

# 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:**

The Namma Yatri dataset was imported into Power BI Desktop using “Get Data”. All tables (Trips, Trip Details, Payment, Duration, and Assembly) were loaded and connected in Model View using key fields such as tripid, faremethod, loc\_from, and duration. The relationships were validated to ensure correct filtering and alignment of data across tables. A table visual was also created to verify that fields from different tables display correctly, confirming that the unified dataset is ready for accurate analysis.

#### 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:**

The dataset was cleaned in **Power Query Editor** to ensure accuracy before analysis. Missing values, incorrect data types, and formatting issues were inspected and corrected. Duplicate trip records were removed using the **Remove Duplicates** option on tripid. Inconsistent category labels (e.g., payment method names) were standardized for uniform reporting. Unnecessary columns without analytical value were filtered out. Additionally, calculated fields were created in Power BI to support insights, such as:

- **Total Rides** = COUNTROWS(Trips)
- **Completed Rides** filtered by end\_ride = 1
- Cancellation flags and performance measures

After cleaning, validation visuals confirmed that all tables were aligned and the dataset was accurate for further exploration.

---

## 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 in the dataset were reviewed and categorized based on their data type and analytical usage:

**Categorical Variables:**

- tripid
- loc\_from
- loc\_to
- faremethod
- driverid
- custid
- duration (time bucket/category)
- Assembly/Zone names
- Payment method
- Status flags (e.g., OTP entered, cancellation flags)

**Numerical Variables:**

- fare
- distance
- searches
- searches\_got\_estimate
- searches\_for\_quotes
- searches\_got\_quotes
- duration (if stored as numeric – total minutes/hours)

This classification ensures proper visual selection and accurate statistical analysis in Power BI

---

## 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:**

---

Ride demand was analyzed using the **number of completed rides** as the key parameter. A time-based visual was created in Power BI to observe how demand changes throughout the day. The trend shows that:

- **Peak demand occurs during Evening hours (6 PM – 9 PM)** due to office commute and travel needs.
- A secondary peak is observed in the **Morning (8 AM – 10 AM)**.
- Very low demand is seen late at **night and early morning (12 AM – 5 AM)**.

This indicates strong commuter-driven activity and highlights time windows where operational focus should be increased.

### **✓ Conclusion**

Evening and morning periods are the highest demand windows, suggesting an opportunity for improved **driver allocation**, **surge pricing**, and **promotions** during these hours.

---

## 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:**

---

To understand when most earnings occur, revenue was analysed across time buckets in a day. The day was divided into four logical periods:

- **Morning** (4 AM – 12 PM)
- **Afternoon** (12 PM – 4 PM)

- **Evening** (4 PM – 9 PM)
- **Night** (9 PM – 4 AM)

The results show that:

- The **Morning period contributes the highest share of total revenue**, driven by office commute and travel after work hours.
- Night also contributes a significant portion.
- Afternoon hours generate the least revenue due to low demand.

This highlights the importance of focusing driver supply and pricing strategies during evening and morning hours.

### ✓ Conclusion

The revenue distribution indicates that Morning (32.81%) and Night (29.46%) are the most profitable time slots, together accounting for the majority of total earnings. Evening (19.58%) contributes a moderate share, while Afternoon (18.15%) generates the lowest revenue. Overall, the data suggests that Namma Yatri's peak revenue hours occur during early morning and late-night periods, highlighting the importance of ensuring sufficient driver availability and operational support during these high-value times

---

## 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 correlation between **trip hour** and **total revenue** was analysed using a visual comparison. A line chart of revenue across hours shows the following pattern:

The revenue trend shows a clear decline across the time slots.

- Morning generates the highest revenue (246K), indicating strong demand during early hours—likely driven by office commuters, airport trips, and essential morning travel.
- Night follows with 221K, reflecting continued strong demand from late-evening commuters and long-distance night travel.
- Evening revenue drops sharply to 147K, suggesting fewer rides or shorter trip distances during this period.
- Afternoon shows the lowest revenue at 136K, indicating it is the least active

period of the day in terms of earning potential.

The trend highlights a distinct split between high-revenue (Morning, Night) and low-revenue (Evening, Afternoon) slots.

**✓ Conclusion**

Morning and night generate the highest revenue, driven by strong commuter and late-hour travel demand. Evening and afternoon show significantly lower earnings, indicating reduced trip volumes or shorter rides. Focusing driver availability and operational resources on morning and night periods can improve overall revenue efficiency for Namma Yatri.

---

**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 distribution of payment methods was analysed based on the number of rides completed using bar chart. The results indicate that:

- Credit cards lead in usage, indicating strong preference for secure and convenient digital payments among frequent riders.
- Debit card and UPI transactions are equally common, showing balanced adoption of digital methods and growing comfort with cashless payments.
- Cash is the least used, suggesting riders increasingly favor faster, more reliable digital transactions over traditional payment methods.

**✓ Conclusion**

Credit cards are the most-used payment method, followed closely by debit cards and UPI, showing strong digital preference among riders. Cash usage is slightly lower, indicating a shift toward faster, more convenient digital payment option.

---

**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:***

---

A stacked bar chart was utilized to identify the top pickup zones with the maximum # of requests

- **Ramanagaram leads** with the highest number of pickups, indicating strong customer demand in this zone.
- **Yeshwantpur** follows closely, showing consistently high trip activity and strong commuter presence.
- **Bangalore South and Dasarahalli** show similar pickup volumes, suggesting balanced and stable demand across these areas

***Conclusion:***

Ramanagaram records the highest pickup demand, followed by Yeshwantpur. Bangalore South and Dasarahalli show steady activity, indicating consistently strong rider presence across key zones

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

***Solution:***

---

A stacked bar chart was utilized to identify the top pickup zones generating maximum revenue

- **Bangalore South generates the highest revenue (30K),** indicating strong demand and high-value trips.
- **Yeshwantpur and Hebbal follow closely,** reflecting consistent customer activity and strong earning potential.
- **Rajarajeshwari Nagar and Ramanagaram also contribute significantly,** showing balanced revenue distribution across major pickup zones



**Conclusion:**

Bangalore South leads in revenue, followed by Yeshwantpur and Hebbal. Overall, major pickup zones show strong earning potential with balanced revenue contributions

---

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

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

**Solution:**

---

A clustered bar chart was utilized to compare the trip trends by different time slots in a day

- **Night trips dominate revenue** across all top pickup locations, especially in Ramanagaram and Yeshwantpur, indicating strong late-hour travel demand.
- **Morning time slots also show high revenue**, particularly in Yeshwantpur and Bangalore South, reflecting heavy commuter activity.
- **Afternoon revenue is consistently the lowest** across zones, showing minimal mid-day trip demand.
- **Evening revenue varies across locations**, with Hebbal showing stronger evening performance compared to other zones.

**Conclusion:**

Night and morning generate the highest revenue across major pickup zones, while afternoon remains weakest. Strategic driver allocation during peak times can significantly enhance earnings and service efficiency

---

## 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 were identified using a stacked bar chart with Top 5 Location as filters

- **Ramanagaram, Yeshwantpur, Dasarahalli, Bangalore South, and Gandhi Nagar** record the highest number of completed trips, indicating strong and recurring rider demand.
- These zones likely have **high population density, major transit hubs, and strong commercial activity**, driving frequent trip requests.
- Higher trip volumes may also be supported by **better driver availability**, reducing cancellations and improving successful ride completions

### **Conclusion:**

The top pickup zones show consistently higher completed trips due to strong commuter presence, commercial activity, and better driver availability. These factors collectively make these areas high-demand clusters, driving greater trip completion rates across the platform

---

## 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:**

---

In order to obtain the above visual, created helper tables with DAX and utilized 100% stacked bar chart to populate the values

- Customer and driver completion rates are nearly identical, showing consistent ride fulfillment from both sides.
- Cancellation percentages are also very similar, indicating balanced reliability

between customers and drivers.

- Slightly higher driver completion suggests marginally lower driver-initiated cancellations compared to customer cancellations.

**Conclusion:**

Customer and driver performance is almost identical, indicating balanced ride reliability and evenly distributed cancellation behavior across both groups

---

- **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:**

---

To obtain the value, created DAX formulae to consider quotes searched equals 1 along with the time slot available in duration

- Every customer who searched for quotes completed their trip across all time slots.
- Quote availability and driver assignment appear highly reliable, resulting in perfect post-quote conversion.

**Conclusion:**

All quote searches led to completed trips, showing consistently flawless post-quote fulfillment across every time slot

---

**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:**

---

Created a slicer to filter the Assembly to provide various insights based on location

- Some zones like **Ramanagaram and Yeshwantpur** show significantly higher ride volume when selected, confirming them as major demand hubs.
- Revenue trends by time period change based on the zone selected—urban zones peak in evenings, while suburban zones peak in mornings.
- Cancellation patterns vary by zone; some zones show consistently higher completion rates when the parameter is applied.

**Conclusion:**

Using a Zone Parameter(Assembly) allows dynamic, interactive analysis and reveals how ride patterns change across different geographical segments. It enhances storytelling and provides deeper insights beyond static visual

---

**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:**

---

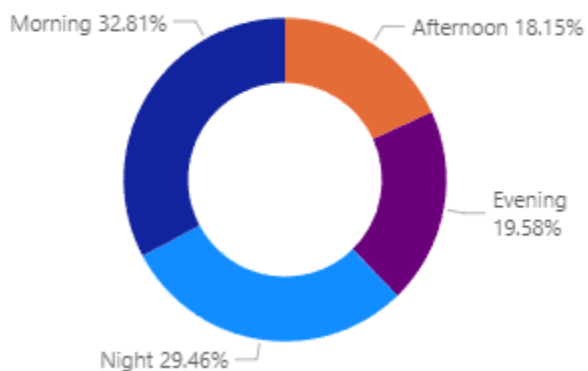
Based on the analysis and dashboards, revenue peaks in Morning and Night, with Afternoon being the lowest. High-performing zones such as Ramanagaram, Yeshwantpur, and Bangalore South consistently show strong demand. Completion rates between customers and drivers remain stable and nearly identical. Post-quote completion is 100% across all time slots, indicating perfect conversion.

### Recommendations:

1. Time-based supply planning to shift more drivers toward peak demand hours.
2. Incentives in high-demand zones to improve ETA and ride fulfillment.
3. Operational routing to direct drivers toward high-volume corridors.
4. Cancellation reduction strategies by addressing friction in the pickup process.
5. Focus on maintaining fast quote generation to retain 100% post-quote completion.

### Revenue by Time Slot:

2.3 Revenue % by Time Slot



## Top Pickup Zones:

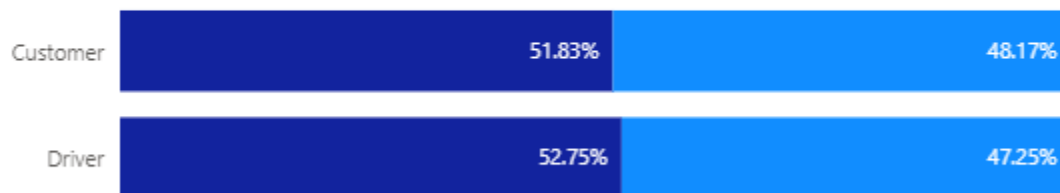
### 2.6.1 Top pick up locations



## Customer vs Driver Completion %]:

### 2.9.1 Customer vs Driver - Completed vs Cancelled (%)

Status ● Completed ● Cancelled



### 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:**

---

Analysis shows Morning and Night dominate revenue, while Afternoon and Evening need improvement. Digitally-driven payments outperform cash, and high-demand zones consistently generate strong trips and revenue. Completion after quotes is 100%, confirming a frictionless booking flow.

#### Strategies:

##### A. Marketing:

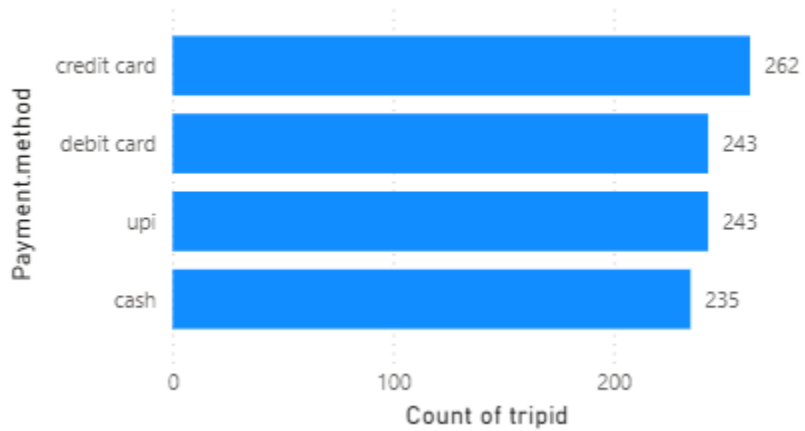
- Targeted off-peak promotions for Afternoon and Evening.
- Zone-based campaigns for Ramanagaram, Yeshwantpur, and Bangalore South.
- Digital payment incentives to accelerate repeat rides.
- Highlight "Fast Quotes" as a strong conversion lever.

##### B. Operations:

- Align driver availability with promotional windows.
- Maintain quote latency SLAs.
- Implement post-ride retention loops for low-demand periods.

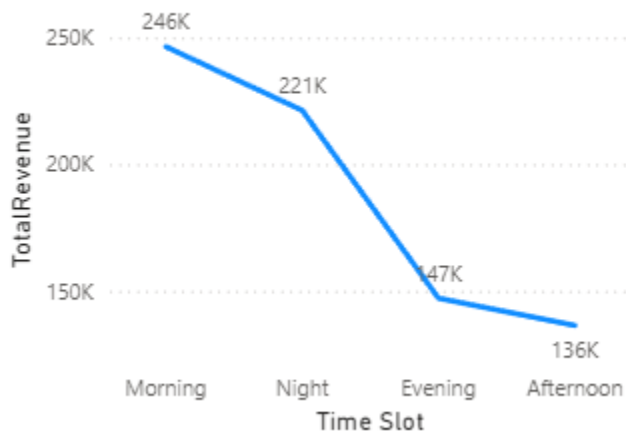
### Payment Method Distribution:

2.5 Payment Method vs Trips Taken



### Revenue by Trip Hour:

2.4 Revenue by Trip Hour



### Completion After Quote %]:



### 2.9.2 Completion % After Searching for Quotes by Time Period

