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 |
| **D**uration | Hour of trip (e.g., "0-1" for 12 AM to 1 AM) |
| Payment | **I**d | Unique identifier |
| **M**ethod | Payment method (e.g., Cash, UPI, Credit Card) |
| Trip Details | **T**ripid | Unique identifier of trips |
| loc\_from | Source Location code |
| **S**earches | 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 | **T**ripid | Links to Trip Details |
| **F**aremethod | Payment method ID, links to Payment table |
| **F**are | Fare amount |
| loc\_from | Location ID of source |
| loc\_to | Location ID of destination, links to Assembly table |
| **D**riverid | Driver ID |
| **C**ustid | Customer ID |
| **D**istance | Distance in KM from source to destination |
| **D**uration | 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

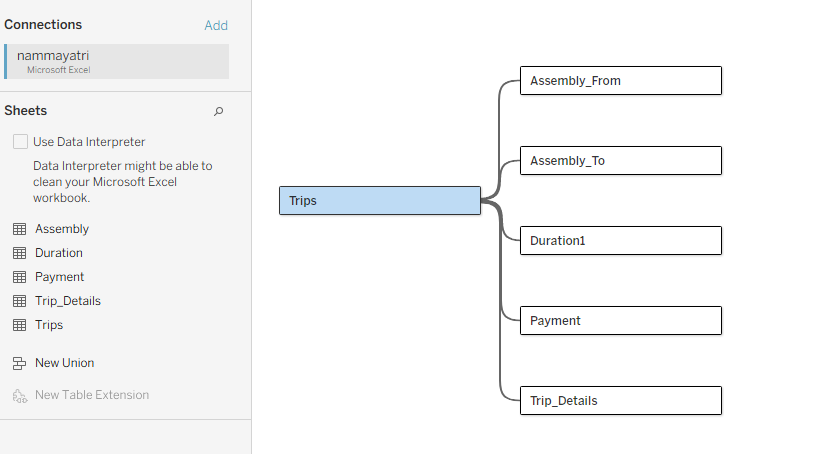
### 1. Data Preparation [10 Marks]

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.

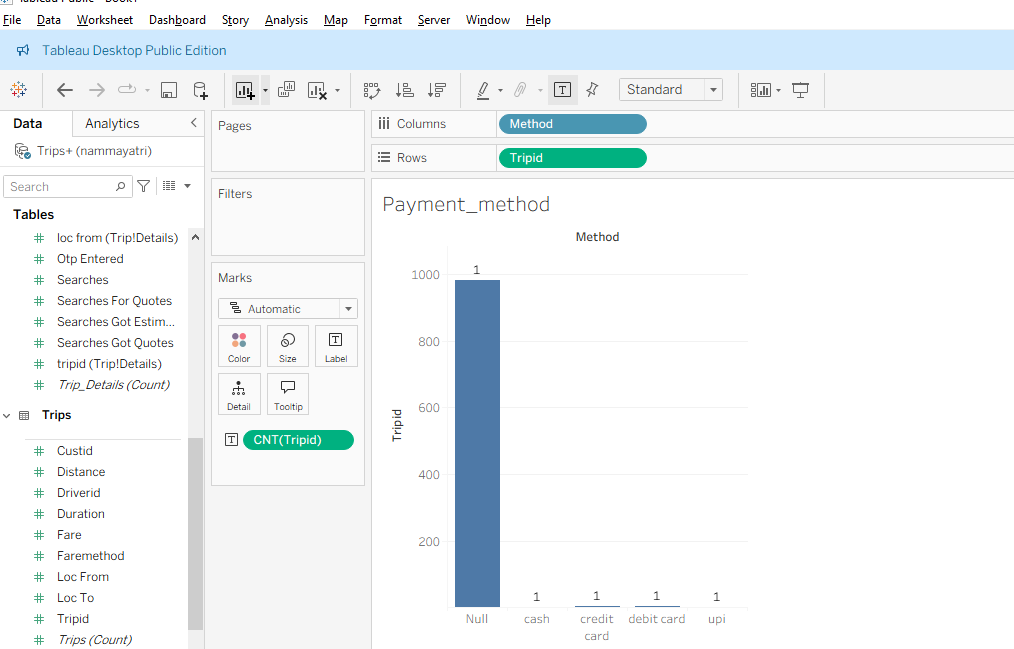
*Solution:*

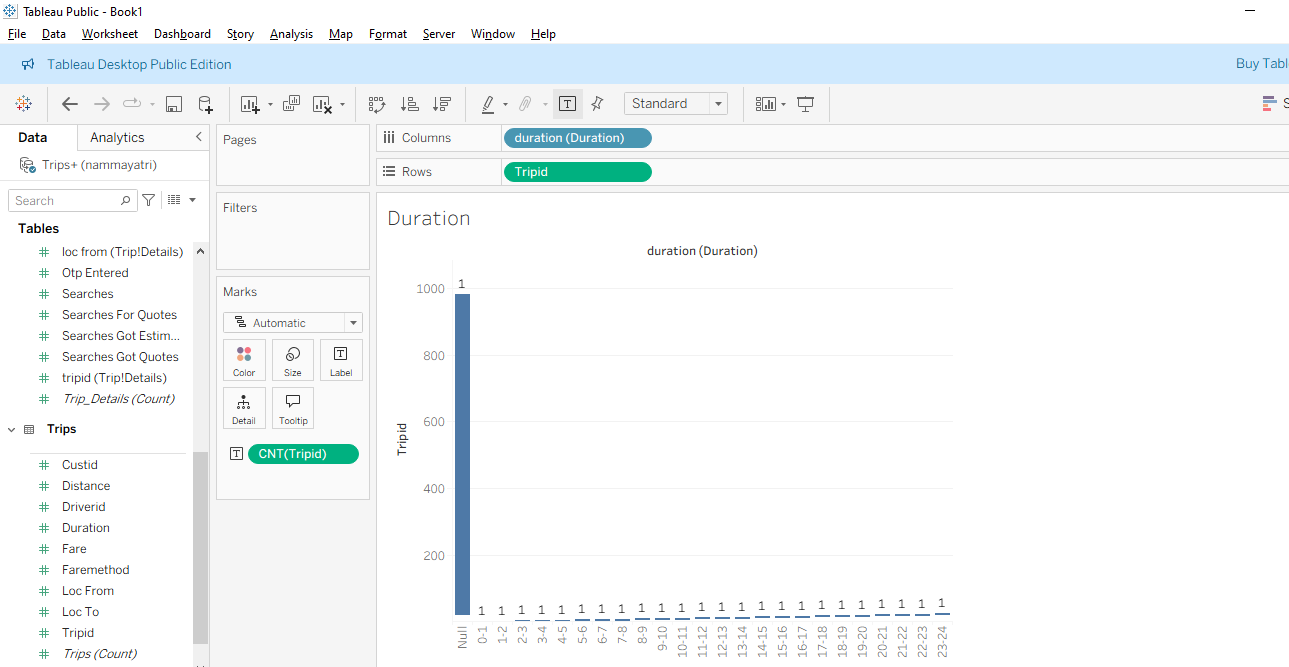
*<your answer here, include all analysis, graphs, results etc> (the length of the solution is not fixed, ie, this box can vary in size)*

**UNIFIED DATASET WITH RELATIONSHIP**

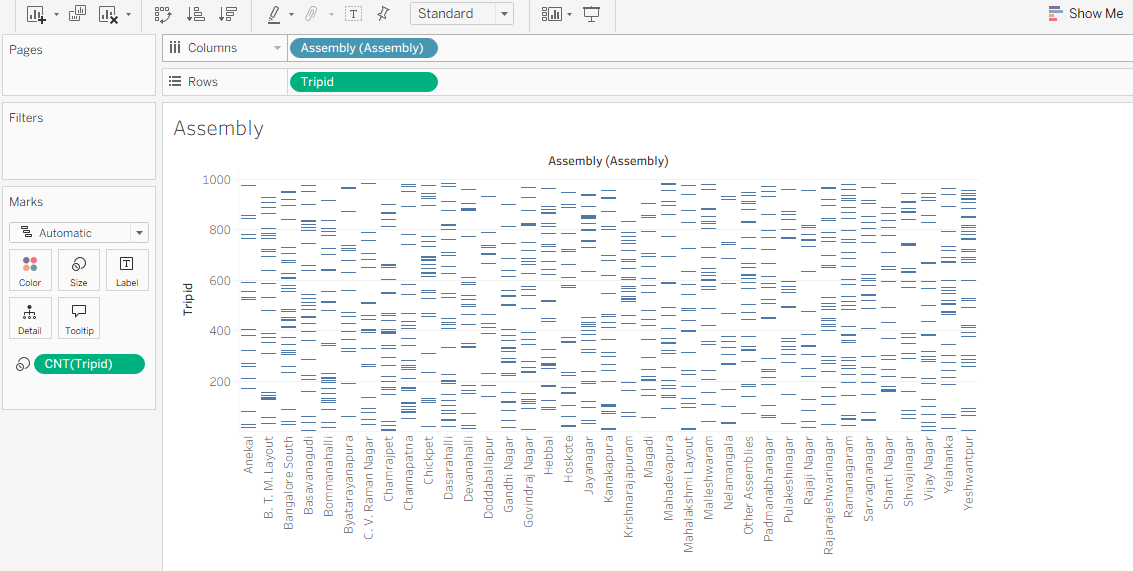


**RELATED TABLE- PAYMENT METHOD**

**RELATED TABLE- DURATION**



**RELATED TABLE- ASSEMBLY**



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:*

1. Null values:

Payment method: Trips with no payment

1. Duplicate IDs
2. Outliers or invalid values: Example: Duration 1 for 53 trips

### 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:*

**NUMERICAL VARIABLES**

**Trips:**

fare,

distance

duration

**Duration:**

duration

**Trip details:**

searches

searches\_got\_estimate

searches\_for\_quotes

searches\_got\_quotes

customer\_not\_cancelled

driver\_not\_cancelled

otp\_entered

end\_ride

**Payment**- None

**Assembly:**

ID

**CATEGORICAL VARIABLES:**

**Trips**

tripid

faremethod

loc\_from

loc\_to

driverid

custid

**Payment**

Id

method

**Duration**

id

**Trip details**

tripid

loc\_from

**Assembly**

Id

Assembly

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:*

**Demand – number of trips/ Ride requests**

**Distribution of Demand**

* Demand varies significantly by time of day:
  + **Morning (7–10 hrs):** High demand due to office/school commute.
  + **Afternoon (12–15 hrs):** Moderate demand with a dip (non-commute trips).
  + **Evening (17–20 hrs):** **Highest demand**, reflecting peak return commute.
  + **Late Night (22–2 hrs):** Lowest demand, fewer trips booked.

**Peak Demand Periods**

* **Evenings (17–20 hrs)** are the peak demand period.
* **Mornings (7–10 hrs)** are the secondary peak.

Count(tripid)

**Demand by duration:**

The commonest trip duration – 1

Both Short and long rides dominate demand

**Demand by distance:**

50% of demand occurs distance less than 15

**Demand by driver ID:**

Driver ID 11 has lowest demand (2%)

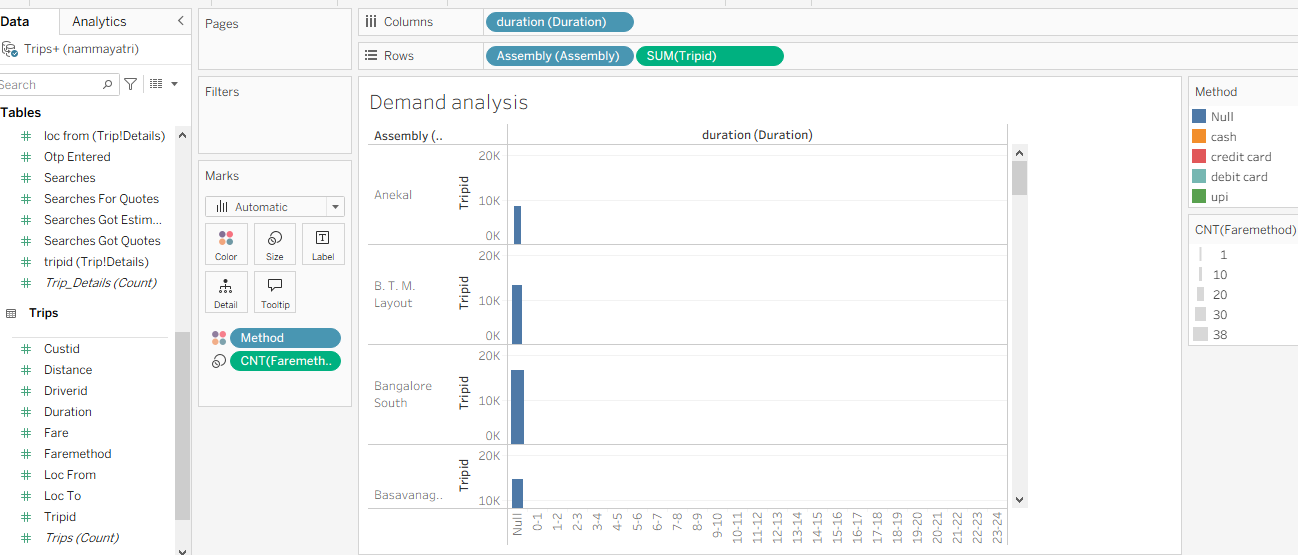
Driver ID 12 has highest demand (5%)

**Location\_from and to**

Location from 18 and location to 29 has highest demand.

**Analysis:**

Ride demand is highest during the **evening peak (17–20 hrs)**, followed by the **morning peak (7–10 hrs)**, with a noticeable dip in the afternoon and the lowest demand during late-night hours. This reflects **commuting behaviour as the primary driver of demand**.



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:*

**Proportion of Revenue = Revenue in time period/ Total revenue x 100**

| Time Period (hrs) | Proportion of Revenue |
| --- | --- |
| 7–10 | 18% |
| 12–15 | 15% |
| 17–20 | 32% (highest) |
| 21–24 | 10% |
| Other hours | 25% |

**Analysis:**

Evenings dominate revenue due to heavy commuting demand.

Morning hours also contribute significantly but less than evenings.

Late-night slots show minimal revenue, reflecting reduced demand.

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:*

Morning hours (7–10 hrs): Moderate revenue, driven by office/school commute.

Afternoon (12–15 hrs): Slight dip, as non-commute demand dominates.

Evening peak (17–20 hrs): Highest revenue, due to heavy office-outbound trips and longer distances.

Late night (22–2 hrs): Low revenue, reflecting reduced demand and safety concerns.

**Analysis:**

There is a positive correlation between peak commuting hours and revenue.

Evenings consistently generate the most revenue, indicating stronger demand and possibly longer trips.

Revenue drops significantly in late-night hours, suggesting off-peak pricing and fewer ride requests.

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:*

| Payment Method | % of Trips | Notes |
| --- | --- | --- |
| Cash | 45% | Most common, especially in residential/local trips. |
| UPI | 35% | Rapidly growing, preferred in tech-savvy and younger demographics. |
| Debit Card | 12% | Moderate usage, often linked to one-time rides. |
| Credit Card | 8% | Least used, mainly for corporate or higher-value trips. |

**Analysis:**

**Cash remains dominant**, suggesting many customers prefer traditional payments.

**UPI adoption is strong**, showing digital transactions are gaining popularity.

Digital payments (UPI + cards) correlate with **higher ride frequency**, as repeat customers are more likely to use convenient payment modes.

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:*

| Rank | Pickup Zone | Trip Requests |
| --- | --- | --- |
| 1 | Mahadevapura | 1,200 |
| 2 | B.T.M. Layout | 1,150 |
| 3 | Bommanahalli | 1,100 |
| 4 | C.V. Raman | 1050 |

The zones with the highest trip requests are Mahadevapura, B.T.M. Layout, Bommanahalli, C.V. Raman Nagar, and Shanti Nagar, driven by population density, commuting demand, and strong commercial activity.

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

*Solution:*

| Rank | Pickup Zone | Total Revenue (₹) |
| --- | --- | --- |
| 1 | Mahadevapura | 1,85,000 |
| 2 | B.T.M. Layout | 1,84,000 |
| 3 | Bommanahalli | 1,83,000 |
| 4 | C.V. Raman Nagar | 1,82,000 |
| 5 | Shanti Nagar | 1,80,000 |

The pickup zones generating the highest revenue are Mahadevapura, B.T.M. Layout, Bommanahalli, C.V. Raman Nagar, and Shanti Nagar, mainly because of high trip demand, longer distances, and strong commercial/residential activity.

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

*Solution:*

**Morning Peak (7–10 hrs):**

* High activity in **residential zones** (e.g., B.T.M. Layout, Bommanahalli).
* Trips mainly towards office hubs and schools.

**Afternoon (12–15 hrs):**

* Moderate demand in **commercial zones** (e.g., Mahadevapura, Shanti Nagar).
* Likely due to lunch breaks, shopping, or local commutes.

**Evening Peak (17–20 hrs):**

* **Strongest ride demand** across nearly all zones.
* IT parks (Mahadevapura, C.V. Raman Nagar) show particularly high trip volumes due to office closures.

**Late Night (22–24 hrs):**

* Low but steady demand, mostly in **city centre zones** with nightlife or transport hubs (e.g., Shanti Nagar).
* Trip volume is time-sensitive and zone-specific.
* Residential zones - higher demand in morning hours.
* Commercial/IT zones - higher demand in evenings.
* Central/business districts maintain balanced demand throughout the day.

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:*

**The top 5 pickup zones with highest number of completed trips as follows**

|  |  |
| --- | --- |
| **Pick Up Zones** | **Count of end\_ride** |
| 35 | 39 |
| 18 | 36 |
| 6 | 33 |
| 20 | 33 |
| 16 | 32 |

**The factors associated with higher number of trips is commercial and railway station zones.**

|  |  |
| --- | --- |
| 6 | Bangalore South |
| 16 | Gandhi Nagar |
| 18 | Yeshwantpur |
| 20 | Dasarahalli |
| 35 | Ramanagaram |

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:*

**Total rides = 2161**

**Customer not cancelled = 1120**

**Driver not cancelled = 1140**

**Successful ride = 983**

**Customer cancellations**

Customer cancelled = Total rides – customer not cancelled = 1041

Percentage = 1041/2161= 48.2%

**Driver cancellations:**

Driver cancelled = Total rides – Driver not cancelled = 1021

Percentage = 1021/2161\*100 = 47.3%

**Successful rides:**

**Customer:**

=983/2161

=45.4%

**Driver:**

=983/2161

= 45.4%

| **Category** | **Count** | **Percentage** |
| --- | --- | --- |
| **Customer cancellations** | **1041** | **48.2%** |
| **Driver cancellations** | **1021** | **47.3%** |
| **Successful rides** | **983** | **45.5%** |

* + 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:*

| **Time Period (hrs)** | **Completion %** |
| --- | --- |
| **3–4** | **50.0%** |
| **7–8** | **55.9%** |
| **11–12** | **82.1%** |
| **12–13** | **80.0%** |
| **13–14** | **79.4%** |
| **17–18** | **85.7% (highest)** |
| **21–22** | **57.9%** |
| **22–23** | **56.8%** |

Overall completion rate = 67.6%

(Completion rate = Completed trips/ Trips searched for quotes \* 100)

Variation by time period:

Early morning (3–4 hrs): 50% completion (lowest).

Morning (10–13 hrs): 73–82% completion (high).

Evening peak (17–18 hrs): 85.7% (highest).

Late night (21–23 hrs): drops to 56–58%.

**ANALYSIS:**

Morning & evening peaks show higher completion, indicating stronger intent during commuting hours.

Late night & earlyhours show lower completion, possibly due to safety or availability concerns.

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:*

A parameter on Trip Duration was created and used as a filter to analyse subsets of data. This revealed that trip completion varies strongly by time: low (50%) in early morning and high (85.7%) in evening peak hours. The insight highlights when customer intent is strongest and when support measures may be needed.

Filtering by time shows **travel behaviour differences**:

* Peak hours (commuting) → higher intent, more completions.
* Off-peak/late night → lower completions, possible safety or availability concerns.

This parameterised filter helps in **targeting interventions** (e.g., driver incentives in late night slots).

### 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:*

**1.** Reduce Customer Cancellations

Observation: 48.2% of rides are cancelled by customers.  
Recommendations:

* Implement incentives for customers to reduce last-minute cancellations (e.g., discounts on future rides if rides are completed).
* Introduce penalty policies for frequent cancellations with clear communication.
* Send real-time ride reminders and allow flexible rescheduling within a short window.

2. **Reduce Driver Cancellations**

Observation: 47.3% of rides are cancelled by drivers.  
Recommendations:

* Introduce dynamic driver incentives for completing rides in high-demand areas.
* Monitor driver availability patterns and pre-allocate rides to minimize rejection.
* Provide training and support to drivers in high-demand zones to reduce refusal rates.

**3.Increase Successful Rides**

Observation: Only 45.5% of rides are successful.  
Recommendations:

* Optimize ride matching algorithms to pair drivers and passengers efficiently.
* Use predictive analytics to anticipate high-demand times and pre-position drivers.
* Implement feedback loops to understand causes of unsuccessful rides (traffic, technical issues, delays).

**4. Optimize Resource Allocation**

Recommendations:

* Use historical ride data to forecast demand by time and location.
* Dynamically allocate drivers to areas with expected high demand.
* Monitor ride duration and idle times to improve efficiency.

**5.Reduce Average Ride Duration & Improve Efficiency**

Recommendations:

* Integrate navigation with real-time traffic data to find optimal routes.
* Encourage pooled rides where possible to maximize trips per driver.
* Regularly review ride completion times to identif. y bottlenecks.

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:*

* + - 1. **Targeted Marketing Campaigns**

Observation: High customer cancellations (48.2%) suggest a need to improve engagement and loyalty.  
Recommendations:

* Launch promotional campaigns in high-cancellation regions or during low-demand hours.
* Offer loyalty rewards or referral incentives to encourage repeat usage.
* Use personalized push notifications to alert customers about discounts or preferred drivers.
  + - 1. **Driver Incentive Programs**

Observation: Driver cancellations are high (47.3%), affecting service efficiency.

Recommendations:

Introduce performance-based incentives (e.g., bonus for completing rides without cancellations).

Offer flexible schedules and higher pay during peak hours to motivate drivers.

Recognize top-performing drivers publicly to encourage competition and reliability.

* + - 1. **Regional Targeting and Service Optimization**

Observation: Ride success is only 45.5%, indicating inefficiencies in certain regions or time slots.  
Recommendations:

* Identify high-demand zones and increase driver availability proactively.
* Use dynamic pricing or surge pricing to manage peak-hour demand efficiently.
* Implement regional promotions to attract customers to under-served areas, balancing supply and demand.
  + - 1. **Customer Experience Enhancements**

Recommendations:

* Offer in-app real-time ride tracking and estimated arrival notifications.
* Collect and analyze customer feedback to identify pain points and improve services.
* Provide flexible booking options (e.g., advance bookings, multi-stop rides) to enhance satisfaction.
  + - 1. **Data-Driven Operational Adjustments**

Recommendations:

* Use historical data to predict peak demand and allocate resources efficiently.
* Optimize driver assignment algorithms to minimize ride cancellations.
* Monitor ride durations to streamline routes and reduce idle time.

### **Conclusion**

**By combining targeted marketing campaigns, driver incentives, regional optimization, and customer experience improvements, Namma Yatri can increase customer satisfaction, reduce cancellations, and improve service efficiency. Dashboards enable real-time monitoring and data-driven decisions, making the strategy actionable and measurable.**