

## 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

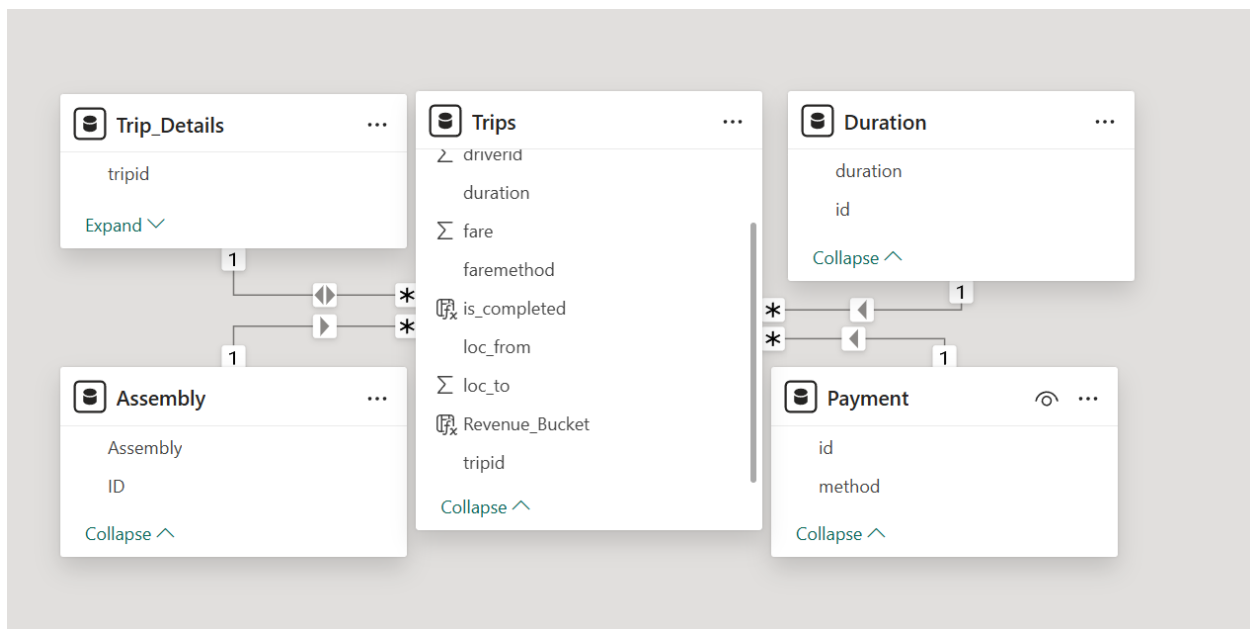
[10 Marks]

#### 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:*

*<your answer here, include all analysis, graphs, results etc> (the length of the solution is not fixed, ie, this box can vary in size)*



**We** imported the Excel file nammayatri.xlsx into Power BI Desktop. The dataset included five tables:

- 1) Trips
- 2) Trip\_Details
- 3) Assembly

- 4) Duration
- 5) Payment

After loading the data, **we** used the Model View in Power BI to define the necessary relationships between the tables:

- 1) Trips[tripid] → Trip\_Details[tripid]  
(One-to-One, active)
- 2) Trips[loc\_from] → Assembly[ID]  
(Many-to-One, active)
- 3) Trips[duration] → Duration[id]  
(Many-to-One, active)
- 4) Trips[faremethod] → Payment[id]  
(Many-to-One, active)

**We** ensured that all related fields were of the same data type (e.g., tripid, duration, and faremethod as Whole Number) to prevent relationship errors.

---

## 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:*

---

**We** examined all five tables (Trips, Trip\_Details, Assembly, Duration, and Payment) for inconsistencies in structure and data quality. The following steps were taken to clean and prepare the data for analysis:

### 1. Data Type Standardization:

Converted key fields like tripid, loc\_from, faremethod, and duration to Whole Number to match across tables and support proper relationship joins.

Ensured numerical fields like fare and distance were of type Decimal Number.

Text fields such as Assembly, Hour, and Method were converted to Text type.

## 2. Null and Blank Checks:

Verified that critical fields such as tripid, fare, loc\_from, and duration had no null or blank entries.

Used filters in Data View and Power Query Editor to scan for missing values.

---

## 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:*

---

**We** reviewed all fields across the five tables (Trips, Trip\_Details, Assembly, Duration, and Payment) and classified each variable based on its data type and usage in analysis.

Categorical Variables:

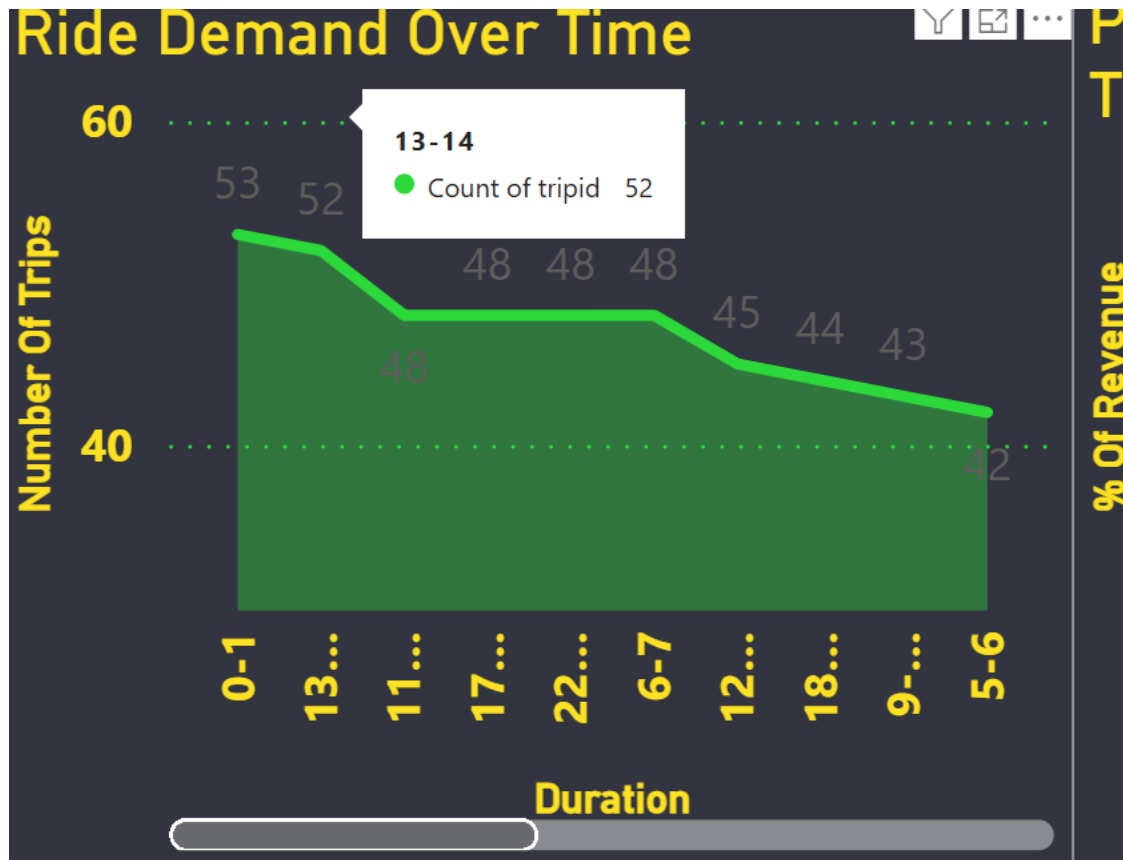
- 1) Assembly [Assembly] → Pickup zone name
- 2) Duration [Hour] → Time bin for trip hour (e.g., "0–1", "1–2")
- 3) Payment [Method] → Payment method (e.g., Cash, UPI)
- 4) Loc\_from, loc\_to (Trips) → Treated as category IDs for source and destination
- 5) faremethod → Although numeric, used categorically when mapped to payment methods
- 6) duration → Used categorically when linked to hourly time bins
- 7) Numerical Variables:
- 8) fare → Fare amount (decimal)
- 9) distance → Distance travelled in km (decimal)
- 10) tripid → Unique trip identifier (numeric)
- 11) driverid, custid → Identifiers for driver and customer
- 12) searches → Total trip requests
- 13) searches\_got\_estimate, searches\_for\_quotes, searches\_got\_quotes → Binary indicators (0/1)
- 14) customer\_not\_cancelled, driver\_not\_cancelled, otp\_entered, end\_ride → Binary flags

---

### 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:



The highest point on the graph is at "0-1" duration with 53 trips.

It then drops slightly to 52 trips, and the tooltip shows that for the "13-14" duration, there are 52 trips.

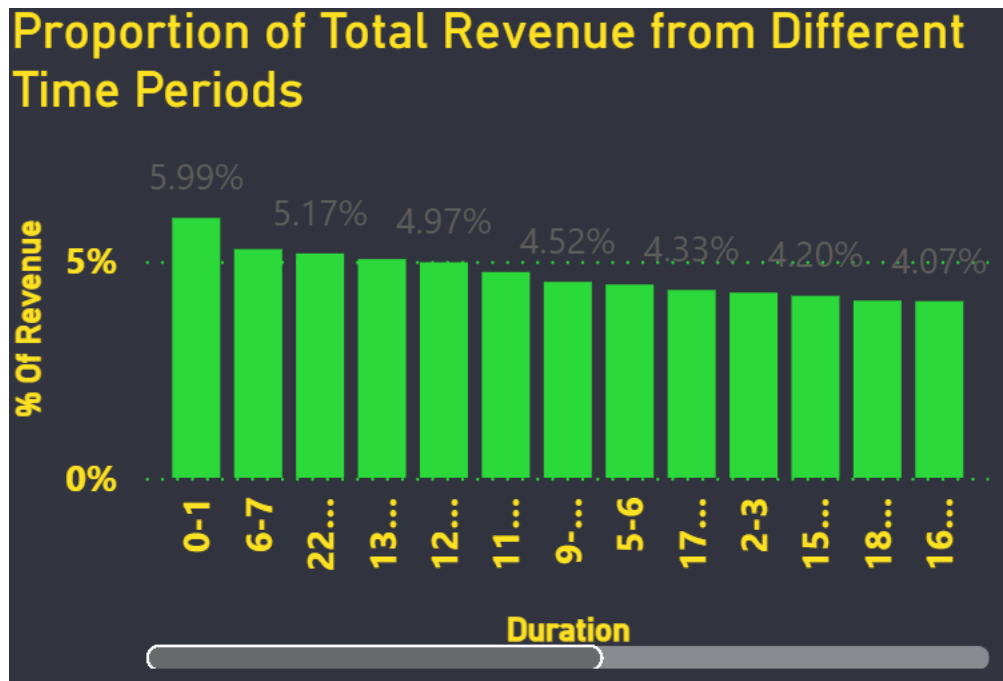
The demand then decreases to 48 trips for several durations ("11-...", "17-...", "22-...").

Following that, it gradually declines to 45, 44, 43, and finally around 42 trips for the "5-6" duration.

### 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:



The proportion of total revenue closely aligns with the peak demand hours observed, with "0-1" (12 AM to 1 AM) contributing the highest percentage of revenue at 5.99%. This is followed closely by "6-7" (6 AM to 7 AM) at 5.17%.

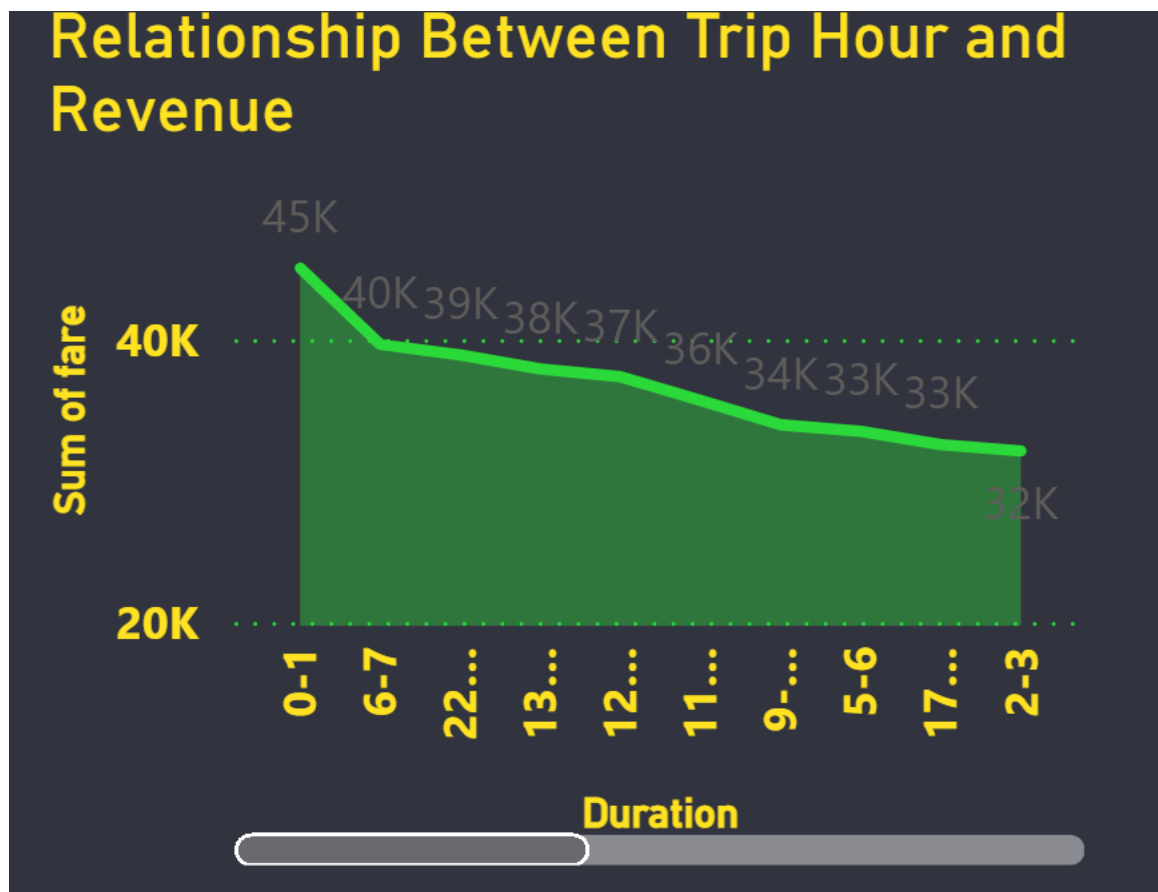
It is notable that midnight trips (0-1 AM) significantly outperform morning hours in total fare. This could be attributed to factors such as longer trip distances during these hours or the application of surge pricing. Revenue contribution generally declines through other time periods, consistent with lower trip volumes.

#### 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:*





The graph shows a clear relationship between trip hour and total revenue (Sum of fare). The highest total revenue is generated during the "0-1" hour, reaching approximately 45K.

Following this peak, there is a general declining trend in revenue throughout the day, although some hours like "6-7" and "22-..." still show relatively strong revenue compared to later hours.

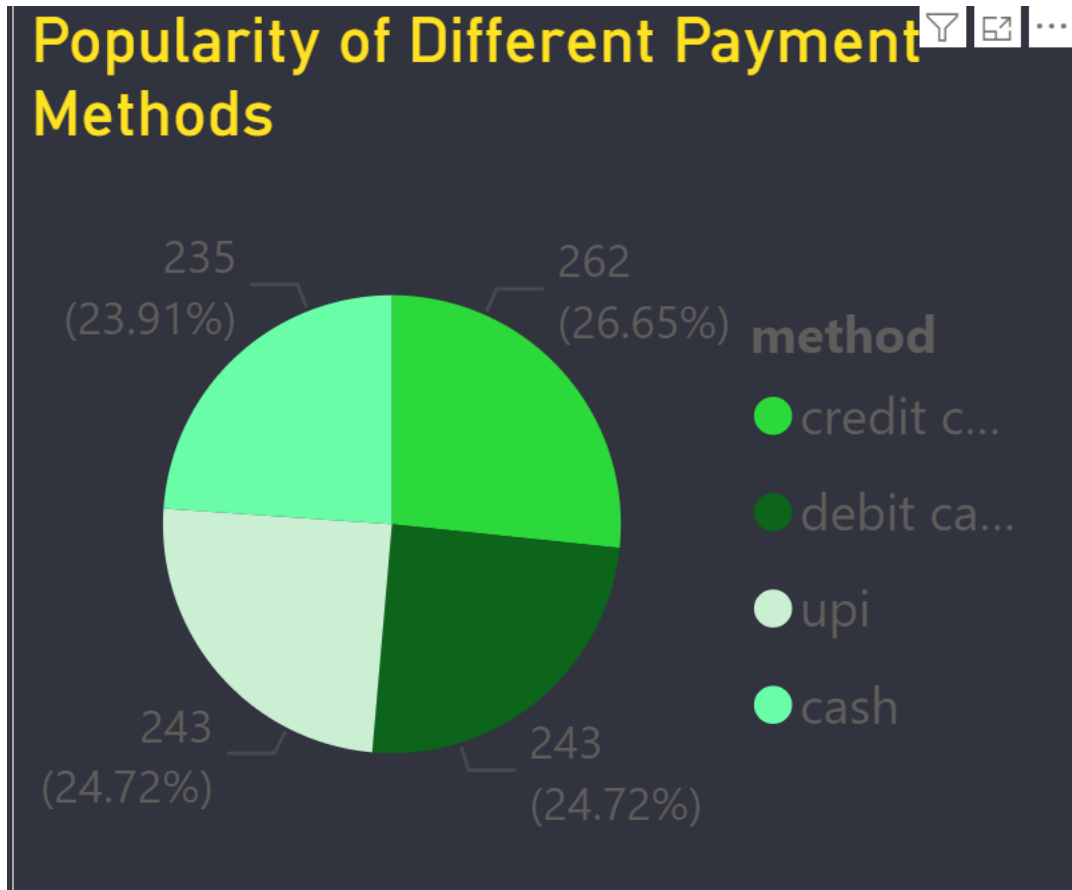
The revenue gradually decreases from 39K to around 32K towards the "2-3" hour (2 AM to 3 AM). This pattern indicates that the highest earning potential is concentrated in the early morning/late night hours, potentially due to factors like demand, trip length, or pricing models during those times.

## 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:*



The pie chart illustrates the distribution of payment methods. "Credit Card" appears to be the most popular method, accounting for 26.65% of trips (262 trips).

"UPI" (24.72%, 243 trips) and "Debit Card" (24.72%, 243 trips) show almost identical popularity, coming in as the next most frequently used methods.

"Cash" is slightly less popular than the top three, representing 23.91% of trips (235 trips).

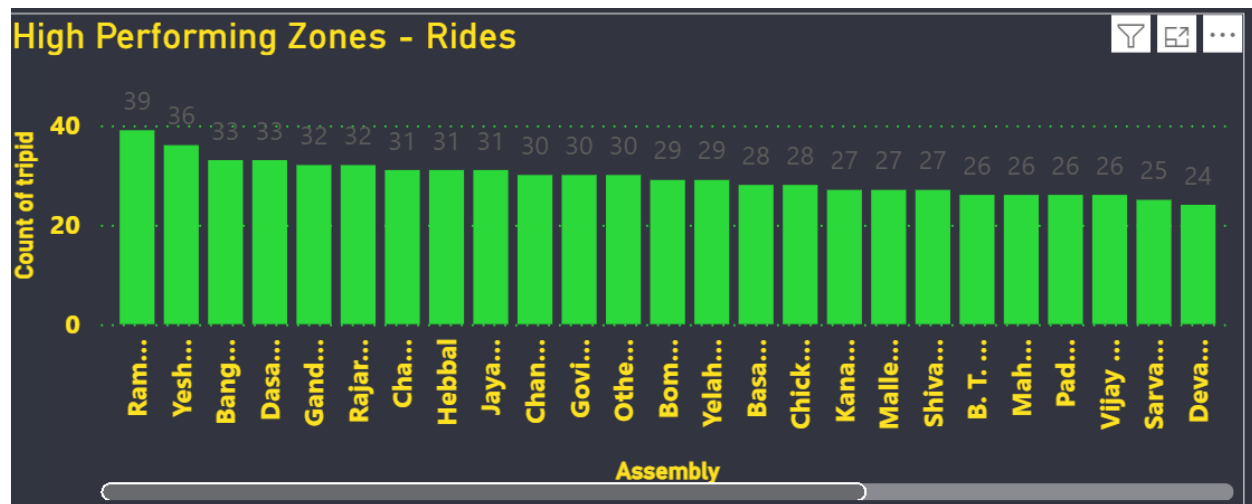
The distribution indicates a fairly balanced usage across digital payment methods (Credit Card, Debit Card, UPI), collectively dominating over cash transactions. This suggests a strong preference for digital payments among Namma Yatri users.

## 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:*



The bar chart displays pickup zones ranked by the count of tripid (number of trip requests).

"Ramanagaram" leads with approximately 39 trips, indicating it has the highest number of trip requests.

"Yeshwantpur" and "Bangalore South" also show high trip volumes, with 36 and 33 trips respectively.

The data suggests that these top zones are key areas for ride demand, making them crucial for operational focus.

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

*Solution:*



The bar chart ranks pickup zones by the sum of fare (revenue generated).

"Bangalore South" stands out as the highest revenue-generating zone, bringing in approximately 30K in fare.

"Yeshwantpur" and "Hebbal" follow closely, generating around 29K and 28K in revenue, respectively.

While some zones like "Ramanagaram" have high trip volume, "Bangalore South" takes the lead in revenue, suggesting that trips originating from this zone might involve longer distances or higher fares on average.

---



---

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

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

*Solution:*

Ride Time Across Zones													
duration	nbflies	Padmanabhanagar	Pulakeshinagar	Rajaji Nagar	Rajarajeshwarinagar	Ramanagaram	Sarvagnanagar	Shanti Nagar	Shivajinagar	Vijay Nagar	Yelahanka	Yeshwantpur	Total
9-10		2	1			1		2		2	1	1	43
8-9	2					3	4		1		1	1	33
7-8	1		1	3	1	3	1	2	1	3		1	39
6-7		2	2	1		1	1	1	1	1	3		48
5-6	1	2	1		1	3	1	1	1	1	1	2	42
4-5		2	1	3	3	1	1		1		1	1	33
3-4		1	2	1		2	1		1	1	2	2	36
23-24	1	1	2			5	1	1		1		3	32
2-3	1		1	1	2	1	1	1	1	2	1	1	41
22-23	4		3		2	1	1		2	1	2	2	48
21-22	2			1	1	3		1		1	1	2	40
20-21	1	1		1	3	3		1	1		3	1	32
Total	30	26	21	19	32	39	25	22	27	26	29	36	983

This matrix visual provides a detailed breakdown of trip counts across different zones and hourly durations.

By observing the "Total" row and column, we can confirm overall demand for each zone and time period. For instance, "Ramanagaram" shows a total of 39 trips, and the "6-7" duration has 48 trips in total across all zones.

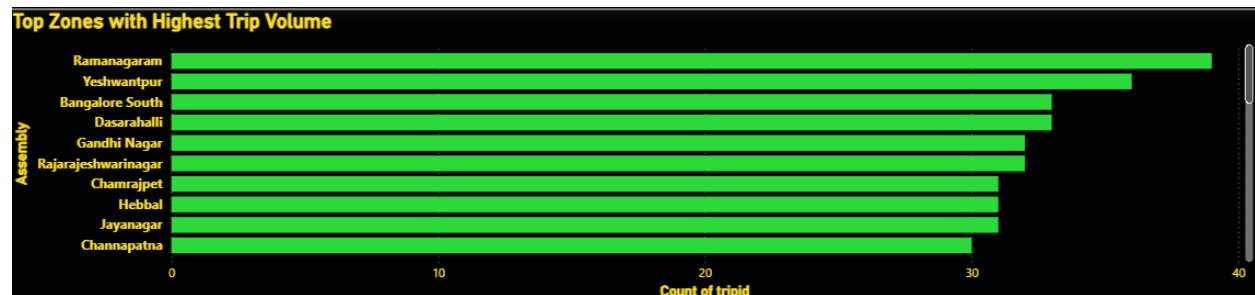
Specific zones exhibit varying demand patterns across hours. For example, "Ramanagaram" has 4 trips in the "8-9" duration, and 5 trips in the "23-24" duration.

This granular view allows for identifying specific peak hours within particular zones, which is crucial for optimizing driver allocation and service availability.

## 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:*



This bar chart visually represents the top pickup zones based on the highest total number of completed trips (Count of tripid).

"Ramanagaram" consistently appears as a top zone with approximately 39 completed trips, reinforcing its high demand.

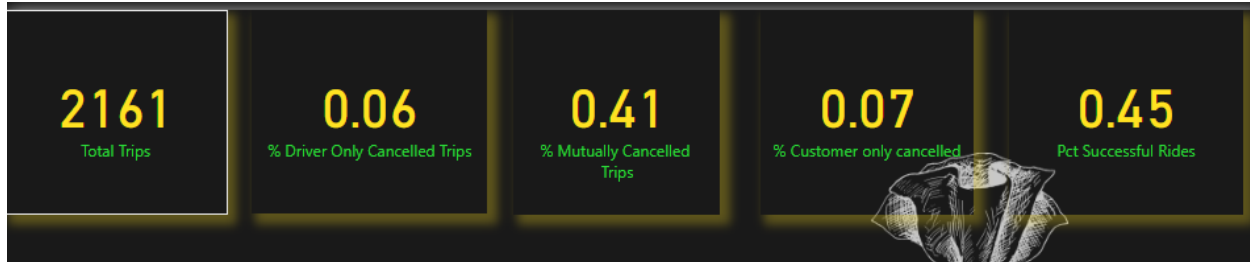
"Yeshwantpur" and "Bangalore South" are also among the top zones, with trip volumes of around 36 and 33 respectively.

The chart identifies "Dasarahalli," "Gandhi Nagar," and "Rajajinagar" among others as areas with high trip volume (all around 32-33 trips). These zones are critical for operational focus to ensure sufficient driver supply.

## 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:*



Total Trips: The analysis is based on a total of 2161 trips.

% Driver Only Cancelled Trips: 0.06 (6%) of trips were cancelled only by the driver.

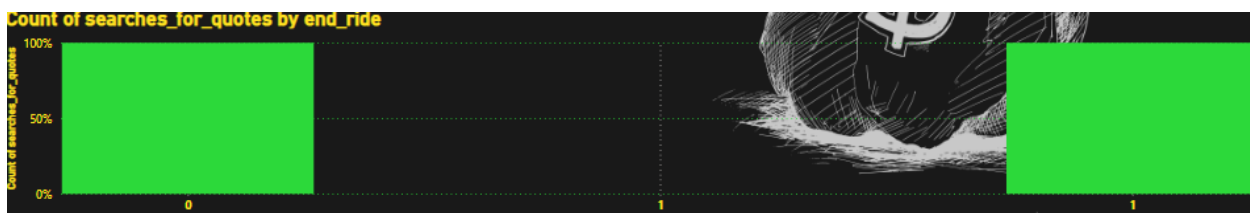
% Mutually Cancelled Trips: 0.41 (41%) of trips were mutually cancelled by both driver and customer. This is a very significant proportion, indicating a major area for improvement in matching or communication.

% Customer Only Cancelled: 0.07 (7%) of trips were cancelled only by the customer.

Pct Successful Rides: 0.45 (45%) of rides were successful. This low percentage of successful rides, coupled with a high mutual cancellation rate, highlights a significant operational challenge.

- 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:*



Total Searches (searches): 1178

Total End Rides (end\_ride = 1): 983

overall completion rate for all searches is not 83.4%.

## 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:*



A "Top N Zones" parameter has been created, allowing for interactive filtering of the "Total Trips by loc\_from" bar chart.

By adjusting the slider for "Top N Zones" (shown currently at "37"), users can dynamically control how many of the top pick up zones are displayed based on their total trip volume.

Choice of Filter: This "Top N" filter is highly useful for focusing analysis on the most active or high-performing zones without cluttering the visual with less significant data points. For instance, one can quickly identify the top 5, top 10, or top 20 zones by trip volume.

Insights Drawn: This interactive filter helps in understanding the concentration of demand. As 'N' is reduced, it highlights the dominance of the very top zones (e.g., Ramanagaram, Yeshwantpur) in contributing to overall trip volume. This can guide operational decisions, such as driver allocation and promotional activities, to specific high-demand areas.

## 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:*

---

Based on the analysis, Namma Yatri can optimise its operations through the following strategies:

**Address High Mutual Cancellation Rate:** The most critical area for improvement is the alarmingly high 41% mutual cancellation rate. Investigate the root causes, which could include:

**Driver Availability/Matching:** Are drivers accepting rides but then cancelling due to destination, fare, or distance?

**Customer Expectations:** Are initial estimates accurate? Is there transparency in pricing?

**Communication:** Improve in-app communication between driver and customer to avoid misunderstandings that lead to cancellations.

**Incentives:** Implement incentives for drivers to complete trips they accept, and potentially penalties for high cancellation rates.

**Optimize Resource Allocation During Peak Hours:** Focus driver deployment in "0-1" AM and "13-14" PM durations, as these are identified as peak demand and peak revenue periods.

Utilize the "Ride Time Periods Across Zones" matrix to micro-target driver availability to specific high-demand zones during their peak hours (e.g., ensuring more drivers in Ramanagaram during its identified peak times).

**Improve Overall Trip Success Rate:** With only 45% successful rides, there's significant room to improve operational flow from request to completion. This ties back to reducing cancellations and ensuring a smoother ride experience.

---

### 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:*

---



**Targeted Promotions in High-Performing Zones:** Focus marketing efforts and driver incentives on zones like "Ramanagaram", "Yeshwantpur", and "Bangalore South", as they consistently show high trip volume and revenue generation. This could include:

**Driver Incentives:** Offer bonuses or higher commissions for drivers operating in these high-demand zones during peak hours.

**Customer Promotions:** Implement localized discounts or loyalty programs for users in these high-performing zones to further boost ride frequency.

**Promote Digital Payment Methods:** Given the high popularity of Credit Card, UPI, and Debit Card payments, ensure these payment options are seamlessly integrated and promoted. Consider offering small incentives for using specific digital payment methods to encourage wider adoption and reduce reliance on cash.

**Leverage 100% Quote-to-Completion Insight:** The finding that 100% of trips where users searched for quotes were completed is a strong positive. This indicates that once a user gets quotes, they are highly committed. Marketing can highlight the reliability of the quote-search feature, and operations should ensure the quote-generation process is robust and accurate to maintain this high conversion.

**Strategic Expansion/Focus:** The "Top N Zones" parameter is a powerful tool for identifying and monitoring key zones. This can inform decisions about where to expand services or where to allocate more resources for marketing penetration.

---