Methodology Report:
Visualisation & Analysis on Namma Yatri Data
By Piyali Mujumdar & Pranav Jadhav

Include your visualisations, analysis, results, insights, and outcomes.

Explain your methodology and approach to the tasks. Add your conclusions to the sections.

Table 1: Data Description

Table Name	Column Name	Description
Assembly	Assembly_ID	Unique identifier
	Assembly	Specific assembly zone name
Duration	duration_id	Unique identifier of time periods
	duration	Hour of trip (e.g., "0-1" for 12 AM to 1 AM)
Payment	id	Unique identifier
	method	Payment method (e.g., Cash, UPI, Credit Card)
Trip Details	tripid	Unique identifier of trips
	loc_from	Source Location code
	searches	Trip request count
	searches_got_estimate	Got an estimated price (1 = user gets an estimate, 0 = does not get an estimate)
	searches_for_quotes	Searched for drivers after estimate (1 - searched, 0 - not searched)
	searches_got_quotes	Got quotes (1 = Driver allotted, 0 = not allotted)
	customer_not_cancelled	Whether customer cancelled or not (1 = Not cancelled)
	driver_not_cancelled	Whether driver cancelled or not (1 = Not cancelled)
	otp_entered	(1 = OTP entered, 0 = not entered)
	end_ride	Whether ride was completed (1 = Completed)
Trips	tripid	Links to Trip Details
	faremethod	Payment method ID, links to Payment table
	fare	Fare amount
	loc_from	Location ID of source
	loc_to	Location ID of destination, links to Assembly table
	driverid	Driver ID
	custid	Customer ID
	distance	Distance in KM from source to destination
	duration	Unique identifier of time periods like duration_id

Points to Note:

1. **Without this methodology document, the other parts of your case study will not be evaluated.**

2. This assignment is different from the ones you have solved before. Make sure that you treat this case study as a storytelling exercise and not an analysis/visualisation one. This will help you be better prepared for the presentations.

3. Once you are done with the analysis and visualisations, there will be many insights at your hand. Make sure that you map the right visuals and takeaways with the right audience since some of these insights might be relevant to one group but not to the other group.

4. **DO NOT** change the text or numbering of any task, as it may cause problems with grading. Write your solutions to a task in the space provided below the respective task.

Tasks to be performed

- Present the overall approach of the analysis.

- Mention the problem statement and the analysis approach briefly.

- To solve a task, you have to create relevant visualisations and derive appropriate insights from the visualisations.

- Add all the plots, insights, calculated field commands, results and outcomes for a task with proper numbering and sequence in the report.

- The scores for all tasks (except conclusions) comprise both analysis work in the visualisation tool and its outcome in the report.

- You will be awarded a score for a task only if the Tableau/PowerBI analysis is correct and is included in the report along with the subsequent insights.

- Finally, draw conclusions based on the analysis.

Scoring:

Report Total Marks: 70
Sections: 3 sections (10 marks + 40 marks + 20 marks)

Analysis and Visualisation

1. Data Preparation [10 Marks]

1.1. Import and Join Tables Correctly [5 Mark]

- Import the Namma Yatri dataset into Tableau/Power BI.
- Ensure that you correctly join all tables to create a unified dataset for analysis.
- Verify the relationships between different tables and confirm that data from various sources is properly aligned for accurate insights.

Solution:
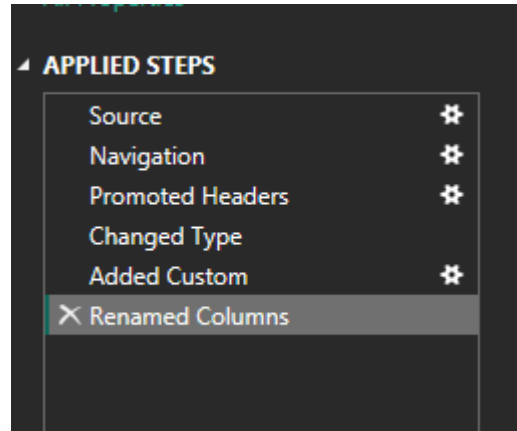
1 We imported the data in the power bi using Excel workbook

2. we transform the data first, by default the data type for some columns were in numeric we converted them into text.

> *E.g. ID column in each sheet, loc_from, loc_to,*

▲ **APPLIED STEPS**

Source	☼
Navigation	☼
Promoted Headers	☼
Changed Type	
Added Custom	☼
✕ Renamed Columns	

3. Created new column in duration sheet of time category which describes whether the triip Is taken 'late night', 'Early morning', 'morning', and so on using M code in power query.

> *= Table.AddColumn(#"Changed Type", "Time_Category", each let*

StartHour = Number.FromText(Text.BeforeDelimiter([duration], "-"))

in

if StartHour >= 0 and StartHour < 4 then "Late Night"

else if StartHour >= 4 and StartHour < 8 then "Early Morning"

else if StartHour >= 8 and StartHour < 12 then "Morning"

else if StartHour >= 12 and StartHour < 17 then "Afternoon"

else if StartHour >= 17 and StartHour < 20 then "Evening"

else if StartHour >= 20 and StartHour <= 23 then "Night"

else "Unknown")

	A^B_C id	A^B_C duration	$^{ABC}_{123}$ Time_Category
1	1	0-1	Late Night
2	2	1-2	Late Night
3	3	2-3	Late Night

5. *We join fact table like Duration, Assembly, Payment with the dimension table Trips with cardinality as one to many and extension of dimension table Trips with its extension Trips detail with it as one to one cardinality*
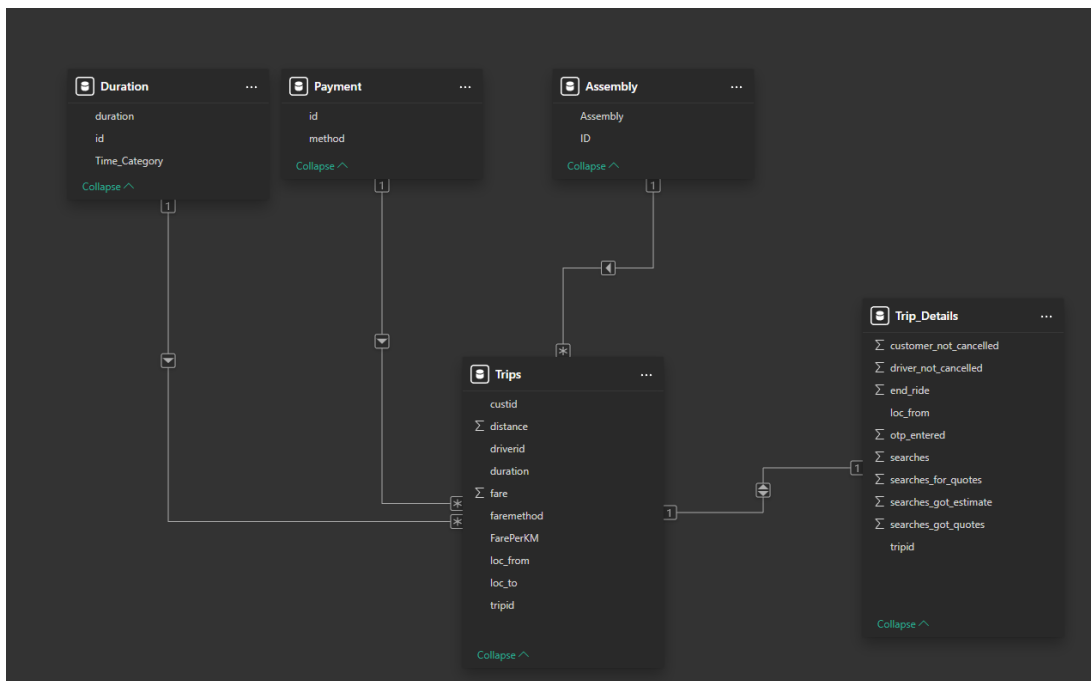
1.2. Find and Resolve Inconsistencies [5 Marks]

- Identify and resolve any inconsistencies or issues in the dataset that might affect the analysis.
- Clean the data to ensure it is structured properly for analysis, removing any irrelevant, duplicate, or erroneous entries.
- While performing the analysis, create calculated fields as needed to ensure the accuracy and relevance of the insights.

Solution:

1. *Number of records in Trip Details is more than that of Trips table, we have ignore the data which is not matched with Trips and kept as it is*
2. *There was one duplicate column in the extension of dimension table Trip Details where it has an extra column as "loc_from" which is deleted from the dataset.*
3. *There were no erroneous entries*
4. *Also there are 2 columns driverId and customerId which are irrelevant for our analysis*
5. *We created Measures like total revenue from the fare column,*
6. *We created the 1 table which is category for the duration which could help us further in the analysis*
7. *We created a calculate column which is price per Km which actually helps us to determine the issue*

2. Exploratory Data Analysis   [40 Marks]

2.1. Classify Variables into Categorical and Numerical [2 Marks]

- Classify all the variables in the dataset into numerical and categorical types.

Solution:

Sheet	Column	Type
Duration	id	Numerical
	duration	Categorical
Assembly	ID	Categorical
	Assembly	Categorical
Payment	id	Categorical
	method	Categorical
Trip_Details	tripid	Categorical
	loc_from	Categorical
	searches	Categorical
	searches_got_estimate	Categorical
	searches_for_quotes	Categorical
	searches_got_quotes	Categorical
	customer_not_cancelled	Categorical
	driver_not_cancelled	Categorical
	otp_entered	Categorical
	end_ride	Categorical
Trips	tripid	Categorical
	faremethod	Categorical
	fare	Numerical
	loc_from	Categorical
	loc_to	Categorical
	driverid	Categorical
	custid	Categorical
	distance	Numerical
	duration	Categorical

2.2. Analyse Ride Demand Over Time [3 Marks]

- Explore the distribution of ride demand over time, including trends across different periods.
- Identify the peak demand periods. Choose an appropriate parameter for demand based on your own understanding.

Solution:
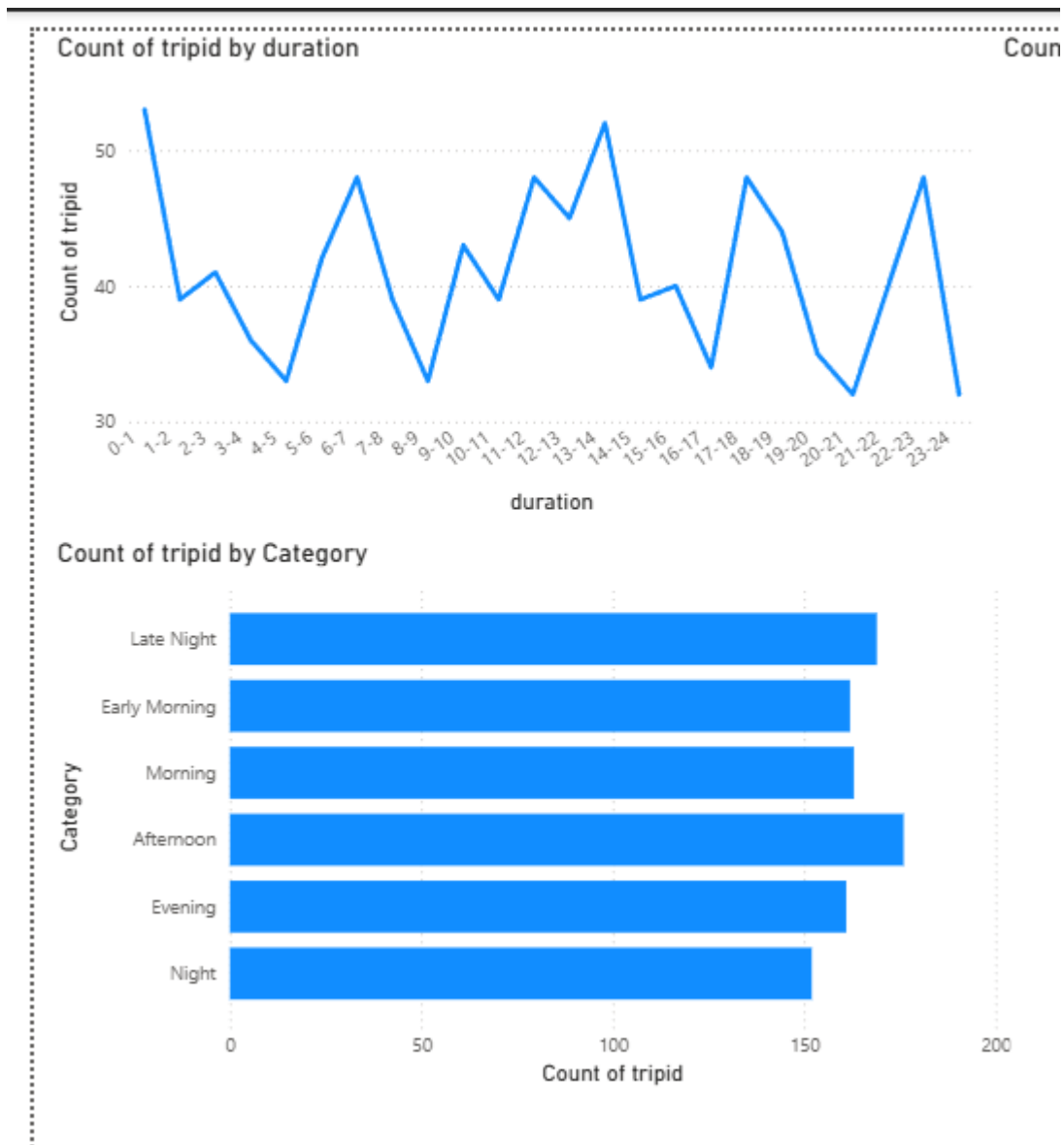
1. We created the line chart with the help of Count of TripIds and Duration
2. We got to know that there were some unusual trends in the line chart where the peak is observed during 0-1 am then 13-14 assuming the working hours
3. 6-7, 17-18, 22-23 these peak hours look Logical

4. *Further we classified the durations in different time periods as in Late Night, Early morning , Morning, afternoon, evening , Night*
5. *We found Afternoon is being highest followed by Late night and Night being lowest of all the time periods where the trips are more.*
6. *It could be drivers availability or Afternoon Heat could also be the reason in Hyderabad.*
7. *This could be the possibly area where Maketing Team could focus on*

2.3. Proportion of Total Revenue from Different Time Periods [3 Marks]

- Calculate the proportion of revenue generated during different time periods and visualise how it contributes to total revenue.

Solution:
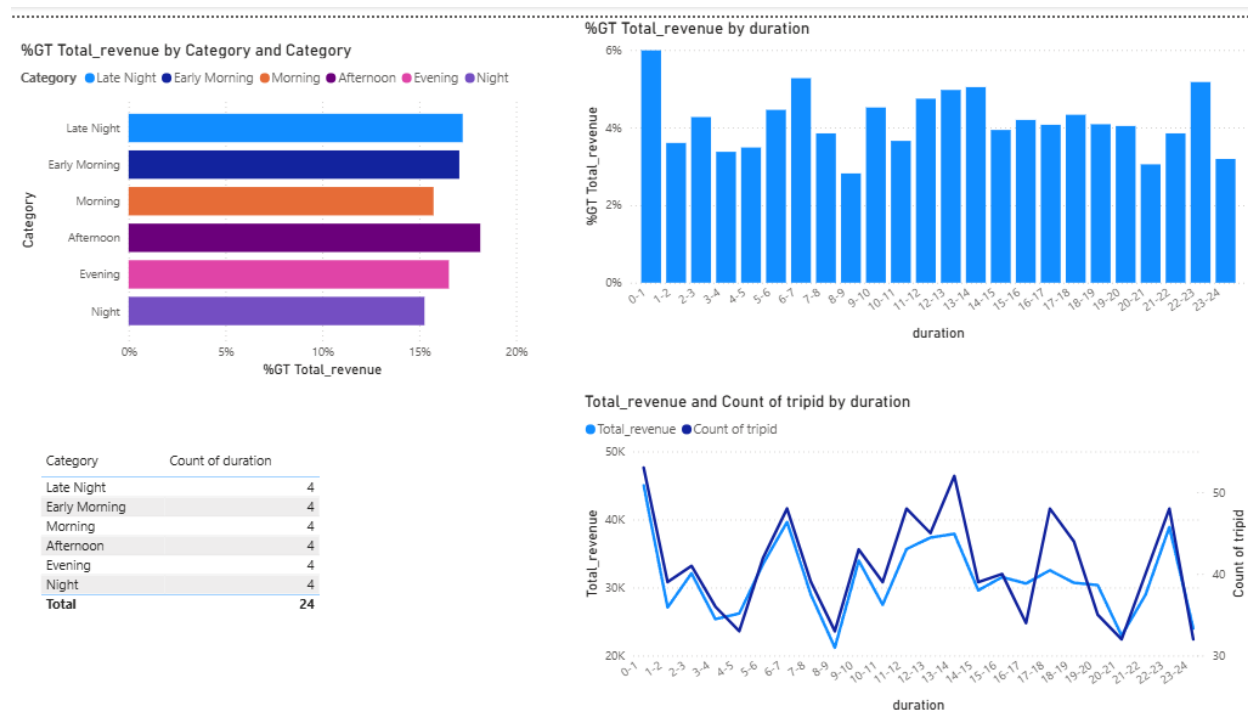
8. *The Total revenue Generated and Count of trips actually traces the similar path in line chart*
9. *Except some time block in Afternoon and Evening*
10. *There could be the rise in the revenue in night and late night specifically 0-1 and 22-23 because of the night rates as it is comparably high.*

2.4. Explore the Relationship Between Trip Hour and Revenue [3 Marks]

- Investigate the correlation between trip hour and total fare.
- Explain any trends or patterns that emerge.

Solution:

TotalFareByHour by duration
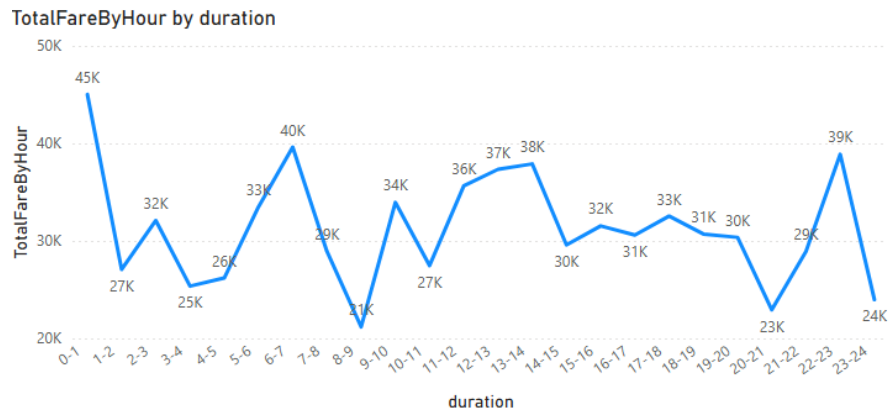
11. We Created a measure for total fare by hour using dax functions and used on the line chart along the duration on x axis.
12. As such there is no positive or negative correlation between the trip hour and revenue
13. But there is trend seen here as the hours are 0-1, 6-7, 22-23 and 13-14. We have seen the peak in the revenue generated
14. The trend can be explained by the Average Fare per km chart below, where the fare is higher due to night charges or high demand and low number or drivers avaiable

AverageFareperkm by duration
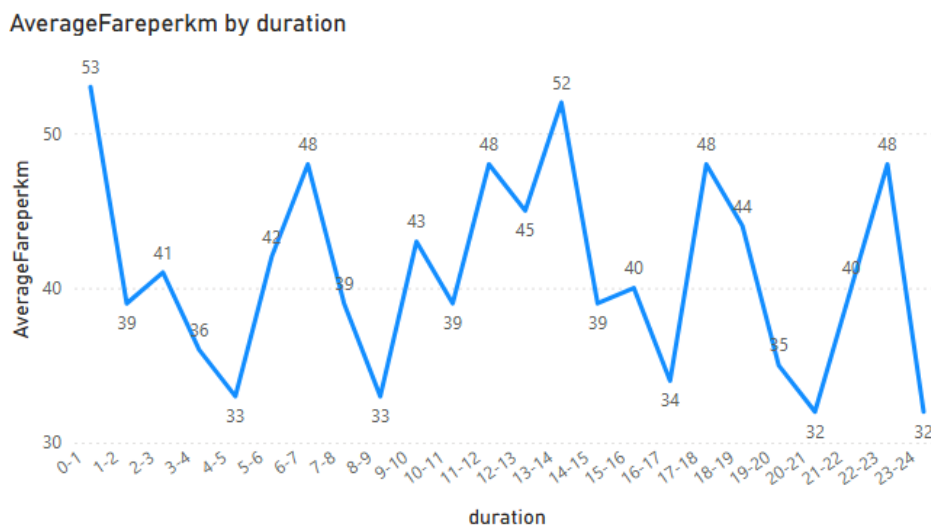
2.5. Examine the Popularity of Different Payment Methods [3 Marks]

- Analyse the distribution of various payment methods used by customers.
- Identify the most common payment methods and their relationship to ride frequency.

Solution:

Count of tripid by method
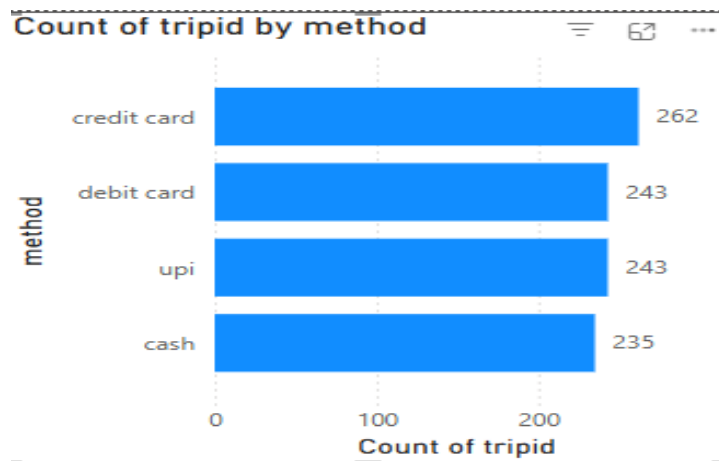
15. With use of Credit card being highest doesn't helps us to find the clear trend in the payment method
16. But the revenue generated would helps us a bit.
17. With maximum revenue generated from Credit card, followed by cash, debit card and upi
18. We Marketing team could promote the use of UPI for the payments as it is the easiest and secure way

Total_revenue by method
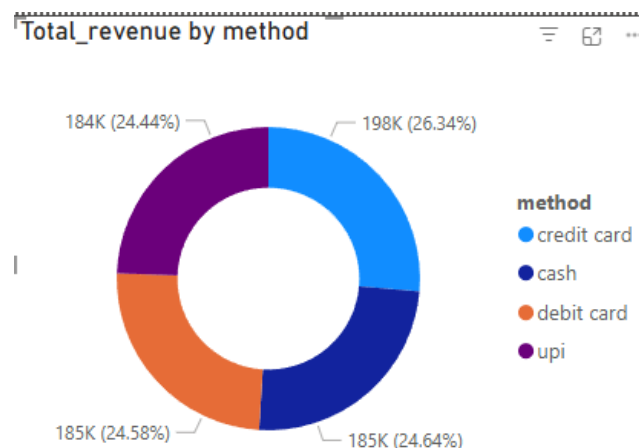
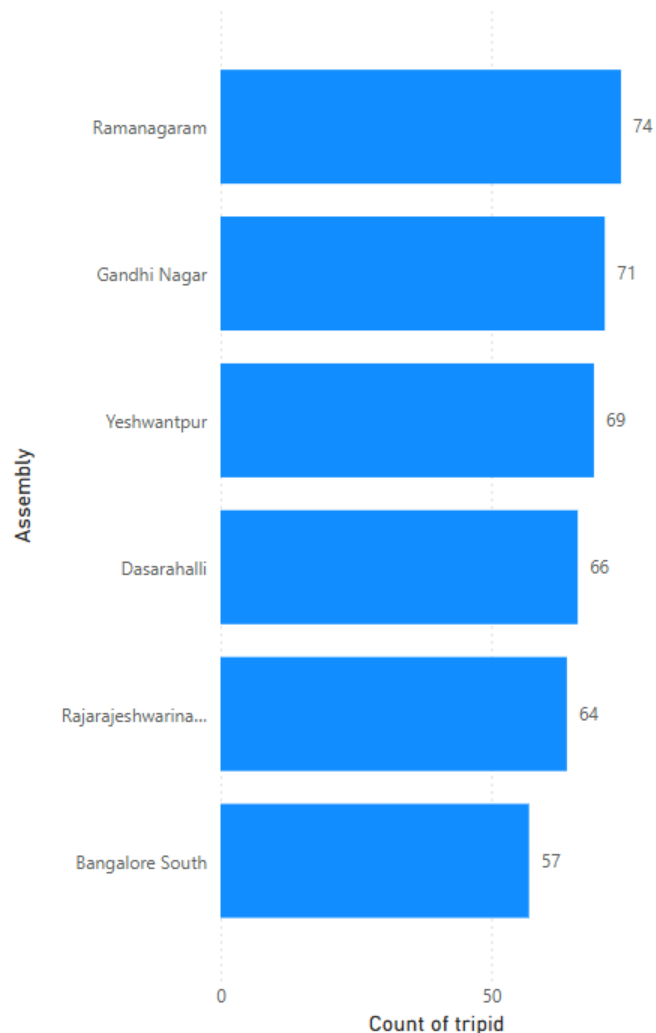## 2.6. Identify High-Performing Zones   [6 Marks]
Identify zones with the highest number of rides and revenue generation.
Analyse factors contributing to their performance:

- **2.6.1. Rides:** Identify pickup zones with the highest number of trip requests.
  [3 marks]

Solution:

19. *The approach was to create a bar chart which would help to visualize the count of rides using assembly and count of tripids*

Count of tripid by Assembly

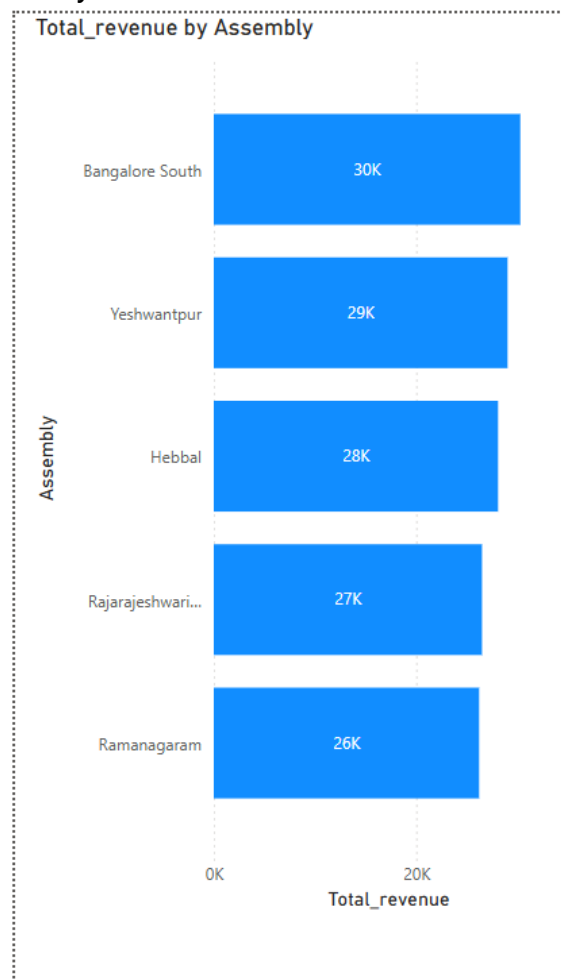| Assembly	Count of tripid
Ramanagaram	74
Gandhi Nagar	71
Yeshwantpur	69
Dasarahalli	66
Rajarajeshwarina...	64
Bangalore South	57

20. *Then applied the advance filter of Top N on the count and produced the visual for top 5 zones with highest number of rides*
21. *Of which Ramanagram being highest followed by Gandhi nagar, Yeshwantpur, Dasarahali , Rajeshwarinagar and Banglore South*

- **2.6.2. Revenue**: Identify pickup zones generating the highest revenue. **[3 marks]**

Solution:

---

22. *The approach was to create a bar chart which would help to visualize total revenue generated using assembly and total revenue measure*

Total_revenue by Assembly

Assembly	Total_revenue
Bangalore South	30K
Yeshwantpur	29K
Hebbal	28K
Rajarajeshwari...	27K
Ramanagaram	26K

23. *Then applied the advance filter of Top N on the total revenue measure and produced the visual for top 5 zones with highest number of rides*
24. *Of which Banglore South being highest followed by Yeshwantpur, Hebbal, Rajeshwari nagar and Ramanagarm*

2.7. Analyse Ride Time Periods Across Zones [4 Marks]

- Compare the trip trends for different time periods across pickup zones.

Solution:

| Assembly	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24
Bangalore South	1	1	1		1		1	1	1		1	2	2	4	1	2	2	2	3	1	1	1	2	1
Chamrajpet	3	1	1	1	1	1	1	1	1		3	5	3	1	1	1	1		1			1	1	2
Channapatna	1	2	2	1	3	2	2	1	1	1	1		1	1			2	1	1	2	1	1	2	1
Dasarahalli	4	1		2	1		1		1	5		4	1	1	3	1	1	3	1	1	1		1	
Gandhi Nagar	3	2	1			1	3			2	1	1	2	1	2	3		3	2	1	1		2	1
Govindraj Nagar		1		2				2	1	1		1	2	2	1	2	4		1	1	1	3	1	4
Hebbal		2	1		1	1	1	1			2	2	3	1	1	3	2	1	2	4	1		1	1
Jayanagar	2	5		2		4	2	2		2					3	1		1	2	1		1	1	1
Other Assemblies	2		1			1		1	2		4	1	2	2	2	1	1		1	1	1	2	4	1
Rajarajeshwarinagar	1	2	2		3	1		1	3			1	3	1	1	2	2		2		3	1	2	1

25. We used the concept of heat map hear, the approach was simple initialy we created a matrix for the 2 categories ie. Assembly and Duration, with value as count of trips

26. Thus we can clearly identify for which zone we are having higher demands at a particular time so we can guide our drivers to that region.

2.8. Top Zones with Highest Trip Volume [3 Marks]

- Identify the top 5 pickup zones with the highest total number of completed trips.
- Analyse factors contributing to the higher number of trips.

Solution:

Count of tripid by Assembly
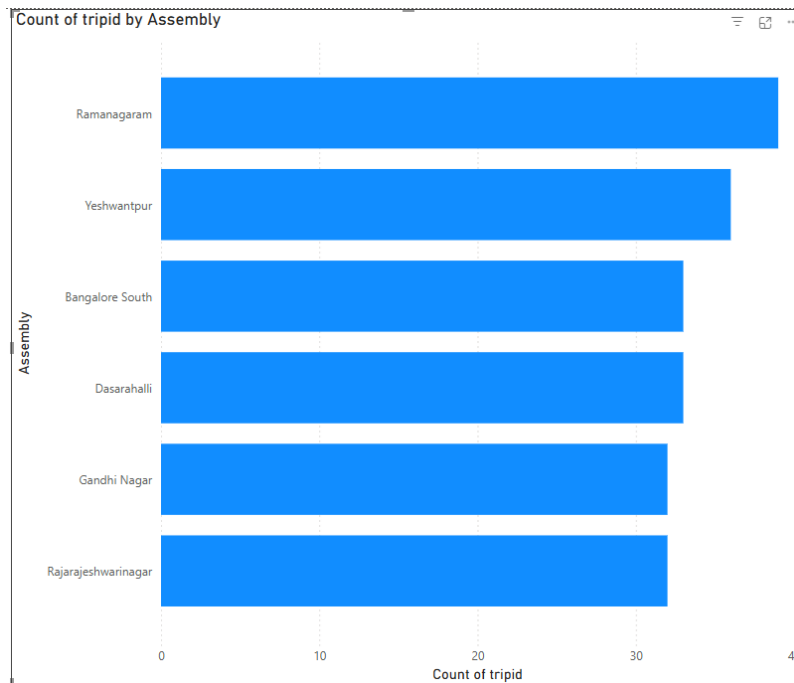
27. *The approach was to create a bar chart which would help to visualize the count of rides using assembly and count of tripids*
28. *Then applied the advance filter of Top N on the count and produced the visual for top 5 zones with highest number of rides*
29. *Of which Ramanagram being highest followed by Yeshwantpur, Banglore South, Dasarahali ,Gandhi nagar, and Rajeshwarinagar*

2.9. Basic Analytical Tasks [8 Marks]

- **2.9.1**
  What are the percentages of cancellations and successful rides by both driver and customer? **[3 marks]**

Solution:
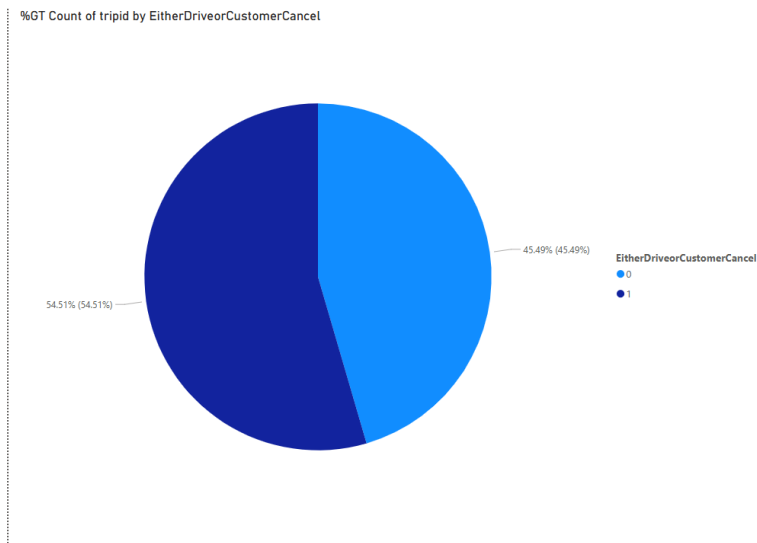
30. *We used the Pie chart for the visuals, to create this visual we created a Colum using if DAX function and Or function where is giving a condition either the driver or customer has cancelled the trip*
31. *We found that almost 55% of the times either the driver or customer has cancelled the trip*
32. *That could be due to time required to reach or fare price, or driver got stuck in the traffic.*

- **2.9.2**

  Analyse the percentage of people who completed trips after searching for quotes. Visualise the variation of this ratio by time periods. **[5 marks]**

Solution:

quotes_searchedcount and completion_ratio by duration
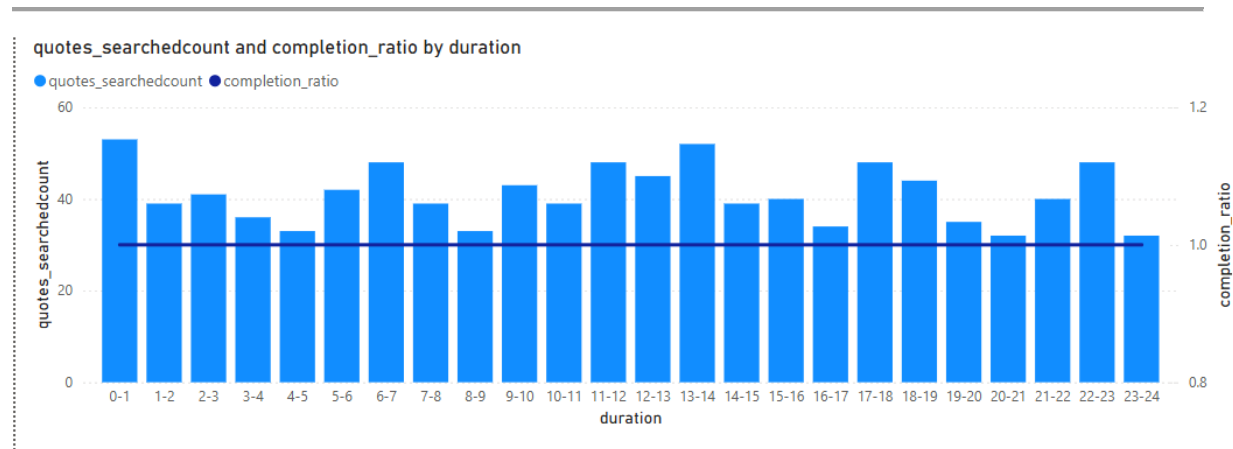
33. *We created a 2 measure quotes_search_count which counts the number or trips which had searched for the quotes if yes and other was tct_After_Search_of_quotes which counts if the both search for quotes and end_ride is yes*
34. *The last measure created was of completion ratio which divides the above 2 created measures and give the ratio*
35. *Then the bar chart was used to visualize Quote searched count and completion ratio*
36. *Here we were amazed that the completion ratio was 100% whenever the quotes were searched for each timeblock.*

2.10. Create a Parameter and Use Filters [5 Marks]

- Create a parameter and use it as a filter on an appropriate subset of the data to interactively analyse and visualise different subsets of the data.
- Explain your choice of filter and insights drawn from this step.

Solution:

37. *We created a parameter of no of trips in percent which is used on the Dashboard which will increase the revenue by same percent as increased in Number of trips (Assuming the 1:1 ratio)*
38. *We have used Filters or Time Period, Payment method, Assembly as pickup point, Driver not cancelled as Driver accepted the ride, Customer not cancelled as Customer Accepted the ride, searched for the quotes and end ride as trip got completed.*
39. *Also created 2 buttons which were linked with the bookmarks which would show 2 vies 1st In Terms of count and 2nd In Terms of Revenue*

# 3. Conclusion   [20 Marks]

3.1. Recommendations for Operational Efficiency   [10 Marks]

- Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
- This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
- Add supporting dashboards.

Solution:

A. *We should Focus on the increasing the number of rides in each area by 10-20% which would help us to reach the revenue figure 750k mark*
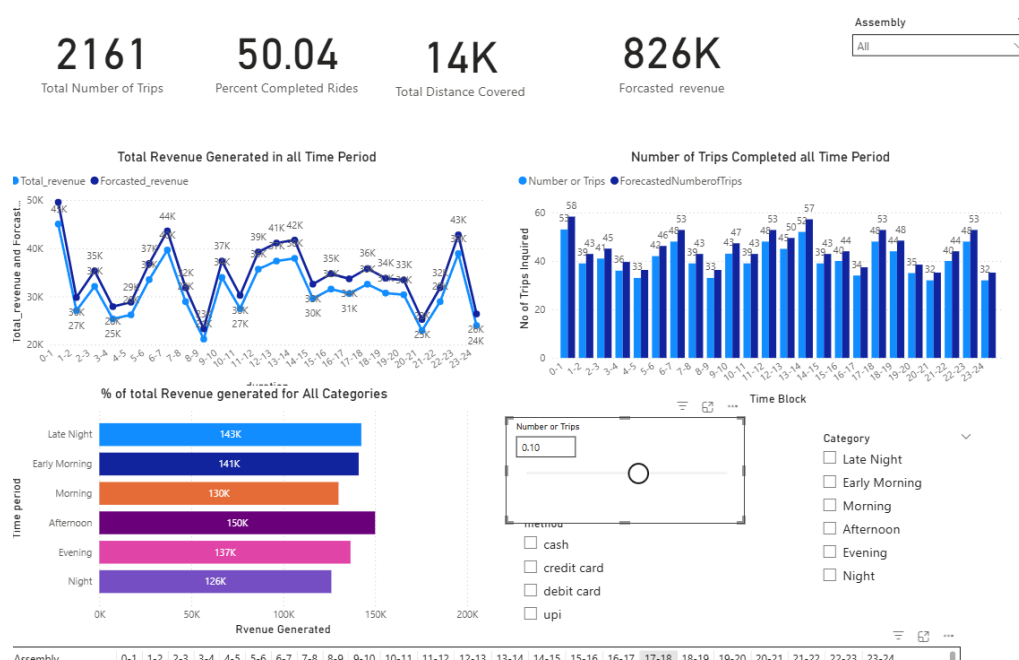
B. Increse the revenue generation across all the zones
   a. The average Fare price for the pickup point is way to high than the Average fare price ie. 62.54 to 53.11 per KM per hr in Anekal
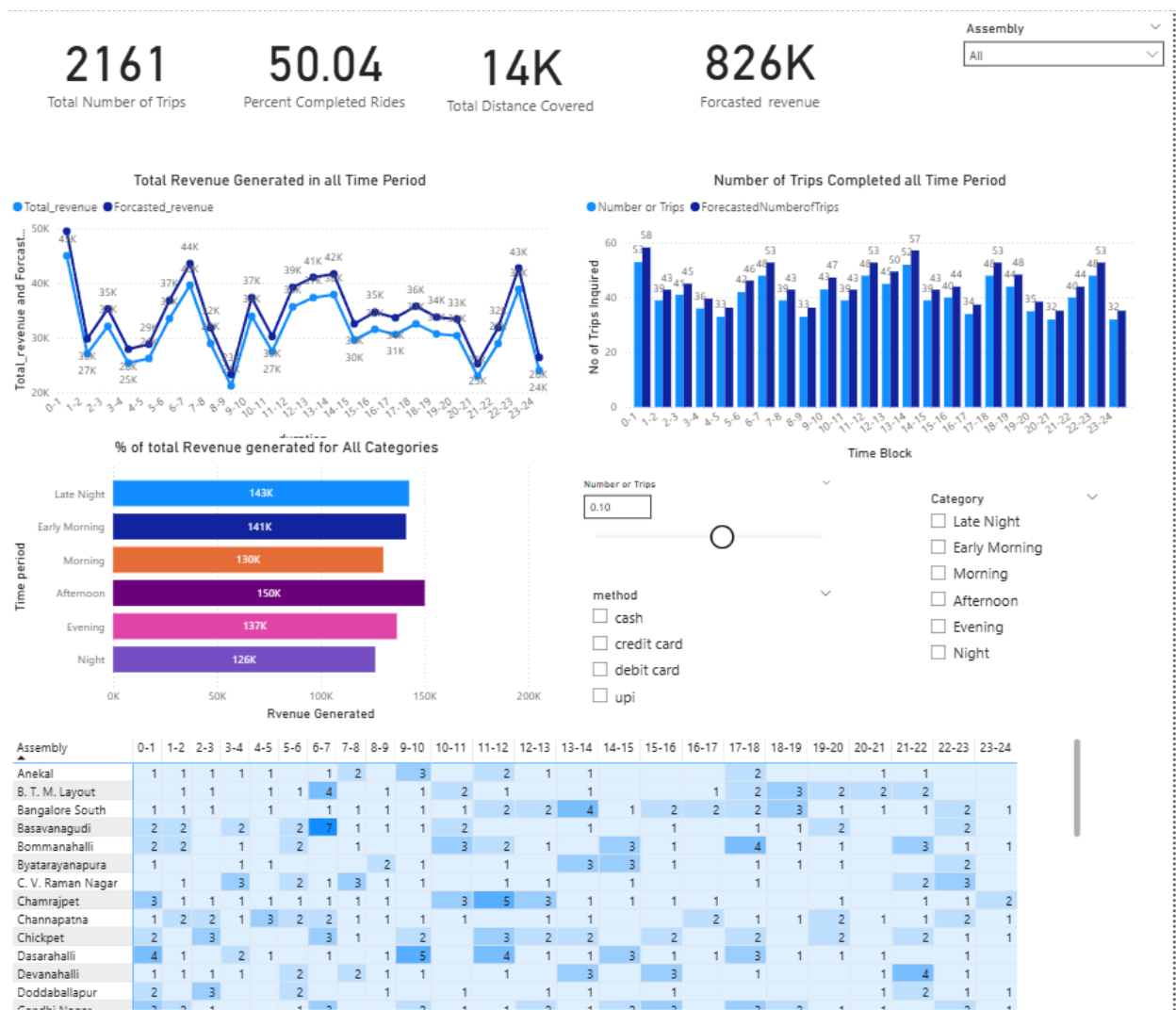   b. The revenue is concentrated in high-demand assemblies
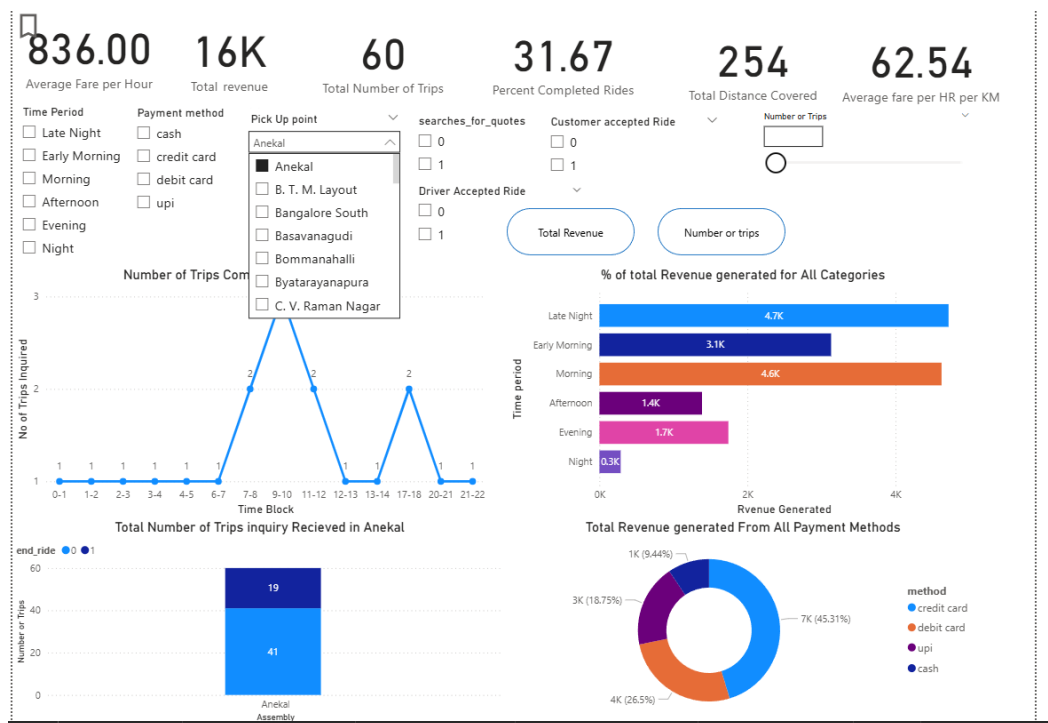
C. Allocation of the Drivers in the specific location in specific time
   a. We have found when there is peak in one zone the other zone is not having that much demand so shifting those drivers to the actual demand location will help improve the operations effectively and help in increase of revenue

D. DASHBOARD

### 3.2. Marketing and Operational Strategy Improvements   [10 Marks]
- Suggest improvements to Namma Yatri's marketing or operational strategies based on your analysis.
- Recommendations could involve promotional efforts, driver incentives, or regional targeting to increase customer satisfaction and service efficiency.
- Add supporting dashboards.

Solution:

A. We need the promotional campaigns to improve the Operations of the Namma Yatri.
   a. There could be various ways to improve including the promotional offers

on rides above particular KM

b. If the Driver is allocated to the client there is 100% completion rate which surely suggest that we have less drivers compared to the demand
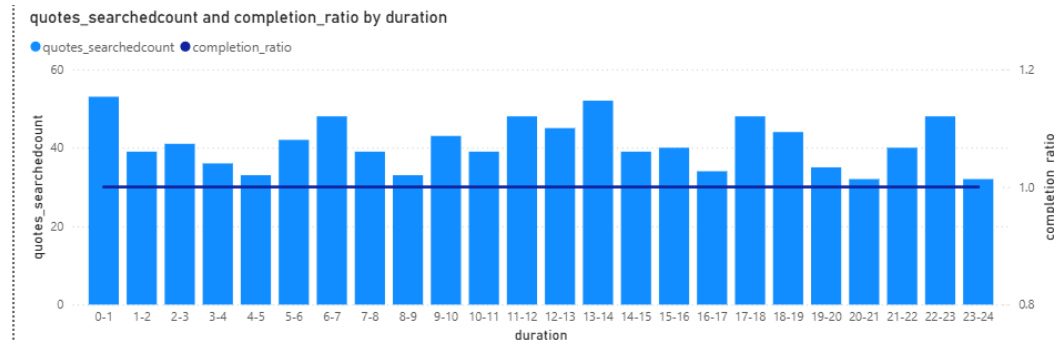
c. We cant control the Client side but there can be control over driver side cancellation which will increase the operational metrics and increase the ride completion rate
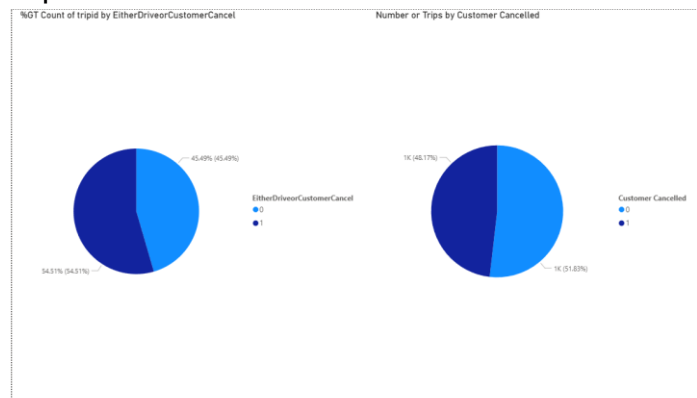
B. We can keep focus on on Top revenue Generating zones, but need to increase the ride counts in Anekal and C V Ramana Nagar. Where the Ride cancelation rate is higher than that of any other city.

Key influencers   Top segments

What influences end_ride to be [0   ⌄] ?

When...	...the likelihood of end_ride being 0 increases by
searches_for_quotes is 0	3.08x
searches_got_estimate is 0	2.27x
Assembly is Anekal	1.26x
Assembly is C. V. Raman Nagar	1.24x

C. DASHBOARD