

## 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	Assembly	Specific assembly zone name		
Duration	duration_id	Unique identifier of time periods		
Duration	duration	Hour of trip (e.g., "0-1" for 12 AM to 1 AM)		
Payment	id	Unique identifier		
1 ayment	method	Payment method (e.g., Cash, UPI, Credit Card)		
	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, o = 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)		
	tripid	Links to Trip Details		
Assembly Duration Payment Trip Details	faremethod	Payment method ID, links to Payment table		
	fare	Fare amount		
Tring	loc_from	Location ID of source		
TTIPS	loc_to	Location ID of destination, links to Assembly table		
Trips	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]

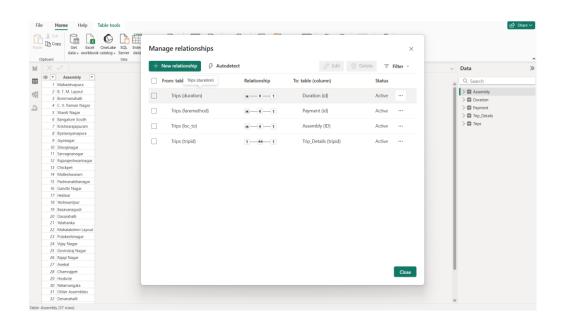
- Import the Namma Yatri dataset into Tableau/Power BI.
- Ensure that you correctly join all tables to create a unified dataset for analysis.
- Verify the relationships between different tables and confirm that data from various sources is properly aligned for accurate insights.

#### Solution:

Steps taken are as follows:

- 1. The Namma Yatri dataset was imported successfully with all its tables-('Trips', 'Trip\_Details', 'Assembly', 'Payment' and 'Duration') in PowerBl.
- 2. Each table was carefully examined to identify the appropriate primary and foreign keys to establish accurate relationships to ensure a unified data model for analysis:
- Trips [tripid] → Trip\_Details [tripid]
- Trips [faremethod] → Payment [id]
- Trips [duration] → Duration [duration id]
- Trips [loc to] → Assembly [Assembly ID]
- 3. Validated each relationship by checking cardinality and referential integrity.
- 4. Checking sample records to confirm data alignment (eg. Cross-table counts like trips with payment method and location).
- 5. Verified data integrity No duplicates and broken line.





## 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:

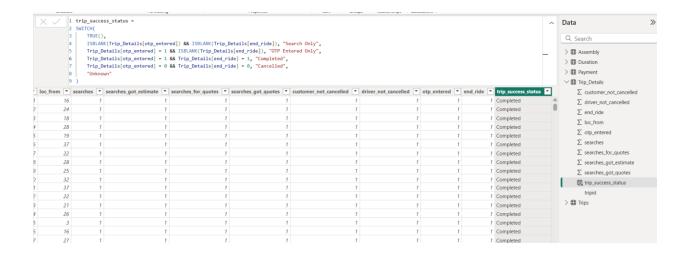
During data preparation, several inconsistencies were identified and addressed to ensure clean and reliable analysis:

## 1. Negative or zero values:

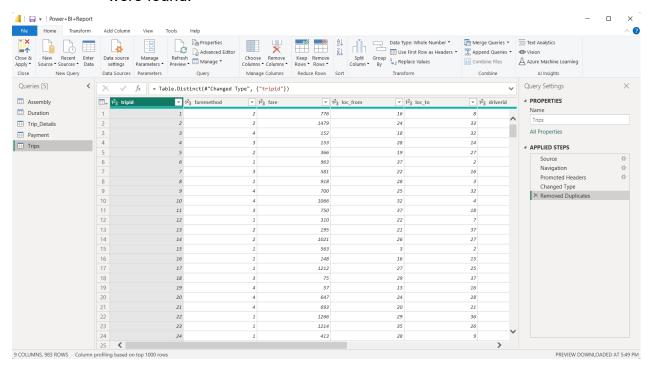
- Confirmed no negative fares or zero-distance trips remained.
- Ensured all payment methods had >0 record.
- **2. Missing or invalid values** in key columns like fare, distance, or duration were checked and none were found.
- **3. Rename Columns for Clarity** improves readability in visuals. Rename cryptic column names in table **Trips** like:
- loc\_from → pickup\_zone\_id
- faremethod → payment\_method\_id



- **4. Calculated columns** were created to support further analysis:
- Trip\_success\_status from flags like end\_ride, otp\_entered



**5. Duplicate tripid entries** were checked to ensure data integrity—none were found.



The dataset was structured for consistency, with irrelevant or inconsistent rows removed to ensure accurate, insight-driven analysis.



## 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:

The variables from all tables were classified as follows:

## ❖ Trips Table

Categorical Variables	Numerical Variables
tripid(identifier)	fare
pickup_zone_id	distance
loc_to (Drop Zone)	duration
pay_method_id	

## ❖ Trip\_Details Table

Categorical Variables	Numerical Variables
tripid (identifier)	None
otp_entered (0/1)	
end_ride (0/1)	
Derived: trip_success_status	
searches_for_estimate (binary (0/1))	
searches (binary (0/1))	
searches_for_qoutes (binary (0/1))	
searches_got_qoutes (binary (0/1))	
driver_not_cancelled (binary (0/1))	
customer_not_cancelled (binary (0/1))	



Above variables from **Trip\_Details** are technically binary categorical variables, though they're encoded numerically (0/1). They might be **numerical** as data type unless they are used for conversion metrics.

## Assembly Table

Categorical Variables	Numerical Variables
ID (Zone ID)	None
Assembly (Zone Name)	None

## ❖ Payment Table

Categorical Variables	Numerical Variables		
Id (identifier for pickup_zone_id)	None		
method (payment method, e.g., Cash, UPI, etc.)			

### Duration Table

Categorical Variables	Numerical Variables		
ld (identifier-duration)	None		
duration ( <i>Time periods (e.g., "0-1", "1-2", etc.))</i>			

#### Notes:

- 1. **ID Fields** are technically numerical but serve as identifiers rather than quantitative measures.
- 2. While not explicitly in the data, we can derive numerical counts from the categorical flags (e.g., total rides completion)
- 3. Duration field can be treated as ordinal categorical since they have a natural order.
- 4. No continuous numerical variables were found like "Trip distance" or "fare" amount in above tables.



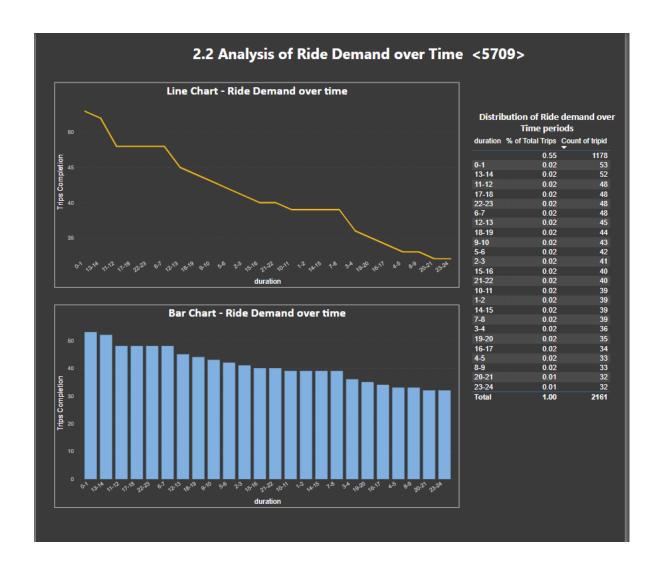
## 2.2. Analyse Ride Demand Over Time [3 Marks]

- 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:

#### **Peak Demand Period:**

- Late Night (12 AM 1 AM): Highest ride demand with duration "0-1".
- Afternoon (1 PM 2 PM): Second highest Duration at "13-14".
- Morning (6-7 AM and 11-12 AM): Morning Ride peak at "6-7 and 11-12".
- Evening (5 PM 6 PM): Evening Ride Peak at "17-18".





#### **Lowest Demand Period:**

- Late Night (11 PM 12 PM): Lowest Ride Demand at duration "23-24".
- Early Morning (4-5 AM and 8-9 AM): Duration at "4-5" and "8-9"
- Late Evening (9 PM 10 PM)

### **Conclusion:**

- 1. For Peak hours:
  - Increase drivers' availability.
  - Use targeted promotion (e.g. discounts for off peak rides)
  - Dynamic pricing.

#### 2. For Low-Demand Hours:

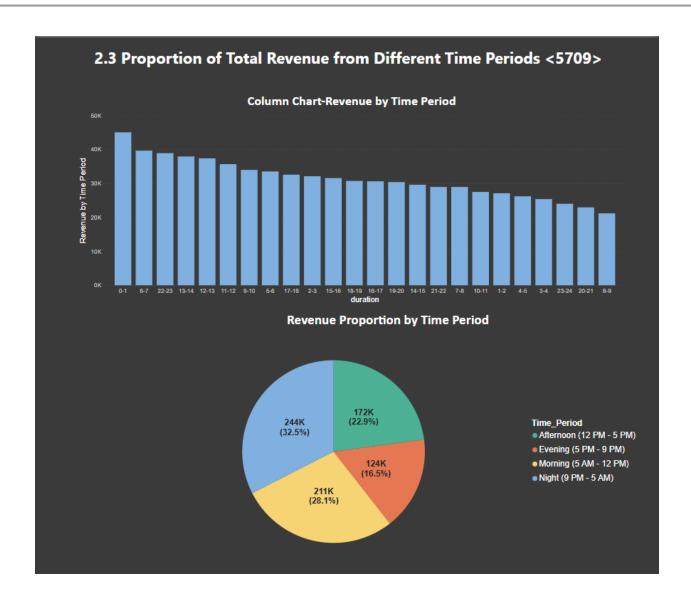
- Offer incentives for driver to work late shifts
- Reduce idle drivers.
- Used possible parameters like "Trip Completion", "Trip Counts (Count of tripid)" and "% of total Trip" for Demand Analysis.

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

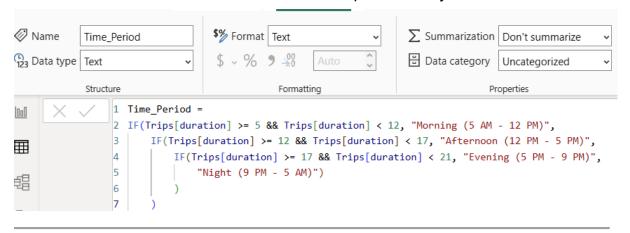
 Calculate the proportion of revenue generated during different time periods and visualise how it contributes to total revenue.

- Majority of revenue was generated during the Night period (9 PM-5 AM).
- Contributing to over 32.5% of the total revenue, suggesting high ride demand after work hours.
- Late Night Hours i.e. 12 AM 1 AM has highest revenue contribution to Total Revenue demand with duration "0-1".
- Increase drivers' availability, use targeted promotion (e.g. discounts for off peak rides) and dynamic pricing.





Created below Calculated measures – "Time\_Period" in Trips Table and used in above Pie Chart for Revenue Proportion Analysis.

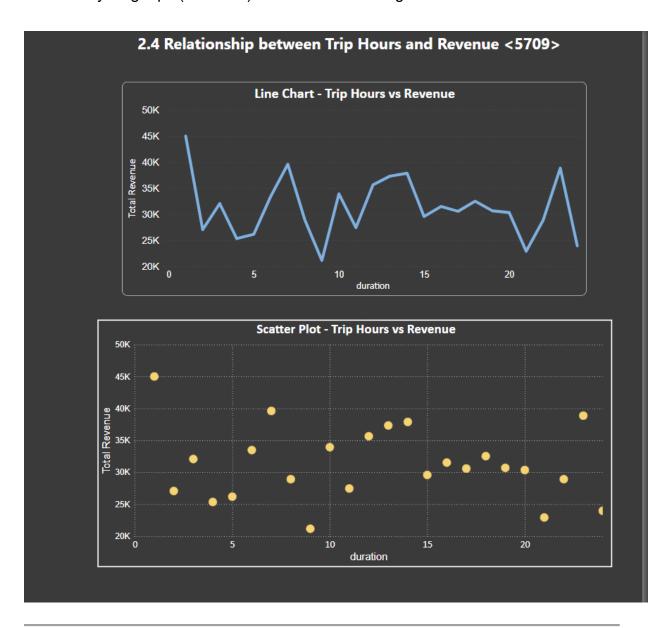




# 2.4. Explore the Relationship Between Trip Hour and Revenue [3 Marks]

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

- ❖ Highest Revenue contribution with Total revenue of 45019 was of 0-1 Trip Hours.
- ❖ Very long trips (>4 hours) are rare but have higher individual fares.

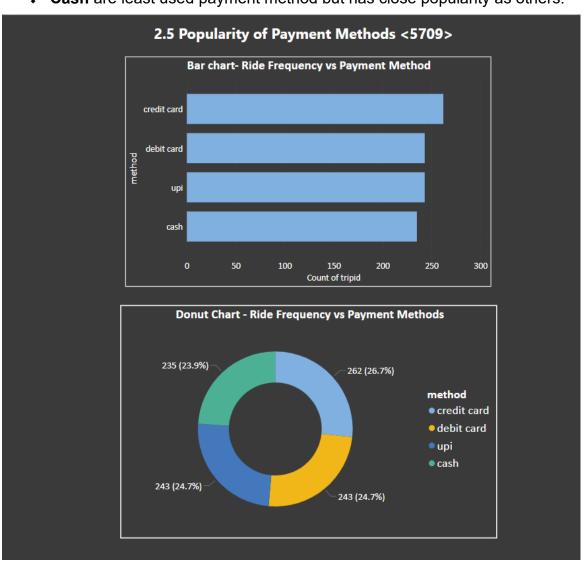




# 2.5. Examine the Popularity of Different Payment Methods [3 Marks]

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

- ❖ Most popular payment method use by customer is Credit Card with 28% of ride frequency following with Second priority i.e. Debit card and UPI with same Ride frequency of around 25% of Total Frequency.
- ❖ Cash are least used payment method but has close popularity as others.



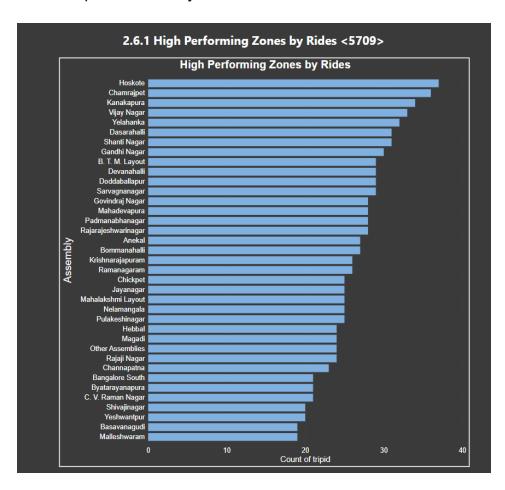


## 2.6. Identify High-Performing Zones [6 Marks]

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]

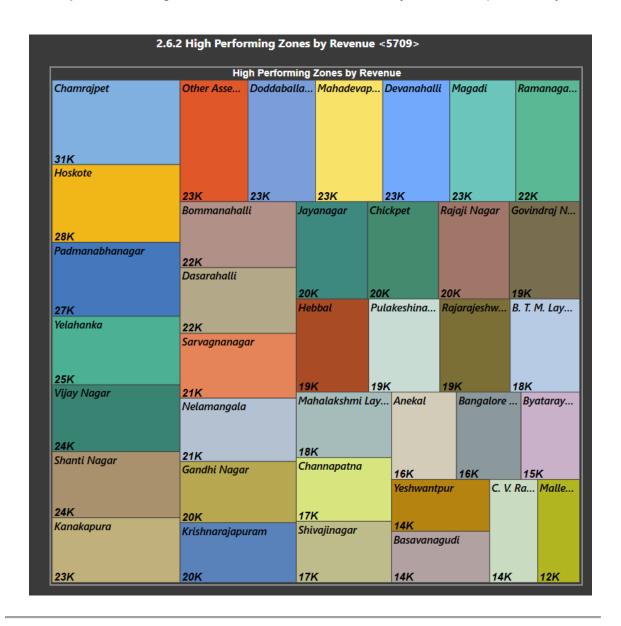
- ❖ The top pickup zones by ride volume are Hoskote (37), Chamrajpet (36), and Kanakapura (34) of total rides, accounting for the majority of trip requests.
- ❖ These areas are likely high-demand due to their commercial presence and dense residential.
- ❖ Targeting these zones for promotional activities and driver allocation can further improve efficiency.





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

- ❖ The top pickup zones with highest revenue are Chamrajpet with fare of total 30617 followed with Hoskote (28209) and Padmanabhanagar (26806) fare.
- ❖ These areas not only have high demand but also longer or higher-priced trips. Prioritizing these zones for driver availability can boost profitability."





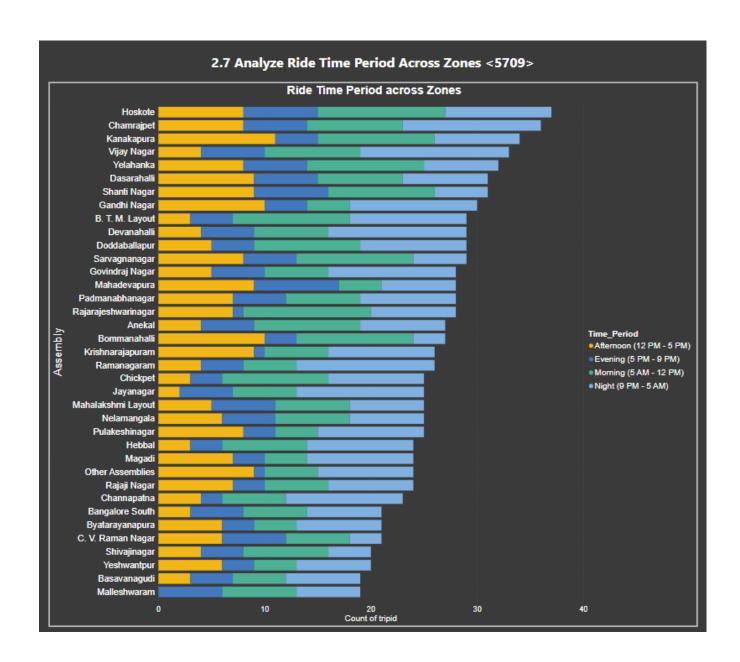
## 2.7. Analyse Ride Time Periods Across Zones [4 Marks]

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

- ❖ The Top pickup zones like Hoskote has peak time hours at Morning (5 AM 12 PM) also its seen throughout all time periods.
- ❖ Zones like Malleshwaram had no Rides in Afternoon (12 PM 5 PM).

			/latrix Table <5709>			
Assembly	Afternoon (12 PM -	5 PM) Eve	ening (5 PM - 9 PM) Morning (5 A	M - 12 PM) Night (9 PM	- 5 AM)	Tota
Hoskote		8	7	12	10	37
Chamrajpet		8	6	9	13	36
Kanakapura		11	4	11	8	34
Vijay Nagar		4	6	9	14	33
Yelahanka		8	6	11	7	32
Dasarahalli		9	6	8	8	31
Shanti Nagar		9	7	10	5	31
Gandhi Nagar		10	4	4	12	30
B. T. M. Layout		3	4	11	11	29
Devanahalli		4	5	7	13	29
Doddaballapur		5	4	10	10	29
Sarvagnanagar		8	5	11	5	29
Govindraj Nagar		5	5	6	12	28
Mahadevapura		9	8	4	7	28
Padmanabhanagar		7	5	7	9	28
Rajarajeshwarinagar		7	1	12	8	28
Anekal		4	5	10	8	27
Bommanahalli		10	3	11	3	27
Krishnarajapuram		9	1	6	10	26
Ramanagaram		4	4	5	13	26
Chickpet		3	3	10	9	25
Jayanagar		2	5	6	12	25
Mahalakshmi Layout		5	6	7	7	25
Nelamangala		6	5	7	7	25
Pulakeshinagar		8	3	4	10	25
Hebbal		3	3	8	10	24
Magadi		7	3	4	10	24
Other Assemblies		9	Ĭ	5	9	24
Rajaji Nagar		7	3	6	8	24
Channapatna		4	2	6	11	23
Bangalore South		3	5	6	7	21
Byatarayanapura		6	3	4	8	21
C. V. Raman Nagar		6	6	6	3	21
Shivajinagar		4	4	8	4	20
Yeshwantpur		6	3	4	7	20
Basavanagudi		3	4	5	7	19
Malleshwaram		J	6	7	6	19
Total		224	161	277	321	983





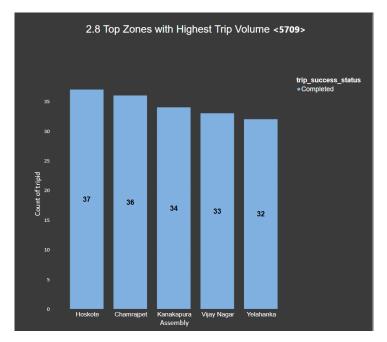


## 2.8. Top Zones with Highest Trip Volume [3 Marks]

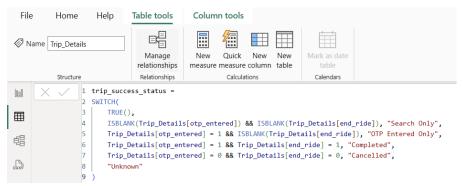
- Identify the top 5 pickup zones with the highest total number of completed trips.
- Analyse factors contributing to the higher number of trips.

#### Solution:

- Top 5 Pickup Zones with Completed Trips are Hoskote, Chamrajpet, Kanakapura, Vijay Nagar and Yelahanka.
- ❖ These zones are densely populated with a mix of residential, commercial, and tech offices, leading to higher ride requests.
- Digital payment adoption, Accessibility and availability of drivers also likely contribute to this high volume.



❖ Below Calculated measure – "trip\_success\_status" was created in Trips\_Details Table for analysis of Ride Completion status.





## 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]

- ❖ Percentage of Customer's Ride Cancellation: 48.2%
- ❖ Percentage of Driver's Ride Cancellation: 47.2%
- ❖ Percentage of Successful Ride Completion: 45.5%
- ❖ These insights help identify lacking points in trip fulfillment and can guide improvements in user experience, driver training, or app functionality.





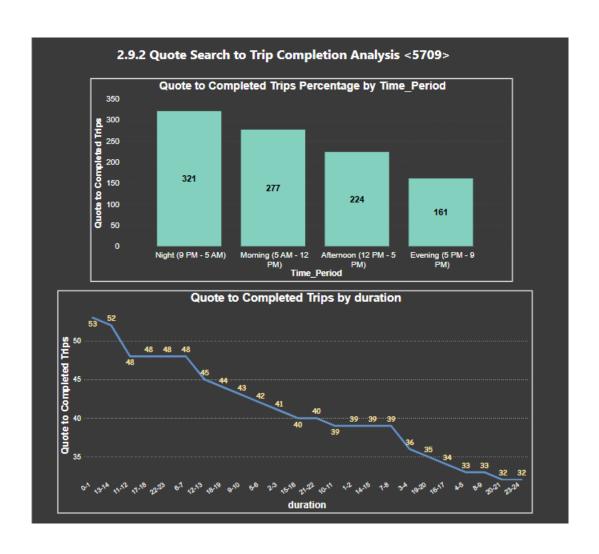
#### 2.9.2

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

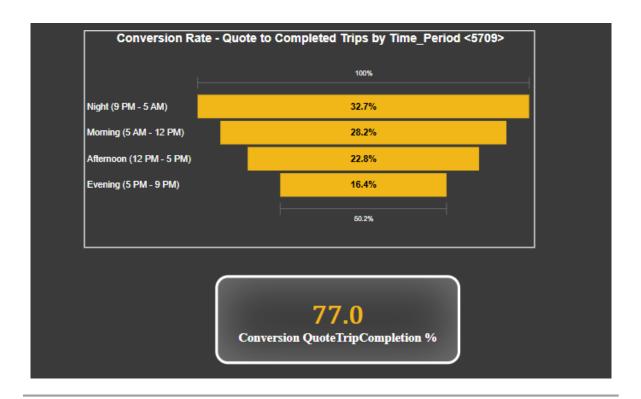
### Solution:

The Analysis showed that:

- Conversion was highest during Night (9 PM 5 AM), possibly due to fewer transport alternatives.
- Lower conversions during Evening (5 PM 9 PM) may suggest higher cancellations due to traffic or availability.







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

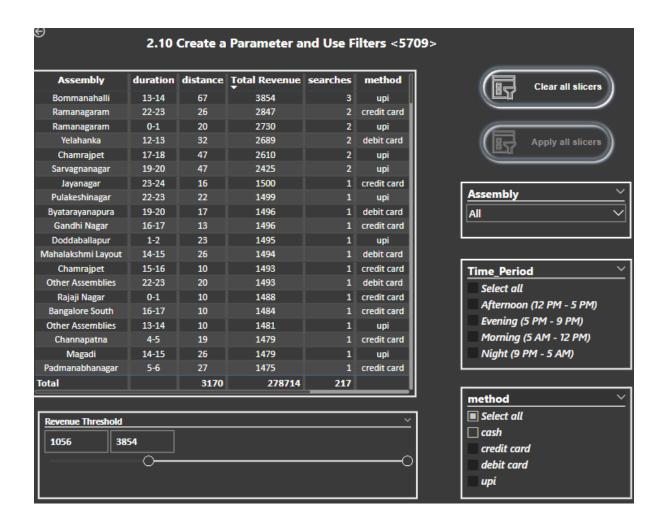
- Create a parameter and use it as a filter on an appropriate subset of the data to interactively analyse and visualise different subsets of the data.
- Explain your choice of filter and insights drawn from this step.

- ❖ A Parameter Revenue\_Thershold with Numeric range was created to filter the data based on Total Revenue. A Numeric range-based slicer was added to the dashboard using this parameter.
- ❖ Another slicer like Time Period, Assembly (Zone area) and Payment Method were used from other existing Tables like Trips, Assembly and Payment to make visual more filtered and interactive.
- ❖ This allows us to interactively explore how metrics such as trip volume and revenue change across different time segments. This enabled a dynamic and intuitive dashboard experience.



## Insights:

- 1. **Night hours** show high demand with **high average fare**, while **evening hours** show fewer trips with **less fares**, possibly due to traffic surcharges.
- 2. **Cash payments** are still dominant in certain zones, but **digital payments** have more popularity for higher completion rates.
- As we keep increasing the **Revenue threshold**, only a few premium zones continue to appear. This indicates that they consistently generate high-paying trips.





## 3. Conclusion [20 Marks]

## 3.1. Recommendations for Operational Efficiency [10 Marks]

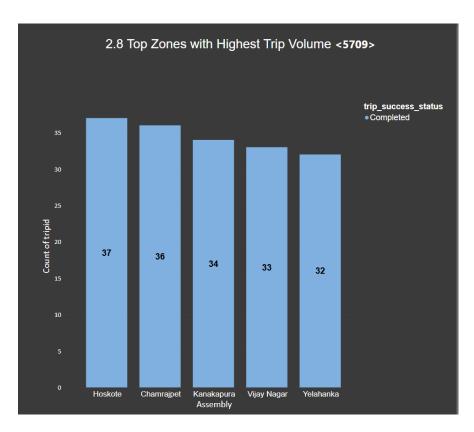
- Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
- This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
- Add supporting dashboards.

#### Solution:

#### **Recommendations:**

## 1. Increase Resource Allocation in High Demand Zones:

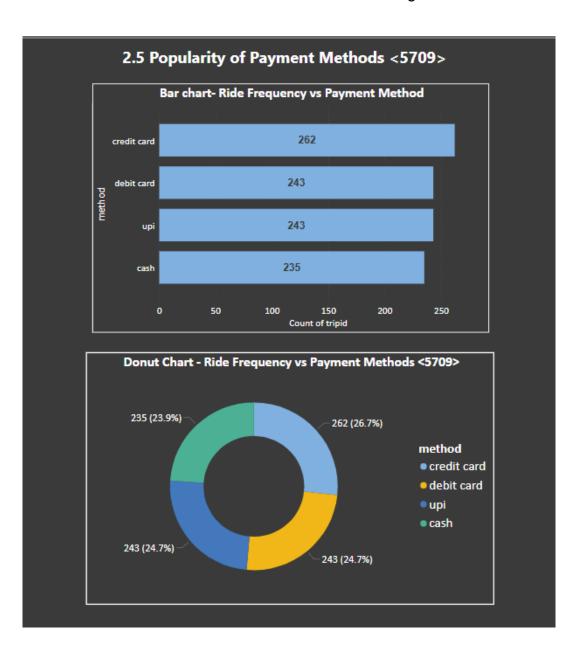
- Focus on Peak Assembly Area: Allocate more drivers in High demand pickup zones with peak ride requests like Hoskote, Chamrajpet, Kanakapura, Vijay Nagar and Yelahanka.
- By encouraging Drivers with good incentive pricing for working proactively during high demands.





## 2. Encourage Digital Payments for Quick Process:

- Customer prefers paying more using UPI, Debit Card and Credit Card more than Cash payments. They had higher Ride Completion Rates.
- Therefore, encouraging usage of Digital payments increases more customer satisfaction and more ride bookings.





#### 3. Reduce Cancellation Rates:

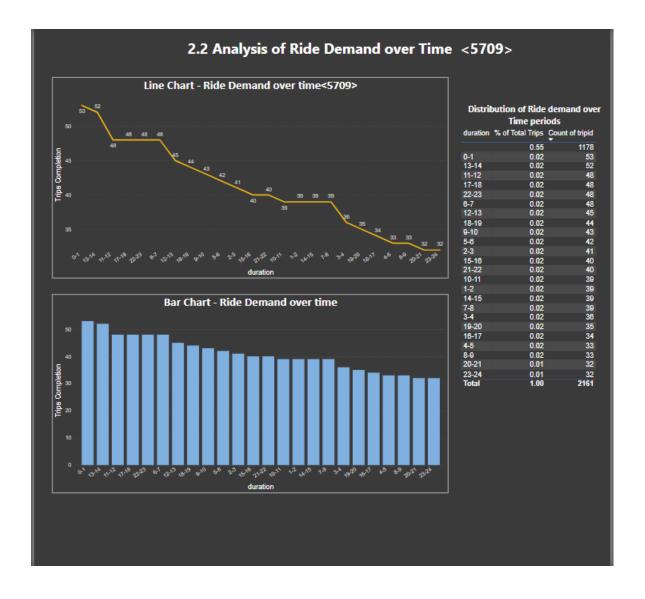
- After performing analysis, we observed significant Cancellation rates by both Drivers and Customers.
- We can improve this by implementing strict **Cancellation policies** with real time alerts for repeated cancellations.
- Also, offering bonus and rewards could encourage drivers to complete rides without cancellations.





## 4. Improve Ride Durations:

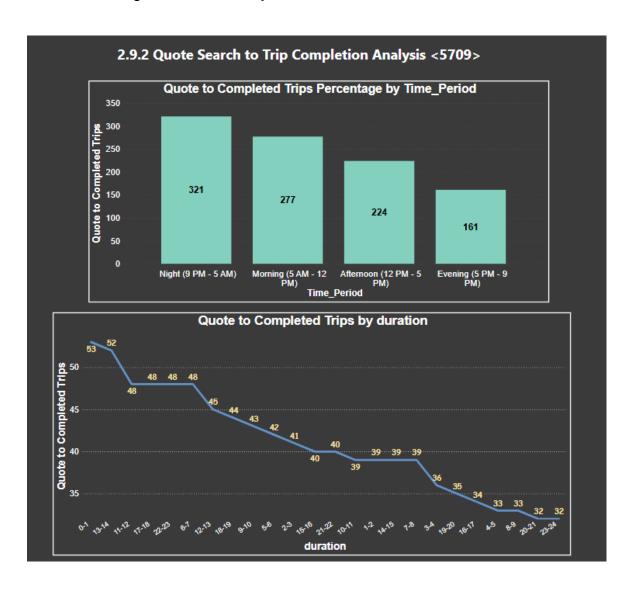
- High Demand for Ride was observed at Night Duration i.e. 12 AM 1 AM,
   this might be due to lack of transportation during night time.
- Offer high discounts and reward points to customer as they contribute in higher completion rates.
- Most trips were completed under 1 hour, especially in city centers.
   Therefore, encourage more short trips during high-demand hours to improve trip frequency and reduce idle driver time.





## 5. Improve Quote – to – Ride Completion Rates:

- A significant gap was identified between users searched for quotes and completed rides.
- To reduce this drop-off, provide easy booking process experience after quote generation.
- Add time limited or instant booking discounts for customers to encourage conversion of quote searches into rides.





## 4. Marketing and Operational Strategy Improvements [10 Marks]

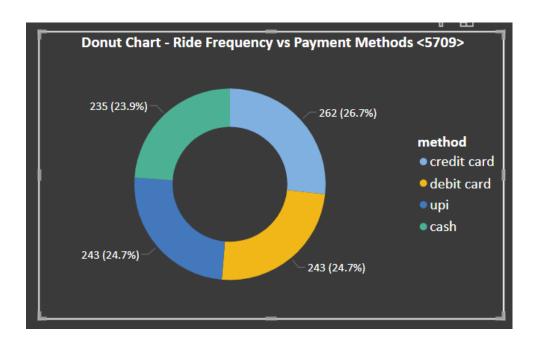
- a. Suggest improvements to Namma Yatri's marketing or operational strategies based on your analysis.
- b. Recommendations could involve promotional efforts, driver incentives, or regional targeting to increase customer satisfaction and service efficiency.
- c. Add supporting dashboards.

#### Solution:

## Operational Strategy Improvements:

## 1. Encourage Adoption for Digital Payments:

- Insights: Digital Payments showed higher reliability and lower dispute rates. Adoption of digital payment can increase and fasten the ride turnaround.
- **Strategy:** Reward customers with **cashbacks** and ride loyalty points for doing payment using **UPI/ Wallet payments**.

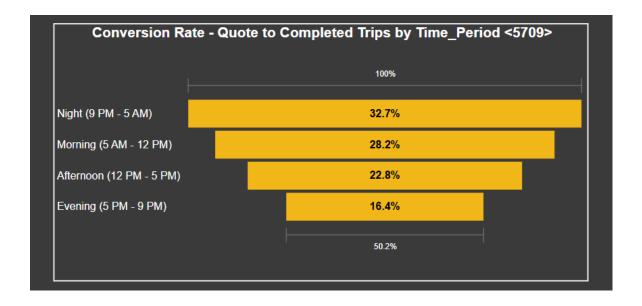


## 2. Optimize Quote - to - Booking Conversion Rate:

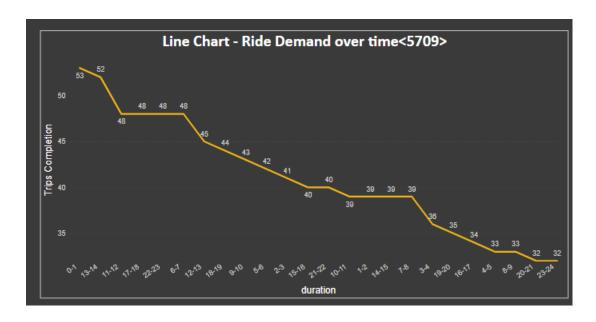
Insights: Significant drop was observed where users search for quotes but do not complete booking ride.



Strategy: Send follow up notification for post quote process to users or limitedtime offers to encourage successful booking completion.



- 3. Incentive for Rides during Off-Peak:
- Insights: A significant drop for rides was seen during off-peak hours.
- ❖ Strategy: Offer high discounts and loyalty points to customer taking rides during off-peak hours to balance demands and resource utility.

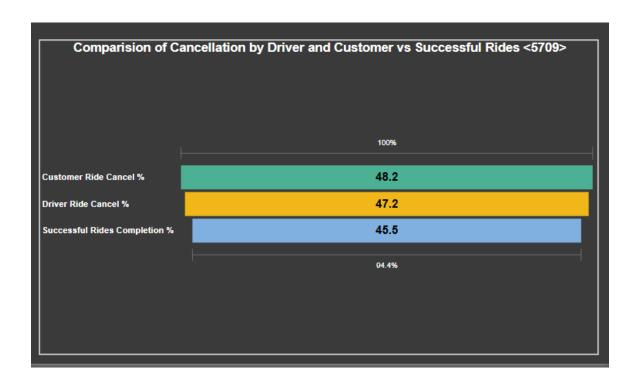




## Marketing Strategy Improvements:

## 1. Incentive Programs for Drivers Performance:

- *Insights:* Some Drivers showed great performance by completing rides successfully with lower cancellation rates.
- **Strategy:** Introduce Reward programs for Appreciating Drivers for their great performance based on customer feedback, ride volumes and completion rates with **offering bonus** and **incentives.** This would encourage them for better services.



## 2. Localized Campaigns to Boost Low-Volume Areas:

- *Insights:* Some pickup zones constantly showed low ride volumes and activity.
- Strategy: Run ads and marketing campaigns in targeted area having lower ride volumes to increase awareness. Provide location-based discounts and first-ride free offers to attract new users.



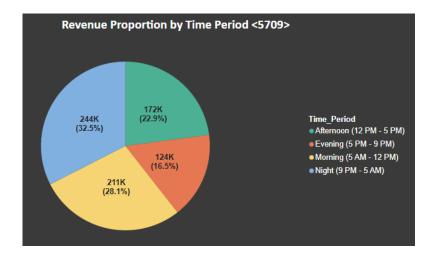
• These would help in focusing on expanding market reach by providing drivers opportunity to work under-areas and reduce their idle time.

			Matrix Table <5709>			
Assembly	Afternoon (12 PM - 9	5 PM) I	Evening (5 PM - 9 PM) M	orning (5 AM - 12 PM)	Night (9 PM - 5 AM)	Total
Hoskote		8	7	12	10	37
Chamrajpet		8	6	9	13	36
Kanakapura		11	4	11	8	34
Vijay Nagar		4	6	9	14	33
Yelahanka		8	6	11	7	32
Dasarahalli		9	6	8	8	31
Shanti Nagar		9	7	10	5	31
Gandhi Nagar		10	4	4	12	30
B. T. M. Layout		3	4	11	11	29
Devanahalli		4	5	7	13	29
Doddaballapur		5	4	10	10	29
Sarvagnanagar		8	5	11	5	29
Govindraj Nagar		5	5	6	12	28
Mahadevapura		9	8	4	7	28
Padmanabhanagar		7	5	7	9	28
Rajarajeshwarinagar		7	1	12	8	28
Anekal		4	5	10	8	27
Bommanahalli		10	3	11	3	27
Krishnarajapuram		9	1	6	10	26
Ramanagaram		4	4	5	13	26
Chickpet		3	3	10	9	25
Jayanagar		2	5	6	12	25
Mahalakshmi Layout		5	6	7	7	25
Nelamangala		6	5	7	7	25
Pulakeshinagar		8	3	4	10	25
Hebbal		3	3	8	10	24
Magadi		7	3	4	10	24
Other Assemblies		9	1	5	9	24
Rajaji Nagar		7	3	6	8	24
Channapatna		4	2	6	11	23
Bangalore South		3	5	6	7	21
Byatarayanapura		6	3	4	8	21
C. V. Raman Nagar		6	6	6	3	21
Shivajinagar		4	4	8	4	20
Yeshwantpur		6	3	4	7	20
Basavanagudi		3	4	5	7	19
Malleshwaram			6	7	6	19
Total		224	161	277	321	983

## 3. Promotions for Low Demand Hours:

- *Insights:* Low Ride Demands were observed during the **Evening** and **Afternoon Hours**.
- **Strategy:** Introduce "**Happy Yatri Hours**" discounts during low demand periods to encourage balance in supplies and demands.
- This will reduce demand fluctuations and maximizing drivers' utilization.





## 4. Optimize Referral Programs for Customers:

- *Insights:* By observing trip hours and ride completion rates, it indicates that majority of customers actively using rides.
- **Strategy:** Launching "Refer a Friend" program where both referrer and referee would get free rides and great discounts on their rides.
- Here, Customers themselves can help us with verbal marketing and is most cost-effective and can boost users' acquisition.

