Unlocking Efficiency: Analyzing Logistics Delay Times

Deepak Jacob Mathew

Table Of Insights:

- Project requirement
- Dataset variables and explanation
- Steps involved
- Pictorial explanation
- Key observations
- Remarks
- Thank note

Project Requirement

- The dataset comprises various variables that capture information regarding the day-to-day logistics activities and associated time delays.
- The dataset includes comprehensive data such as dates, locations, transporter and freight forwarder information, weight, rates, invoice status, as well as the time periods of delays across multiple locations.
- Our objective is to identify the diverse variables contributing to delays and draw conclusions that can help in resolving them.

Dataset variables and explanation

- Entry ID: Identifier for the entry (e.g., "DL-1").
- Freight forwarder: The freight forwarder involved in the logistics (e.g., "Freight forwarder ABC").
- Transporter: The transporter involved in the logistics (e.g., "Transporter A").
- Depart From: The departure location (e.g., "Mogadishu").
- Arrive To: The arrival location (e.g., "Victoria").
- Loading Date: The date when the freight was loaded.
- Offloading Date: The date when the freight was offloaded.
- Client Invoice status: The invoice status with the client (e.g., "Invoiced", "Paid").
- Transporter Invoice status: The invoice status with the transporter (e.g., "Invoiced").
- Service / Product: The service or product involved in the logistics (e.g., "TRANSPORTATION").

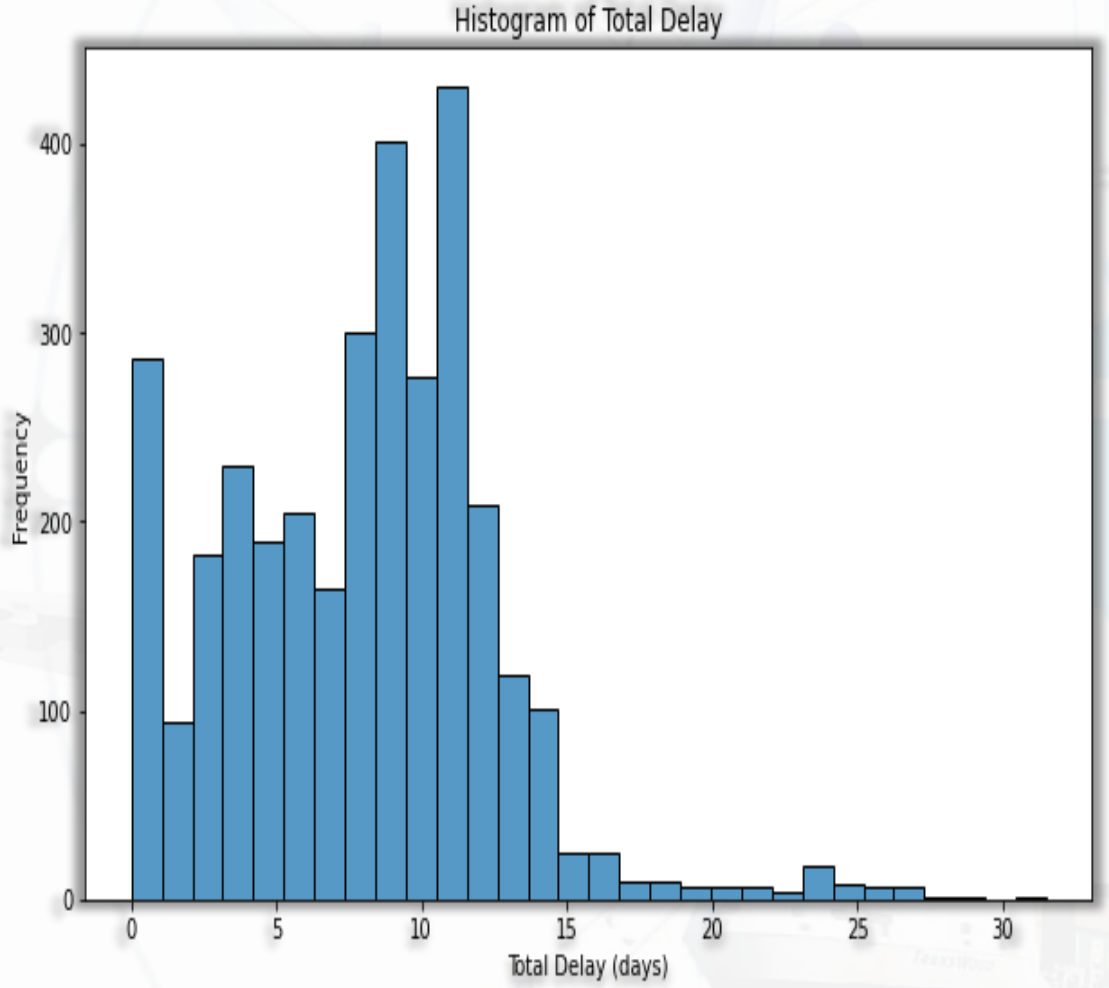
Dataset variables and explanation

- Gross Weight(t): The gross weight of the freight in ton.
- Client Rate per t: The client rate per ton.
- Client Currency: The currency of the client's payment (e.g., "USD").
- Supplier Rate per t: The supplier rate per ton.
- Supplier Currency: The currency of the supplier's payment (e.g., "USD").
- POD Attached: Whether a Proof of Delivery (POD) is attached.
- Delay at toll gate in days: The delay at the toll gate in days.
- Delay at workshop in days: The delay at the workshop in days.
- Delay at border in days: The delay at the border in days.
- Transit time: Time from the origin to the destination

Steps involved

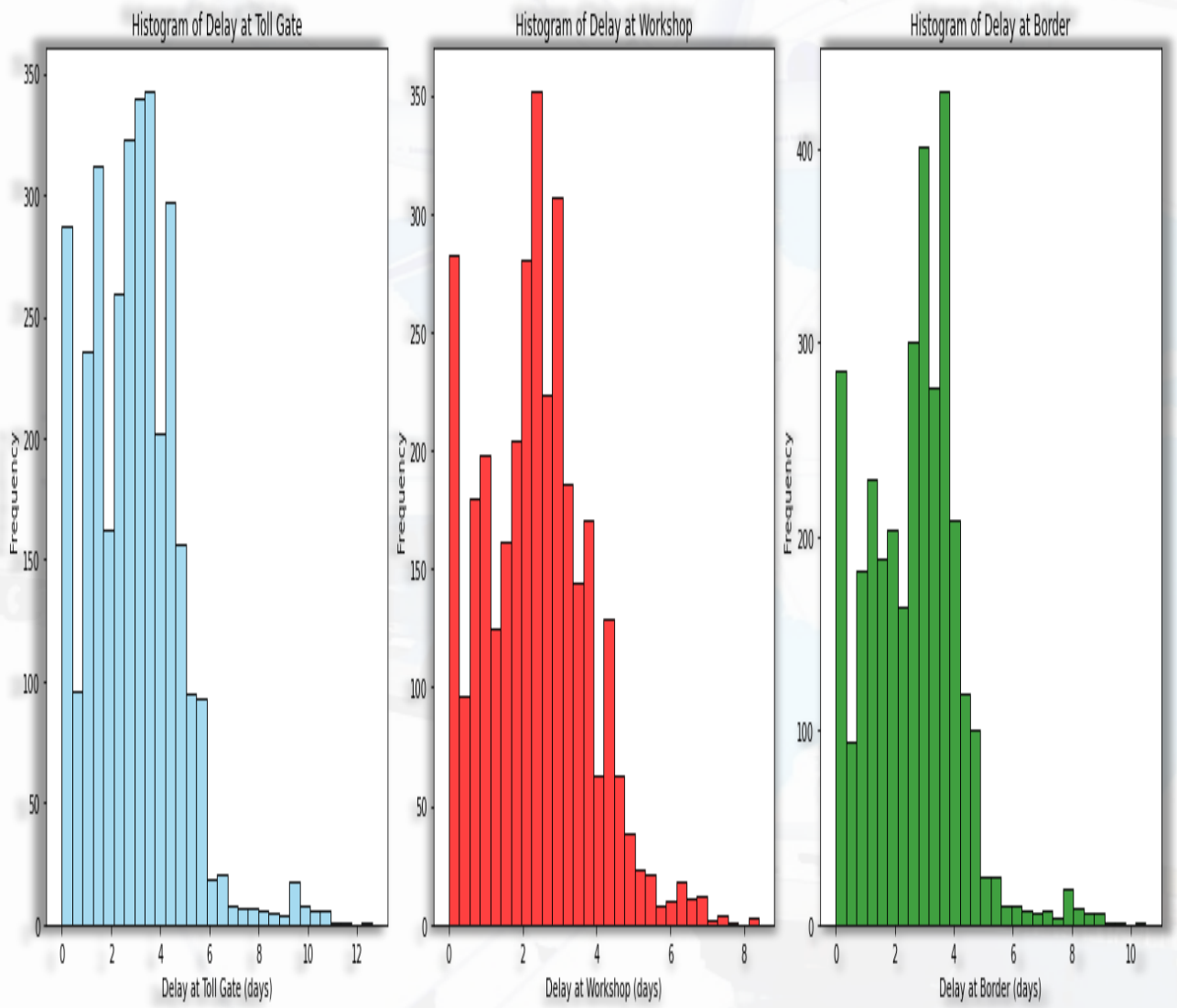
- The dataset and its variables were initially verified and the meaning of each variable was established using Python in Jupyter Notebook.
- The required libraries for data exploration and plotting were imported.
- The unique values of each variable were examined to determine which variables would be considered for the analysis. Variables that contained only a single value in all cells were identified and can be excluded from further analysis.
- A variety of visualizations and analyses were performed, and for each insight, an explanation was provided based on statistical analytics.
- Lastly, an automated dashboard was created in Power BI to provide a comprehensive summary of the dataset.

Histogram of total delay in days



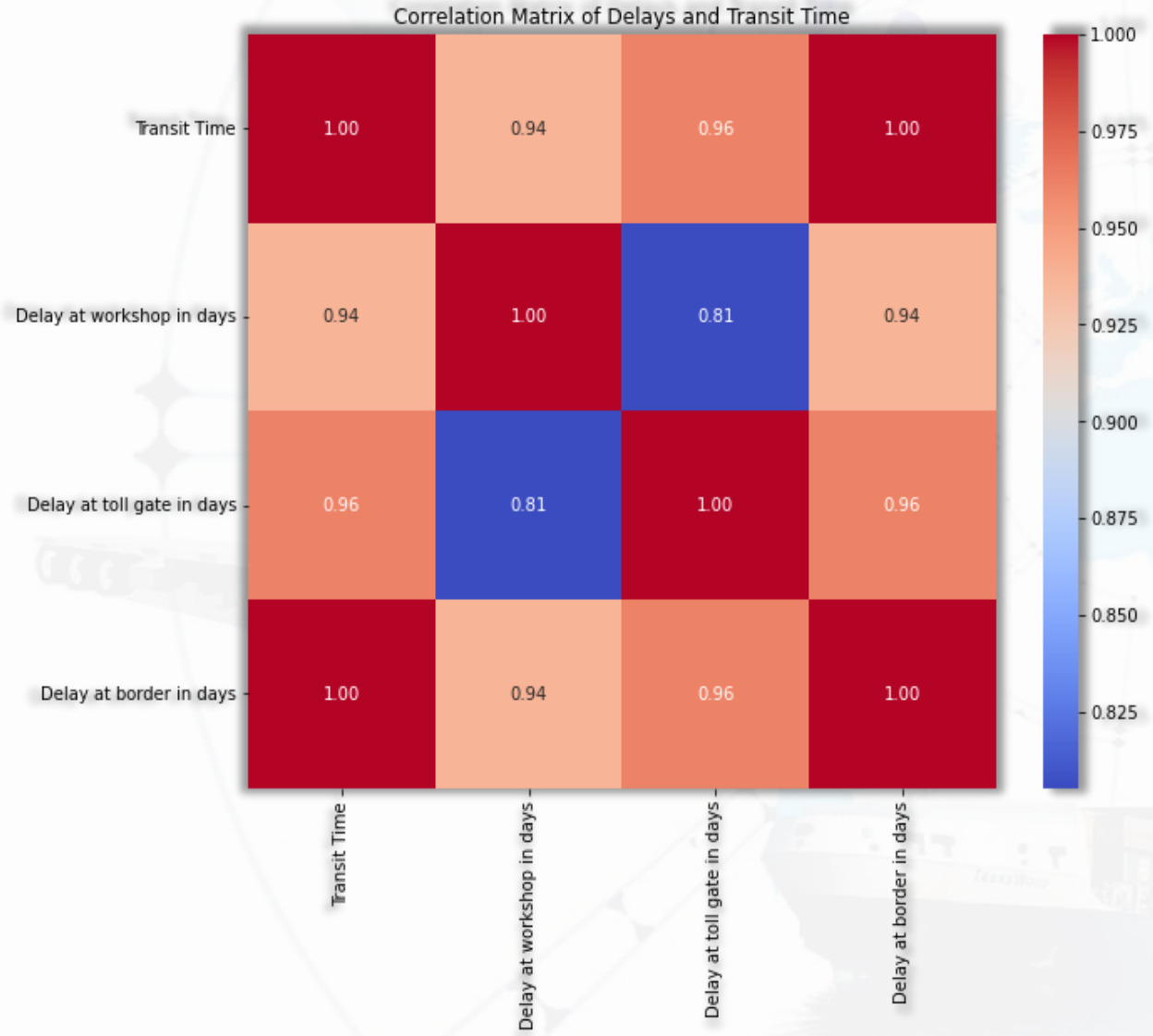
- Total delay is a new variable created from the addition of total delays in days at workshop, tollgate and border.
- A histogram was generated by dividing the data into 30 bins and plotting the frequency of total delays in each bin. This allowed for a clear visualization of the distribution of delays and provided insights into the concentration of delays within specific ranges.
- The histogram of total delay shows that the majority of shipments experience a delay between 4.5 to 10.5 days.
- There are very few shipments with extremely high delay times, indicating that such cases might be outliers or exceptional situations.

Histogram of total delay at each location



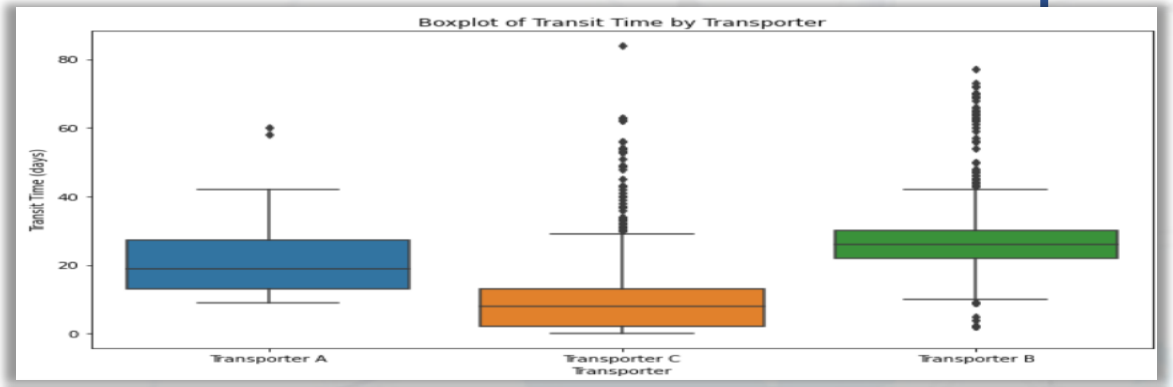
- A histogram was created to visualize the distribution of delays at different locations, including the workshop, tollgate, and border. This allowed for a better understanding of the frequency and magnitude of delays at each specific location.
- Also, statistical analysis to find the mean, median, standard deviation, Q1, Q3 etc.
- Delays follow a somewhat similar distribution, with the majority of delays falling within a certain range and fewer instances of extreme delays. This suggests that while delays do occur, they are generally kept within a manageable timeframe.
- The histogram of delays at the toll gate, workshop, and border show that these three factors contribute to the overall delay, and their distributions are quite similar. The majority of delays are concentrated around 2-4 days for each factor.

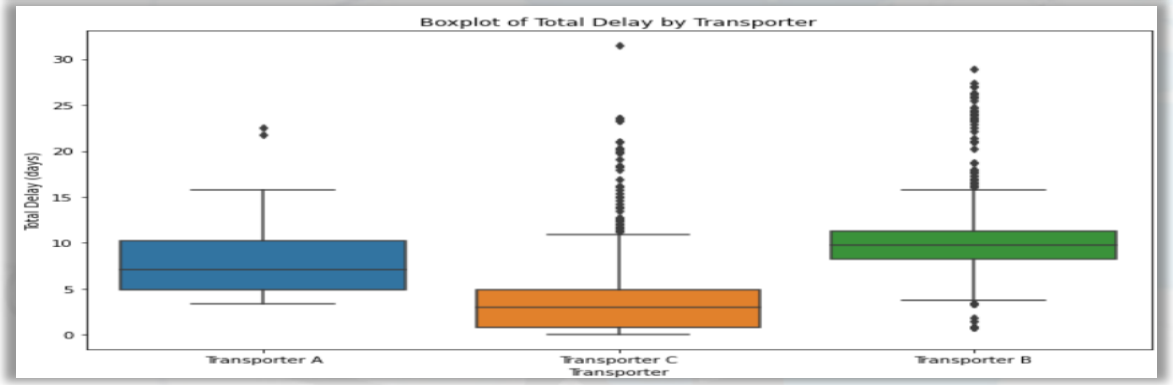
Correlation matrix



- A correlation matrix was computed to analyze the relationships between variables, including transit time (the difference between offloading and loading dates in days) and delays at locations such as the workshop, tollgate, and border. This matrix provided a comprehensive overview of the correlations between these variables, allowing for the identification of any significant associations or dependencies.
- The heatmap of correlations shows that the transit time is positively correlated with delays at the toll gate, workshop, and border. This means that as the delay in any of these areas increases, the overall transit time tends to increase as well.
- Slightly there is less correlation between the workshop delay and toll gate delay from the given dataset.
- There is also a strong correlation between the different types of delays, indicating that when there is a delay in one area (e.g., the workshop), there tends to be a delay in other areas as well (e.g., toll gate or border). This could suggest systemic issues in the logistics process that cause cascading delays.

Box and Whisker: Transporter vs Delay and Transit time

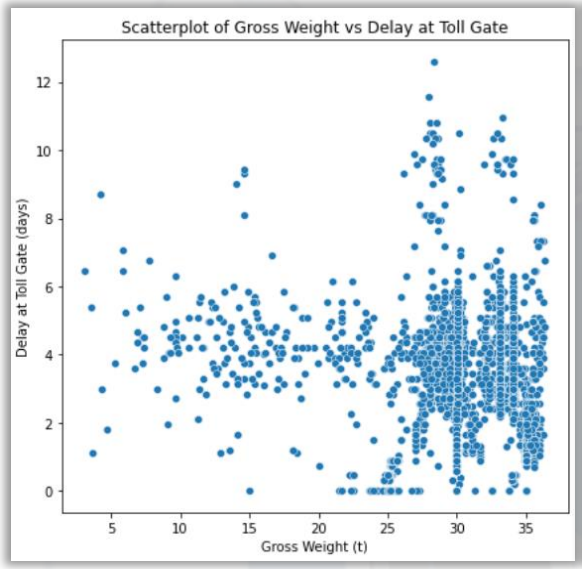


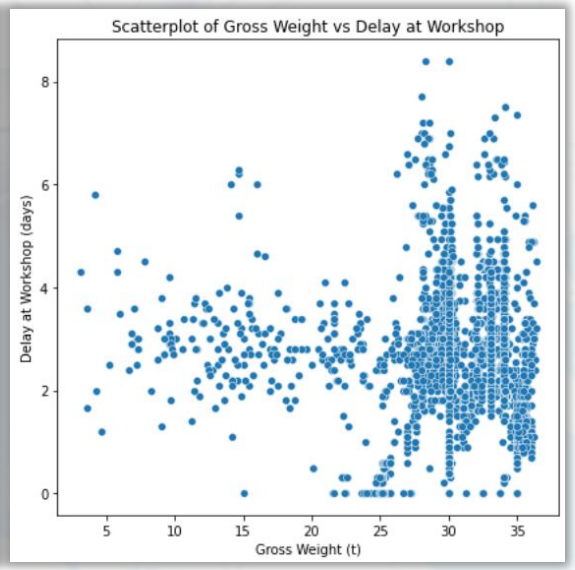


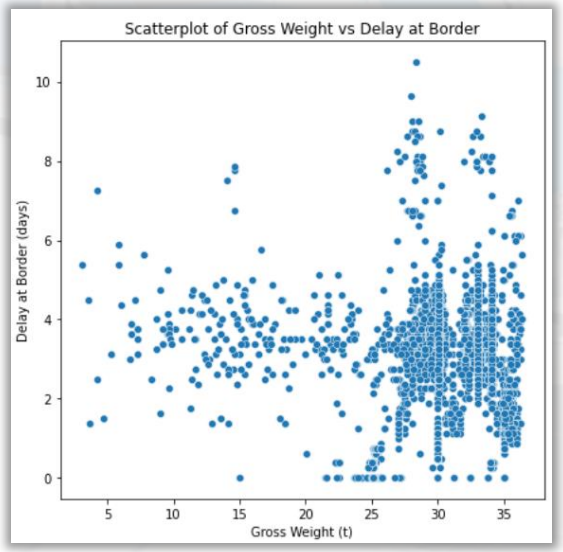
Transporter	count	mean	std	min	25%	50%	75%	max
Transporter A	128.0	7.643555	3.398742	3.375	4.875	7.125	10.21875	22.500
Transporter B	2113.0	9.987636	3.357281	0.750	8.250	9.750	11.25000	28.875
Transporter C	1078.0	3.757305	4.087599	0.000	0.750	3.000	4.87500	31.500

- A boxplot was generated to visualize the relationship between the Transporter and Transit time and Transporter and Total delays in days.
- Total delay and transit time varies between different transporters. Some transporters experience longer delays, while others have shorter delays. This could be due to various factors, such as the efficiency of the transporter, the routes they use, their load management practices, and so on.
- The variation in transit times within each transporter (as indicated by the height of the boxes and the length of the whiskers) also suggests that there are factors other than the transporter that affect transit time.
- Transporter C has less delay time compared to other 2.
- Standard deviation of all 3 transporters seems to be closer with each other, but the median of transporter B is higher than others.

Gross weight and Delay at workshop, tollgate and border:

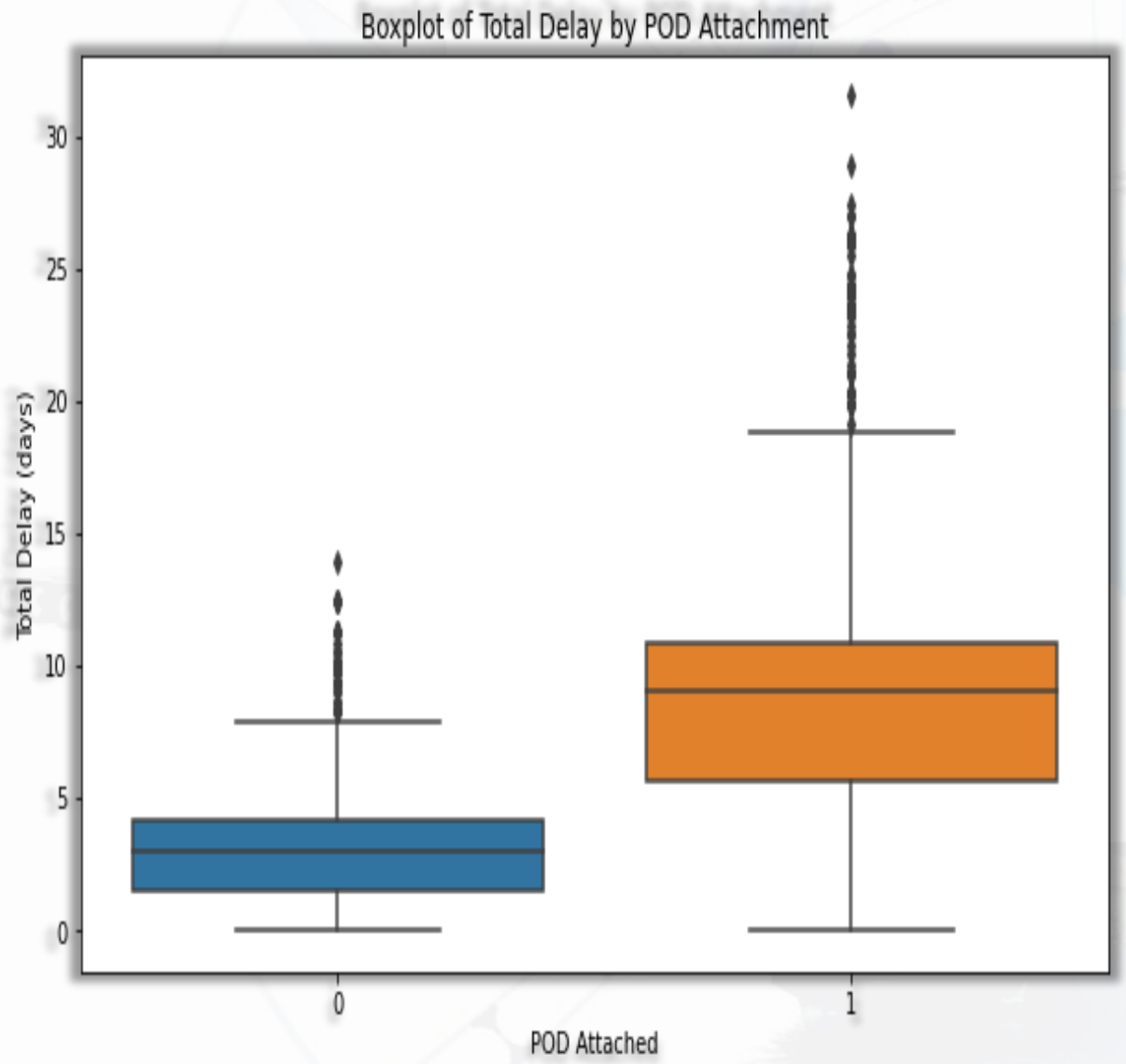






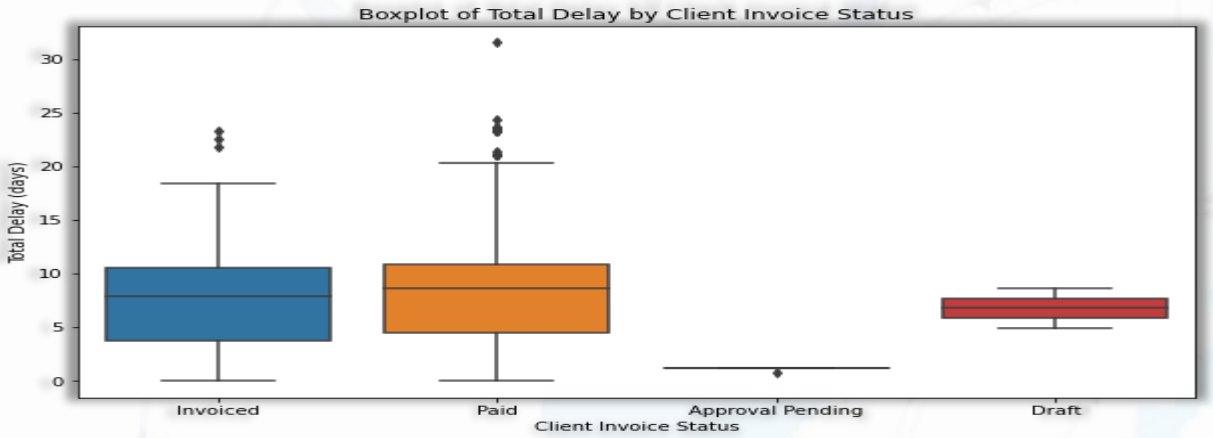
- A scatter plotting was generated to visualize the relationship between the gross weight and total delays in days at workshop, tollgate and border.
- The scatterplots show that there is no clear relationship between the gross weight of the freight and the delays at the toll gate, workshop, or border. **This suggests that the weight of the freight does not directly influence these types of delays.**
- Despite the lack of a clear trend, there is a spread of delay times across different weights, indicating variability in the logistics process. This lack of consistency could be due to other factors, such as the specific route taken, the efficiency of the transporter, or external factors like weather or traffic conditions.

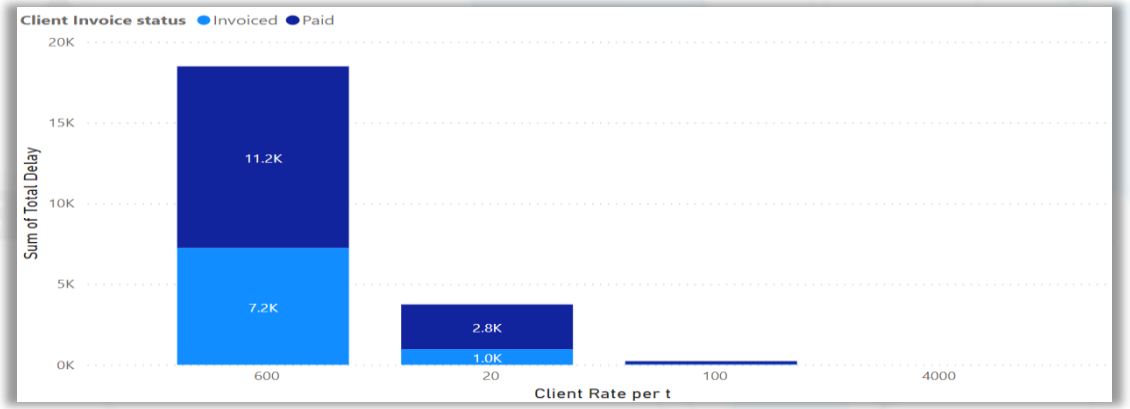
Box and Whisker: POD attached and total delay:



- A boxplot and statistical analysis were generated to visualize the relationship between Proof Of Document attached and total delays in days.
- In the context of the dataset, the value 0 is designated to represent shipments with Proof of Delivery (POD), while the value 1 is assigned to represent shipments without Proof of Delivery (POD).
- The boxplot shows that there is a significant difference in total delay between shipments with a Proof of Delivery (POD) attached and those without. Shipments without a POD tend to have longer total delays. **This could suggest that the lack of a POD may lead to inefficiencies or complications in the logistics process that result in delays.**
- The range of total delays is also larger for shipments without a POD, indicating a higher level of variability in these cases.

Box and Whisker: Client invoice status and total delay:

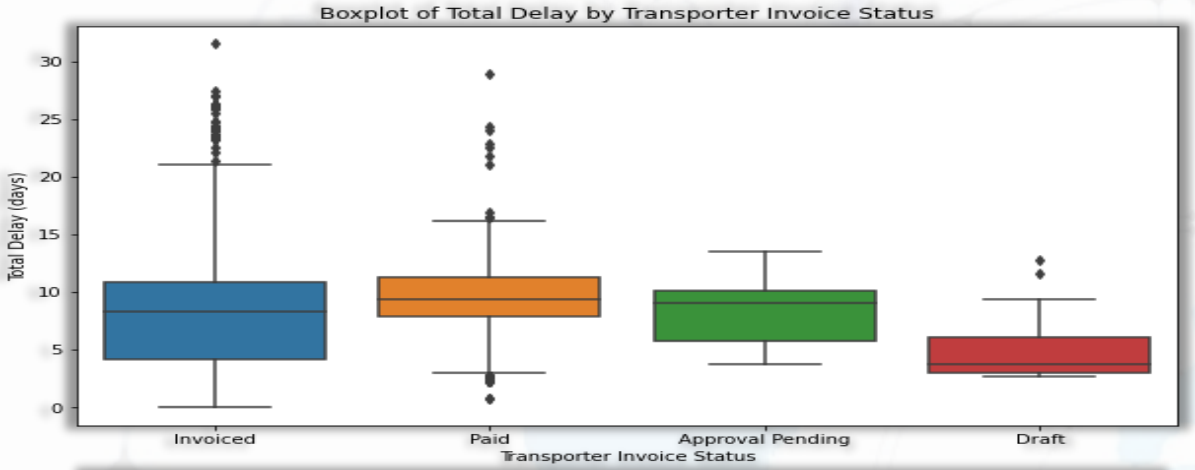


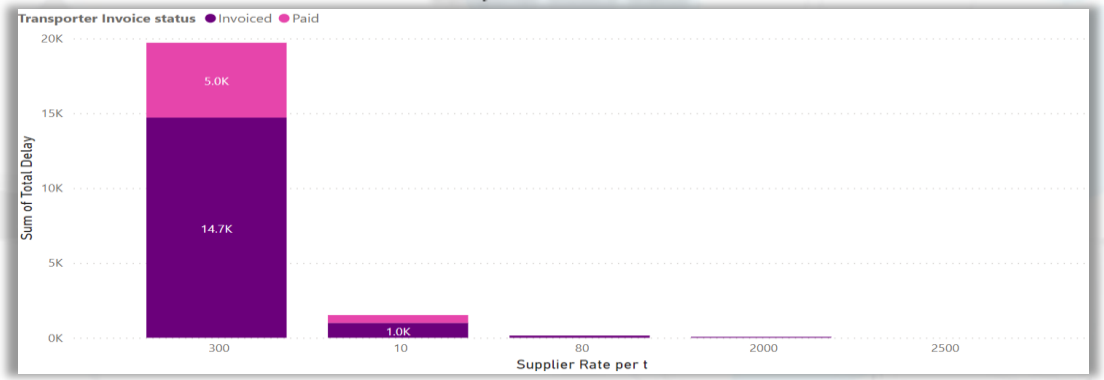


Client Invoice status	count	mean	std	min	25%	50%	75%	max
Approval Pending	5.0	1.050000	0.167705	0.750	1.1250	1.125	1.1250	1.125
Draft	2.0	6.750000	2.651650	4.875	5.8125	6.750	7.6875	8.625
Invoiced	1147.0	7.198235	4.145124	0.000	3.7500	7.875	10.5000	23.250
Paid	1799.0	7.893552	4.578301	0.000	4.5000	8.625	10.8750	31.500

- A boxplot and statistical analysis were generated to visualize the relationship between the invoice status and total delays in days.
- The boxplot shows some variation in total delay depending on the client invoice status. Shipments with an invoice status of 'Invoiced' and 'Paid' seems to be almost same to each other in mean, median and standard deviation. It is unable to predict whether the status of 'Paid' makes it easy for the faster delivery.
- There is also a significant amount of variability in total delay within each invoice status category, as shown by the height of the boxes and the length of the whiskers. This indicates that while invoice status may influence delay times, there are also other factors at play.
- **Median of Paid is slightly higher than Invoiced, which means the Paid status may or may not influence the delay time.**

Box and Whisker: Transporter invoice status and total delay:

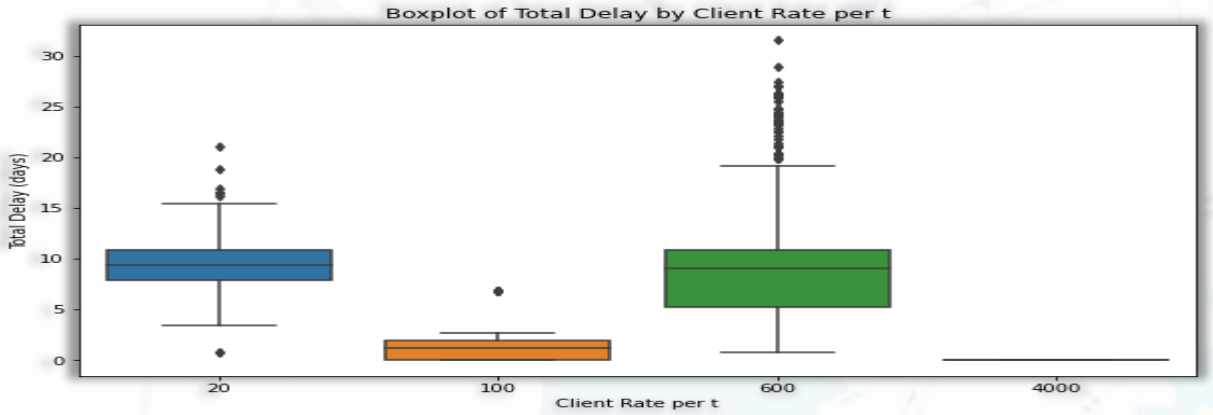


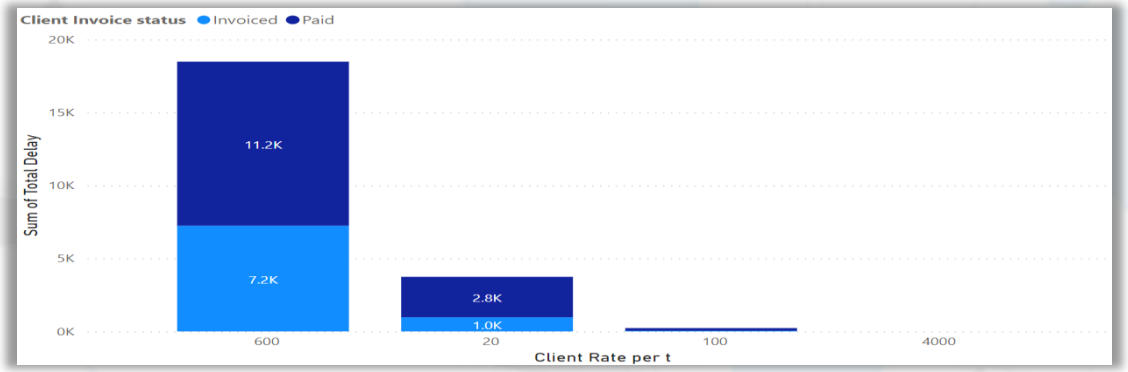


Transporter Invoice status	count	mean	std	min	25%	50%	75%	max
Approval Pending	102.0	8.297794	2.427797	3.750	5.71875	9.000	10.125	13.500
Draft	32.0	4.968750	2.710865	2.625	3.00000	3.750	6.000	12.750
Invoiced	2045.0	7.782396	5.029992	0.000	4.12500	8.250	10.875	31.500
Paid	601.0	9.195300	3.494147	0.750	7.87500	9.375	11.250	28.875

- A boxplot and statistical analysis were generated to visualize the relationship between the transporter invoice status and total delays in days.
- From the analysis, apart from 'Draft' all others seems to have closer median values.
- 'Paid' seems to be having outliers on both sides. The presence of outliers on both sides can impact the interpretation of the data. It indicates that there are values that are significantly different from the majority of the data points, and they can potentially affect the overall statistics and, conclusions drawn from the data.
- The low variability indicated by the close values of the mean, median, and standard deviation of all 3, allows for more accurate predictions and estimations based on these variables.
- **Although the median of Paid and Invoice are closer to each other, making the status Paid may or may not affect the delay.**

Box and Whisker: Client rate per ton and total delay:

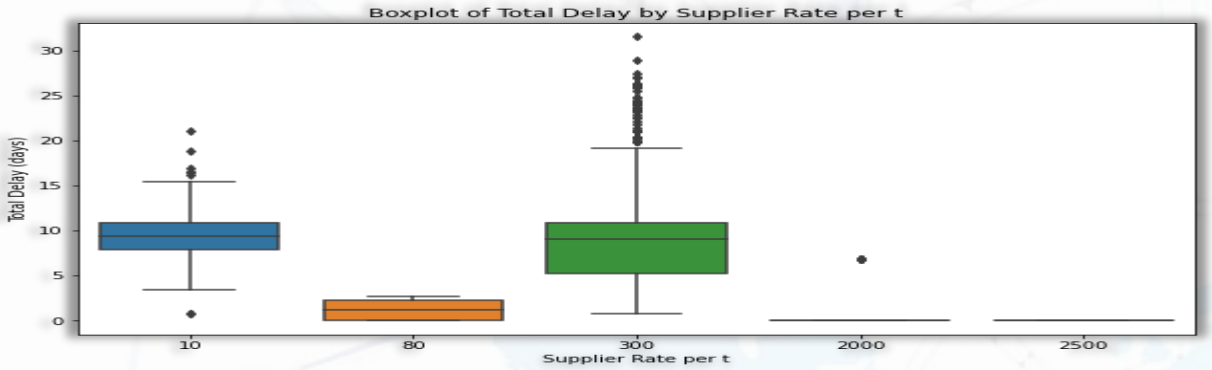


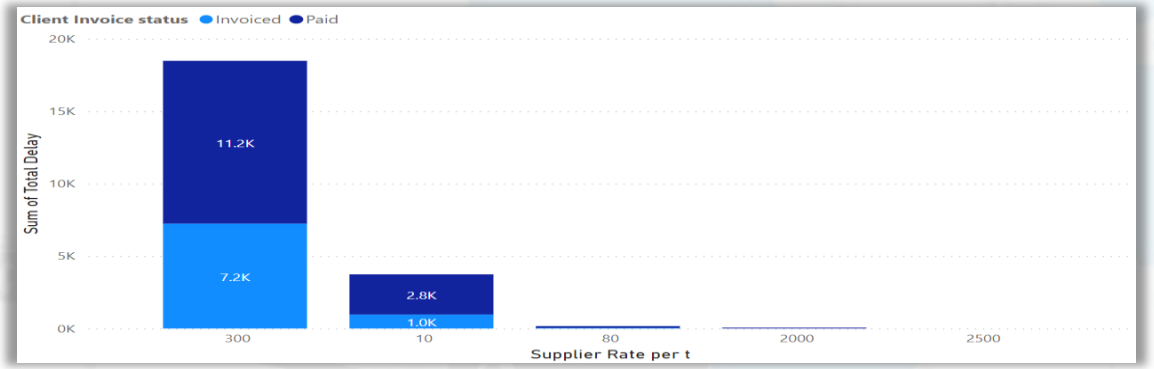


Client Rate per t	count	mean	std	min	25%	50%	75%	max
20	473.0	9.241808	2.566250	0.75	7.875	9.375	10.875	21.00
100	196.0	1.161352	1.530262	0.00	0.000	1.125	1.875	6.75
600	2468.0	8.725132	4.253289	0.75	5.250	9.000	10.875	31.50
4000	182.0	0.000000	0.000000	0.00	0.000	0.000	0.000	0.00

- A boxplot and statistical analysis were generated to visualize the relationship between the client rate per ton and total delays in days.
- The client rate per ton with '600 USD' seems to have higher median compared to others. Hence reducing the price could be one of the ways to resolve this issue. In the case of 600 USD, we have more outliers but the data is symmetric.
- With a smaller standard deviation, the values of 20 USD are more tightly clustered around the mean, indicating a higher level of consistency and predictability. On the other hand, the larger standard deviation of 600 USD suggests a wider range of values and greater variability, making it more difficult to accurately predict future values.
- Therefore, based on the information provided, the **predictability of 20 USD is expected to be better than that of 600 USD due to its lower standard deviation and tighter distribution of values.**

Box and Whisker: Supplier rate per ton and total delay:

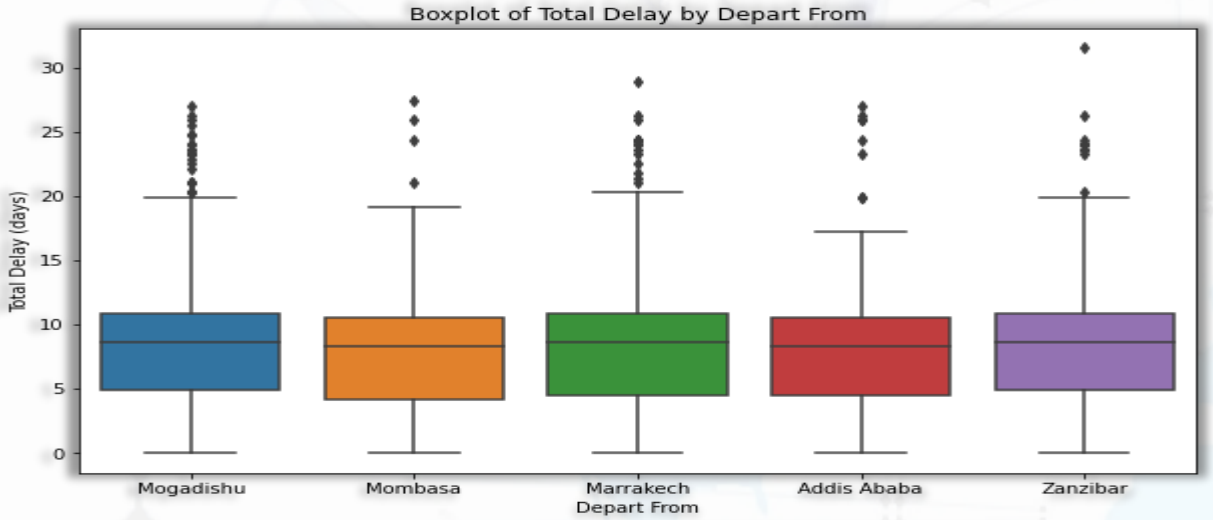


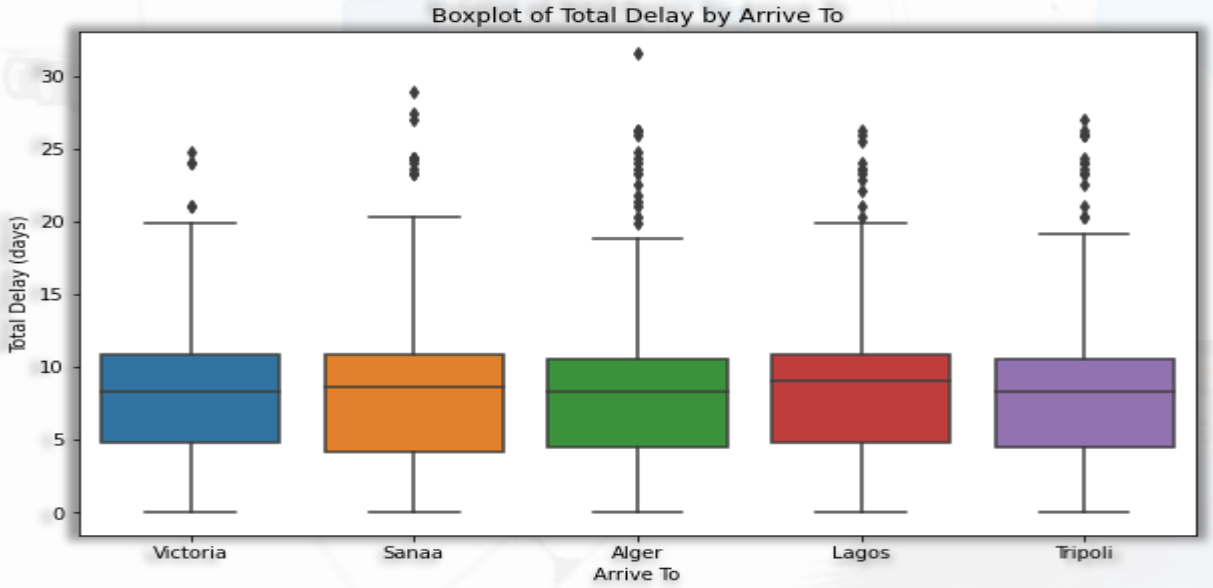


Supplier Rate per t	count	mean	std	min	25%	50%	75%	max
10	473.0	9.241808	2.566250	0.75	7.875	9.375	10.875	21.000
80	148.0	1.127534	0.913857	0.00	0.000	1.125	2.250	2.625
300	2468.0	8.725132	4.253289	0.75	5.250	9.000	10.875	31.500
2000	48.0	1.265625	2.662489	0.00	0.000	0.000	0.000	6.750
2500	182.0	0.000000	0.000000	0.00	0.000	0.000	0.000	0.000

- A boxplot and statistical analysis were generated to visualize the relationship between the Supplier rate per ton and total delays in days.
- Rates of 10 USD and 300 USD exhibit higher delay times, while the rate of 80 USD shows significantly lower delay times. Identifying the specific reasons behind these variations may be challenging due to the lack of clear patterns. However, it is worth noting that the standard deviation of 300 USD is higher compared to 10 USD, indicating a wider spread of delay times around the mean for 300 USD.
- Based on this observation, it can be concluded that the delay times for 10 USD are more predictable than those for 300 USD. **Therefore, the data suggests that delay times associated with the rate of 10 USD are more reliable and predictable compared to those associated with the rate of 300 USD.**

Box and Whisker: Depart from & Arrive to and total delay:





- A boxplot and statistical analysis were generated to visualize the relationship between the Depart from and Total delays in days, and Arrive To and Total delays in days.
- **Both variables 'Arrive To' and 'Transport From' seems to have same statistics, hence shall be avoided from the analytics.**

Key Observations

- Total delays range from 4.5 to 10.5 days.
- Increased delays at workshop, tollgate, and border locations contribute to longer transit times.
- Transporters A and C experience higher delays compared to other transporters.
- The weight of the shipment does not have a direct impact on delays.
- Delays tend to increase in cases where Proof of Delivery (POD) is not provided.
- The paid status of invoices, both for the client and transporter, does not significantly affect the delay according to the dataset.
- Higher delays are observed for client rate 600 USD and supplier rate 300 USD. Reducing these rates may potentially decrease the delay times
- .Surprisingly, the paid status of client invoices does not show lower delays in client rates.
- However, the paid status of transporter invoices has a positive effect in reducing the delay time.

Remarks

- Reasons for Outliers: Outliers can occur due to various factors, such as different routes taken, unfavorable weather conditions, breakdowns, or inconsistencies in the data.
- These outliers can significantly influence the mean of the data, causing it to be either lower or higher than the median. As a result, the median is often used as a more robust measure of central tendency for observations.
- Skewness: In our dataset, we observe a presence of outliers beyond the maximum value. However, upon conducting statistical analysis, we find that the data exhibits a certain level of symmetry. Although there may be a slight skew to the left or right at times, this skewness is minimal and does not have a significant impact on the central tendency of the data.
- In summary, while outliers are present in the dataset, they do not heavily impact the central tendency, and the data generally demonstrates a symmetric distribution.

Deepak Jacob Mathew