Optimizing Public Transit Operations By Metro Move

An Exploratory Data Analysis of public Transportation

Optimizing Public Transit Operations By Metro Move

An Exploratory Data Analysis of public Transportation

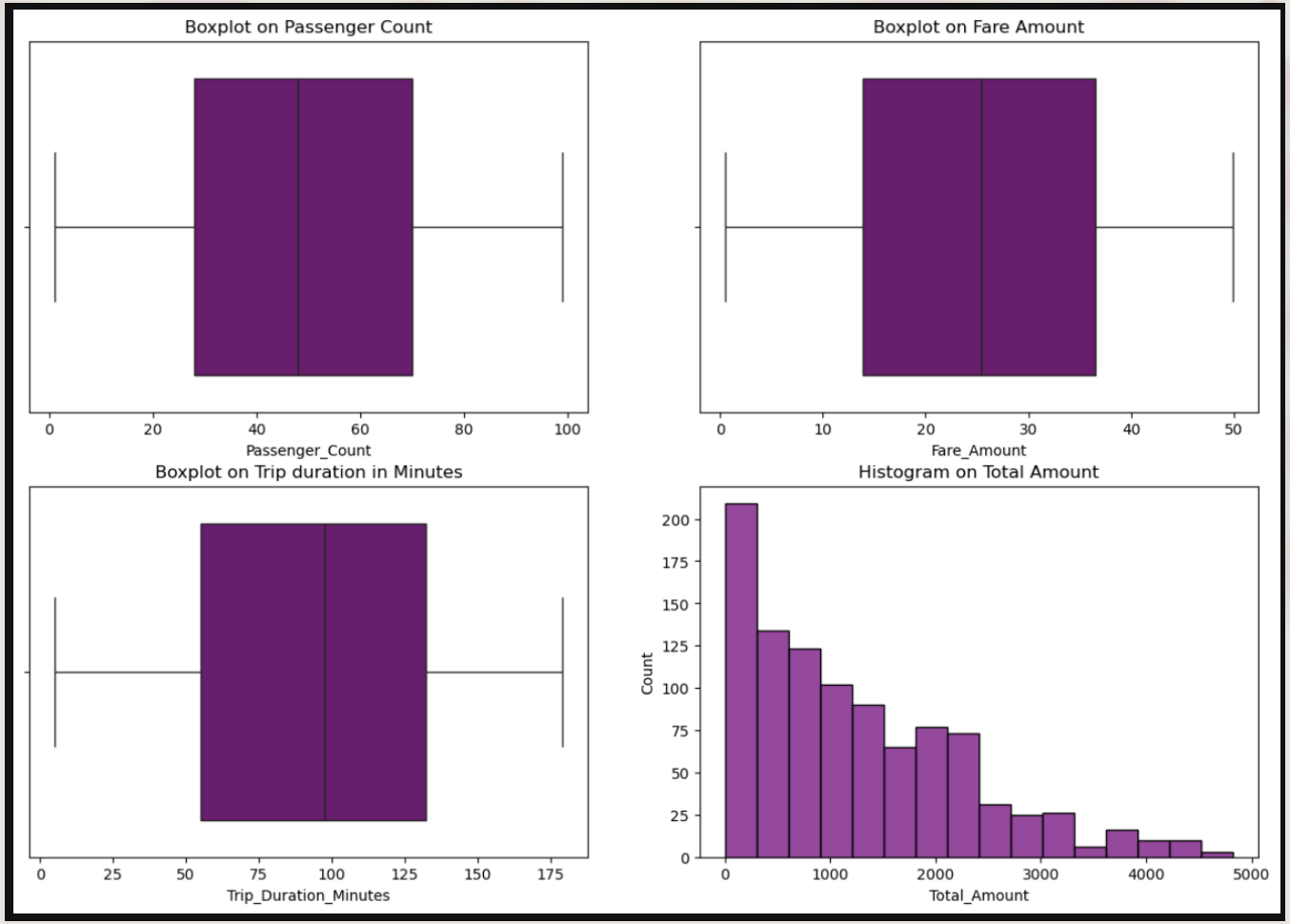
Evaluating The PC, WA, TD, TA Numerical Values through Univariate Analysis

The concentration/number of passengers count that usually commute through any of the mode of transport ranges between 28 and 70.

Majority of the passengers pay an amount between 12 and 38$ to get to their location.

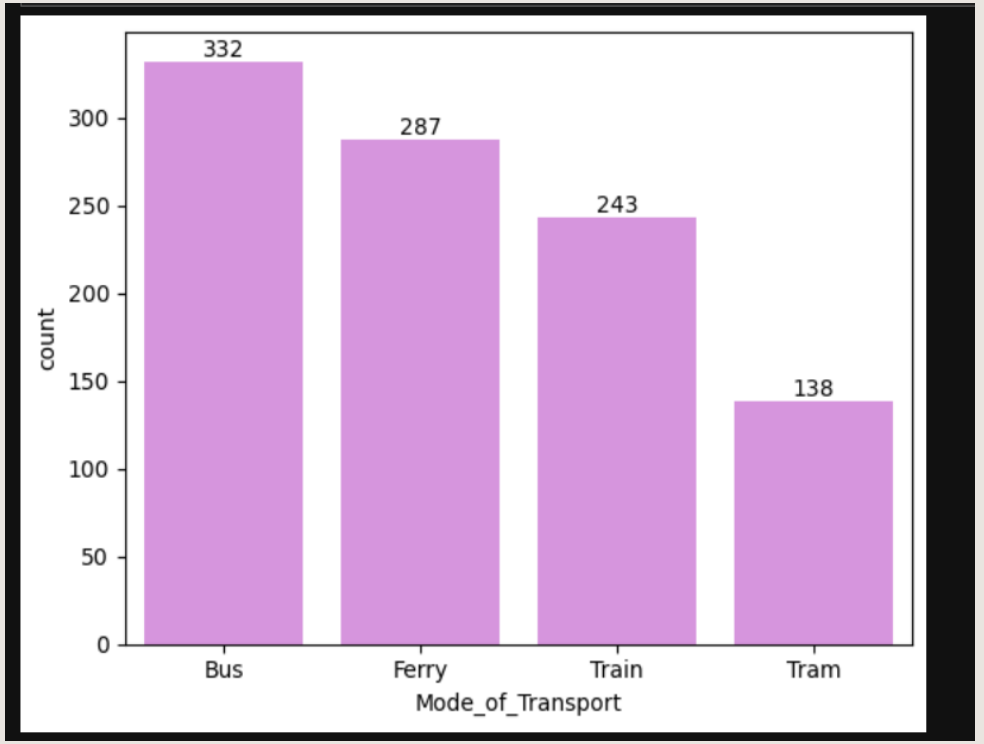
Majority of the trip duration in minutes was between 52 minutes and 2 hours 10 minutes.

The histogram shows that there is an outlier in the data, The distribution is right skewed and the highest total amount made is about 4800, this means some passengers paid higher fare.



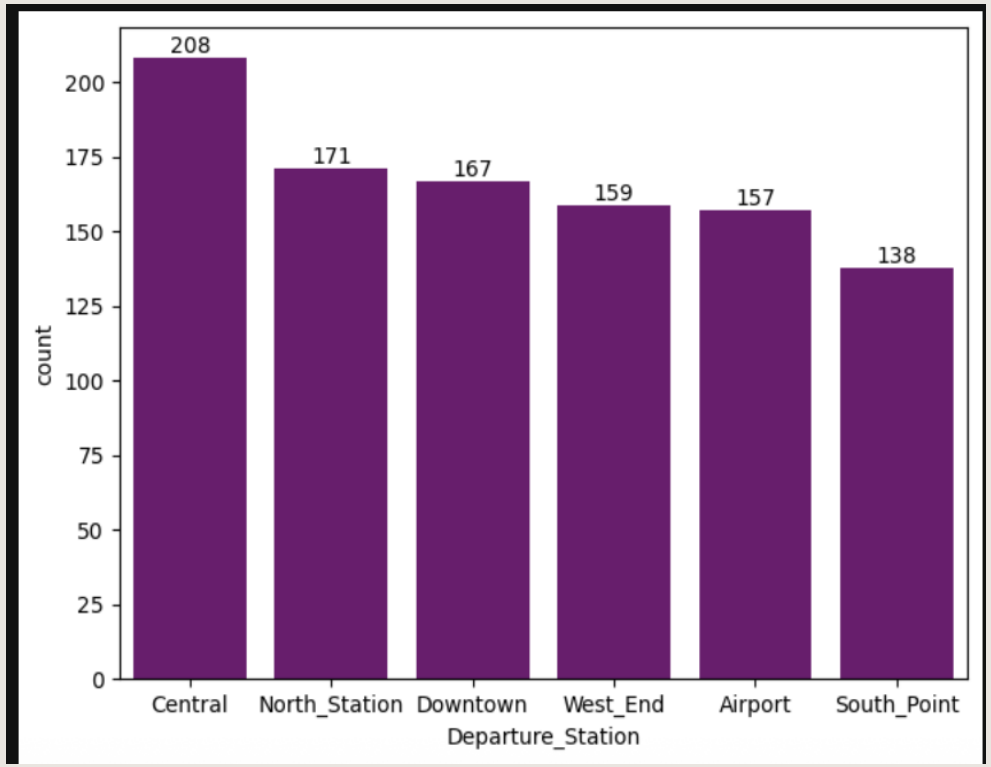
Evaluating Categorical Data Visualization through Univariate analysis

The highest mode of transportation used is the Bus, this means that the number of times bus was used as mode of transportation is more than others



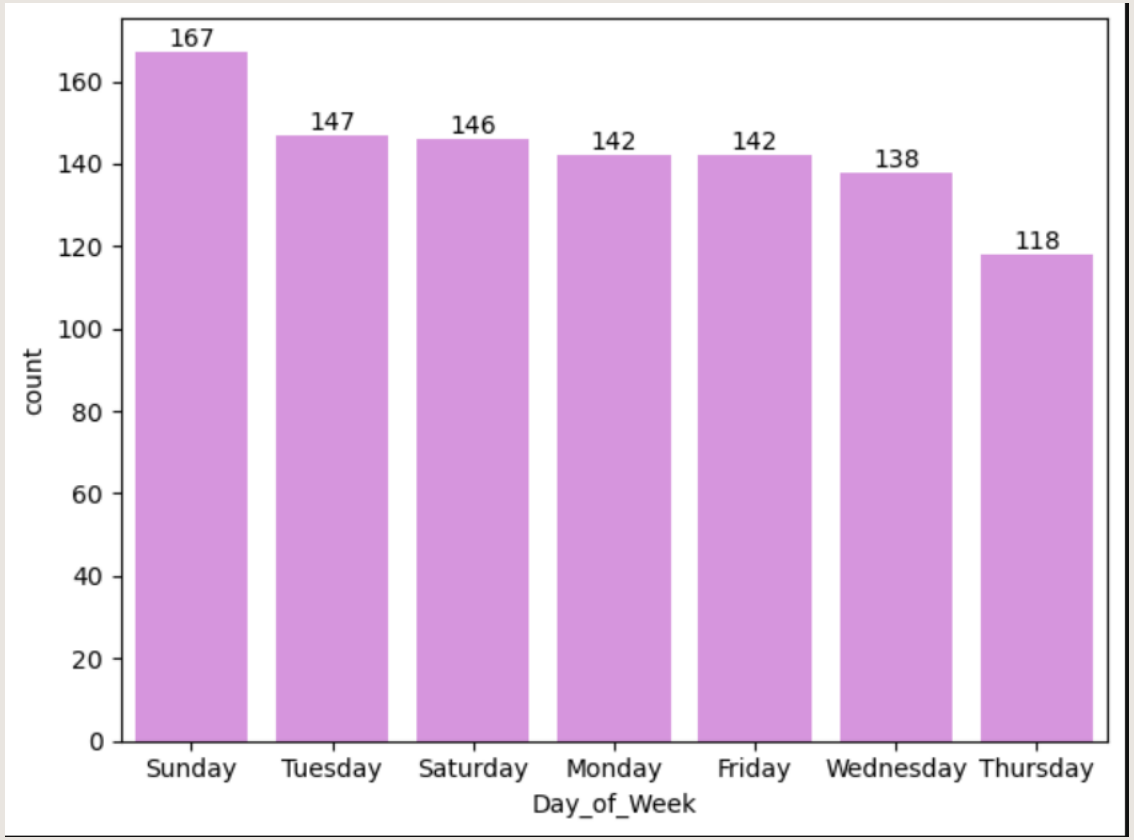
Evaluating Categorical Data Visualization through Univariate analysis

The most used departure station is the Central station because it has the highest count with 208.



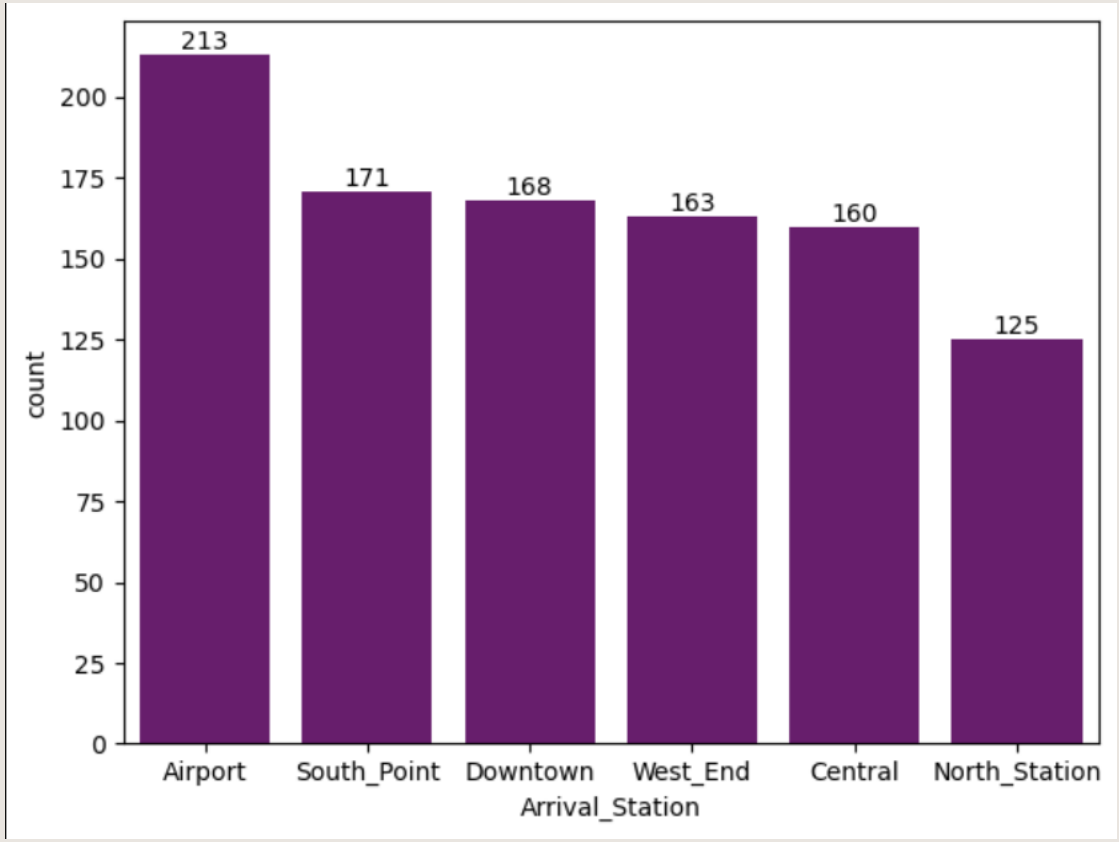
Evaluating Categorical Data Visualization through Univariate analysis

Observation shows that passengers commute more on Sundays than any other day with 167 counts.



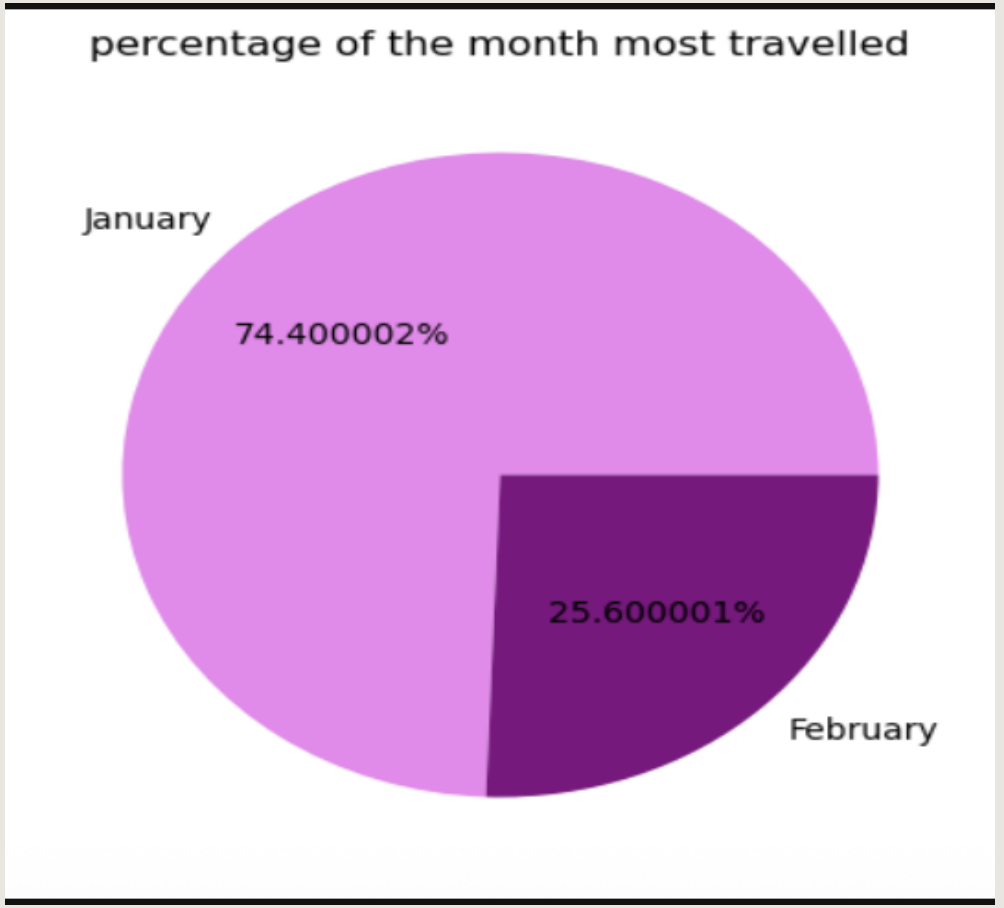
Evaluating Categorical Data Visualization through Univariate analysis

Observation shows that the most frequent arrival station used is the Airport Arrival Station with a count of 213. This station must be very close to the Bus park.



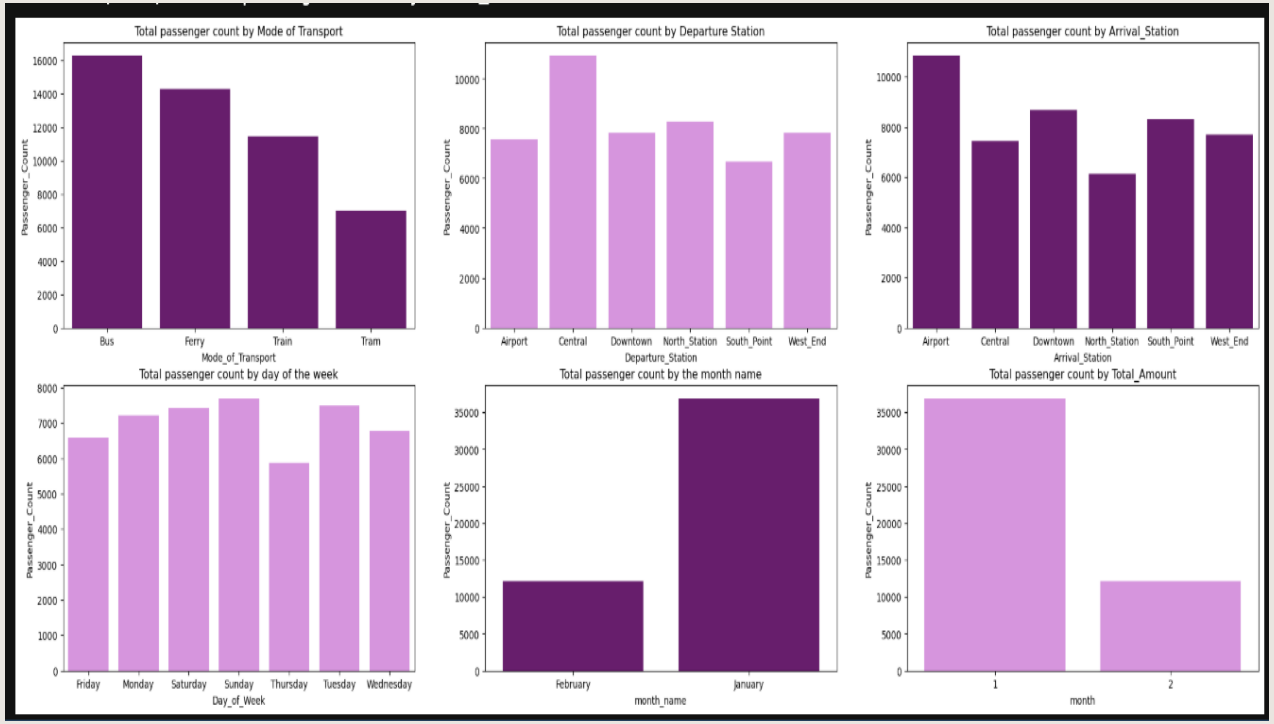
Evaluating Categorical Data Visualization through Univariate analysis

This shows that people commute more in January than February as it is in this chart.



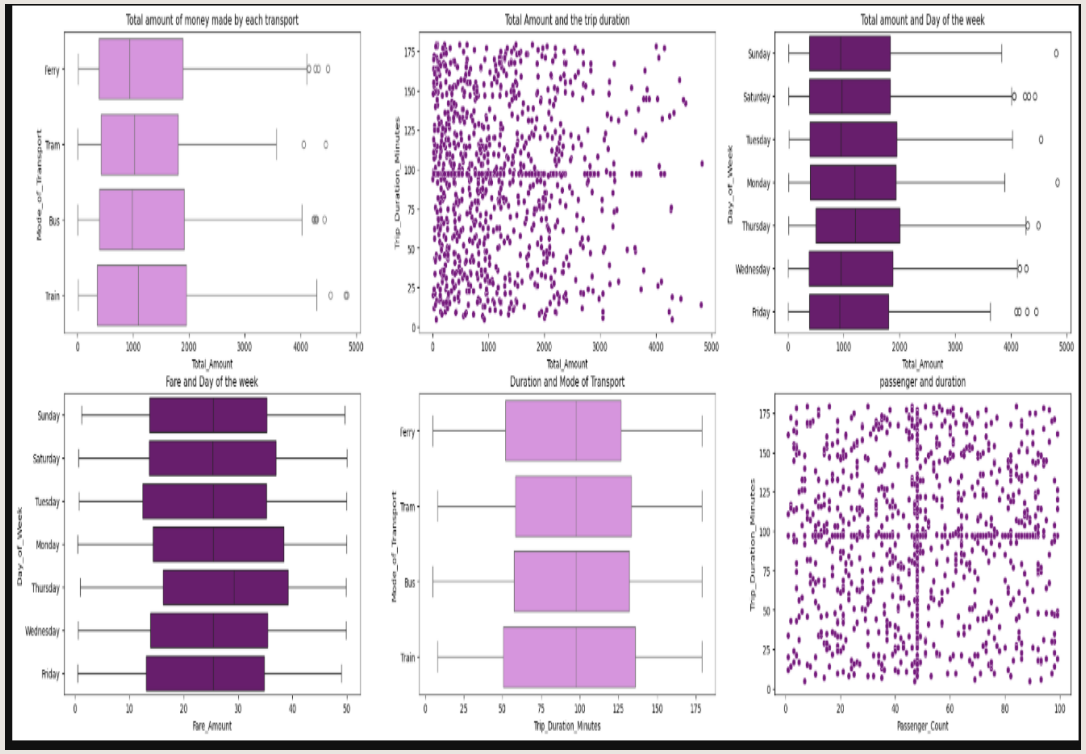
Evaluating Categorical /Numerical Data Visualization through Bivariate analysis

- The passenger count by Mode of transportation in the 1st graph shows that passengers commute more by Bus than ferry or Train
- The passenger count by Departure station in the 2nd graph shows that passengers frequently use Central as their departure station
- The passenger count by Arrival station in the 3rd graph shows that passengers arrive frequently around the airport area
- The passenger count by day of the week in the 4th graph shows that passengers travel more on Sundays followed by Tuesdays
- The passenger count by month name in the 5th graph shows that passengers commute mostly in January when compared with February in 2024



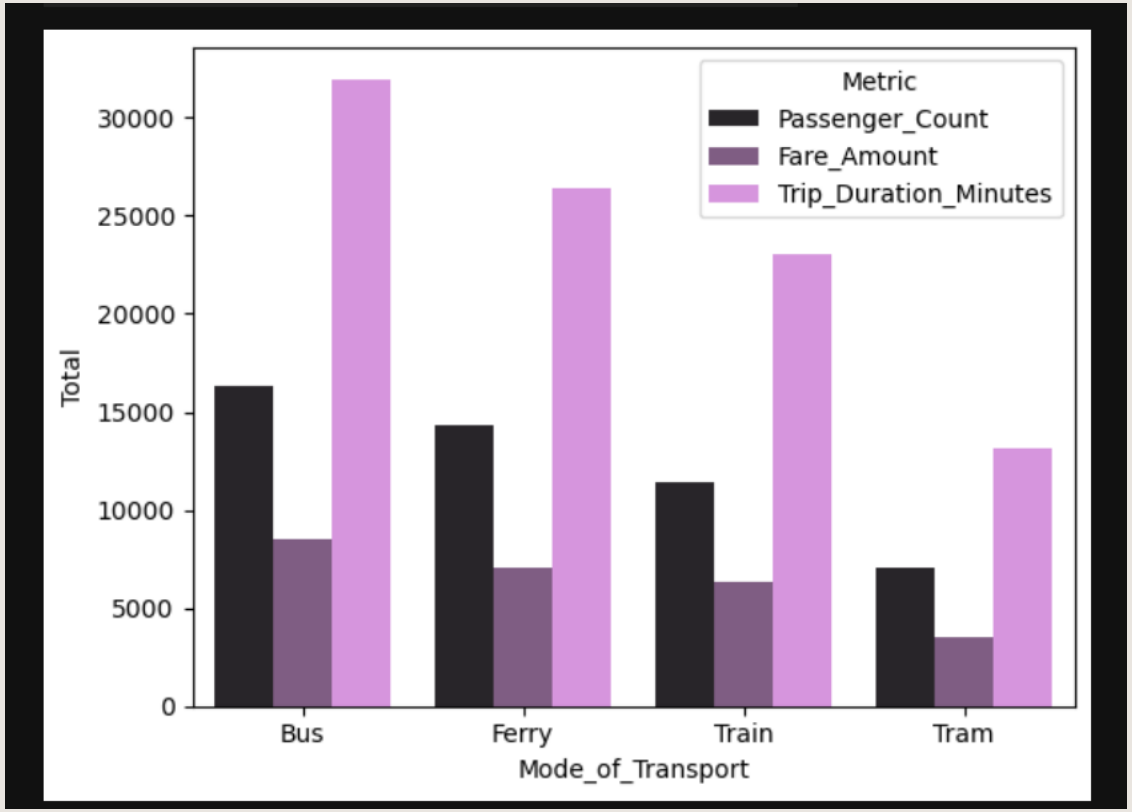
Evaluating Categorical Data Visualization through Bivariate analysis

- The first box plot shows the relationship between mode of transport and total amount made and the concentration is between 300 and 1800, this means that the total amount paid by either bus or Ferry or train and tram is within 1800$ however there are outliers which means that the total amount paid for train transport was up to 4800 while for ferry and bus, it was almost 4500.
- The 2nd graph shows that there are people who spent close to 100 minutes on their trip to their destination irrespective of the total amount.
- Looking at the 3rd graph that shows relationship between day of the week and total amount, it is a bit unusual on Thursday because total amount made varies between 800 and 1800 while other mode of transportation is between 500 and 1500 dollars. This means total amount made daily is between 500 and 1800
- On the 4th graph, it means on Mondays, passengers can pay up to 38$, while on Sundays, passenger pays between 12 and 35 dollars, While the most fare amount is between 12 and 35, on Thursday, passengers still pays up to 39 dollars
- On the 5th graph, mode of transport and trip duration in minutes, ferry takes between 53 and 126 minutes while Train takes a minimum of 51 minutes and a max of 130 minutes. This gives us a clue into how long stays on any mode of transportation. This means that the minimum any passenger can use on transport is 51 minutes and the maximum they can spend is 130 minutes
- On the 6th scatter plot, at least 48 passenger count spends about 98 minutes while on commute to their destinations.



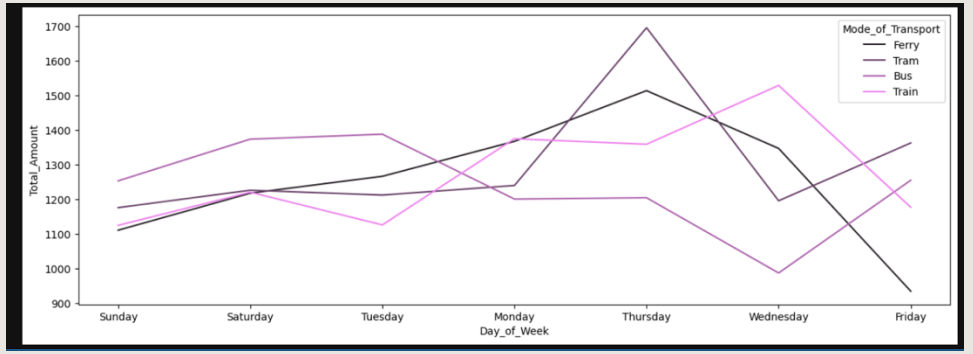
Evaluating Categorical Data Visualization through Multivariate analysis

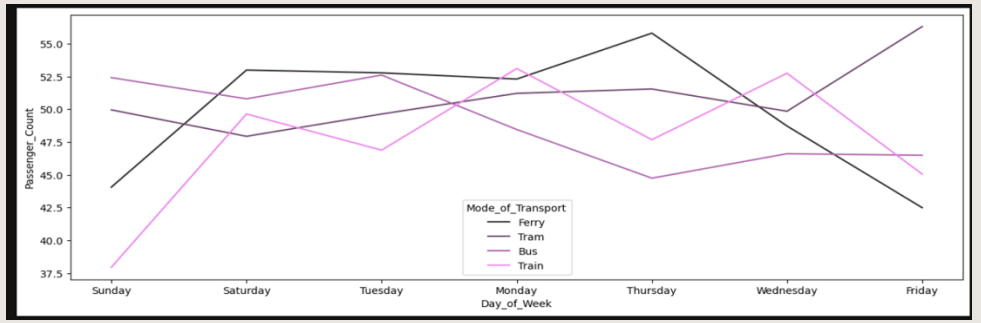
- Using mode of transport to define relationship between number of passengers, fare amount and trip duration.
- The bus mode of transportation is the first choice for passengers, although its fare amount is higher than other mode of transportation, and despite its long duration, passengers still prefer it. It is probably a Bus used to commute when people are going for excursions and vacation.
- The Ferry type of transportation is the 2nd most favorite mode of transportation for people because despite its long duration, and fare amount, people prefer it to the train and the Tram and it could be for sight seeing so people pay to be on it for long hours
- The train's mode of transportation has the 3rd rank mode of transport due to its count of passengers, although the fare amount and the trip duration in minutes is lower than Ferry and the Bus, people still prefer it to the Tram.
- Tram is the lowest in trip duration and even fare amount, yet people do not really fancy it. This could be due to its short duration. Since they are for short distance, it defeats the leisure and excursion purposes. Hence people are not willing to use it, even though its the shortest and lowest fare amount.
- It shows that the set of passengers using this transport mode are people who want to socialize, people who wants to network, play, relax, see some nice places and have fun.



Evaluating Categorical Data Visualization through Multivariate analysis

- The trendline shows that the Tram makes the highest total amount on Thursdays , it also has the highest passenger count on Fridays as well and has the shortest trip duration.
- Tram has the highest passenger count on Fridays which is the most travelled day followed by the Ferry which happens on Thursdays
- the train has the lowest passenger count on Sundays
- On Mondays, the train has the highest passenger count
- while passenger counts has begun to reduce for Ferry, Bus and Train, it has started to peak for the Tram





Evaluating Categorical Data Visualization through Multivariate analysis

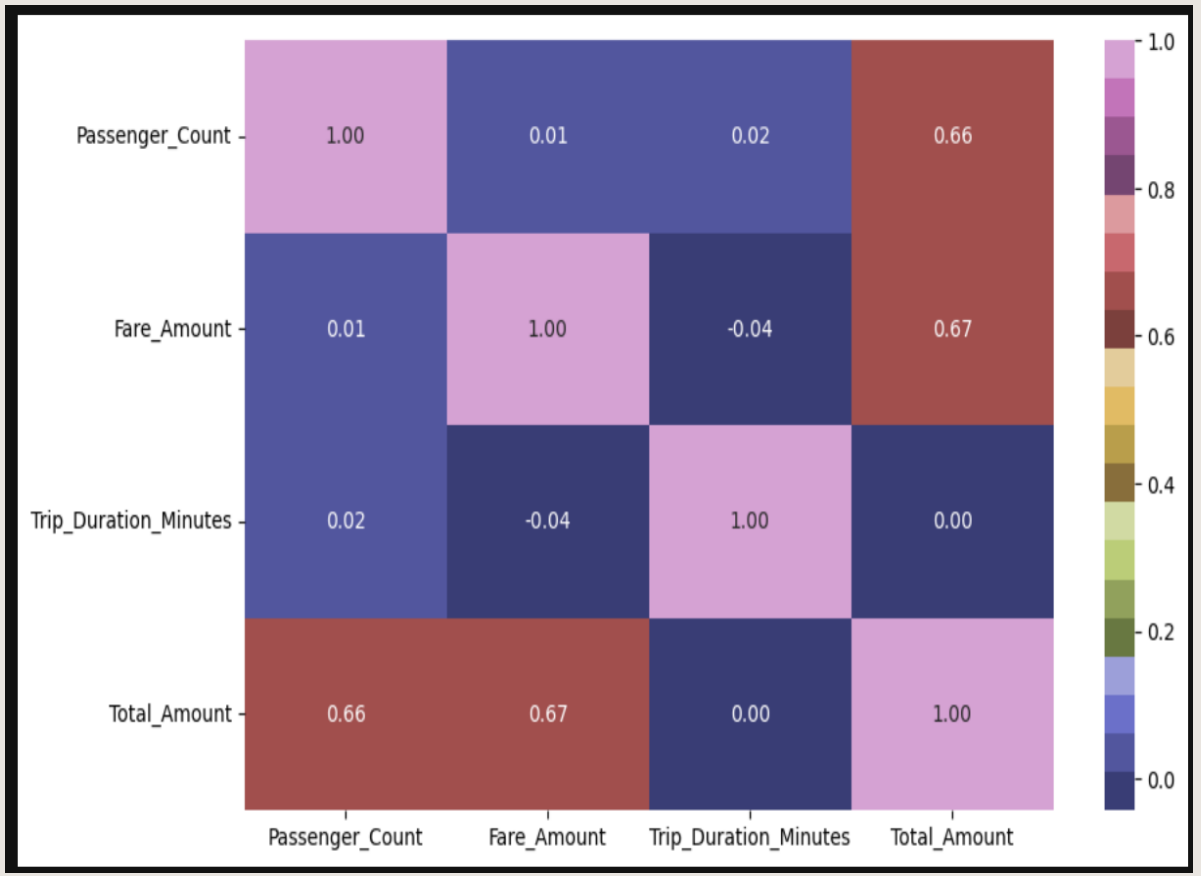
There is a strong correlation between passenger count and Total amount at 0.66,

This means passenger count determines the total amount.

There is a weak correlation/ almost no correlation between fare amount and trip duration with a value of -0.04,

This means fare amount does not really determine the trip duration

there is a weak positive correlation between trip duration and passenger count.



Insights and Recommendation

From the analysis, passengers use the commute for different reasons.

Passengers use the Tram for short distance or if they are going to places that would not take them long arrival time. Since Trams are short form of trains , they can only take a little number of passengers at a time. The Trams is most likely used by professionals and career driven people.

Bus mode of transport seems to be the highest count but that doesn't mean it has the lowest trip duration, or highest passenger count.

Recommendation

- Tram mode of transportation should be made more effective as it accounts for the most used during the week due to its short distance and path.
- Pricing adjustment that better reflects operational costs tied to trip duration should be considered because it seems that the fare amount does not align with the pricing logic.
- Trip duration can be reduced when shorter routes are determined by Metro Move solutions operations in order to increase the number of passengers
- Promotions or service enhancements that encourage higher passenger numbers per trip should be considered, because this is directly tied to revenue.

Index

- Outliers in the data that was filled up through the median.
- Trip date that was converted to months, year, month name
- Trip ID that was dropped due to its insignificance to the data
- Inconsistencies that were handled in the data to ensure accuracy.