upGrad

Methodology Report:
Visualisation & Analysis on Namma Yatri Data

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)

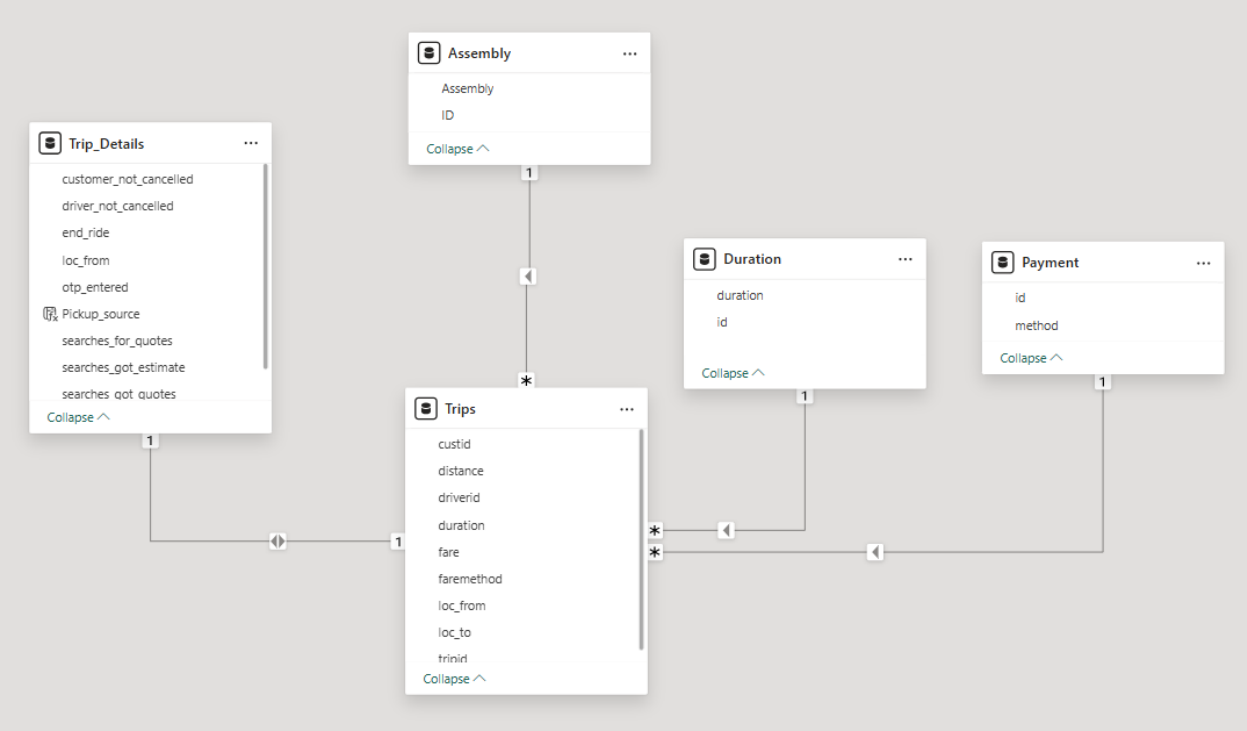
Analysis and Visualisation

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:



We imported the Namma Yatri dataset into Power BI and successfully joined all the tables by using appropriate keys to ensure a unified database. Below are the relationship between these tables:

- Trips (tripid) → Trip_Details (tripid): One-to-One
- Trips (faremethod) → Payment (id): Many-to-One
- Trips (duration) → Duration (id): Many-to-One
- Trips (loc_to) → Assembly (id): Many-to-One

These relationships were created in Model view. All relationship are active, and referential integrity is visually verified. This unified model allows us to analyze data

across multiple dimensions like pick up zones, payment method, and trip hour without inconsistenies. No duplicated or mismatched keys were found, therefore no concerns related to relationship issues were raised.

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:

- After checking the datasets, there are no issues related to missing/null values, duplicated entries or outliers.
- There is just one concern related to data format .i.e some columns are unnecessarily summarized. Thus, we need to hange the data format of those columns from 'sum' to "don't' summarize" in the column tool tab to return the data back into right format.
- Remove "search" column as there is 100% response "1".

---

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:

Numerical variables:

- Fare (from Trips)
- Distance (from Trips)

Categorical variables:

- Duration (from Trips, Duration)
- Custid (Trips)
- Driverid (Trips)
- Faremethod (Trips)

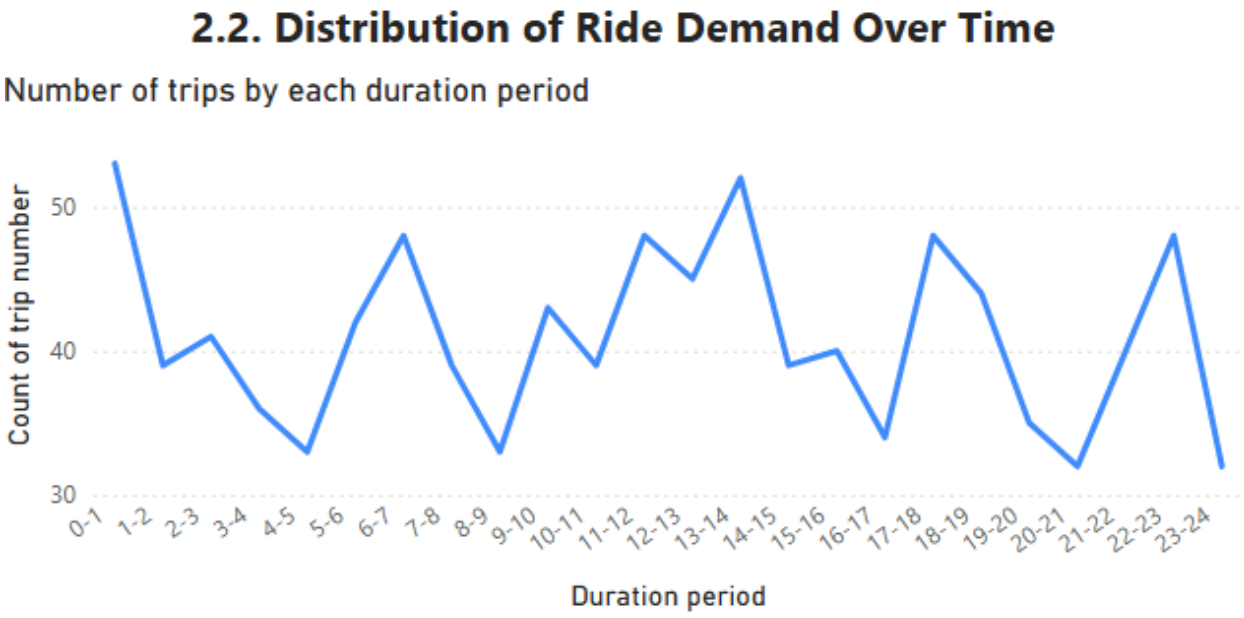
- Loc_from, loc_to (Trips)
- Tripid (Key in multiple tables, categorical for joining)
- Assembly (Assembly)
- Method (Payment method)
- Search_for_quotes, Search_got_estimate, Search_got_quotes, Customer_not_cancelled, driver_not_cancelled, otp_entered (from Trips_Details – binary categories)

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:

To analyse the ride demand, we need to refer to number of trips (count tripid) across all time period throughout the day.



Ride demand fluctuates over time throughout the day. The ride demand picks up significantly during 0-1, 6-7, 13-14, 17-18, 22-23. Low periods are 4-5, 8-9, 16-17, 20-21.

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

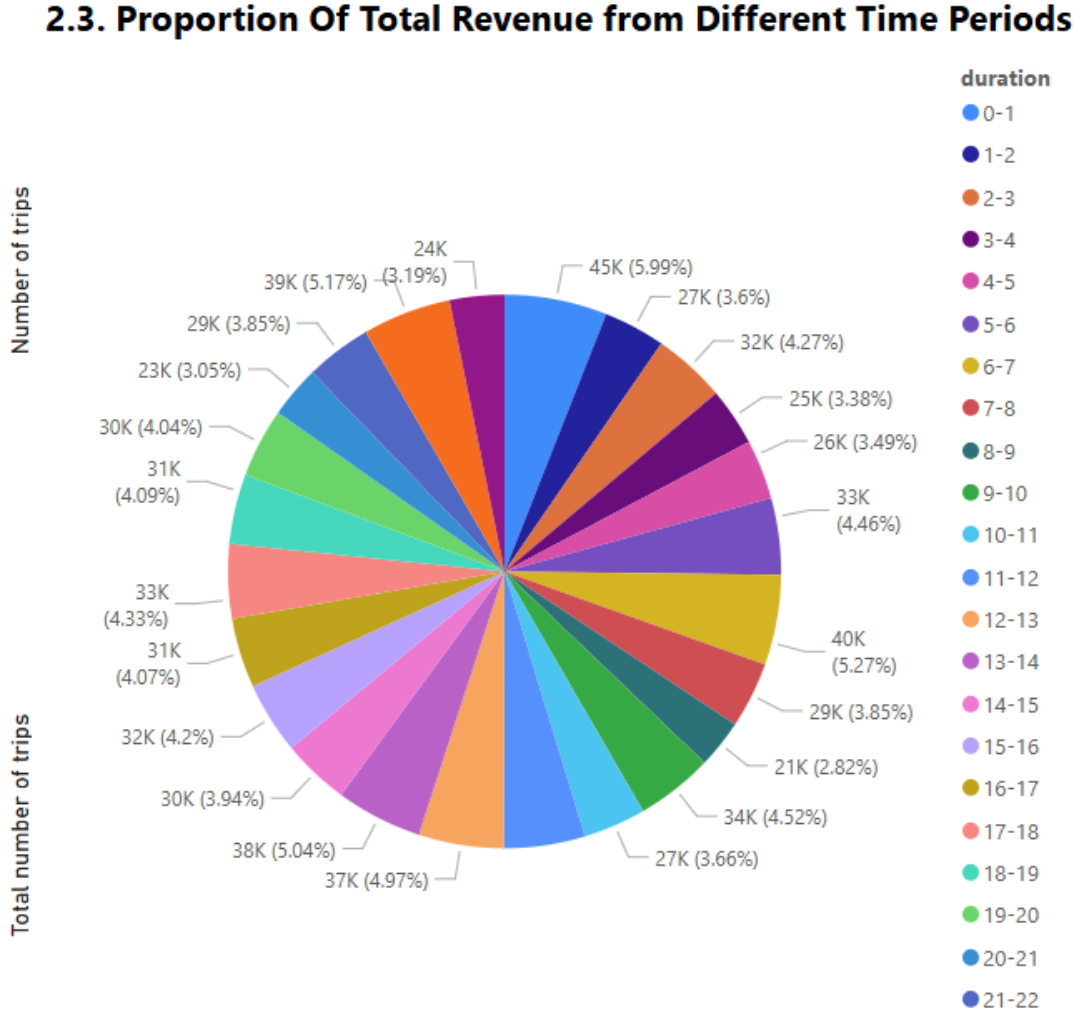
- Calculate the proportion of revenue generated during different time periods and visualise how it contributes to total revenue.

Solution:

Total revenue = total trip fare

Proportion of revenue generated during different time periods = % of total trip fare generated on each time period.

To calculate the proportion of revenue, we use the fare column from Trips table and group the data by duration and visualized by a pie chart.



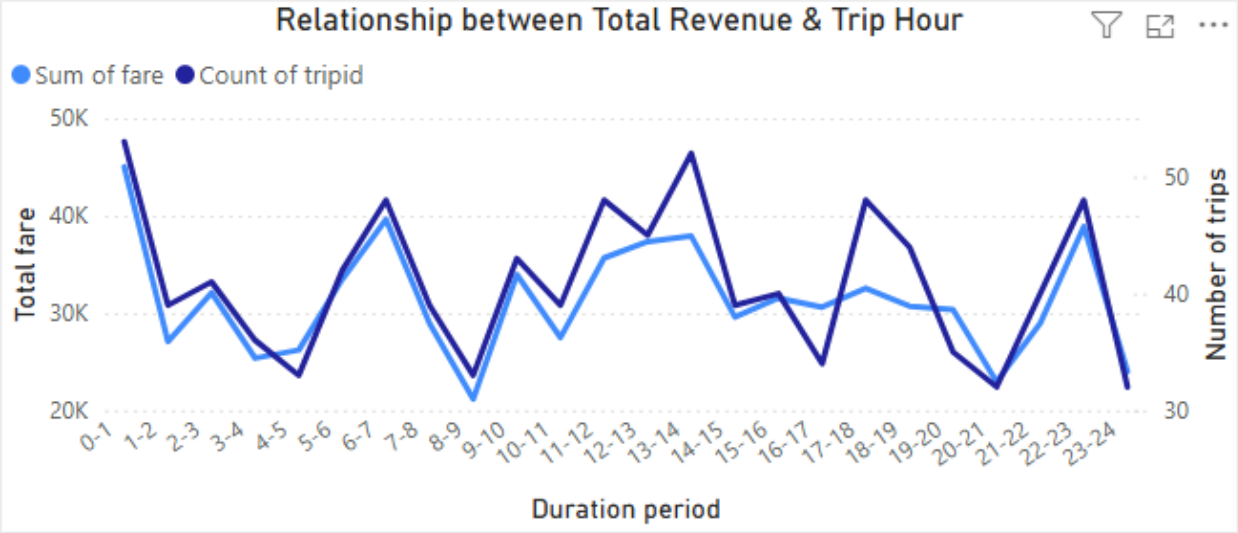
2.4. Explore the Relationship Between Trip Hour and Revenue [3 Marks]

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

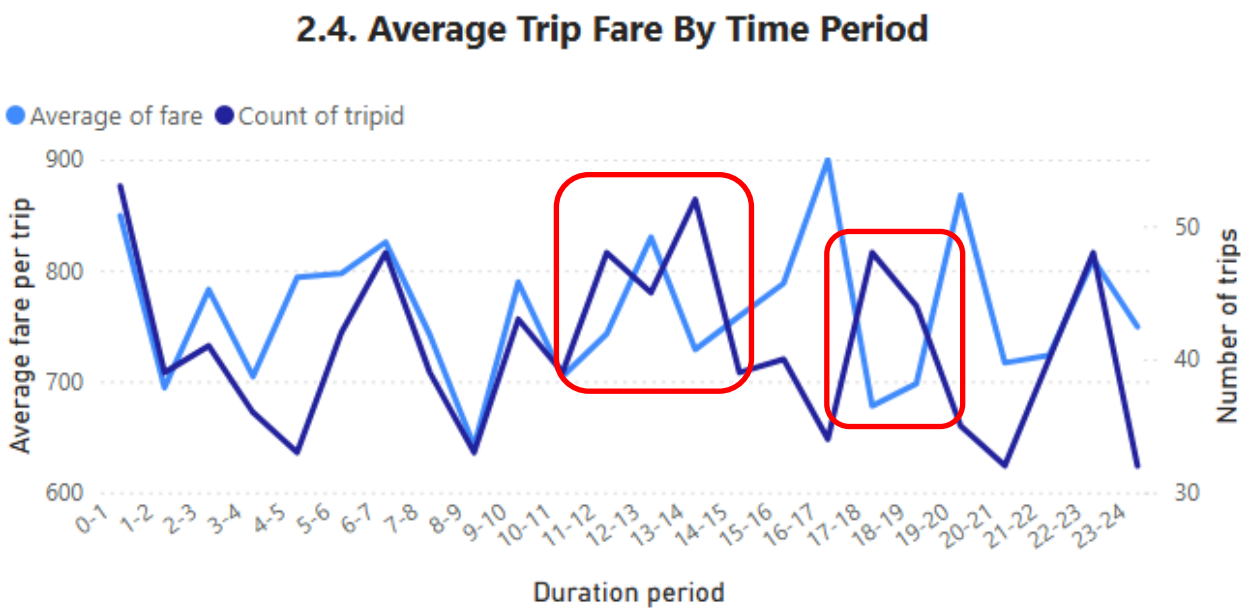
Solution:

To investigate the correlation between trip hour and total fare, we need to run total trip fare with total number of trips to see the relationship between revenue and trip hour.

Then we also need to run the average trip fare with trip number to see if there is any unreasonable pricing in any peak duration.



There is a strong relationship between trip hour and revenue. However, there are some certain trip hours when the revenue is not as high as expected due unreasonable fare charged: 11-14, 17-19 periods.



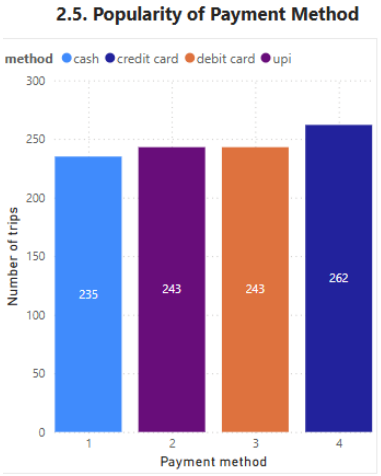
For those periods, though ride demand is quite high however trip fare was charged quite low, leading to lower revenue than expected.

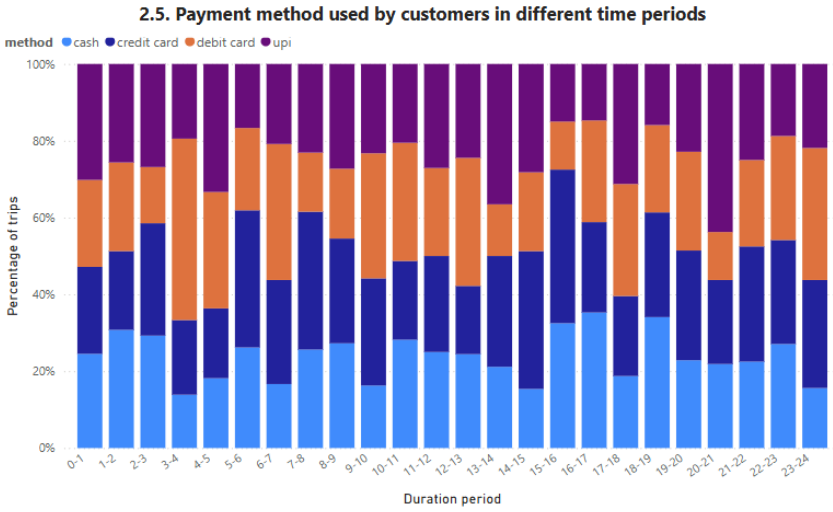
2.5. Examine the Popularity of Different Payment Methods [3 Marks]

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

Solution:

A simple column chart was created to demonstrate the popularity of different payment methods.





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]

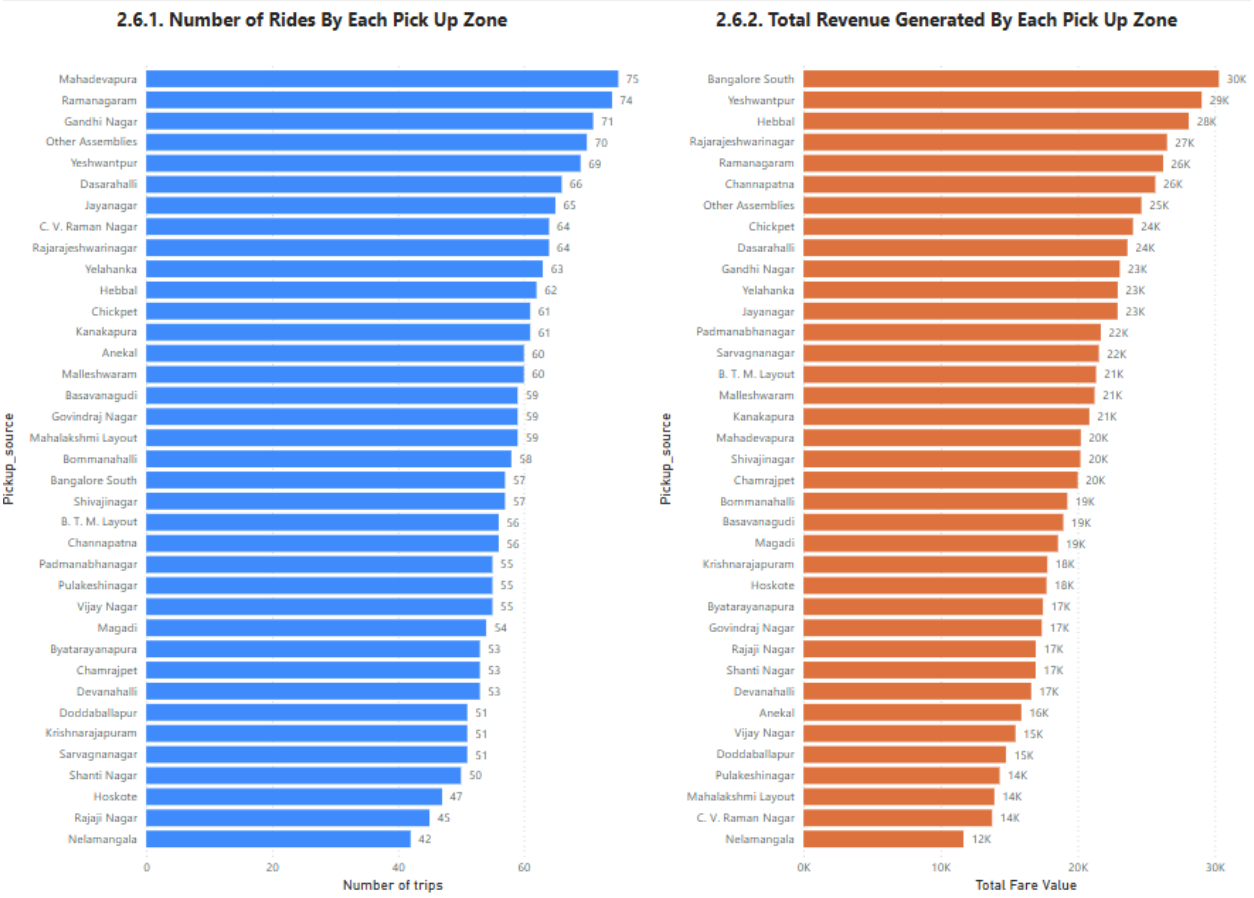
Solution:

To identify high-performing zones, a bar chart were created to demonstrate the number of rides of each pick up zone. We also use filter to pick up top 5 high performing zones in terms of number of trips (change basic filter into Top N filtering).

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

Solution:

We do similar steps as in point 2.6.1 for total fare to pick up the top 5 high performing zones in terms of revenue.



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

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

Solution:

A matrix data table was created to visualise the number of trips booked by different time periods across pickup zones.

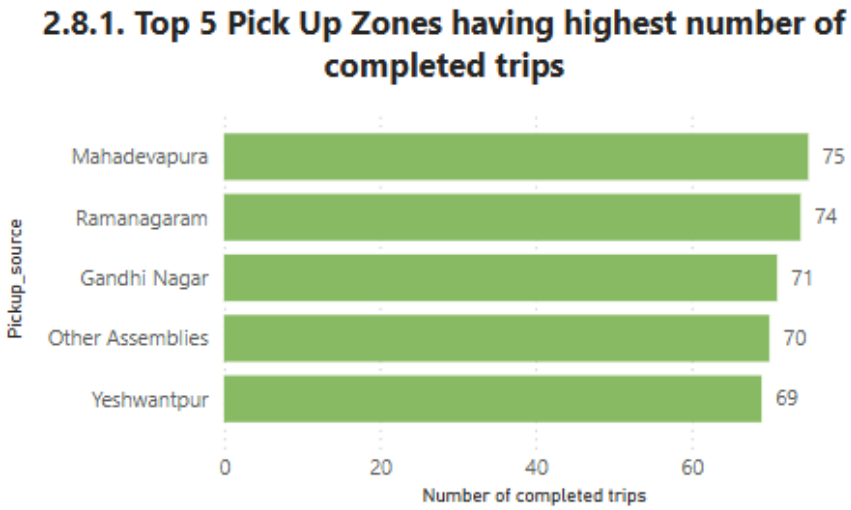
2.7. Ride demand by pick up zones

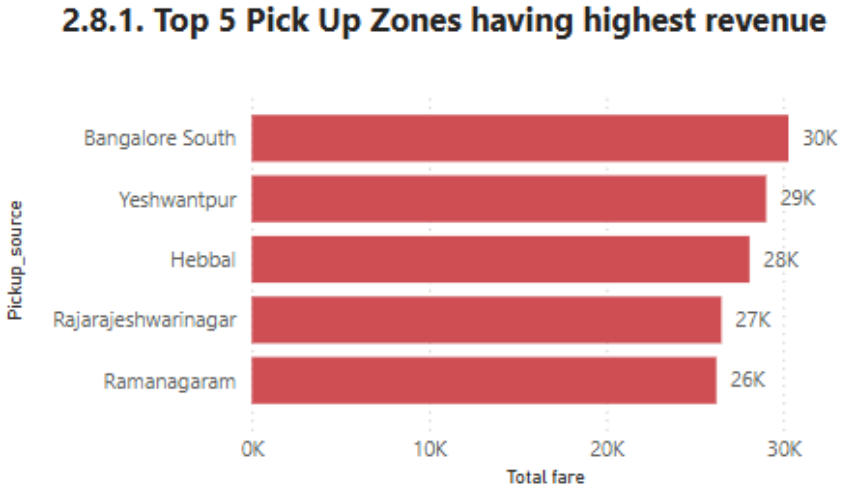
Assembly	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	Total
Anekal	1	1	1	1	1	1	3	2	1	1	1	1		3				1	1	3	1	1	1	1	27
B. T. M. Layout	1	3	2	1	2	1		2	2	2	2		1	1		1		2	1	1		1	2	1	29
Bangalore South	1	1	2	1			2		2	1	1		2		1		4			1	1		1		21
Basavanagudi	1			2	1	2	1		1				1	1		1	1	1		2	1	1	1	1	19
Bommanahalli					2	2	2	1		3	1	3	1	4	1	1	1	1		1		1	1	1	27
Byatarayanapura	1		1	1	1		1		2			2	1	2		1			1	2		2	2	1	21
C. V. Raman Nagar			1	1	1	1		2		2		1	2		2	1	1	3	2				1		21
Chamrajpet	5	2		2		2	2	2		1	2	1		1	5	1	1	4		1	1	1	2		36
Channapatna	1	3	2		1			2	1	1	1			1	2	1		2			2		1	2	23
Chickpet		2		2	2	4	1			2	1		1	1		1		2		1		1	3	1	25
Dasarahalli	1	1		1		3	1	2	1	1		2	2	2	3		1	3	2			2	2	1	31
Devanahalli	2		1	2		2	2	1	1		1		1	2		1		1	1	3	2	2	2	2	29
Doddaballapur		2	1	3		3	2	2	2		1	1	1		1	2	1	2	1			1	1	2	29
Gandhi Nagar	3	1	3	3			1		1	1	1	1	4	3		2	1	1	2				2		30
Govindraj Nagar	1	1	4	1			1	2	2		1	1	1	3			1		3	1		1	4		28
Hebbal	2	1	3		1		1	1		2	3		2	1			2	1				3		1	24
Hoskote	1		2		3	2	4		1	1	1	4	1	2		1		3	2	2	2	2		3	37
Jayanagar	1	2	2			1		2	1	1	1		1			1		4	1		1	3	2	1	25
Kanakapura	1	2	1		1	1	2	1	3	2	1	3	2	4		2		2	2		4				34
Krishnarajapuram	2	1	1	1				1	1	3	1	4		1	3	1	1					1	3	1	26
Magadi	3	1	2	1	1	1	1		1			2	1	1	3				1	2	2	1			24
Total	**53**	**39**	**41**	**36**	**33**	**42**	**48**	**39**	**33**	**43**	**39**	**48**	**45**	**52**	**39**	**40**	**34**	**48**	**44**	**35**	**32**	**40**	**48**	**32**	**983**

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:

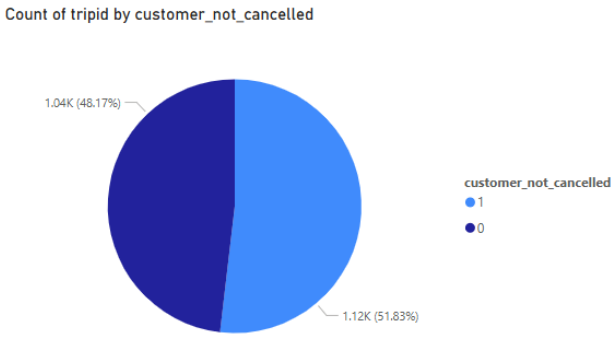


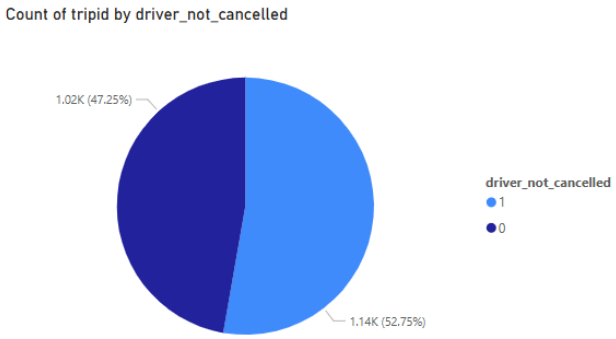


2.9. Basic Analytical Tasks [8 Marks]

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

Solution:



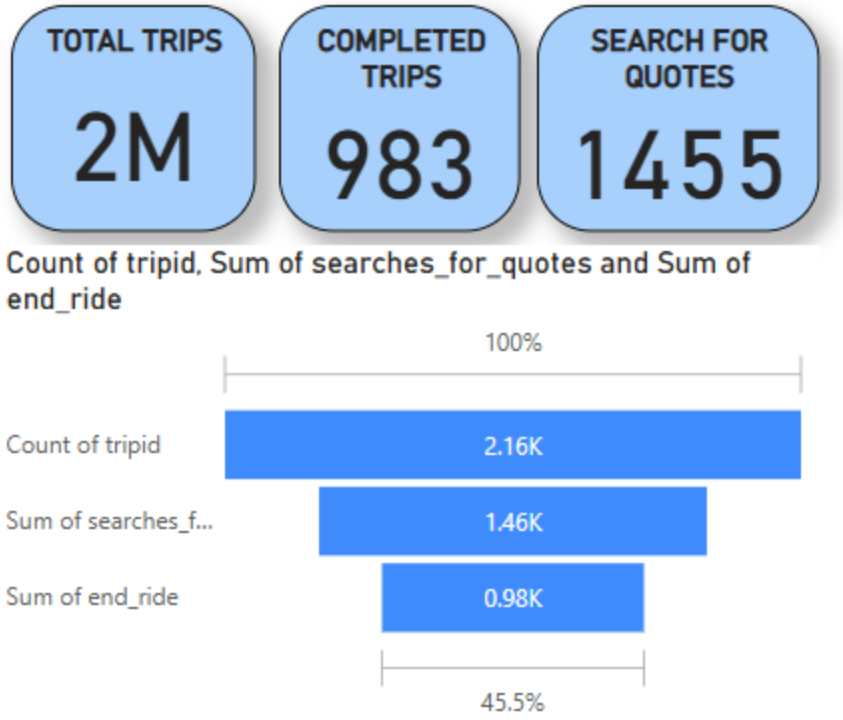


The ratio of cancellation derived from customers and drivers are quite similar.

- **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:

2.9.2. Ratio of completed trips after searching for quotes

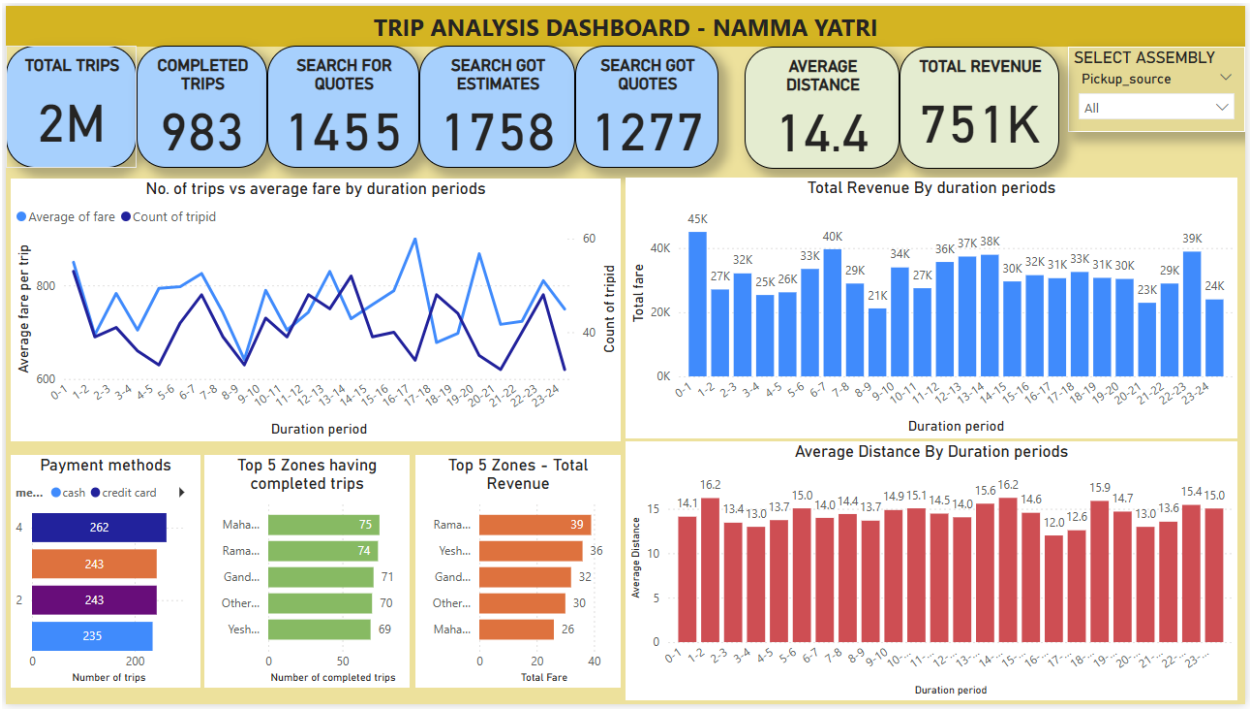


2.10. Create a Parameter and Use Filters [5 Marks]

- 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:

We choose “Assembly” as a parameter for filter across all key metrics to create a dashboard.



3. Conclusion [20 Marks]

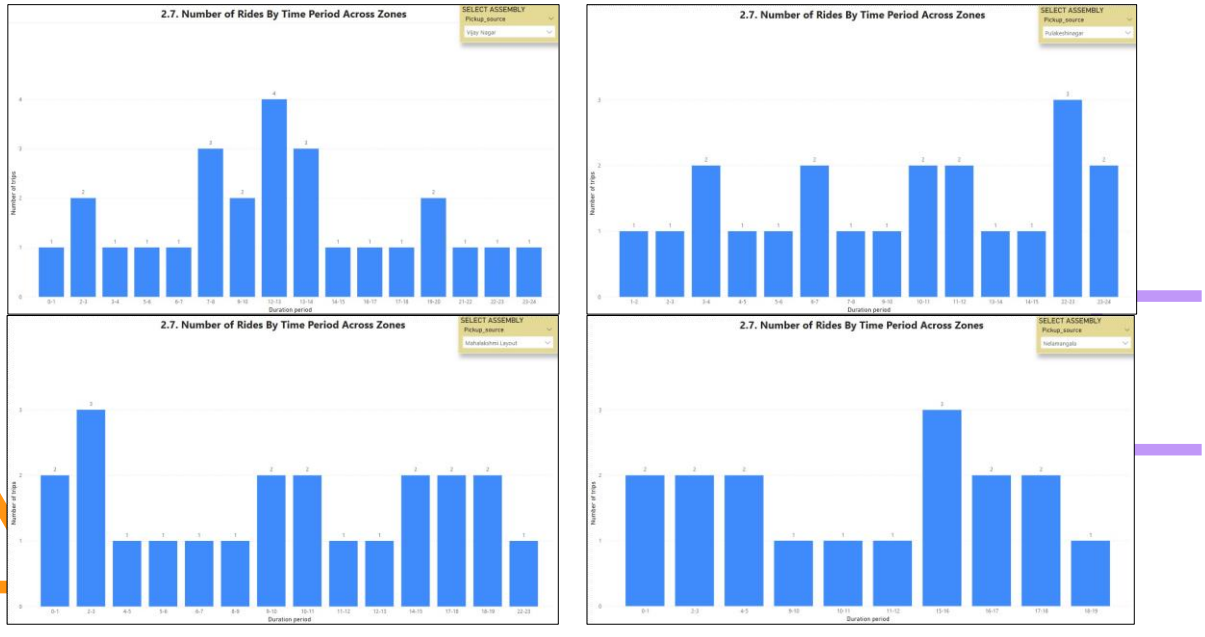
3.1. Recommendations for Operational Efficiency [10 Marks]

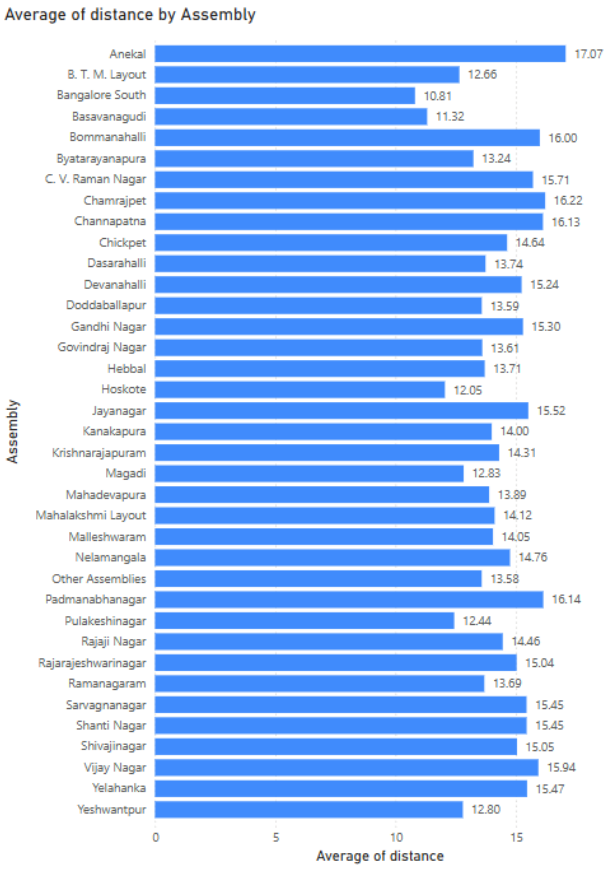
- Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
- This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
- Add supporting dashboards.

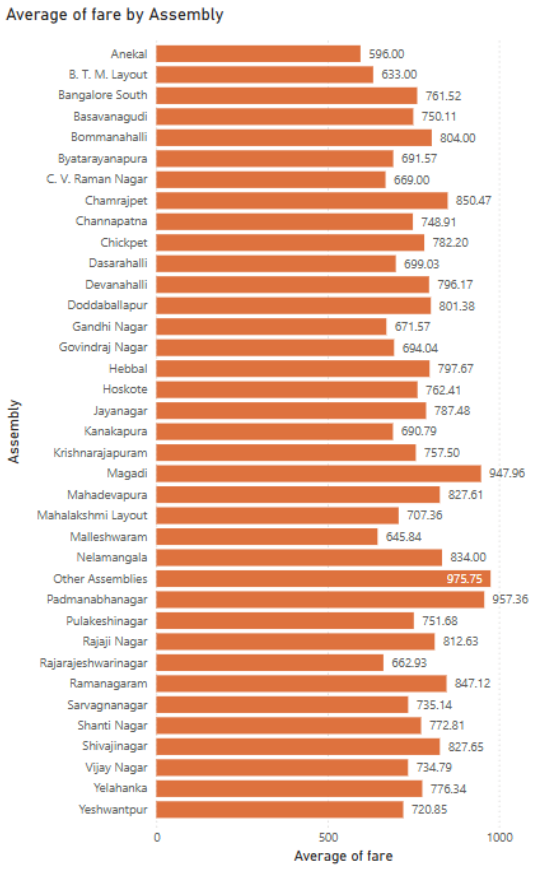
Solution:

- There are certain periods of the day that ride demand increase significantly, thus, the company should boost more drivers to meet that hide demand.
- The peak demand of different locations are quite different, there is a need to optimize the resource efficiently to either capture the high demand in peak time and allocate drivers across locations reasonably.
- Review pricing strategy across locations to ensure a fare charge for drivers or incentivize drivers in lower charge locations to encourage drivers not to cancel the trip.
- Pricing strategy is also need to adjusted to ensure a competitiveness and fairness across locations to trigger the service usage of low demand locations where we observes significantly higher fare charged i.e. Doddaballapur.

Ride Time Periods Across Zones (1)







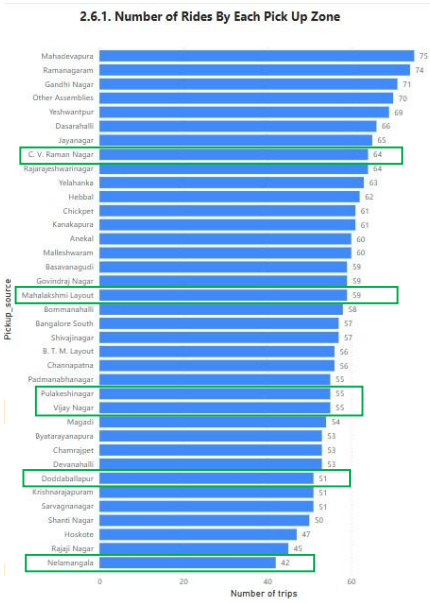
3.2. Marketing and Operational Strategy Improvements [10 Marks]

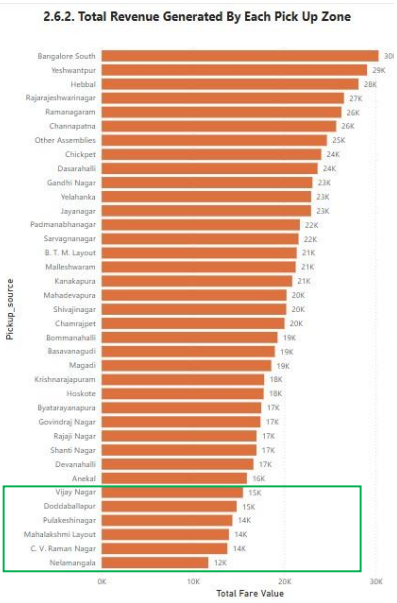
- Suggest improvements to Namma Yatri's marketing or operational strategies based on your analysis.
- Recommendations could involve promotional efforts, driver incentives, or regional targeting to increase customer satisfaction and service efficiency.
- Add supporting dashboards.

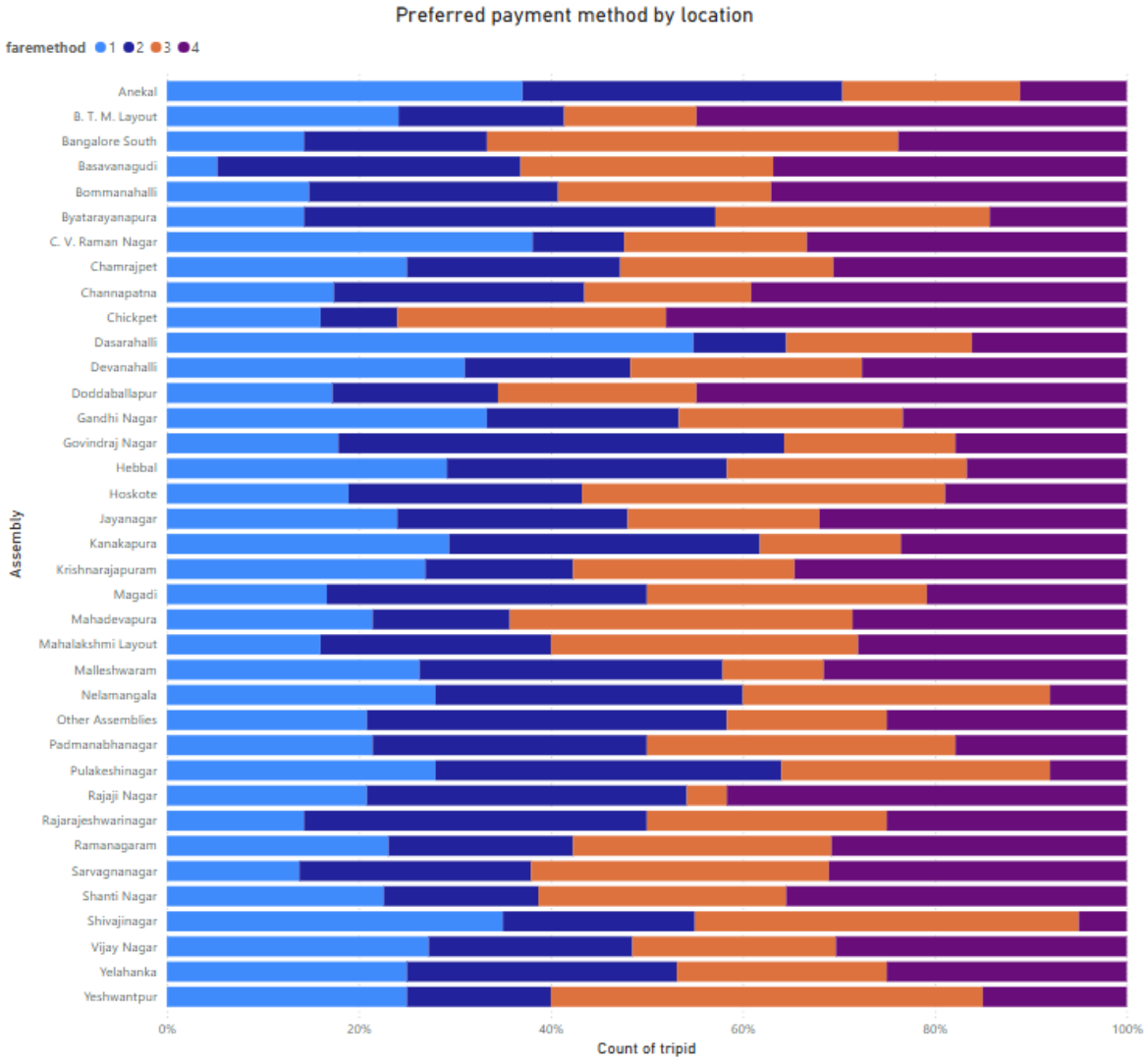
Solution:

- Launch more marketing campaign in the low demand locations (where having lower number of trips i.e. Vijay Nagar, Pulakeshinagar, Mahalashmi Layout & Nelamangaia, Doddaballapur) to enhance the brand awareness about Namma Yatri company in customers' mind in those location.
- Offer good price to trigger first customers trying service of the company.
- Given credit card is the most popular payment method, Namma Yatri can offer some promotion campaigns to offer better price for customers to trigger demand in low period of the day.
- Different locations have different preferred payment method. Depending on customers' preference of each location, we provide relevant offers to trigger customers book our ride service.

Low performing zones in total revenue






Preferred payment method by location
faremethod 1 2 3 4
Anekal
B. T. M. Layout
Bangalore South
Basavanagudi
Bommanahalli
Byatarayanapura
C. V. Raman Nagar
Chamrajpet
Channapatna
Chickpet
Dasarahalli
Devanahalli
Doddaballapur
Gandhi Nagar
Govindraj Nagar
Hebbal
Hoskote
Jayanagar
Kanakapura
Krishnarajapuram
Magadi
Mahadevapura
Mahalakshmi Layout
Malleshwaram
Nelamangala
Other Assemblies
Padmanabhanagar
Pulakeshinagar
Rajaji Nagar
Rajarajeshwarinagar
Ramanagaram
Sarvagnanagar
Shanti Nagar
Shivajinagar
Vijay Nagar
Yelahanka
Yeshwantpur
Assembly
0%
20%
40%
60%
80%
100%
Count of tripid