA ANALYSIS (EDA)**

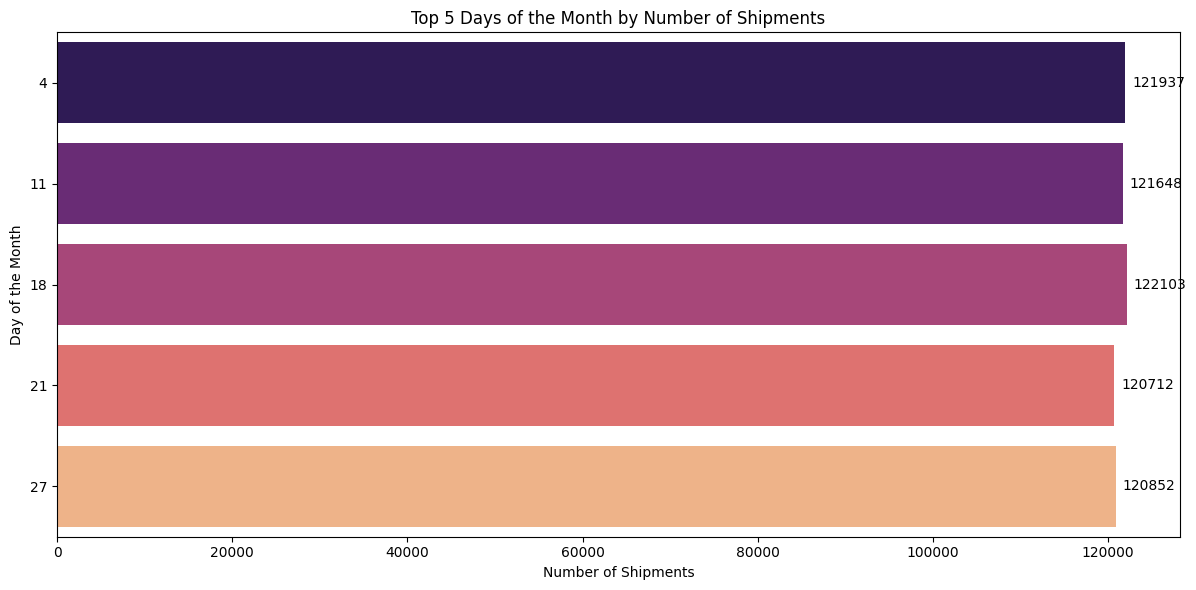
**1. TIME AND DEMAND ANALYSIS**

* **Number of Shipments per Day of the Week**

****

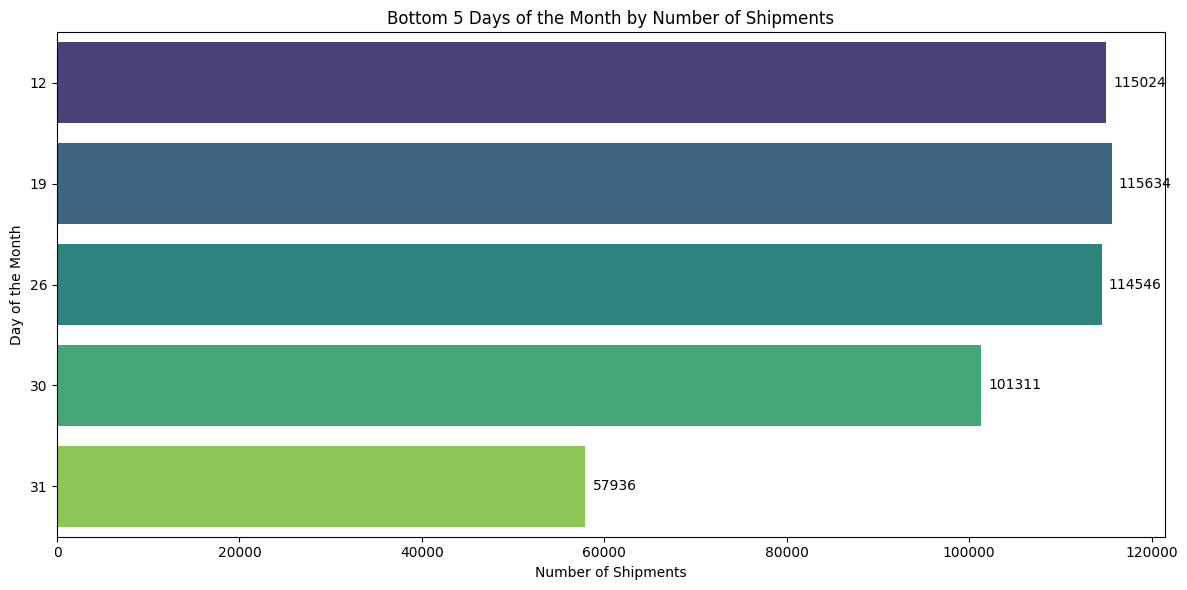
The above plot shows how shipment volumes vary across the days of the week.This is based on no of shipment placed, It shows the shipments are highest during weekdays, possibly due to people ordering during the weekdays to get their orders by weekends. This helps supply chain companies to manage their resources efficiently, like more delivery resources are going to be required for weekends while packaging and other verification resources are required for weekdays.

* **Top 5 Days of the Month by Number of Shipments**



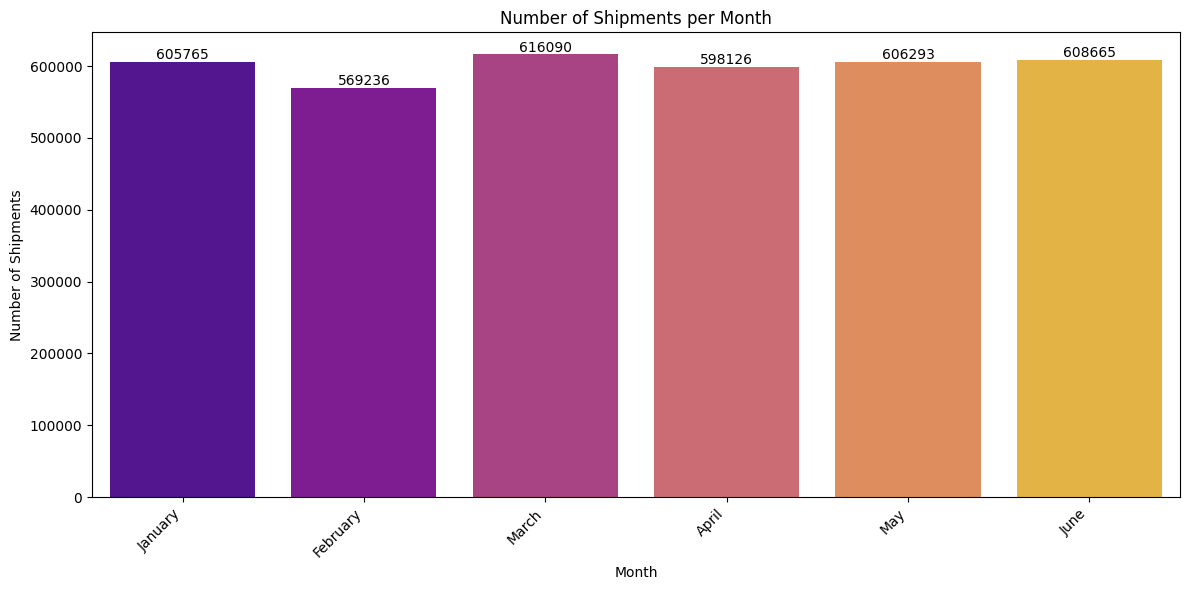
This horizontal bar plot shows the top 5 days of the month with the highest shipment volumes, i.e. the days when the shipments are maximum. We try to understand that if there is any pattern in customer spending on Orders as they reach near month ends salaried class tend to spend less but since orders contain all types of categories like essential parcels and giftings, so there is no proper trend in ordering on days of month.

* **Bottom 5 Days of the Month by Number of Shipments**



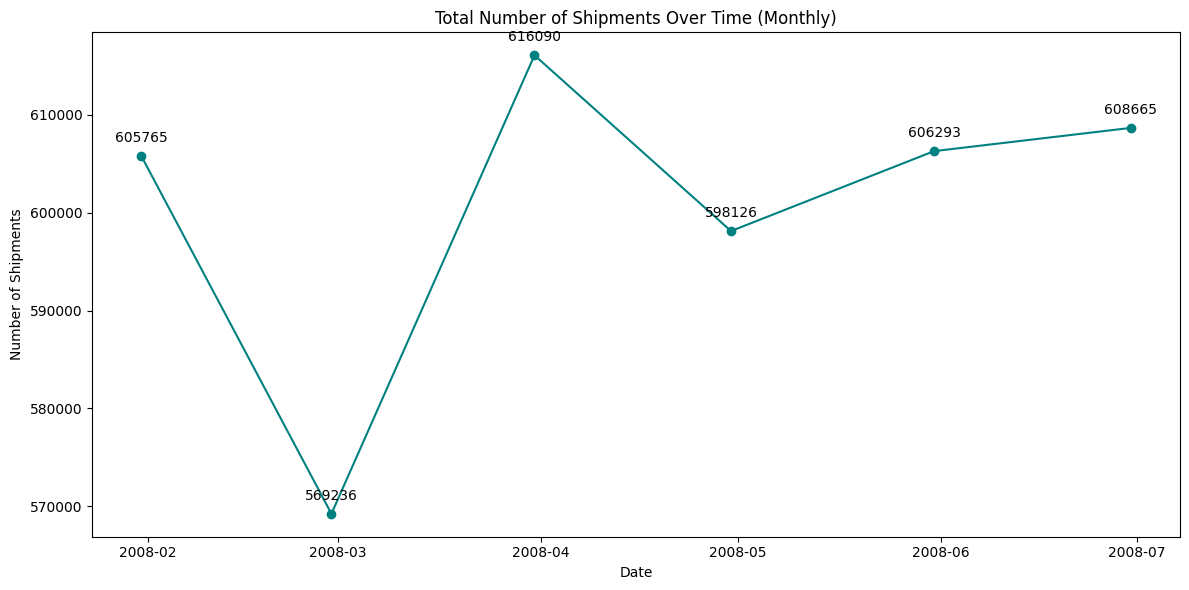
This plot shows that the shipment activity tends to **decrease near the end of the month**, possibly because of inventory reconciliation, financial closures, or fewer new orders being processed. Or it could be that people tend to want their orders at the start of the month when they have more money. From this it is clear that on the last two days of month people are spending less on Orders. Especially due to salaried class people and their spending behaviour during Month ends.

* **Number of Shipments per Month**



The above plot shows the consistency in monthly shipment counts reflects **steady logistics performance** and **balanced demand across quarters**. This plot helps us in understanding that there is any pattern in monthly data while placing orders. There is little possible decline in the month of February because a lot of Orders tend to be placed during Christmas and New Year along with essentials and Gifts, so following month tend to see a decline in orders then again orders can be seen with the same momentum.

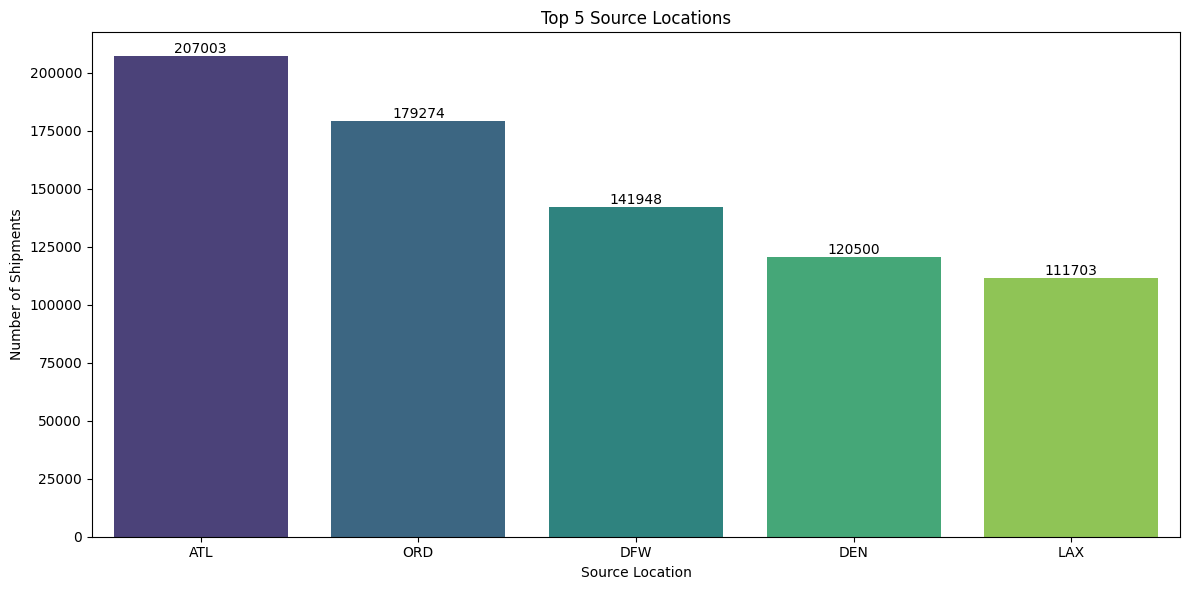
* **Total Number of Shipments Over Time (Monthly)**



This time series plot shows the total number of shipments on a monthly basis. We can observe the trends in the shipment volume over the recorded period. The spike in March likely represents increased business activity at the end of the first quarter. Overall, the trend line shows consistency with mild month-to-month variations, reflecting a well-balanced shipment operation through mid-year.

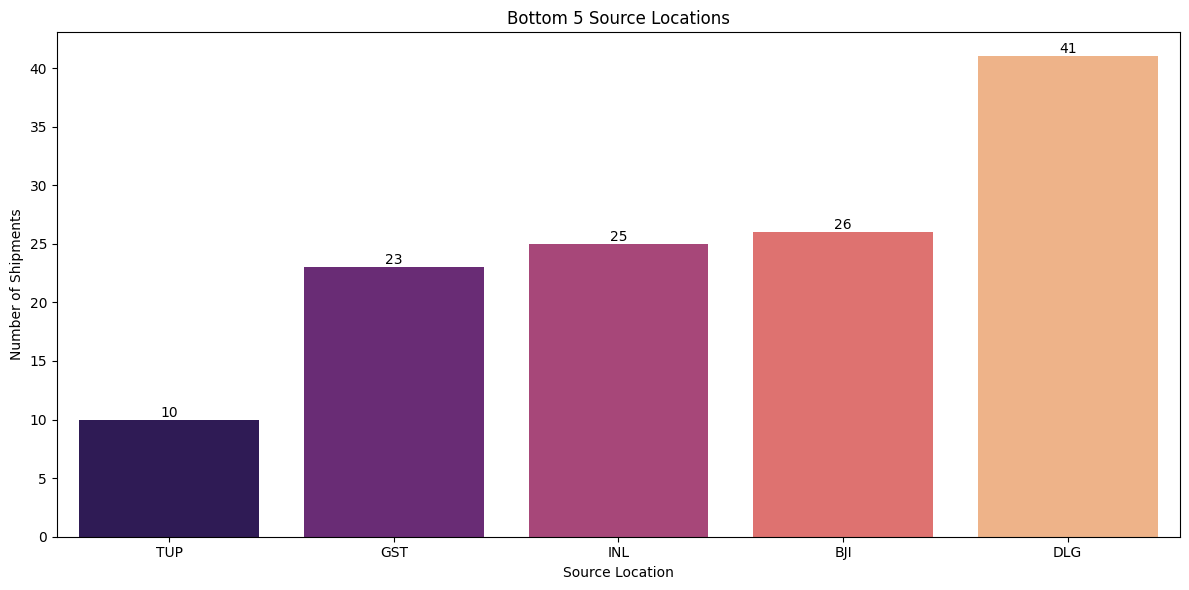
**2. GEOGRAPHICAL ANALYSIS**

* **Top 5 Source Locations**

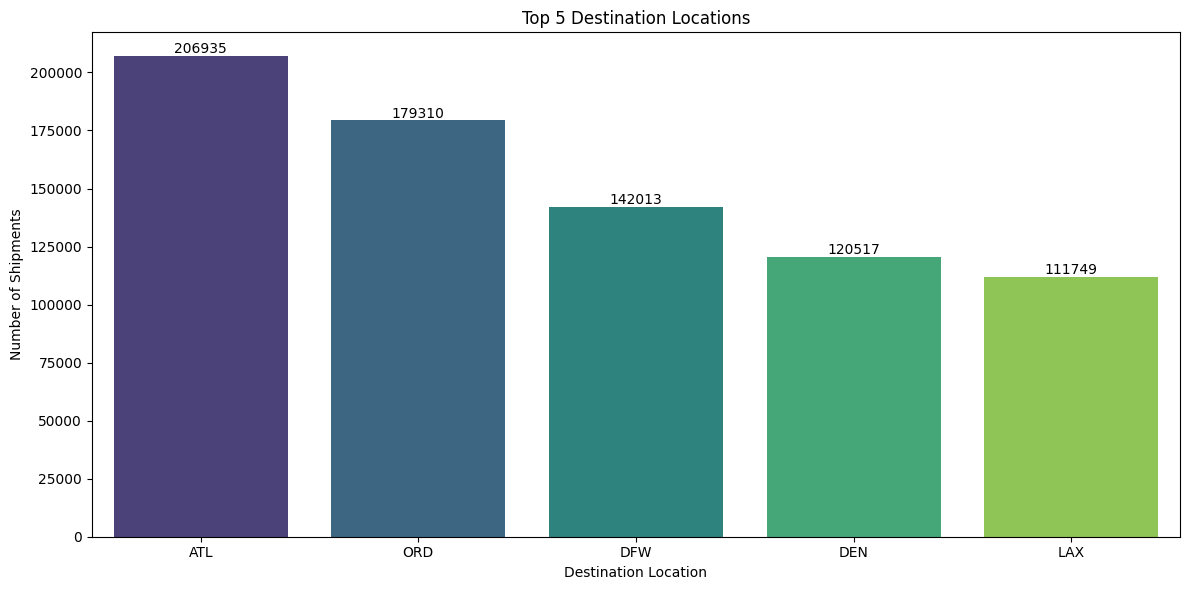
****

This bar plot shows the top 5 source locations with the highest number of outgoing shipments. These locations represent major hubs for shipments and **ATL dominates** as the primary source location, suggesting it’s a central hub for shipment distribution. This helps in understanding the geographical pattern that which region has the most no of orders , so we can utilize more resources in those regions for faster delivery. This is for the source location from where the most no of orders are being shipped.

* **Bottom 5 Source Locations**

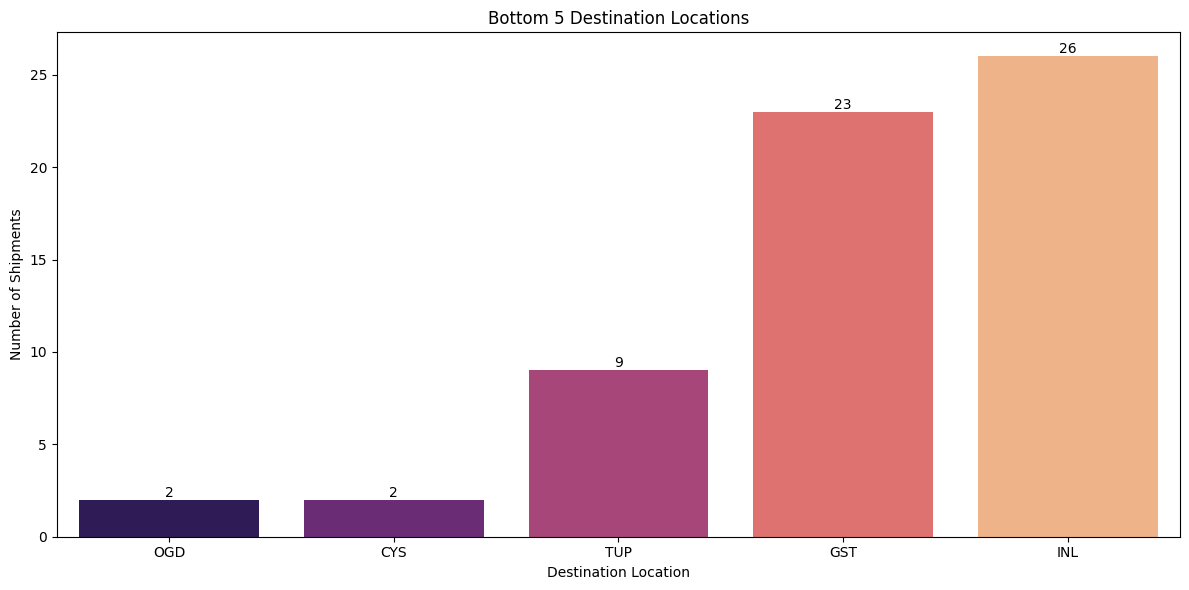
****This bar plot shows the bottom 5 source locations with the lowest number of outgoing shipments. These locations contribute negligibly to the overall shipment volume compared to top hubs (like ATL). This helps in understanding that from which region we have least no of source locations, so we focus more on these regions, and increase the number of orders from source locations.

* **Top 5 Destination Locations**

****

This bar plot shows the top 5 destination locations with the highest number of incoming shipments. The distribution pattern of top destinations is quite similar to that of the top source locations, the same cities dominate both ends of the shipment network. Our Source and Destination both have ATL as highest order placed and delivered , it shows that we have most of the operations coming from ATL because of smoother delivery. This helps us in understanding the methods or routes used in ATL so we can utilize that in other cities also.

* **Bottom 5 Destination Locations**



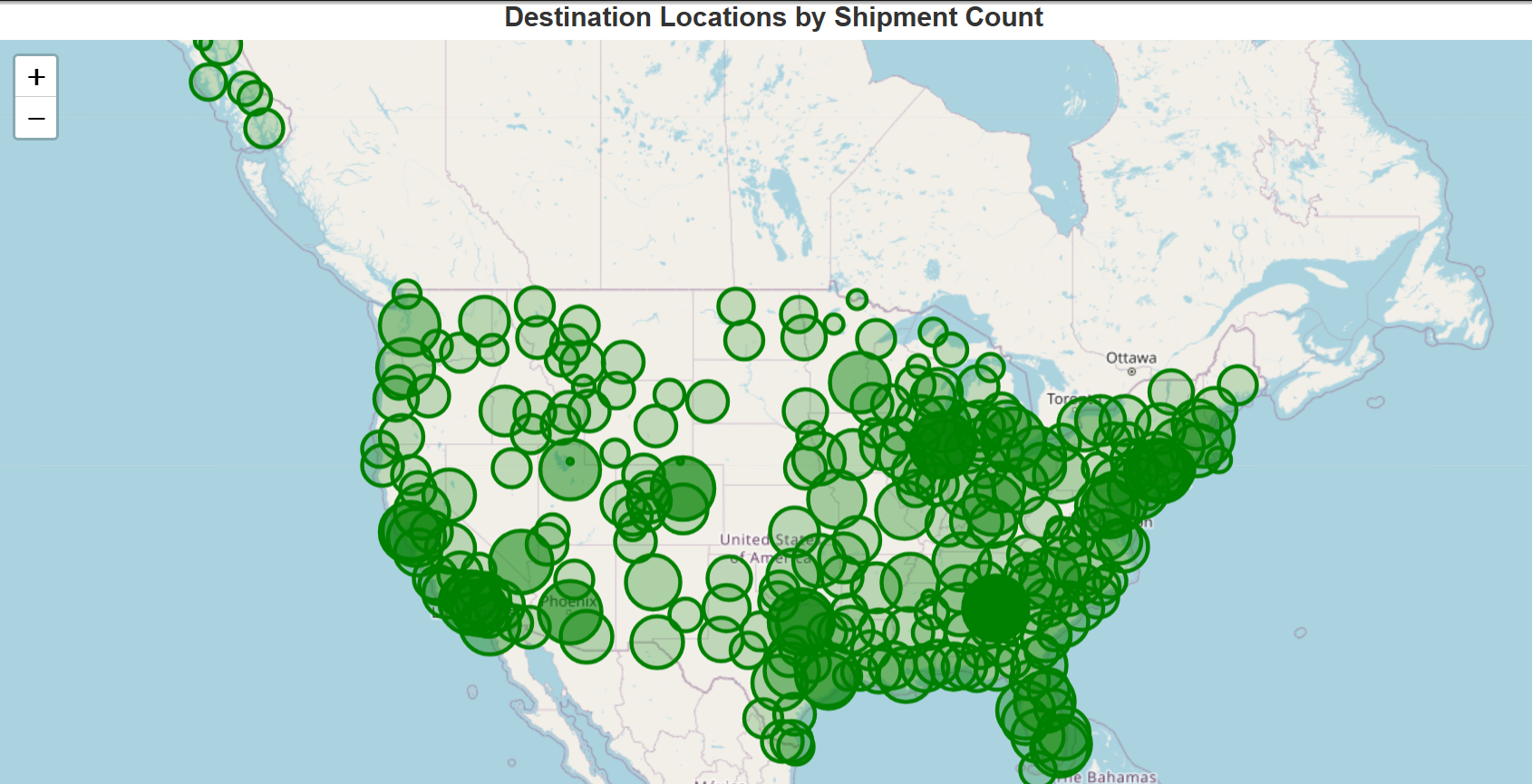
This bar plot shows the bottom 5 destination locations with the lowest number of incoming shipments. These locations have minimal inbound shipment activity, suggesting they are low-demand or remote destinations in the network. This helps shipment businesses to identify the reasons why we have very less orders for those regions. It is because of less connectivity to cities or other brands catering in that region.

* **Source Locations by Shipment Count**



This map shows the regions from which fedex orders are placed as source operations which is mostly in the USA and Canada and some of its neighbouring parts. This shows fedex operations are limited in those regions and possible extension to other regions.

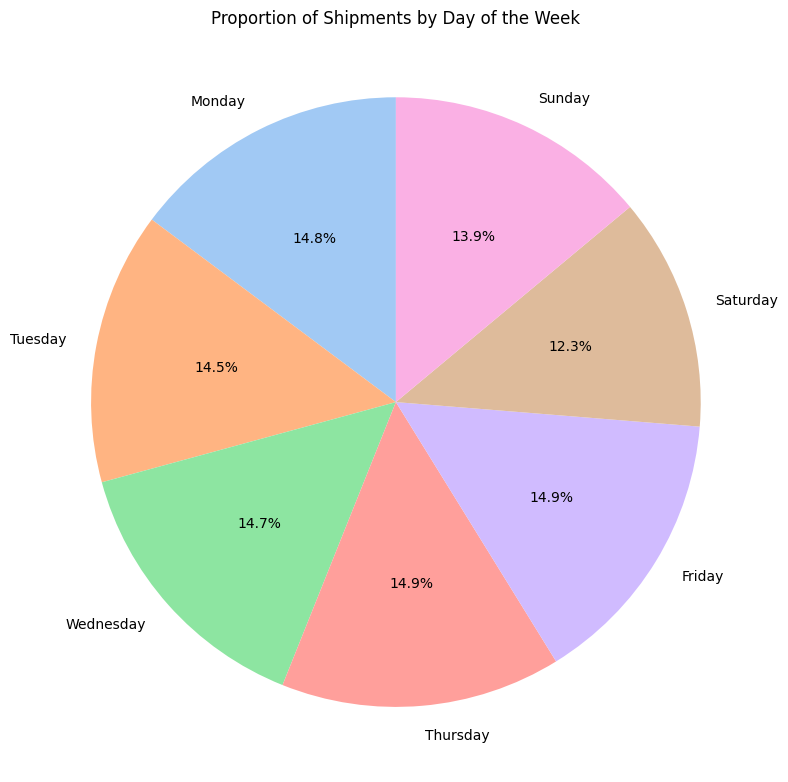
* **Destination Locations by Shipment Count**



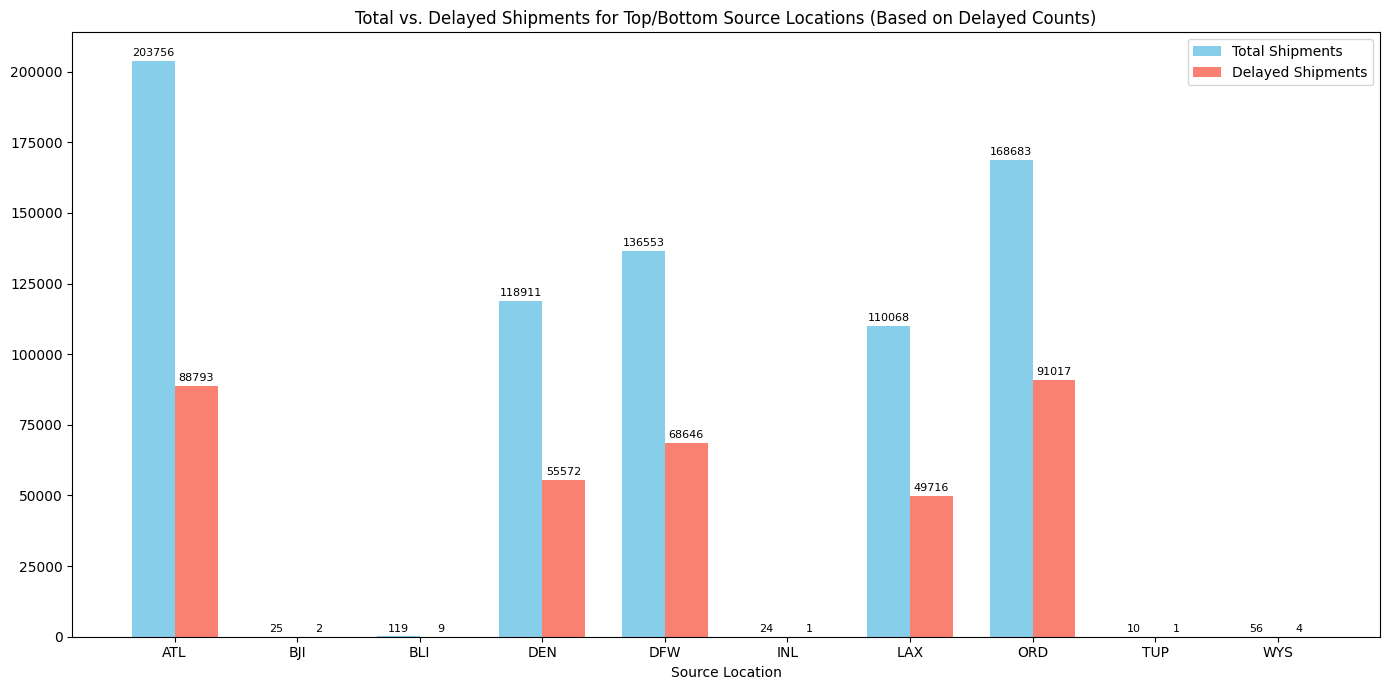
This map shows the regions from which fedex orders are to be delivered as fedex is operating mostly in the USA and Canada and some of its neighbouring parts. The graph shows a different pattern: for Source fedex orders are from many places around the USA but for destination is limited to the central part of the USA.

**3. SHIPMENTS ANALYSIS**

* **Proportion of Shipments by Day of the Week**

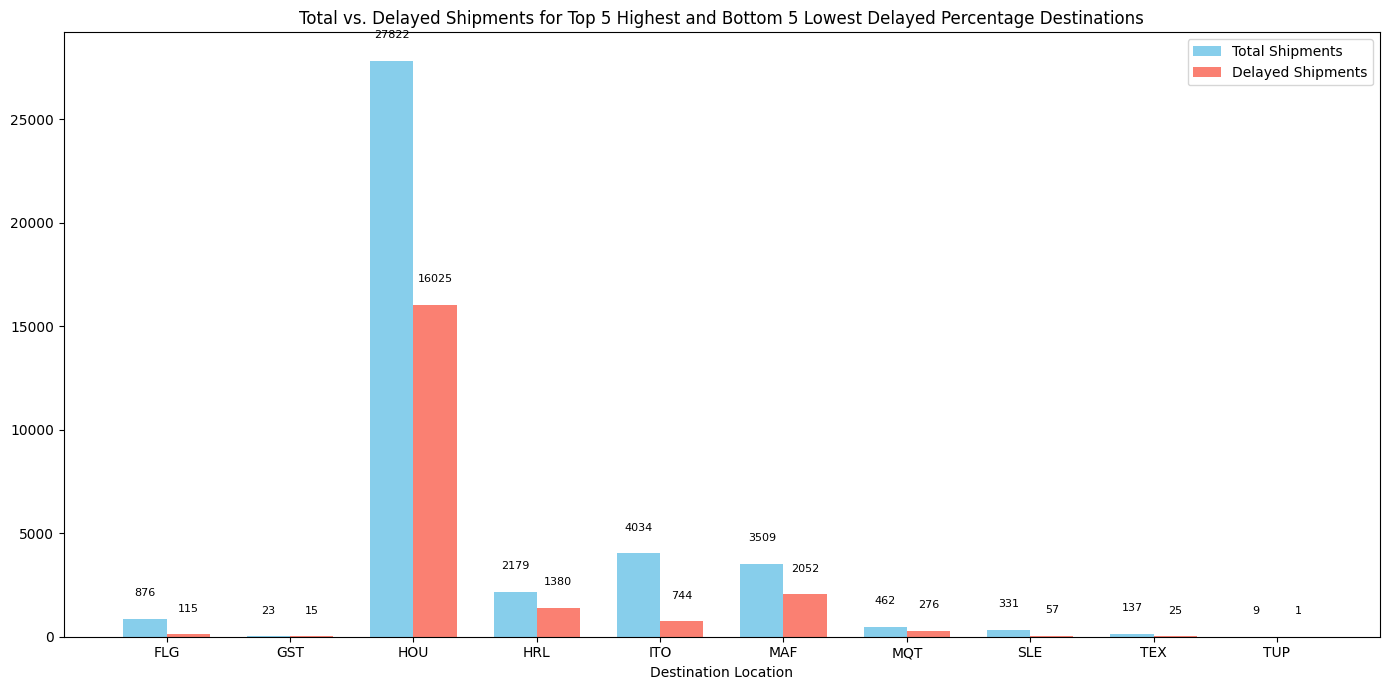
This pie chart titled “Proportion of Shipments by Day of the Week” shows how shipments are distributed across different days. The majority of shipments occur Monday through Friday, ****showing a weekday-dominant shipping pattern.

* **Total vs. Delayed Shipments for Bottom Source Locations**



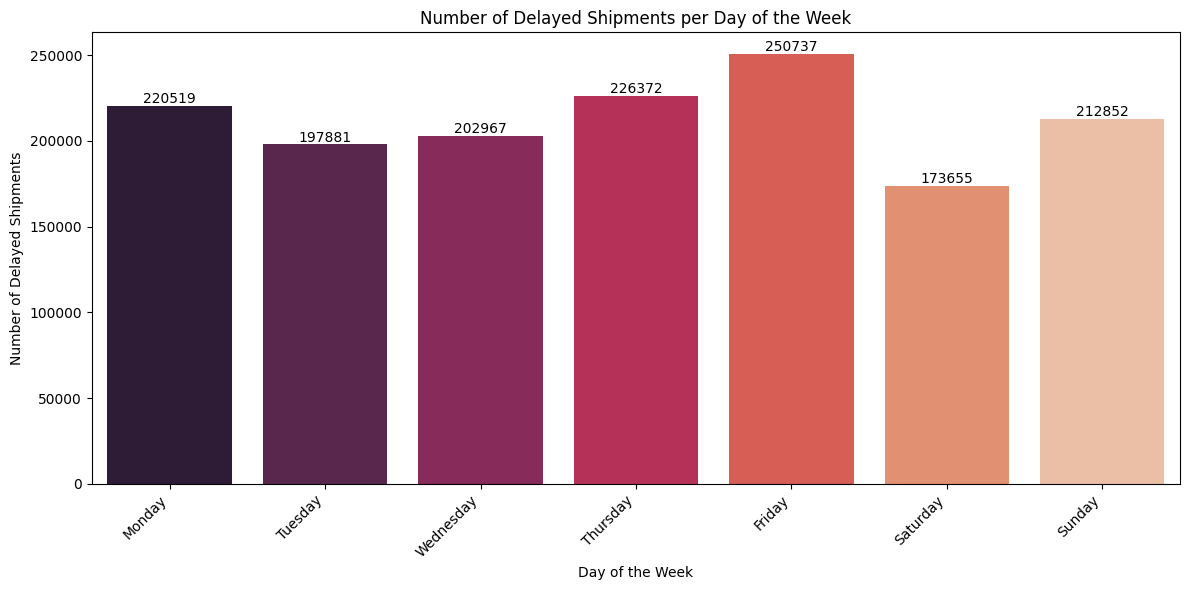
The majority of shipment delays occur at high-volume source locations, indicating that **delay frequency is correlated with shipment volume**. These are the least with the least number of delayed shipments, although these numbers seem like a lot but it's common in the shipment industry. ATL has the lowest percentage of delayed orders which explains the better connectivity, service which leads to the highest number of orders from ATL.

* **Total vs. Delayed Shipments for Bottom Destinations**

****

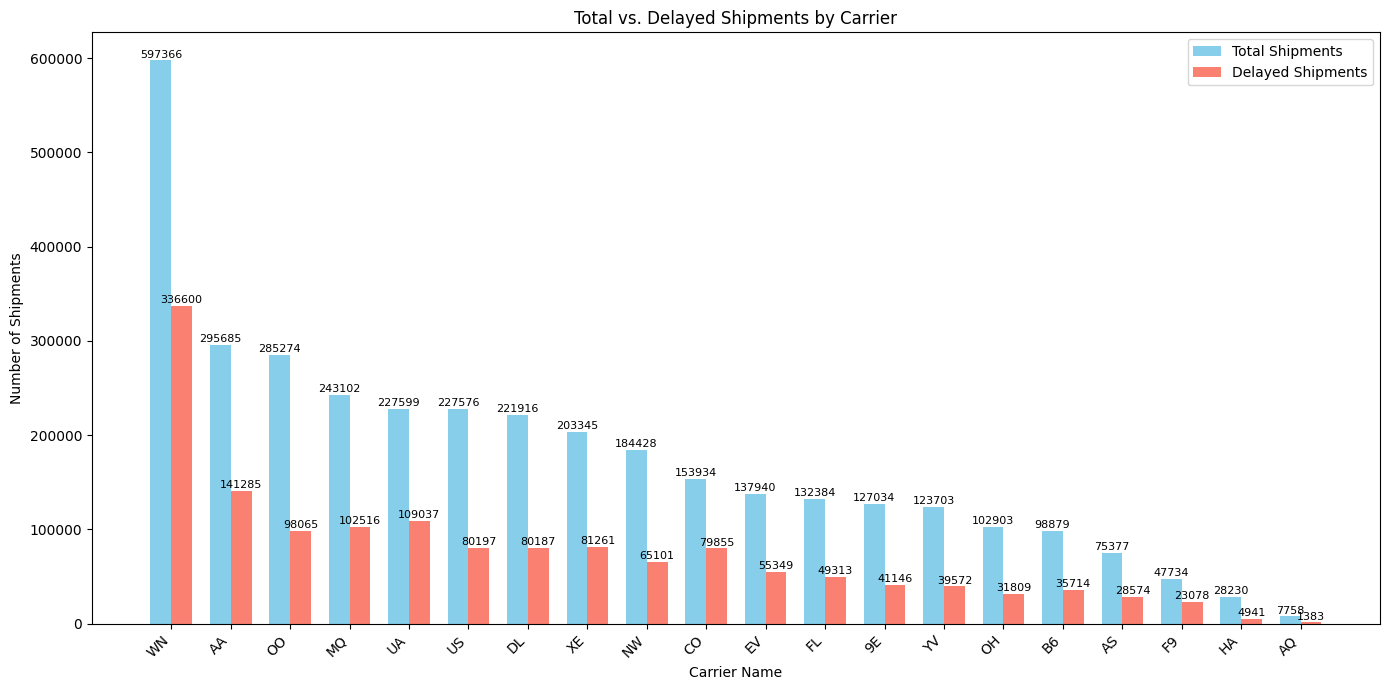
The delay frequency correlates with shipment volume, but the delay percentage still varies widely, indicating that certain destinations . This is the plot for the destination with maximum % of delays , we can see that in most of the places the number of orders are also very less , it indicates that those regions might have less connectivity or longer distance or not very smooth operations of fedex.

* **Number of Delayed Shipments per Day of the Week**

****

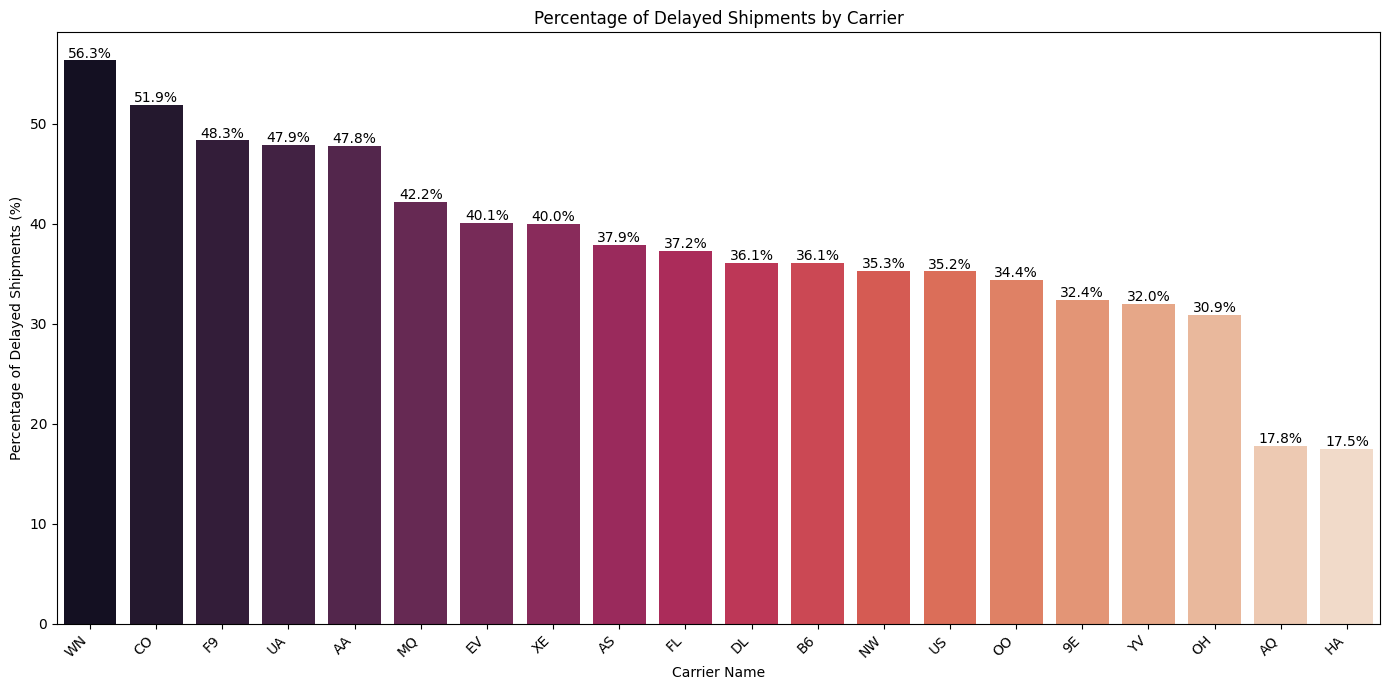
This bar plot shows the number of delayed shipments per day of the week. Delays are most frequent on Fridays and least frequent on Saturdays, indicating a weekly operational cycle where shipment delays peak toward the weekend.

* **Total vs. Delayed Shipments by Carrier**

****

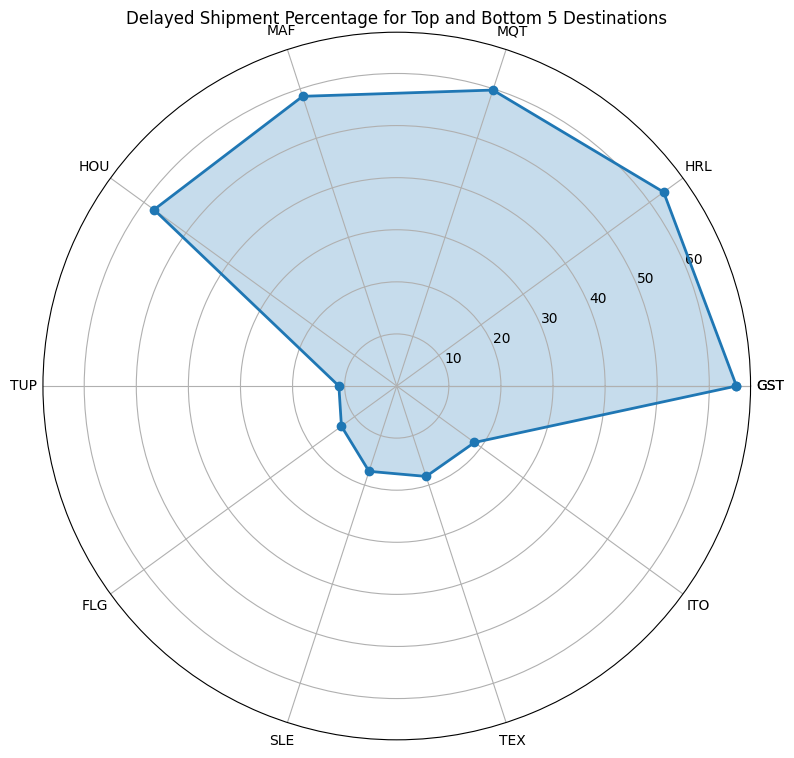
The above plot highlights significant differences in total and delayed shipments among various carriers. The number of delayed shipments generally correlates with total shipment volume, meaning that larger carriers experience more absolute delays due to higher activity. This helps package business to understand that from all available carriers which carriers have the most number of delayed packages, they can find the root cause of delays and fix it.

* **Percentage of Delayed Shipments by Carrier**

****

This plot illustrates the proportion of delayed shipments relative to total shipments for each carrier. Large carriers tend to have more delays both in number and percentage, likely due to heavy workloads. Smaller carriers generally perform better in keeping shipments on time. WN is carrying most of our orders , it has most % of the delays. This explains problems in delays. We can optimize this carrier or shift a few orders through other carriers to reduce load and no delays from WN carriers .

* **Delayed Shipment Percentage for Top and Bottom 5 Destinations**

****

The radar chart visualizes the **Delayed Shipment Percentage** for the ten destinations identified as the Top 5 (highest delay) and Bottom 5 (lowest delay) performing destinations.

**Destinations:** These are the labels (MAF, MQT, HRL, GST, ITO, TEX, SLE, FLG, TUP, HOU) set as the angular axis tick labels. They form the spokes of the wheel.

**Percentage:** These are the values plotted radially outward from the center. They are mapped to the concentric circles labeled 10, 20, 30, 60 The distance from the center represents the magnitude of the delayed shipment percentage.

### **High Delay Destinations (Top 5)**

These destinations form the large, outward-stretching part of the blue polygon, indicating a high proportion of delayed shipments.

* **GST:** Has the highest delayed percentage, extending past the 60% mark.
* **HRL, MQT, and MAF:** All show consistently high percentages, falling between 55% and 60%.
* **HOU:** Also shows a high delay percentage, close to the 50% mark.

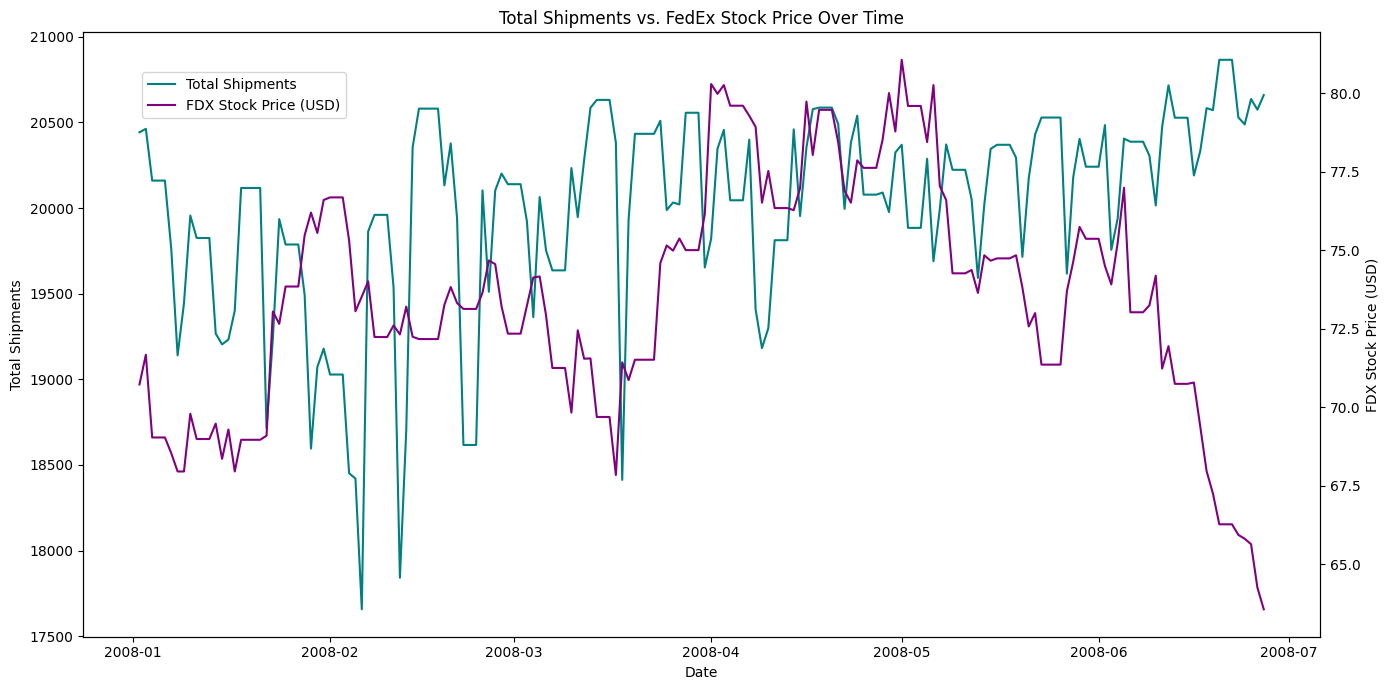
### **Low Delay Destinations (Bottom 5)**

These destinations form the inward-denting part of the blue polygon, indicating a low proportion of delayed shipments.

* **TEX, SLE, and FLG:** These destinations have the lowest delay percentages, with points closest to the center, likely in the 10% to15% range.
* **TUP and ITO:** Show slightly higher percentages than the lowest three, likely in the 20% to 30% range.

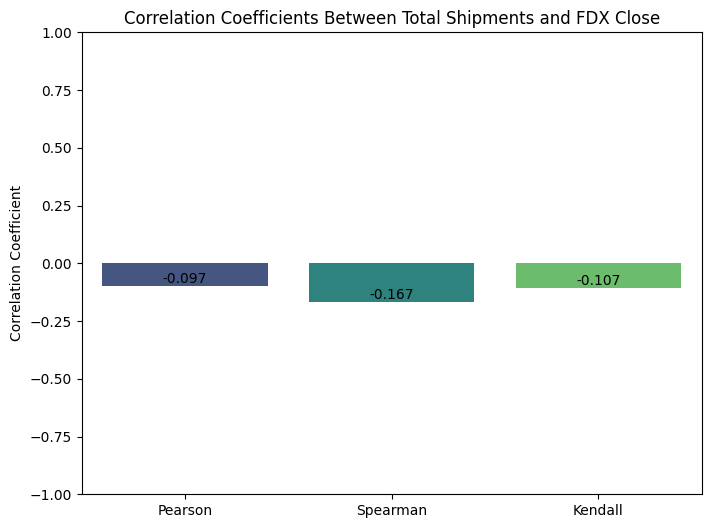
**4. STOCK DATA ANALYSIS**

* **Total Shipments vs. FedEx Stock Price Over Time**

****

The above plot shows that there is no clear or consistent pattern between shipments and the stock price. While shipments tend to stay within a certain range, the stock price first rises until around May and then declines sharply in June – July. Hence, the stock price movements were not strongly driven by the shipment volumes during this period, there could be other market or economic factors which likely had a larger influence.The possible reasons are because everyday the no of orders is not displayed or open to public, so there is no possibility of Fedex to mimic stock price or correlate apart from times other than festive season(because everyone can understands the reason of increased no of orders- gifts and others) or any bulk industry others. Sharp decline in stock price at the end due to the stock market crash in 2008.

* **Correlation Coefficients Between Total Shipments and FDX Close**

****

**Interpretation of Correlation Coefficients**

The bar chart shows the Pearson, Spearman, and Kendall correlation coefficients between the total number of shipments and the FedEx closing stock price.

\*   **Pearson correlation**: Measures the linear relationship between two variables. A value close to 0 suggests a weak linear relationship. The Pearson correlation coefficient indicates a very weak negative linear relationship between total shipments and FDX close. This means that as total shipments slightly increase, the FDX closing price tends to slightly decrease, but this relationship is not strong or consistent in a linear fashion.

\*  **Spearman correlation**: Measures the monotonic(having the property either of never increasing or of never decreasing as the values of the independent variable or the subscripts of the terms increase) relationship between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. A value close to 0 suggests a weak monotonic relationship. The Spearman correlation coefficient indicates a weak negative monotonic relationship. This means that there is a slight tendency for the FDX closing price to decrease as the total shipments increase, even if the relationship is not strictly linear.

\*  **Kendall correlation**: Also measures the monotonic relationship between two variables, based on the number of concordant and discordant pairs. A value close to 0 suggests a weak monotonic relationship. The Kendall correlation coefficient also indicates a weak negative monotonic relationship, similar to the Spearman correlation, confirming the slight tendency for FDX to decrease as shipments increase.

All three correlation coefficients are close to zero and negative, suggesting a very weak to weak negative relationship between the total number of shipments and the FedEx closing stock price during this period. It's important to note that correlation does not imply causation. Other factors likely have a stronger influence on the stock price.

**RESULTS AND KEY FINDINGS**

The analysis of FedEx’s supply chain and stock market data provided several meaningful insights into delivery performance, shipment trends, and their possible links with financial performance.

**1. Consistent Shipment Volume Across Time**

The total number of shipments remained steady across months, showing only slight fluctuations. March recorded the highest shipment volume, while February saw a dip, likely due to its shorter duration. This consistency indicates stable operational capacity.

**2. Weekly Demand Patterns**

Shipments were highest on weekdays, with Thursday and Friday being the busiest. Saturday and Sunday had fewer deliveries, reflecting reduced operational activity during weekends.

**3. Monthly Demand Trends**

Each month showed similar shipment levels. There were no large seasonal spikes. This suggests that FedEx manages a consistent workload through the year.

**4. Day of Month Effects**

Most shipments happened around the middle of the month. Fewer shipments went out at the end of each month. This may be linked to internal scheduling or billing cycles.

**5. Geographical and Delivery Insights**

Average delivery times varied between destinations, but most locations maintained predictable and timely deliveries. Longer-distance routes showed slightly higher delays, as expected in large-scale logistics networks.

**6. Shipment Delay Distribution**

Most of the shipments were delivered close to their planned times and the delays occurred mainly in longer distances.

**7. Operational Efficiency and Distance Relationship**

Delivery times varied slightly by location. Urban areas had faster deliveries, while distant regions took a bit longer. The variation stayed within expected limits.

**8. Link between Operations and Stock Performance**

There was a weak short-term correlation between FedEx’s shipment volume and its stock price. The operational consistency helps maintain stability, but stock performance depends on multiple external and financial market factors beyond logistics efficiency.

**CONCLUSION**

We analysed the Fedex dataset From Year 2008, we used this dataset because a limited number of supply chain dataset is available along with company names and stocks details. Our dataset contains 36 lakhs+ rows, and analyzes the supply chain of Fedex. First based on our dataset Fedex operations are limited to the USA, Canada and neighboring areas. There is scope of expansion in other continents and huge business oppurtunities. Diving deeper into data Even Many parts of USA is also still not using Fedex. We did geographical analysis of our data.There are a total of 297 source destinations and 297 destination stations on which Fedex operates. These are the top 5 source destinations from which most orders are coming (ATL-207003,ORD-179274,DFW-141948,DEN-120500,LAX-111703) and top 5 destination stations are (ATL-206935,ORD-179310,DFW-142013,DEN-120517,LAX-111749).

These are destinations with top 5 best(lowest) average delivery time, Negative indicates that orders are reached at destination before estimated time.

|  |  |
| --- | --- |
| **Destination** | **ActualDeliveryTime** |
| **PSE** | -235.850000 |
| **DHN** | 315.063676 |
| **BQN** | 443.903394 |
| **IPL** | 602.445131 |

These are destinations with 5 worst average delivery times. This miscalculated using how much order is delayed from estimated delivery time. (BLI-1400.10, ALO-1395.08, INL- 1383.16, CMX-1334.55, SCE- 1321.81). This will help Fedex to improve operations and delivery efficiency in these regions.

In Timeframe analysis we analyse each granular level of data while monthly data does not have any significant difference but when we move to weekly data we find out that most of the orders are being placed during weekends in datewise also the last two days of every month have least number of orders this shows a customer spending pattern specially from salaried class people spending less on weekends.

We also did Carrier analysis , for all the carriers which are carrying orders to analyze which carrier is most efficiently ordered within estimated time, WN carrier is handling most of the orders and also have the most number of delayed package this shows that a lot of stress is on one particular carrier, we need to divert few orders through other carriers and improve the efficiency of carriers after finding root causes of delivery delays.

In the end we did the correlation analysis of FedEx stock data with their daily no of orders, we applied three different correlation techniques and all correlations give results in negative this shows there is no direct correlation between both parameters. The possible reasons are because everyday the no of orders is not displayed or open to the public, so there is no possibility to investors to invest on basis of no of orders on daily basis to mimic stock price or correlate apart from times other than festive season(because everyone can understands the reason for the increased no of orders- gifts and others) or any bulk industry deliveries or companies quarterly result reason. Sharp decline in stock price at the end due to the stock market crash in 2008.

Altogether this project shows the areas of improvement for Fedex in operations, efficiency and also possibility of expansion to other countries. This project is a detailed operational analysis of Fedex by keeping together all Geographical analysis, TimeSeries analysis, Carrier Analysis, Delivery analysis and Correlation analysis showing the business gaps and areas to work on.