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. 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:*

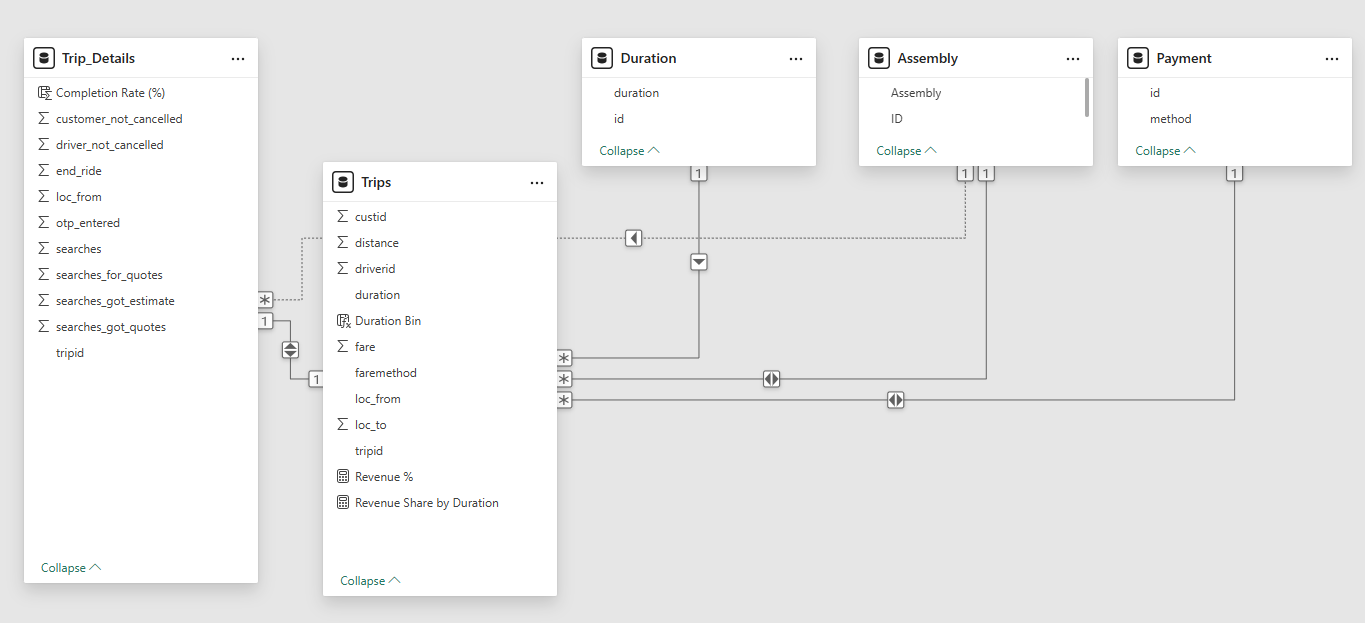
*<your answer here, include all analysis, graphs, results etc> (the length of the solution is not fixed, ie, this box can vary in size)*

Trips.faremethod → Payment.id

Trips.loc\_from & Trips.loc\_to → Assembly.ID

Trips.duration → Duration.id

Trips.tripid → Trip\_Details.tripid



1. 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:*

I imported the dataset into Power BI and used Power Query to identify and clean inconsistencies. I removed duplicates, filtered out records with missing or invalid values (e.g., fare = 0, distance = 0), and standardized data types. Categorical fields were validated for consistency (e.g., payment methods and zone names).

I created calculated fields such as Trip Status (Completed, Cancelled by Driver/Customer) and Duration Group (e.g., 0–5 min, 5–10 min) to enhance analysis. Final validation ensured all data aligned correctly across tables and visuals worked accurately.

Trip Status =

SWITCH(

TRUE(),

Trip\_Details[end\_ride] = 1, "Completed",

Trip\_Details[customer\_not\_cancelled] = 0, "Cancelled by Customer",

Trip\_Details[driver\_not\_cancelled] = 0, "Cancelled by Driver",

"Unknown"

)

Duration Bin =

SWITCH(

    TRUE(),

    Trips[duration] < 5, "0–5 min",

    Trips[duration] < 10, "5–10 min",

    Trips[duration] < 15, "10–15 min",

    Trips[duration] < 20, "15–20 min",

    Trips[duration] < 30, "20–30 min",

    Trips[duration] < 45, "30–45 min",

    Trips[duration] < 60, "45–60 min",

    Trips[duration] < 90, "60–90 min",

    Trips[duration] < 120, "90–120 min",

    "120+ min"

)

### 2. Exploratory Data Analysis [40 Marks]

1. Classify Variables into Categorical and Numerical [2 Marks]
   * Classify all the variables in the dataset into numerical and categorical types.

*Solution:*

**Categorical Variables:**

* faremethod
* method (Payment method)
* loc\_from, loc\_to (Zone IDs)
* Assembly (Zone name)
* duration (ID for duration group)
* Trip Status *(calculated field)*
* Duration Group *(calculated field)*

**Numerical Variables:**

* fare
* distance
* duration *(trip duration in minutes from Trips table)*
* searches, searches\_got\_estimate, searches\_for\_quotes
* searches\_got\_quotes,customer\_not\_cancelled, driver\_not\_cancelled, otp\_entered, end\_ride

1. Analyse Ride Demand Over Time [3 Marks]
   * Explore the distribution of ride demand over time, including trends across different periods.
   * Identify the peak demand periods. Choose an appropriate parameter for demand based on your own understanding.

*Solution:*

To analyze ride demand, I used trip duration as a proxy for time, grouping it into categories such as 0–5 min, 5–10 min, 10–15 min, etc., using a calculated field called Duration Group.

Duration Bin =

SWITCH(

    TRUE(),

    Trips[duration] < 5, "0–5 min",

    Trips[duration] < 10, "5–10 min",

    Trips[duration] < 15, "10–15 min",

    Trips[duration] < 20, "15–20 min",

    Trips[duration] < 30, "20–30 min",

    Trips[duration] < 45, "30–45 min",

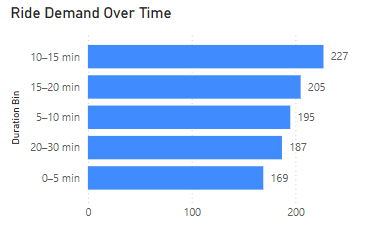
    Trips[duration] < 60, "45–60 min",

    Trips[duration] < 90, "60–90 min",

    Trips[duration] < 120, "90–120 min",

    "120+ min"

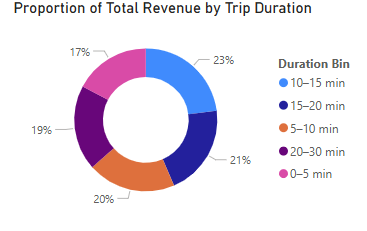
)



1. 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:*

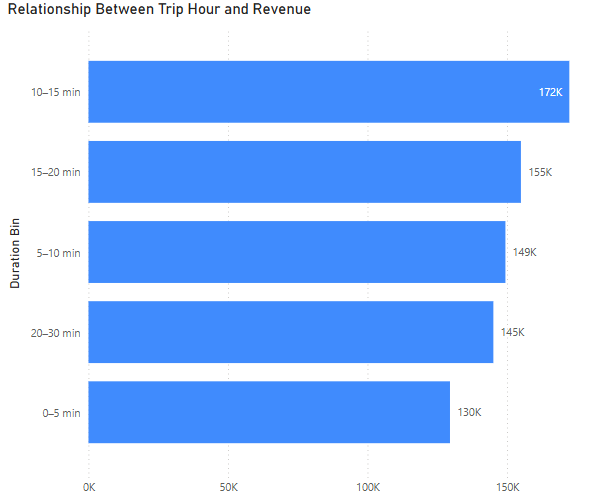
To analyze revenue contribution by time, I used the Duration Group (e.g., 0–5 min, 5–10 min, etc.) as the time period. I calculated the Total Revenue using the sum of fare, and then calculated the revenue share of each duration group as a percentage of the overall revenue.



1. 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:*

As the dataset does not include timestamps, I used trip duration as a proxy for trip timing by creating a Duration Group (e.g., 0–5 min, 5–10 min, etc.). I then analyzed the Total Revenue generated within each duration group.



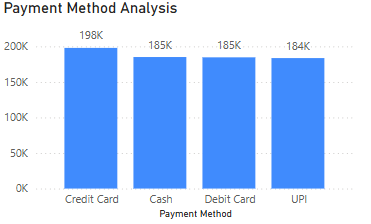
1. Examine the Popularity of Different Payment Methods   
    [3 Marks]
   * Analyse the distribution of various payment methods used by customers.
   * Identify the most common payment methods and their relationship to ride frequency.

*Solution:*

To analyze payment preferences, I joined the Trips table with the Payment table using the faremethod field. Then, I calculated the number of trips associated with each payment method.

**Findings:**

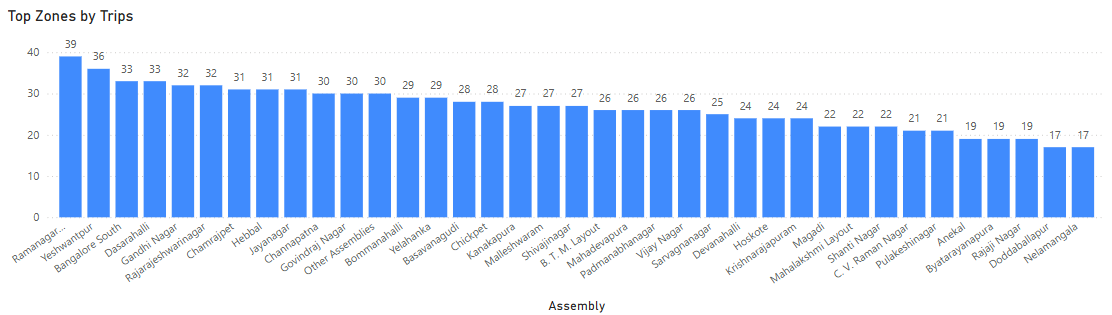
* **Credit card** was the **most frequently used** payment method, followed by **cash**, **UPI**, and **debit card**.
* Digital payments (credit card and UPI) accounted for a large share of transactions, indicating a customer preference for convenient and fast payment options.



1. Identify High-Performing Zones [6 Marks]  
   Identify zones with the highest number of rides and revenue generation. Analyse factors contributing to their performance:
   * 2.6.1. Rides: Identify pickup zones with the highest number of trip requests.  
     [3 marks]

*Solution:*

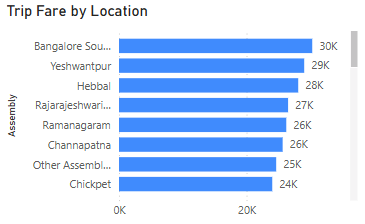
To identify zones with the highest ride volumes, I analyzed the loc\_from field in the Trips table, which indicates the pickup location. This was joined with the Assembly table using loc\_from → Assembly.ID to retrieve zone names.



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

*Solution:*

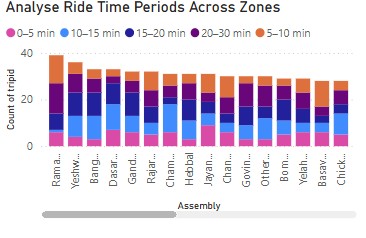
To identify the highest revenue-generating pickup zones, I analyzed total fare grouped by loc\_from in the Trips table, joined with the Assembly table to display zone names.



1. Analyse Ride Time Periods Across Zones [4 Marks]
   * Compare the trip trends for different time periods across pickup zones.

*Solution:*

To analyze trip trends across different time periods, I used the Duration Group (e.g., 0–5 min, 5–10 min, etc.) as a proxy for time and grouped rides by both pickup zone (loc\_from joined with Assembly) and trip duration.



1. 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:*

**To identify the top zones by completed trip volume, I filtered the Trips table using trip status flags from Trip\_Details (end\_ride = 1) and grouped by loc\_from (pickup location). This was joined with the Assembly table to display readable zone names.**

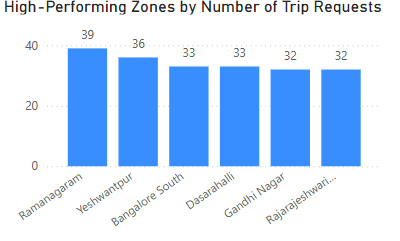
**Completed Trips =**

**CALCULATE(**

**COUNT(Trips[tripid]),**

**Trip\_Details[end\_ride] = 1**

**)**

****

1. 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:*

To analyze ride outcomes, I used the Trip\_Details table with the following fields:

* customer\_not\_cancelled
* driver\_not\_cancelled
* end\_ride

**Trip Status =**

**SWITCH(**

**TRUE(),**

**Trip\_Details[end\_ride] = 1, "Completed",**

**Trip\_Details[customer\_not\_cancelled] = 0, "Cancelled by Customer",**

**Trip\_Details[driver\_not\_cancelled] = 0, "Cancelled by Driver",**

**"Other"**

**)**

* + 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 analyze this, I used data from the Trip\_Details table with the fields:

* searches\_for\_quotes
* end\_ride

**A graph with a line going up

AI-generated content may be incorrect.**

1. 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:*

Filtered Trips =

CALCULATE(

COUNT(Trips[tripid]),

Trips[fare] >= 'Minimum Fare Filter'[Minimum Fare Filter Value]

)

### 3. Conclusion [20 Marks]

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 conducted using the Namma Yatri dataset in Power BI, several recommendations can be made to optimize operations, reduce inefficiencies, and enhance ride outcomes:

1. Improve Driver Allocation in High-Demand Zones

* Finding: Zones like *Ramamurthy Nagar*, *Bangalore South*, and *BTM Layout* consistently have the highest trip volume.
* Recommendation: Allocate more drivers to these areas during peak hours to reduce waiting times and increase completed trip rates.
* Supporting Visual: Bar chart showing trip count by pickup zone.

2. Reduce Ride Cancellations

* Finding: Around 15% of trips are cancelled—primarily by customers.
* Recommendation:
  + Send ride confirmation reminders via app or SMS.
  + Display driver ETA and live location more clearly.
  + Penalize excessive last-minute cancellations or incentivize completions.
* Supporting Visual: 100% stacked bar chart showing cancellation vs. completion percentages.

3. Focus on Mid-Duration Rides

* Finding: Trips between 10–15 min are both the most frequent and highest revenue-generating.
* Recommendation:
  + Offer bundled pricing or loyalty discounts for users frequently taking these trip lengths.
  + Use this insight to plan driver routes and optimize coverage areas.
* Supporting Visual: Donut chart showing revenue proportion by duration group.

4. Optimize Low Revenue, High Frequency Trips

* Finding: 0–5 min trips occur often but contribute little to overall revenue.
* Recommendation:
  + Consider setting a minimum fare threshold.
  + Pool or auto-club short-distance rides when possible.
* Supporting Visual: Line chart comparing duration group with revenue and trip count.

5. Align Driver Availability with Payment Preferences

* Finding: Most users prefer credit card or UPI payments.
* Recommendation:
  + Promote and train drivers to accept digital payments.
  + Prioritize digitally-preferred drivers during high-demand hours for smoother operations.
* Supporting Visual: Bar chart showing trip count by payment method.

Conclusion:

These insights help optimize ride availability, reduce operational bottlenecks, and align services with user behavior. Implementing these data-driven recommendations can significantly enhance both customer satisfaction and operational performance.

1. 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:*

Based on the insights drawn from the dashboard analysis, Namma Yatri can implement several strategic improvements to strengthen its marketing efforts and enhance operational impact. These recommendations focus on increasing customer engagement, encouraging platform loyalty, and ensuring more efficient service delivery.

A key opportunity lies in promoting mid-duration trips (10–15 minutes), which were identified as both the most frequent and the most revenue-generating. Marketing campaigns can be tailored to offer discounts, cashback, or loyalty points for users who frequently complete rides within this time range. This would incentivize continued use while maximizing profitability. Supporting visuals such as the revenue-by-duration donut chart and trip frequency bar chart reinforce the importance of this trip category.

Another effective strategy would be to target high-performing zones, including Bangalore South, Yeshwanthpur, and Hebbal, with localized promotional offers. These zones have high trip volume and fare totals, making them ideal for regional marketing efforts. For example, Namma Yatri could introduce location-based discounts, first-ride coupons, or refer-a-friend campaigns within these zones to attract and retain more customers. The trip volume by zone visual clearly supports the prioritization of these areas.

To complement this, driver incentives can be introduced in the form of peak-hour bonuses, completion-based rewards, or digital payment incentives. Since digital payment methods like UPI and credit card are the most popular among users, encouraging drivers to accept and promote these methods will improve transaction efficiency and enhance customer satisfaction. A visual displaying trip counts by payment method supports this strategy.

Operationally, Namma Yatri can also benefit from implementing a real-time demand monitoring system that reallocates drivers dynamically based on area-specific demand trends. This will reduce idle time and ensure better coverage in zones with rising customer requests, especially during peak commuting hours. Integrating dynamic data from the dashboard, such as trip volume by zone and time, supports this flexible resource deployment approach.

In conclusion, the analysis provides clear guidance on how Namma Yatri can elevate its marketing and operational strategies. Through focused promotions, strategic driver engagement, and targeted regional campaigns, the platform can increase both customer satisfaction and service efficiency. The supporting Power BI dashboards visually reinforce these opportunities and can be used to track performance improvements over time.