

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:

Imported Namma yatri dataset into PowerBI and joined all the tables using appropriate keys to create a unified data model.

Trip_Details(tripid) and Trips(tripid) tables are joined (one to one)

Trips(duration) is merged to Duration(id) (many to one)

Trips(faremethod) is merged to payment (id) (many to one)

trips(loc_from) is merged to Assembly(id) (many to one)

Validated that no duplicate or mismatched keys were causing relationship issues.

<input type="checkbox"/> From: table (column) ↓	...	Relationship	To: table (column)
<input type="checkbox"/> Trips (tripid)		1 —> 1	Trip_Details (tripid)
<input type="checkbox"/> Σ Trips (loc_from)		* —> 1	Assembly (ID)
<input type="checkbox"/> Σ Trips (faremethod)		* —> 1	Payment (id)
<input type="checkbox"/> Trips (duration)		* —> 1	Duration (id)

The unified model helped to analyze data without inconsistencies.

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:

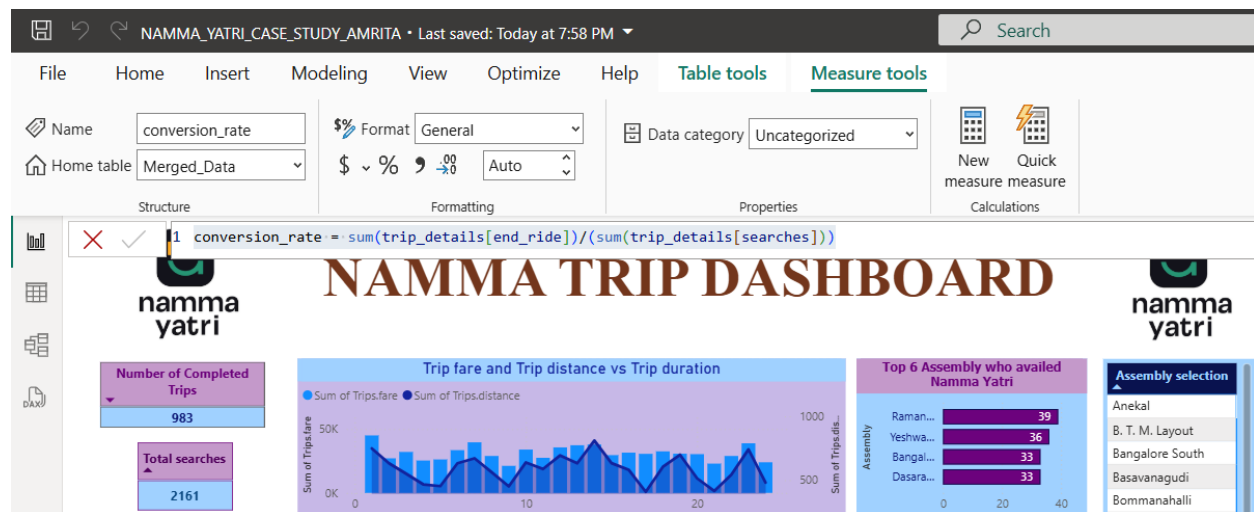
Observed no major inconsistencies in the dataset that might affect. The data is well-structured and ready to be used for analysis. Missing or null values in key fields such as tripid, duration, fare etc. was checked and none was found. Considering negative fares or durations were not found hence we can say no incorrect or outlier values are present. Data types of the columns in all tables were consistent. No duplicate entries were found in primary key columns.

present in the data.

The dataset-maintained integrity across all tables which ensures reliable results and accurate insights during analysis.

Created calculated field :

$$\text{conversion_rate} = \frac{\text{sum}(\text{trip_details}[\text{end_ride}])}{(\text{sum}(\text{trip_details}[\text{searches}]))}$$



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:

- duration (Duration)
- fare, distance, duration (Trips)
- searches, searches_for_quotes, searches_got_quotes (Trip_Details)

Categorical Variables:

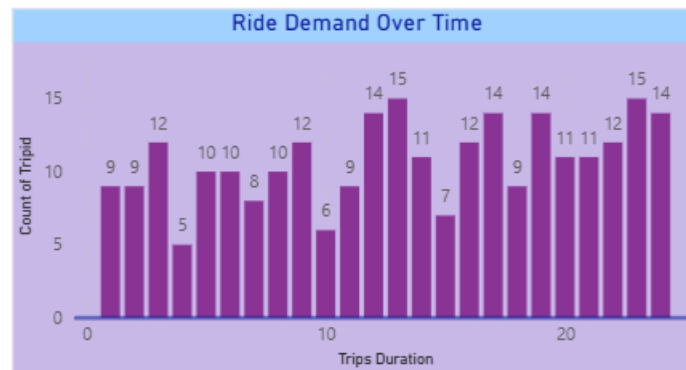
- custid, driverid, faremethod (Trips), loc_from, loc_to (Trips)
- tripid (multiple tables, categorical for joining)
- Assembly (Assembly table)
- method (Payment method)
- customer_not_cancelled, searches_got_estimate (Trip_Details)

Categorical variables represent data that can be classified into groups or categories, while numerical variables represent data that can be measured or counted. Categorical variables are also known as qualitative variables that describe qualities whereas Numerical variables are also known as quantitative variables that represent quantities.

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:



Namma Yatri ride demand exhibits peak periods during commute hours 9AM-2PM (commute to work) and 7PM-12AM (commute to home). This displays correlation between demand and daily routines. A key parameter for assessing demand could be the number of trips as available on the app within a specific time frame.

For analysis we plotted count of Tripid against duration to get the distribution across 24hours interval.

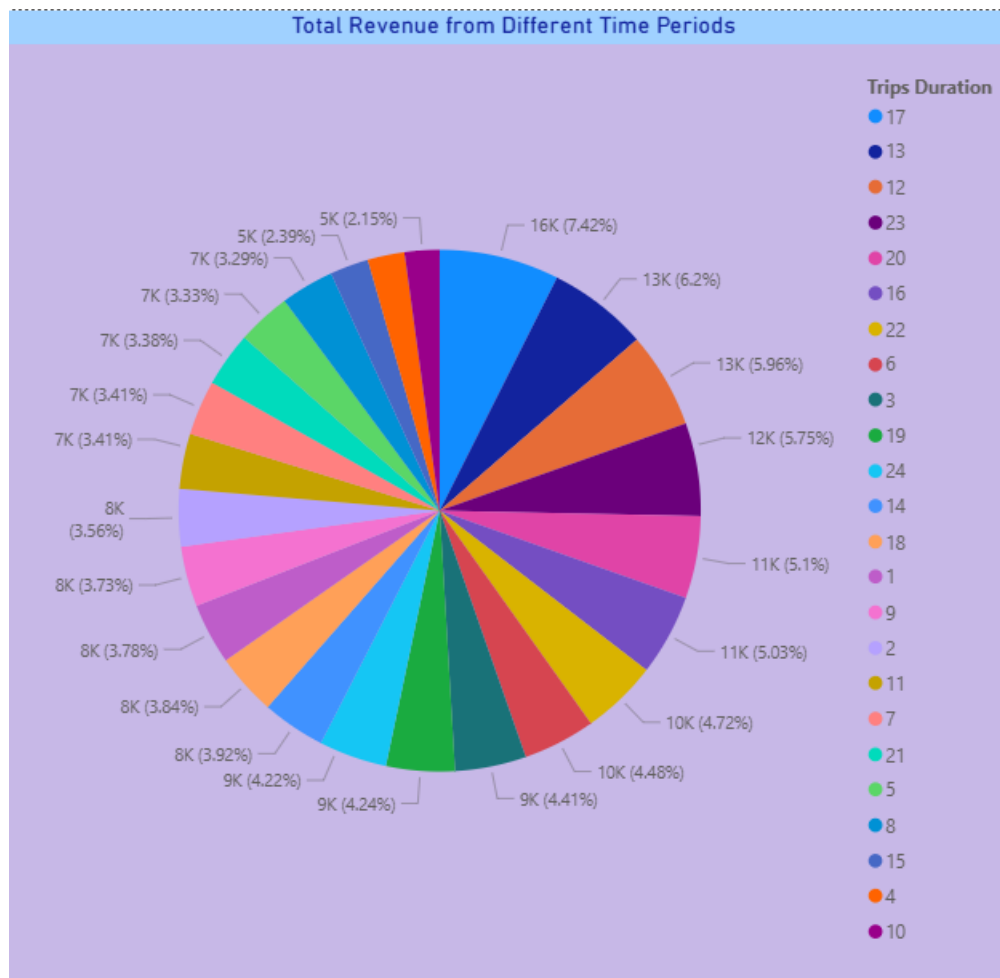
2.3. Proportion of Total Revenue from Different Time Periods [3 Marks]

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

Solution:

At 5pm total trip fare collected is maximum amount of Rs 15921 which is 7.42% of total revenue followed by 1pm.

We plotted a pie chart with Legend as trip duration and values as sum of fare. This gave us a view of how much revenue each time period contributes to. There is a strong correlation between ride demand and revenue generation by hour. Time slots with more rides naturally contribute more to total revenue.



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

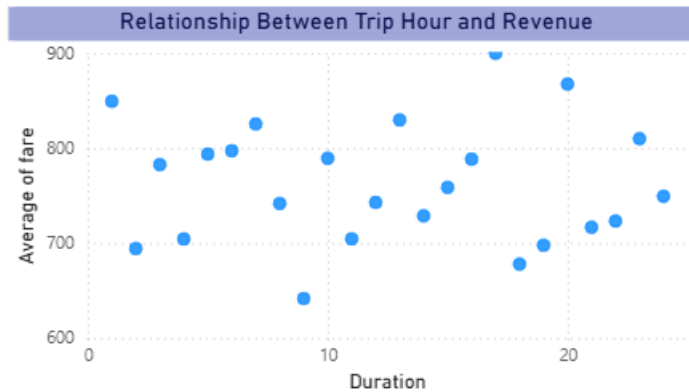
- Investigate the correlation between trip hour and total fare.
- Explain any trends or patterns that emerge.

Solution:

Trips during peak hours (e.g., morning commute to office, evening rush from office) shows a higher average fare compared to off-peak hours. This is because demand is typically higher during these times and increased fares. This could be attributed to

increased demand and potentially higher traffic.

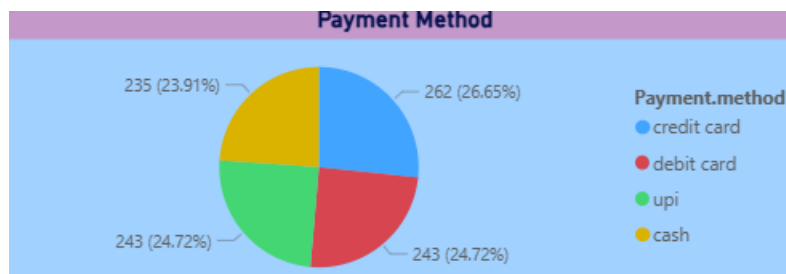
Created a scatter plot to analyze the relationship between ride duration and average fare. It shows moderate positive correlation.



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:



Credit card is the most popular payment method (26.65% of total payments) followed by Debit card and UPI. This shows higher trend of usage of digital payment methods. The frequency of rides is positively correlated with the availability of digital payment options.

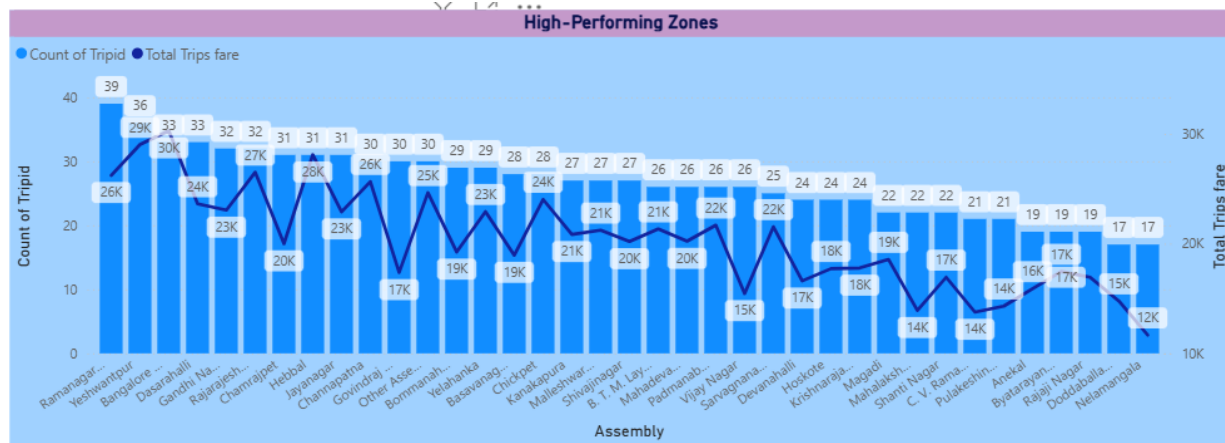
2.6. Identify High-Performing Zones [6 Marks]

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:

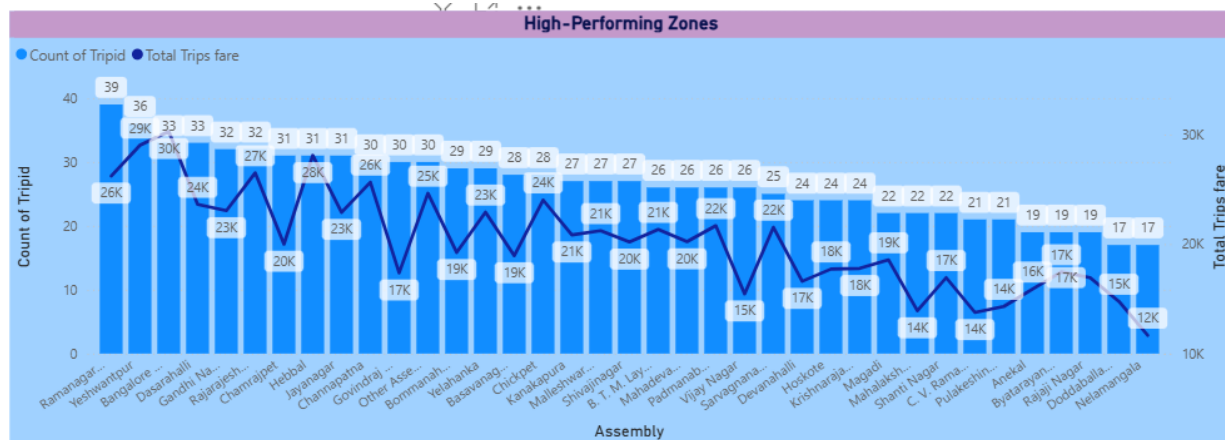
Ramanagaram has the highest number of trip requests (39).



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

Solution:

Bangalore South generates the highest revenue (30295) followed by Yeshwantpur (29035)

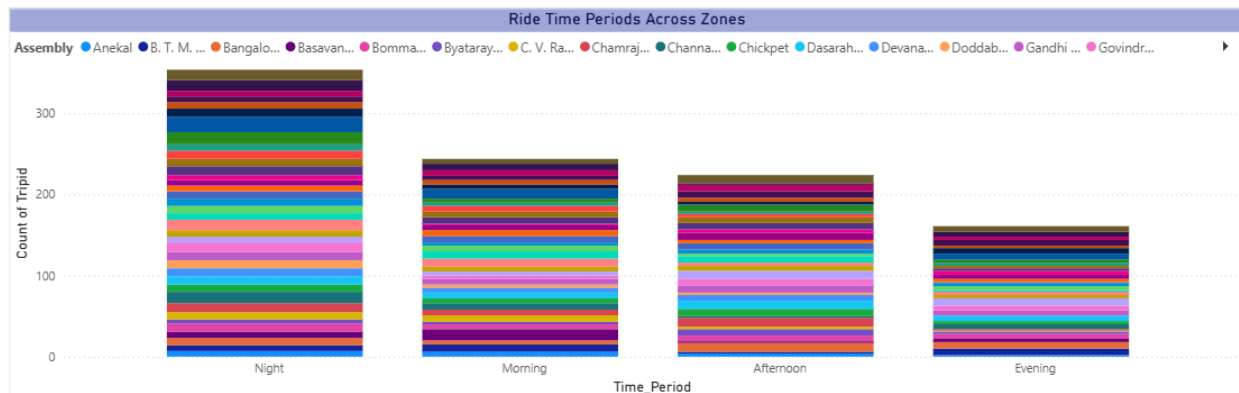


Bangalore South generates the highest revenue.

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

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

Solution:



File Home Help Table tools Column tools

Name: Time_Period Format: Text Summarization: Don't summarize Data category: Uncategorized

Structure: 1 Time_Period = 2 VAR hour = VALUE(Duration[id]) -- assuming Duration[id] is a number or a number-formatted string 3 RETURN 4 SWITCH(5 TRUE(), 6 hour >= 6 && hour < 12, "Morning", 7 hour >= 12 && hour < 17, "Afternoon", 8 hour >= 17 && hour < 21, "Evening", 9 "Night" 10)

Formatting: Properties: Sort by column: Sort Data groups: Groups Manage relationships: Relationships New column: Calculations

id	duration	Time_Period
1	0-1	Night
2	1-2	Night
3	2-3	Night

The above stacked column chart was created which reveals that Night hours had the highest number of trips, followed by Morning, Afternoon, and Evening. Zone-wise distribution is clearly distinguishable through color-coded stacks.

Bangalore South, Channapatna and Chamrajpet has trips at almost all hours of the day. At 7AM, Basavanagudi has the most trip and at 12pm Chamrajpet and Yeshwantpur has the most number of trips.

At 12am Ramanagaram has the most number of trips and at 10am Dasarahalli has the most number of trips.

Trips duration	Assemblies	Padmanabhanagar	Pulakeshinagar	Rajaji Nagar	Rajarajeshwarinagar	Ramanagaram	Sarvagnanagar	Shanti Nagar	Shivajinagar	Vijay Nagar	Yelahanka	Yeshwantpur	Total
1	2	3				1	2	3	1	1	1	2	53
2			1	1	1	2	1	2	3		1	1	39
3	1			1	1	2	1	1	1	2	1	1	41
4			1	2	1		2	1		1	1	2	36
5			2	1	3	3	1	1		1		1	33
6	1		2	1		1	3	1	1	1	1	2	42
7			2	2	1		1	1	1	1	3		48
8	1			1	3	1	3	1	2	1	3	1	39
9	2					3	4			1		1	33
10			2	1			1		2		2	1	43
11	4		1	2				2		1		2	39
12	1			2		1		1	1	2			5
13	2		1		2	3			3	4			45
14	2		2	1		1			2	3	3	1	52
15	2		2	1		1	1	2		1			39
16	1		2		1	2		1				1	40
17	1				1	2	4	2		1	1	1	34
18						1	3		1	2	1	1	48
19	1			2		2	1	2	1	3		3	44
20	1		3		1		1		1	1	2	1	35
21	1		1		1	3	3		1	1		3	32
22	2				1	1	3		1		1	1	40
23	4			3		2	1	1		2	1	2	48
24	1		1	2		1	5	1	1		1		32
Total	30	26	21	19		32	39	25	22	27	26	29	983

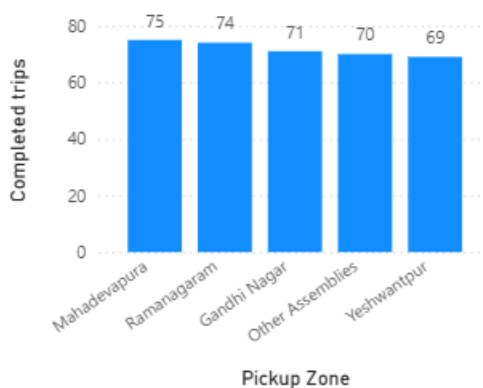
2.8. Top Zones with Highest Trip Volume [3 Marks]

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

Solution:

Factors contributing to higher trip numbers in general might include population density, proximity to commercial areas and the presence of transportation hubs. Popular events or attractions can also temporarily increase trip numbers in certain areas. Income levels and household size can also influence trip generation.

Zones with Highest Trip Volume



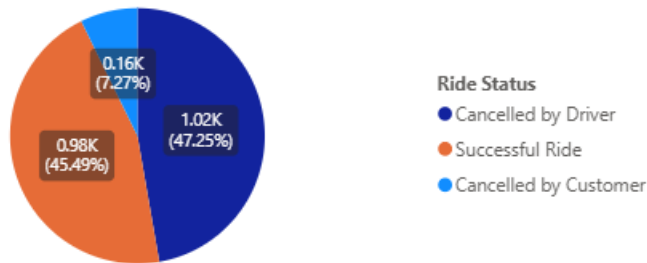
Mahadevapura, Ramanagaram , Gandhi Nagar, Other Assemblies and Yeshwantpur are the top five pickup zones with the highest number of completed trips.

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:

Percentages of cancellations and successful rides by both driver and customer



We used the Trip_Details table, considering:

- customer_not_cancelled (1 = not cancelled, 0 = cancelled)
- driver_not_cancelled (1 = not cancelled, 0 = cancelled)
- end_ride (1 = ride completed, 0 = not completed)

Created a new calculated column:

File Home Insert Modeling View Optimize Help Table tools

Name Ride Status Format Text

Data type Text

Summarization Don't summarize

Data category Uncategorized

Structure Formatting Properties

```
1 Ride Status =  
2 SWITCH(  
3 TRUE(),  
4 Trip_Details[driver_not_cancelled] = 0, "Cancelled by Driver",  
5 Trip_Details[customer_not_cancelled] = 0, "Cancelled by Customer",  
6 Trip_Details[end_ride] = 1, "Successful Ride",  
7 "Other"  
8 )
```

From Pie chart obtained result as Cancelled by Driver: 47.25% ,
Cancelled by Customer: 7.27% , Successful Ride: 45.49%.

- **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:

```
1 Time_Period =
2 VAR hour = RELATED(Duration[id]) -- 'id' should be the hour of the trip
3 RETURN
4 SWITCH(
5     TRUE(),
6     hour >= 6 && hour < 12, "Morning",
7     hour >= 12 && hour < 17, "Afternoon",
8     hour >= 17 && hour < 21, "Evening",
9     "Night"
10 )
```

```
1 % Completed After Quote Search =
2 DIVIDE(
3     CALCULATE(
4         DISTINCTCOUNT(Merged_Data[Trips.custid]),
5         NOT(ISBLANK(Merged_Data[end_ride])),
6         Merged_Data[searches_for_quotes] > 0
7     ),
8     CALCULATE(
9         DISTINCTCOUNT(Merged_Data[Trips.custid]),
10        Merged_Data[searches_for_quotes] > 0
11    ),
12    0
13 )
```

Percentage of people who completed trips after searching for quotes by
Time_Period

Time_Period	Sum of searches_for_quotes	Completed_Trips	% Completed After Quote Search
Afternoon	224	224	1.00
Evening	161	161	1.00
Morning	244	244	1.00
Night	826	1532	1.00
Total	1455	2161	1.00

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:

```
MinQuoteSearches = GENERATESERIES(0, 10, 1)
```

```
1 % Completed After Quote Search (filtered) =  
2 VAR MinQuotes = SELECTEDVALUE(MinQuoteSearches[MinQuoteSearches]) RETURN  
3 DIVIDE(  
4     CALCULATE(  
5         DISTINCTCOUNT(Merged_Data[Trips.custid]),  
6         NOT(ISBLANK(Merged_Data[end_ride])),  
7         Merged_Data[searches_for_quotes] >= MinQuotes  
8     ),  
9     CALCULATE(  
10        DISTINCTCOUNT(Merged_Data[Trips.custid]),  
11        Merged_Data[searches_for_quotes] >= MinQuotes  
12    ),  
13    0  
14 )
```

MinQuoteSearches



This parameterized filter lets us interactively explore the behavior by user engagement:

Low filter (MinQuotes = 0): Shows average user conversion.

Higher filter (MinQuotes = 2+): Reveals how more interested or serious users behave.

We might find that users who search multiple times are much more likely to convert, or the opposite.

3. Conclusion

[20 Marks]

3.1. Recommendations for Operational Efficiency [10 Marks]

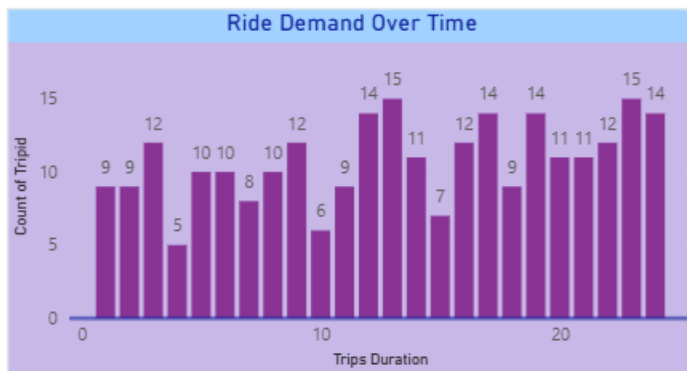
- Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
- This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
- Add supporting dashboards.

Solution:

Optimize Driver Availability by Time Period

Completion rates are highest during the Afternoon and Evening, indicating peak demand.

Using dynamic driver incentives or driver shifts to increase coverage during underperforming time slots might help. Allocating more drivers to zones during peak periods to reduce wait times and cancellations would help.



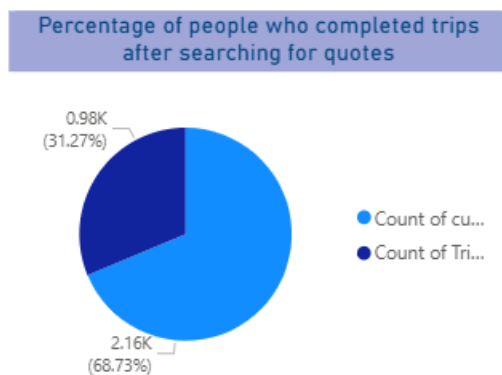
Focus on High-Intent Users to Improve Completion Rates

Users with higher quote search counts have a significantly higher trip completion rate. Implement targeted promotions (e.g., discounts or reminders) for users who have searched for quotes multiple times but haven't booked yet.

Simplify Quote-to-Booking Experience

Many users search for quotes but don't complete bookings.

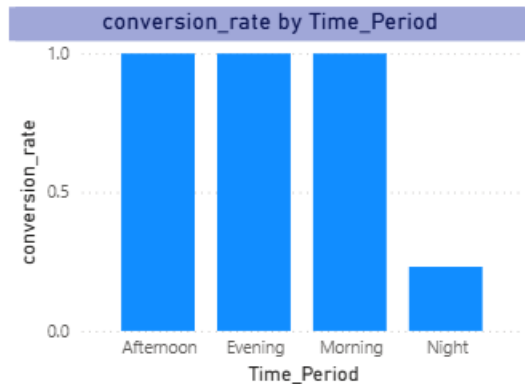
Introducing a simplified booking interface and reduce time between quote and ride confirmation to improve conversion.



Reduce Cancellations in Low-Conversion Time Periods

Night time has a lower completion rate, likely due to driver unavailability or safety concerns.

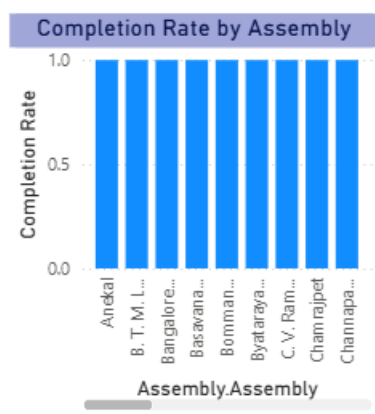
Introducing driver incentives for night shifts and implement cancellation prediction to pre-emptively suggest alternatives to riders.

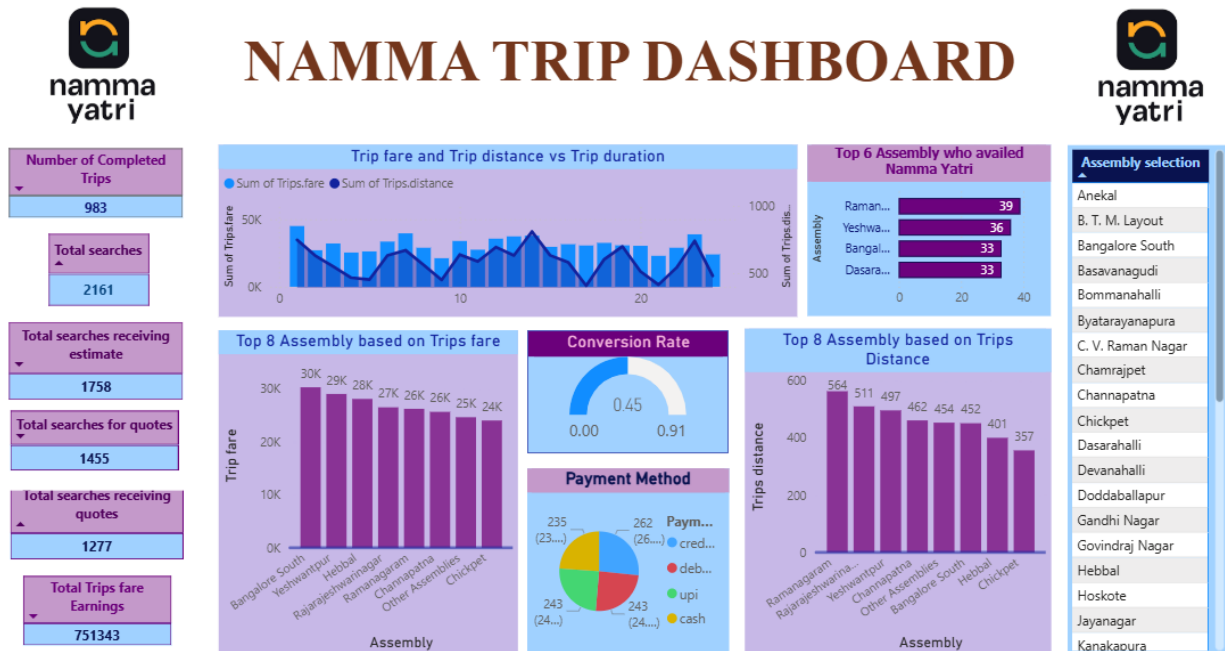


Zone-Based Allocation Strategy

Certain zones consistently show higher or lower completion rates.

Reallocating driver distribution based on historical completion performance by zone to balance supply and demand.





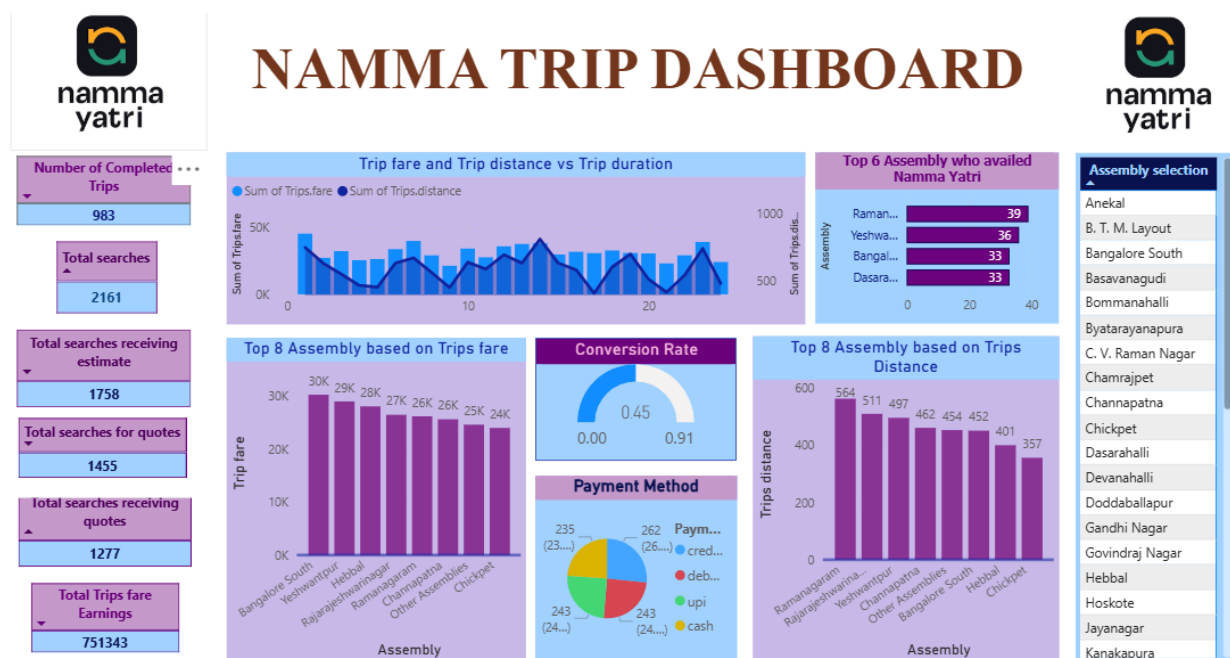
3.2. Marketing and Operational Strategy Improvements [10 Marks]

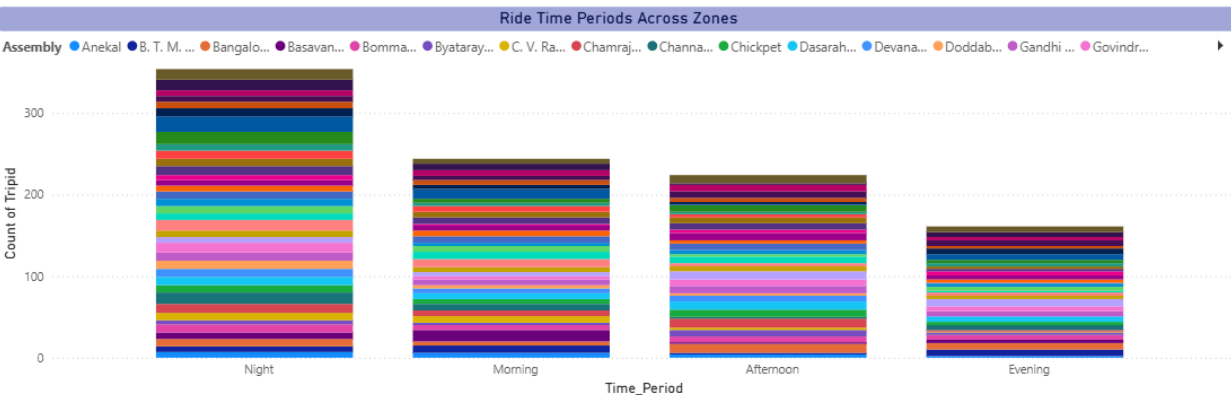
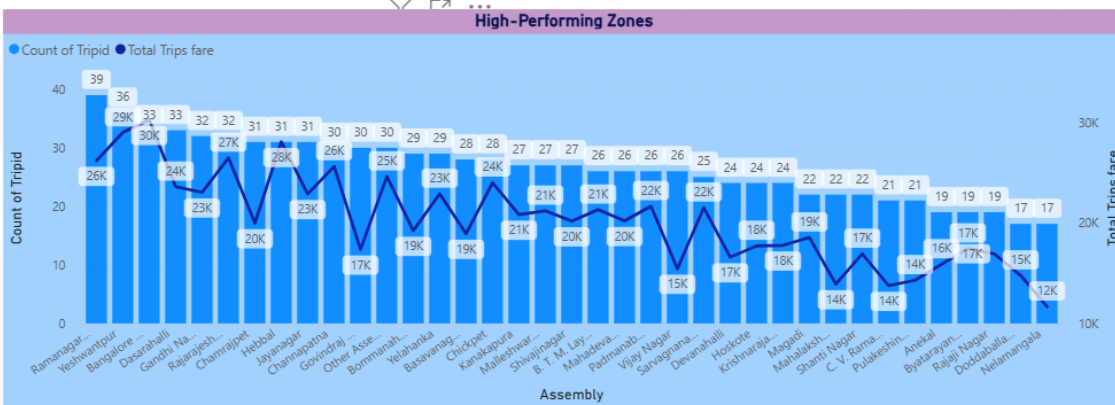
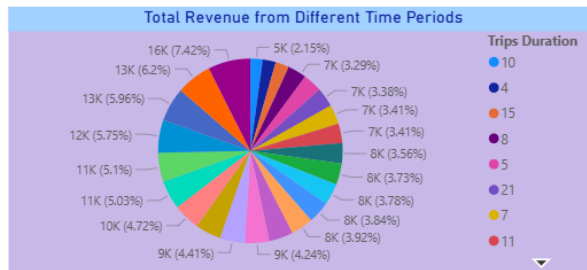
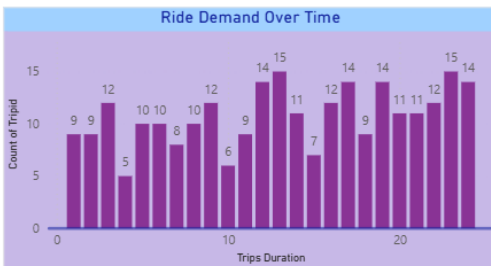
- Suggest improvements to Namma Yatri's marketing or operational strategies based on your analysis.
- Recommendations could involve promotional efforts, driver incentives, or regional targeting to increase customer satisfaction and service efficiency.
- Add supporting dashboards.

Solution:

1. Promotional efforts in low-demand zones and off-peak hours
 - Offer discounts or cashback during early mornings and in low-demand areas.
 - Increases ride requests and improves driver utilization.
2. Incentivize high-performing drivers
 - Give bonuses to drivers with high ride completion rates and low cancellation rates.

- Improves service reliability and reduces cancellation instances.
3. Target high-cancellation zones with marketing campaigns
- Run localized ads and awareness drives in areas with more cancellations.
 - Builds trust and increases service usage in underperforming zones.
4. Optimize routing and driver deployment during peak hours
- Adjust driver distribution and routes to reduce ride delays and improve efficiency.
 - Enhances punctuality and user satisfaction.
5. Promote digital payment options
- Encourage UPI and card payments with exclusive offers.
 - Reduces cash handling and improves operational efficiency.





Percentage of people who completed trips after searching for quotes by Time_Period

Time_Period	Sum of searches_for_quotes	Completed_Trips	% Completed After Quote Search
Afternoon	224	224	1.00
Evening	161	161	1.00
Morning	244	244	1.00
Night	826	1532	1.00
Total	1455	2161	1.00