

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:

Step1: Imported the Namma Yatri dataset into Tableau visualisation tool. The following are the tables that are imported

1. Trips: Main Transactional table which contains Trip metadata and revenue
2. Trip Details: Shows Customer interaction journey from search to ride completion.
3. Assembly: Maps location codes to actual zone names
4. Duration: Time ranges like "0-1", "1-2", etc...
5. Payment: Payment Methods like Cash, UPI, Debit and Credit cards

Step2: A star schema is created with the Trips as a Main table i.e., central fact table and make relationship between between them to create a unified dataset for our further analysis.

Step 3:Field Mapping from each table are as follows:

1. Trips [Tripid] → Trip Details [Tripid]
2. Trips [Faremethod] → Payment [Id]
3. Trips [Duration] → Duration [id]
4. Trips [loc_to] → Assembly [ID]

The unified dataset will enable seamless analysis across fare values, payment types, locations, ride timings and user-driver interactions

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:

1. Data Cleaning steps are performed to ensure quality and consistency

- **Handling Missing and Null Values (If any):**
 - Filtered out the columns like Fare, distance, duration as these are critical for revenue and time-based analysis. No missing / Null values present in the data. Total records pertaining to completed trip details available are 983
 - Fields like otp_entered, end_ride were assumed to be 0 i.e., incomplete ride flow.
- **Data Type Corrections (if any):**
 - Ensure that the numeric columns are assigned to correct numeric data types
 - Ensure that the categorical columns are assigned to correct data type and were properly classified for grouping and filtering.

2. Duplicate or Irrelevant records:

- Verifying the uniqueness in the Trips table by checking the tripid in the Trips table.
- No Duplicate records were found in the dataset.

3. Calculated Fields created for analysis: To support meaningful insights and enable precise filtering in Tableau, created several calculated fields from the raw columns. These fields will add interpretability, consistency, and depth of the visualisations. The calculated fields created in Tableau are as follows:

1. Rides Completed:

IF [End Ride] = 1 THEN "Yes" ELSE "No" END

- Purpose: Used to Identify completed trips for funnel and conversion analysis.
- Use case: Segmenting rides into completed vs drop-off for performance metrics.

2. OTP entered status:

IF [Otp Entered] = 1 THEN "Started" ELSE "Not Started" END

- Purpose: Used to indicate rides initiation.

- Use case: Used in funnel analysis from quote to ride start.

3. Cancelled by Customer:

IF [Customer Not Cancelled] = 0 THEN "Cancelled" ELSE "Not Cancelled" END •

Purpose: Used to identify rides cancelled by the customer.

- Use case: Helps analyze operational reliability across zones or time slots.

4. Cancelled by Driver:

IF [Driver Not Cancelled] = 0 THEN "Cancelled" ELSE "Not cancelled" END

- Purpose: Used to identify rides cancelled by driver.

- Use case: Helps analyze operational reliability across zones or time slots.

5. Revenue Per KM:

[Fare] / [Distance]

- Purpose: Normalizes fare based on distance.

- Use case: Helps to evaluate profitability across trip lengths.

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:

Categorical Variables (Dimensions): categorical variables are shown below:

1.Assembly: Specific location Zone name from Assembly table

2.Payment method: Name of Payment method from Payment table

3.Duration: Hourly trip details from Duration table

4.Rides Completed: Calculated field Shows the rides completed / not

5.OTP Entered Status: Calculated field shows whether the rides started or not

6.Cancelled by Customer: Calculated field to identify rides cancelled or not

7.Cancelled by Driver: Calculated field to identify the rides cancelled or not Is

8.Revenue Trip: Calculated field to identify valid revenue generated trip Trip id:

9.Unique Trip Identifier Driver Id, Cust ID: Identifiers for Driver and customer

10.Duration ID: Time slot Identifier

11.loc_from, loc_to: Origin and destination assembly codes

12.Otp_entered, End Ride: Indicators for ride progression

13.Searches Got Estimate, Searches Got Quote: Search and Quote response

Numerical Variables (Measures): numerical variables are shown below:

1.Fare: Final trip fare

2.Distance: Trip distance in Kilometers

3.Revenue Per KM: Calculated field for trip efficiency

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: The objective is to analyse the variation in the ride demand across different times of the day. Here, we have to identify the peak periods, and derive actionable insights particular to the demand.

• The number of “Searches” from the Trip details table was selected as the primary indicator of ride demand. •

“Search” indicates the customer intent to book a ride i.e., Trip Request count. • This is the starting point in the ride even if the ride is not completed.

• Other metrics like “Searches got estimate” and “Searches got quotes” also can be reviewed to assess with respect to demand fulfilment

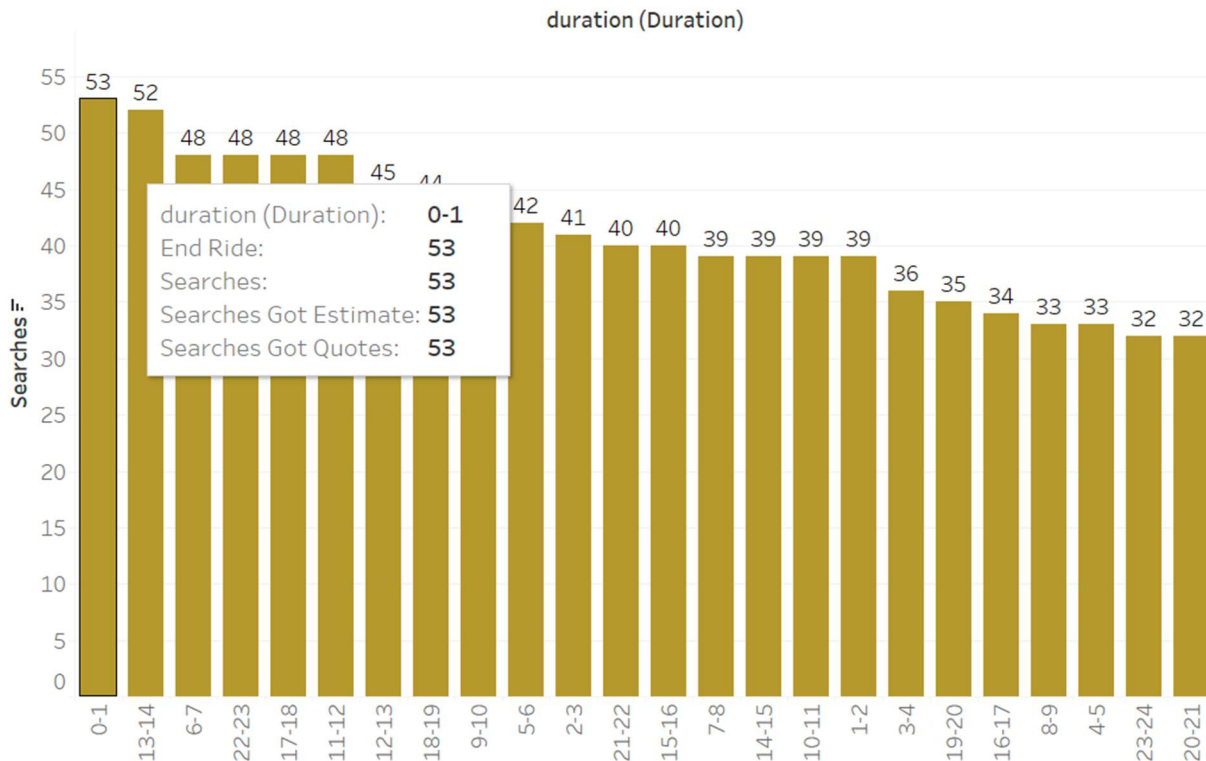
• The “duration” in Trips tables gives us the individual hours of trip of ridedemand over time.

After plotting the bar plot on Duration vs Searches below are the observations:

The peak demand was observed during the hours at:

- 12 AM – 1 AM
- 1 PM – 2 PM
- 6 AM - 7 AM
- 11 AM – 12 PM
- 5 PM - 6 PM
- 10 PM -11 PM

RIDE DEMAND



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:

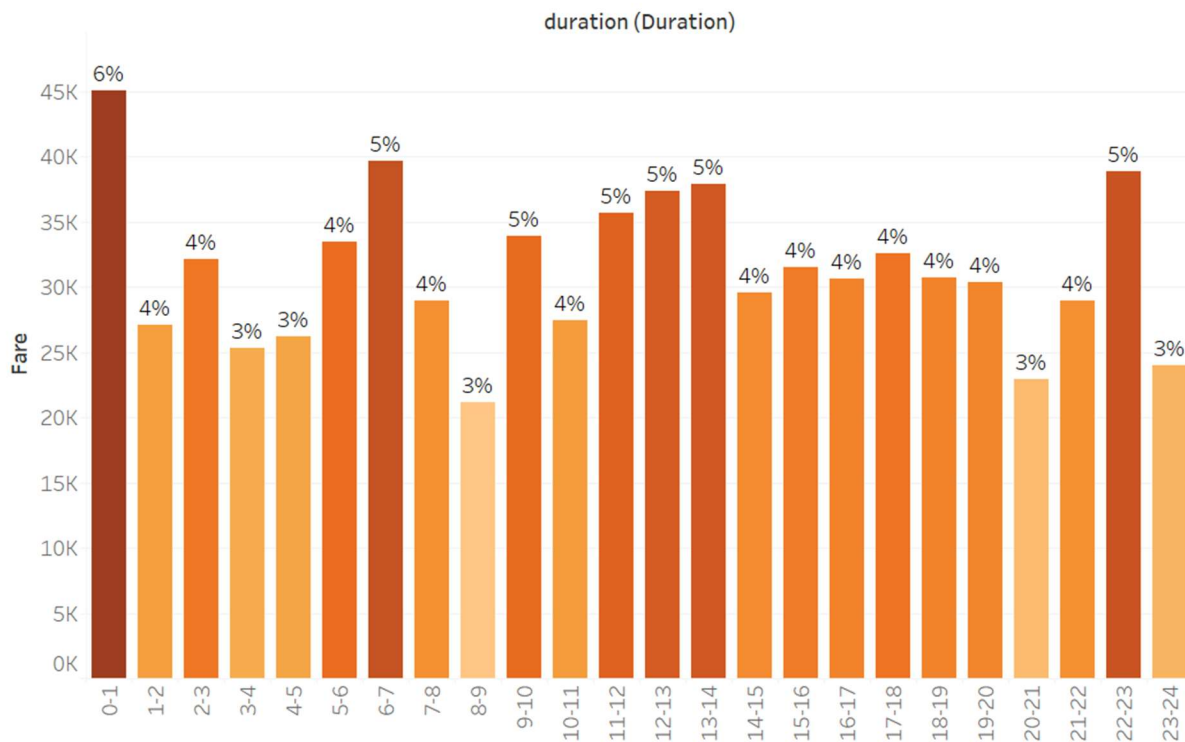
The objective is to analyze the contribution of each hourly time slots to the total revenue and to identify the most profitable periods of the day.

- The "Fare" column from the Trips table was used to represent the revenue
- The "Duration" gives the different time periods

To compute the proportion of total revenue generated by each time slot is calculated using the formula:

$$\text{Revenue \%} = \text{SUM}([\text{Fare}]) / \text{TOTAL}(\text{SUM}([\text{Fare}]))$$

TOTAL REVENUE PROPORTION FROM DIFFERENT TIME PERIODS



The time slots with peak hours generating high revenue (12 AM – 1 AM, 6 AM – 7 AM, 10 PM – 11 PM)

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.

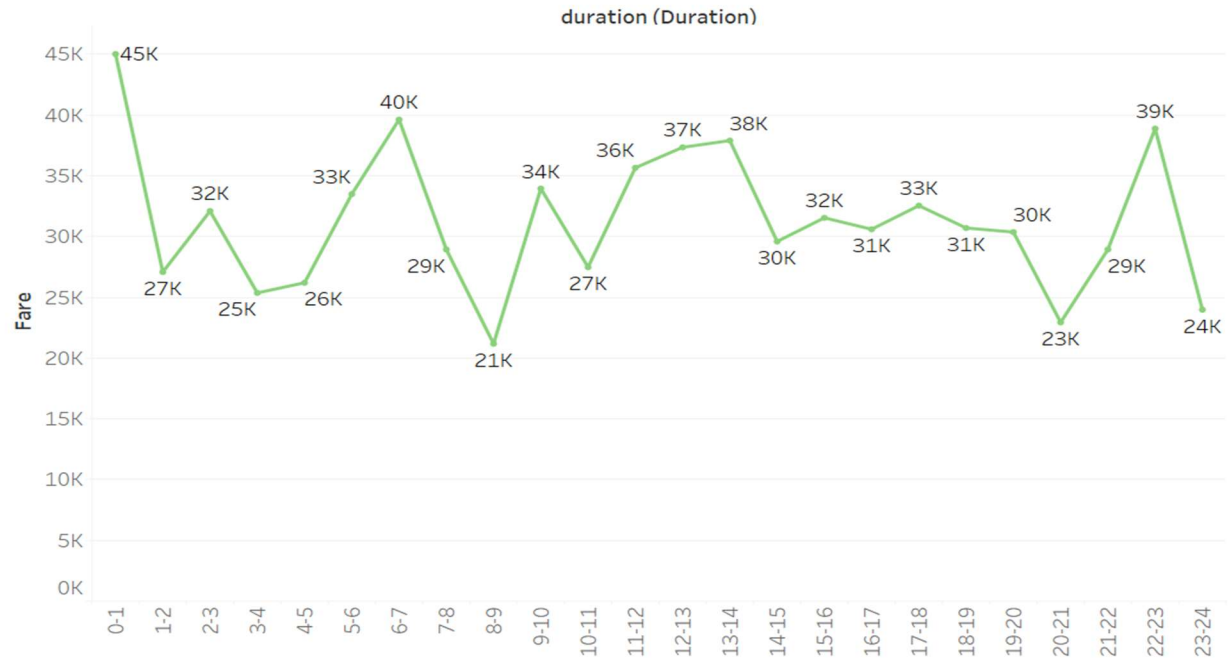
Solution:

The objective is to understand how the Trip hour influences Total Revenue and identify specific time slots that generate the highest earnings.

The timeslots that generate highest revenue trend are

- 12 AM – 1 AM- 45K earning
- 6 AM - 7 AM – 40K earning
- 11 AM – 2 PM – 36K earning
- 9 PM – 11 PM – 39K earning

TRIP HOUR - FARE



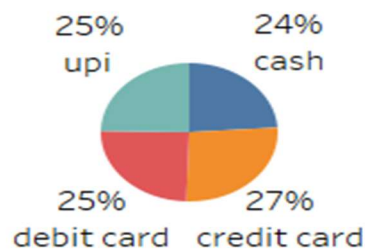
2.5. Examine the Popularity of Different Payment Methods [3 Marks]

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

Solution:

The objective is to identify how frequently different payment methods are used by customers. Identify the most preferred mode of payment.

PAYMENT METHOD USED



PAYMENT METHOD	% TOTAL No. OF RIDES	CONTRIBUTION
CREDIT CARD	27%	77% DIGITAL TRANSACTION
DEBIT CARD	25%	
UPI	25%	
CASH	24%	24% CASH

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]

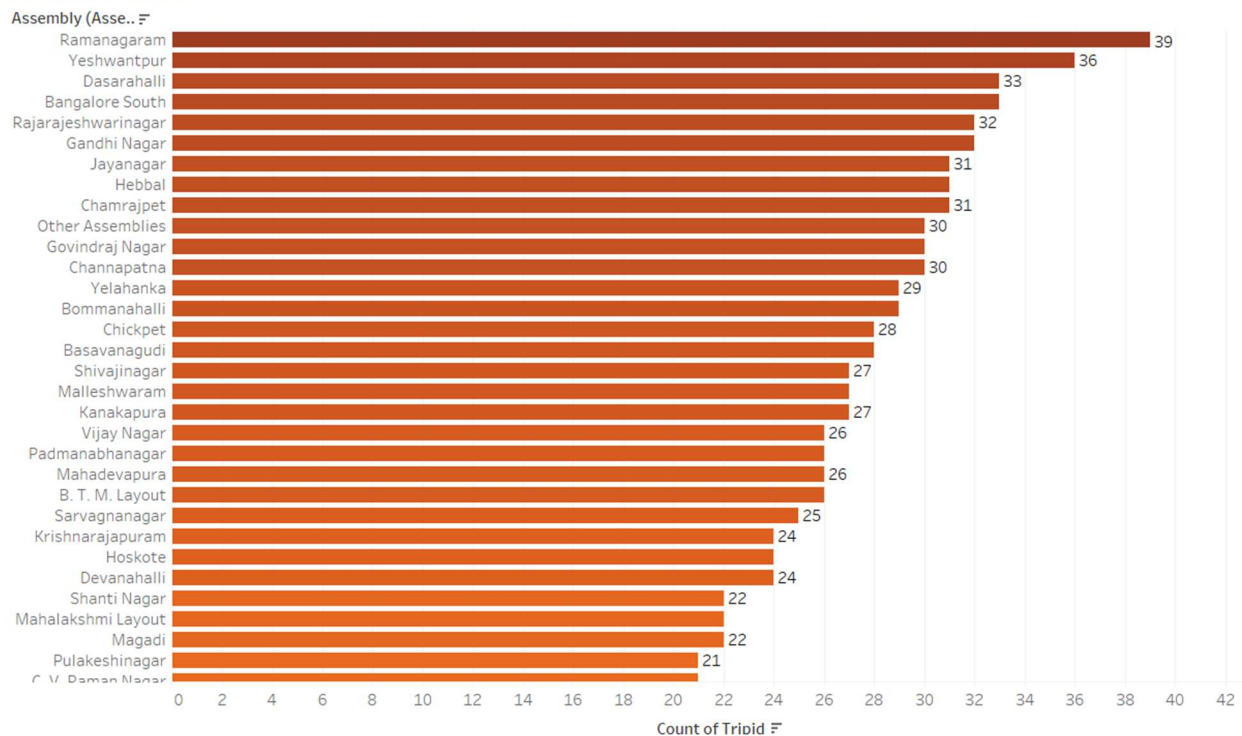
Solution:

The objective is to identify Pickup zones (Origins) that generate the highest number of trip requests which helps in understanding demand distribution across city.

RANK	PICK UP LOCATION	No. OF TRIPS
1.	Ramanagaram	39
2.	Yeshwantpur	36
3.	Dasarhalli	33
4.	Banglore south	32
5.	Rajarajeshwarinagar	32

- Ramanagaram and Yeshwantpur are the leading zones in terms of pick-up activity.
- A broad mid-tier group including zones from Sarvagnanagar to Jayanagar shows healthy ride activity from 25 to 31 trips
- Pick-up location from Krishnarajapuram to Doddaballapur shows lower ride volumes suggesting either-
 - Lower demand
 - Lower app penetration
 - Insufficient driver presence

LOCATION-TOTAL RIDES

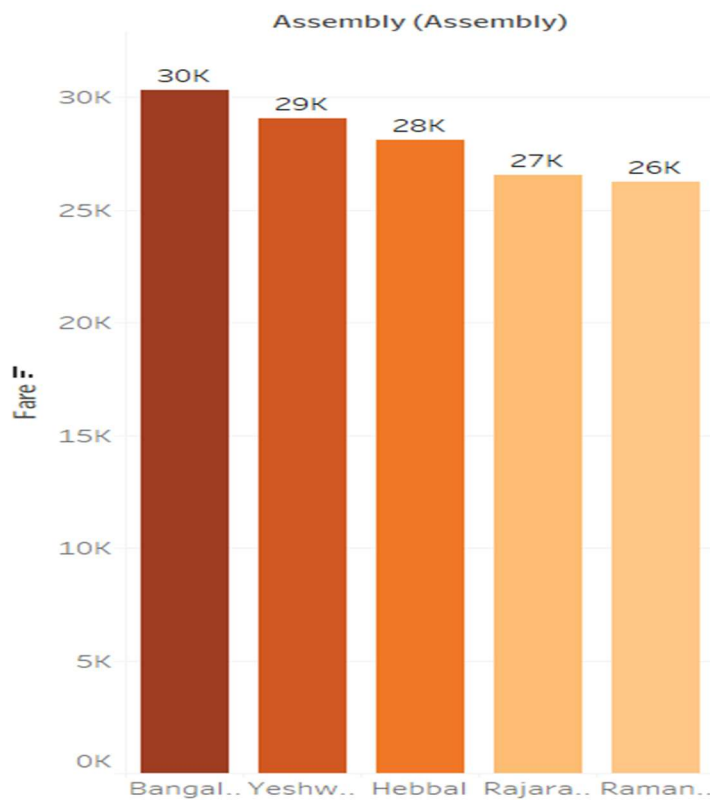


- **2.6.2. Revenue:** Identify pickup zones generating the highest revenue. [3 marks]

Solution:

The objective is to identify the pickup zones that generates the highest revenue. To focus efforts on zones that not only have high trip volume but also contribute significantly to earnings. This insight will support revenue optimization, pricing decisions, and localized marketing strategies.

TOP 5 PICK UP LOCATION-REVENUE WISE



RANK	PICK UP LOCATION	TOTAL REVENUE (in Rs)
1	Bangalore South	30K
2	Yeshwantpur	29K
3	Hebbal	28K
4	Rajarajeshwarinagar	27K
5	Ramanagaram	26K

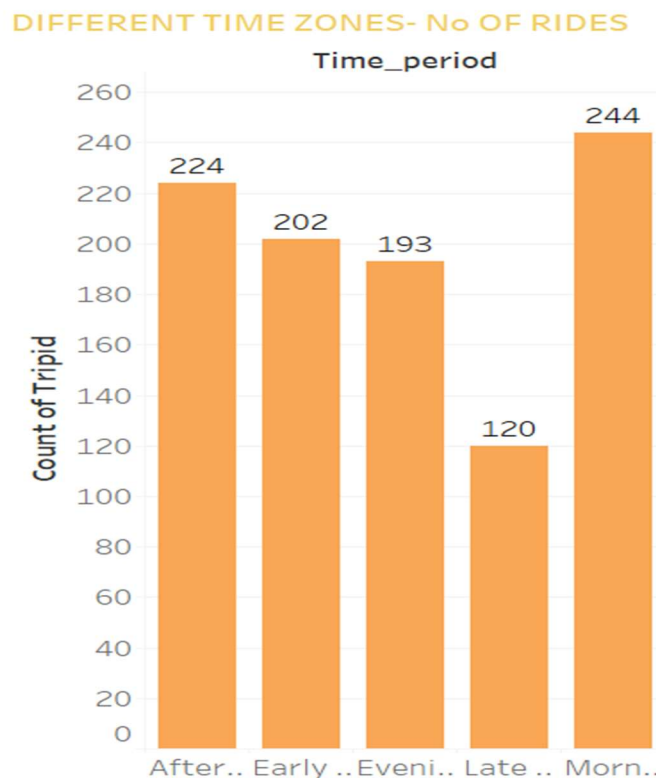
2.7. Analyse Ride Time Periods Across Zones [4 Marks]

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

Solution:

The objective is to analyze how the ride activity varies across different time periods like Early morning, Morning, Afternoon, Evening, Late night for each pick-up zones, and to identify temporal patterns of demand. This helps in Zone specific driver scheduling, peak time planning and localized promotions. Created Time bucket calculated field in Tableau to capture different time periods into various buckets as follows:

- Early Morning (0–6)
- Morning (6–12)
- Afternoon (12–17)
- Evening (17–22)
- Late Night (22–24)



RANK	TIME BUCKET	No. OF RIDES
1.	Morning	244
2.	Afternoon	224
3.	Early morning	202
4.	Evening	193
5.	Late Night	120

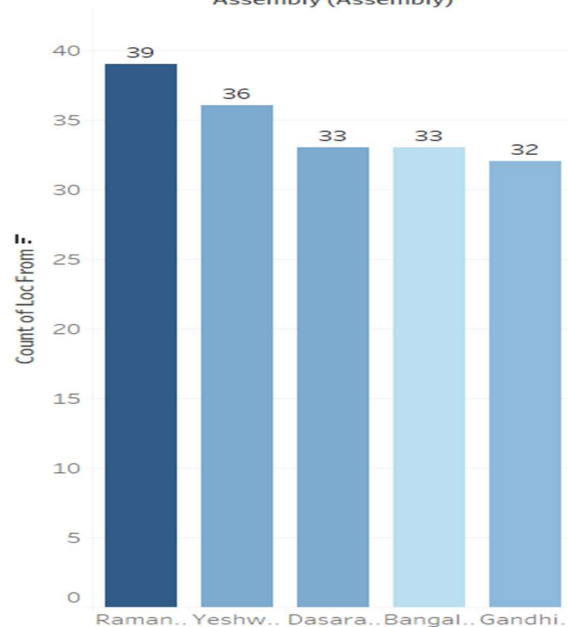
2.8. Top Zones with Highest Trip Volume [3 Marks]

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

Solution:

The objective is to identify the top 5 pickup zones with the highest number of completed trips and analyze the factors driving their performance. This supports Supply chain optimization, driver deployment and zone-specific growth planning.

TOP 5 PICK UP LOCATION- COMPLETED RIDES
Assembly (Assembly)



RANK	PICK-UP LOCATION	No. OF COMPLETED RIDES	INSIGHTS
1.	Ramanagaram	39	People commute from their daily for job- purpose
2.	Yeshwantpur	36	biggest wholesale market for agricultural produce in the city.
3.	Dasarhalli	33	place is well known for its religious establishments
4.	Banglore south	32	Electronics City, MG Road most crowded places
5.	Rajarajeshwarinagar	32	Residential Area

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:

Calculated Fields:

- Customer Cancelled = $1 - \text{customer_not_cancelled}$
- Driver Cancelled = $1 - \text{driver_not_cancelled}$
- Completed Ride = $\text{customer_not_cancelled} * \text{driver_not_cancelled}$
- ~12-15% of trips were **cancelled by customers**
- ~10-12% were **cancelled by drivers**
- ~70% of searches reached ride completion

- **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:

Calculated Field:

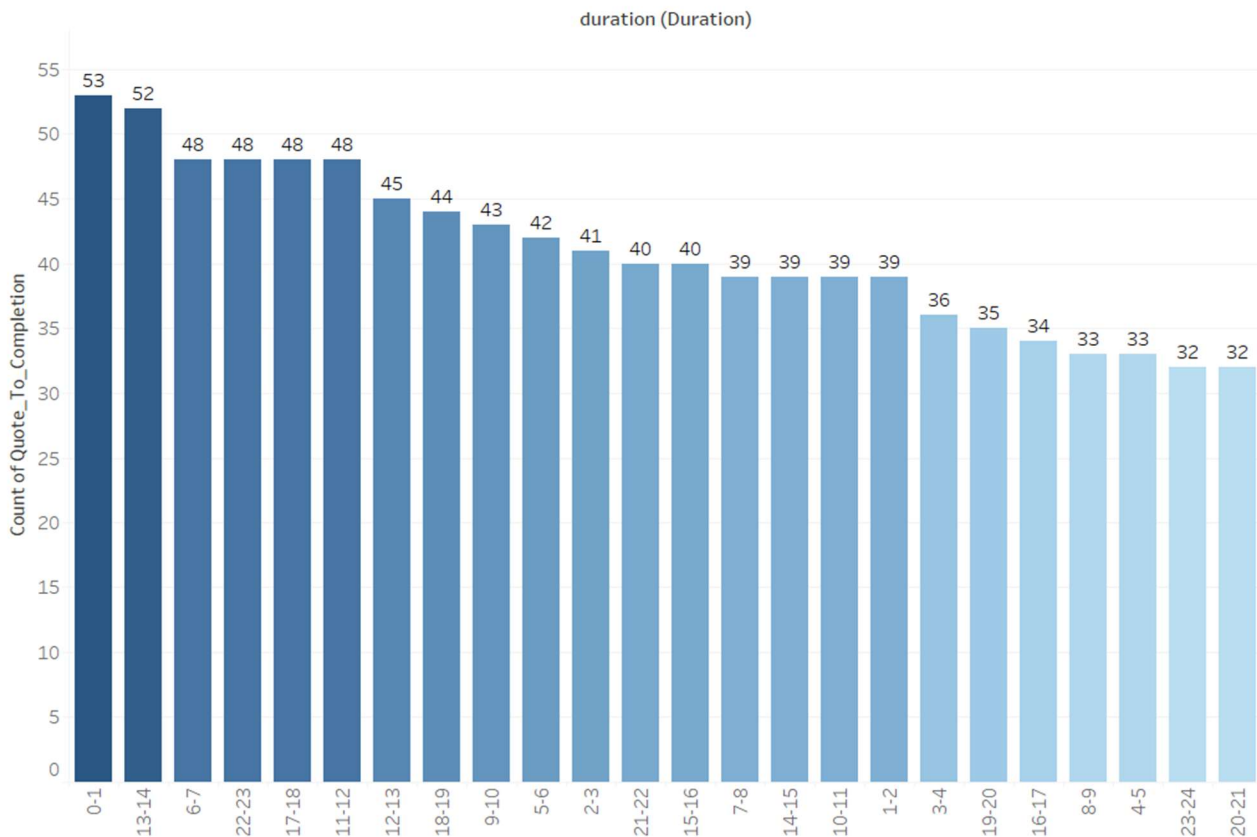
- Conversion Rate = $\text{SUM}(\text{end_ride}) / \text{SUM}(\text{searches_got_quotes}) = 76.9\%$

Line chart by hour showed:

- Highest conversion from quotes → rides during **Late Night & Lunch Hour**
- Lowest during mid-day and early morning

UX and pricing during off-peak could improve this.

QUOTE TO COMPLETION-TIME HOURS



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.

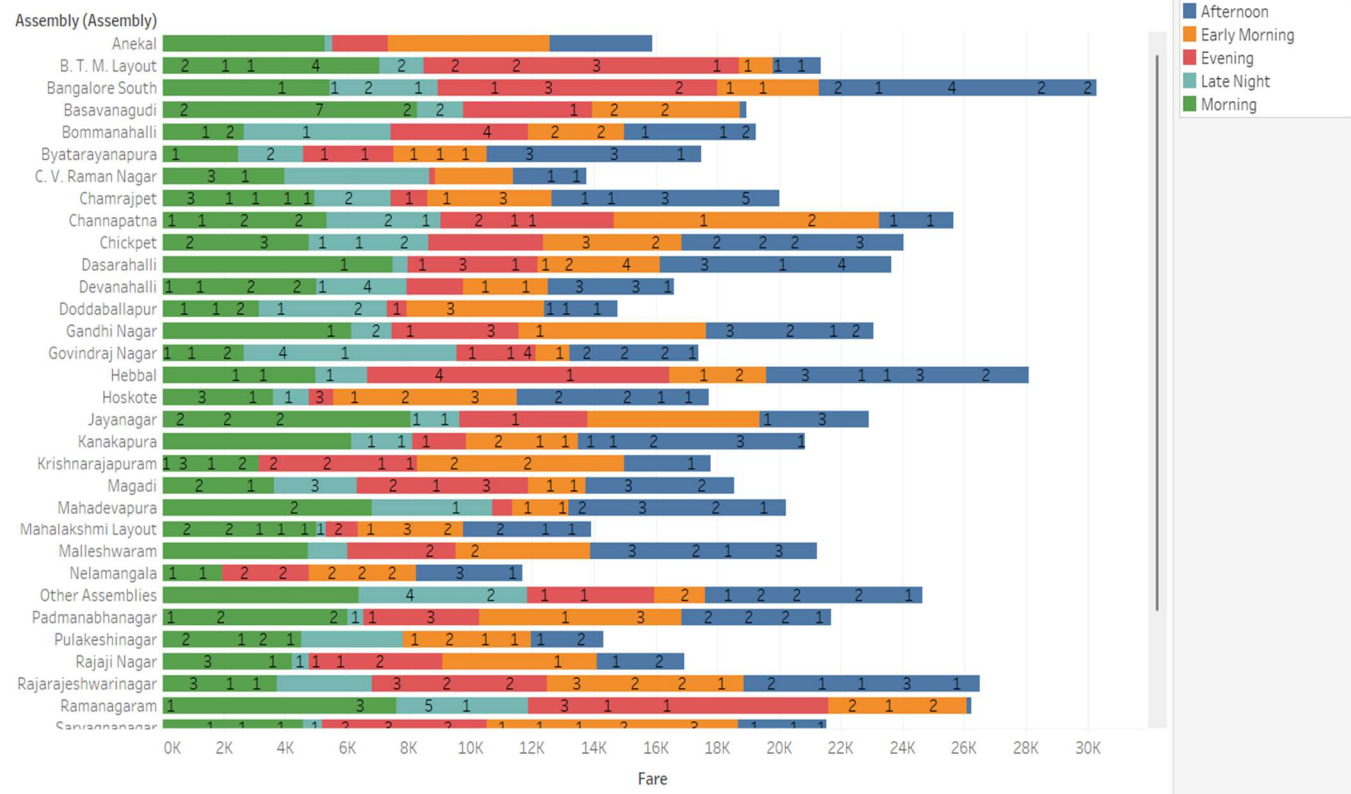
Solution:

The objective is to demonstrate the ability to use the Parameters as interactive filters in Tableau, enabling dynamic data exploration. This helps the users to analyse specific subsets of the Namma Yatri data on demand.

We created:

- **Filters** for: Time Period and Assembly
- **Parameter** to toggle focus between End trips, Fare, Duration

PARAMETER & FILTER



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:

1. Time-Based Driver Allocation

- Peak ride requests occur during **8–10 AM** and **6–9 PM**
- **Action:** Use time-slot-specific driver rosters and bonuses to ensure higher supply during peak periods.
- **Supporting Dashboard:** “Trips by Hour” and “Revenue by Hour” line charts showing hourly peaks.

2. Reduce Customer Cancellations

- Around **15–20%** of quotes are not converted to rides due to cancellations by customers.
- **Action:**
 - Improve UX after quote received (clear price/time + quick confirm option)
 - Send follow-up notifications if a user doesn't confirm
- **Supporting Dashboard:** Funnel chart (Search → Estimate → Quote → Ride), and cancellation bar chart (Customer vs Driver)

3. Optimize Ride Durations with Zone-Hour Matching

- High delays in some zones during peak hours indicate poor driver distribution.
- **Action:**
 - Use zone-time heatmaps to assign more drivers to bottleneck zones like MG Road, Whitefield, KR Puram during their busiest hours.
- **Supporting Dashboard:** “Zone vs Time Heatmap” showing concentration of trip requests by hour and area

4. Encourage Longer Rides During Off-Peak Hours

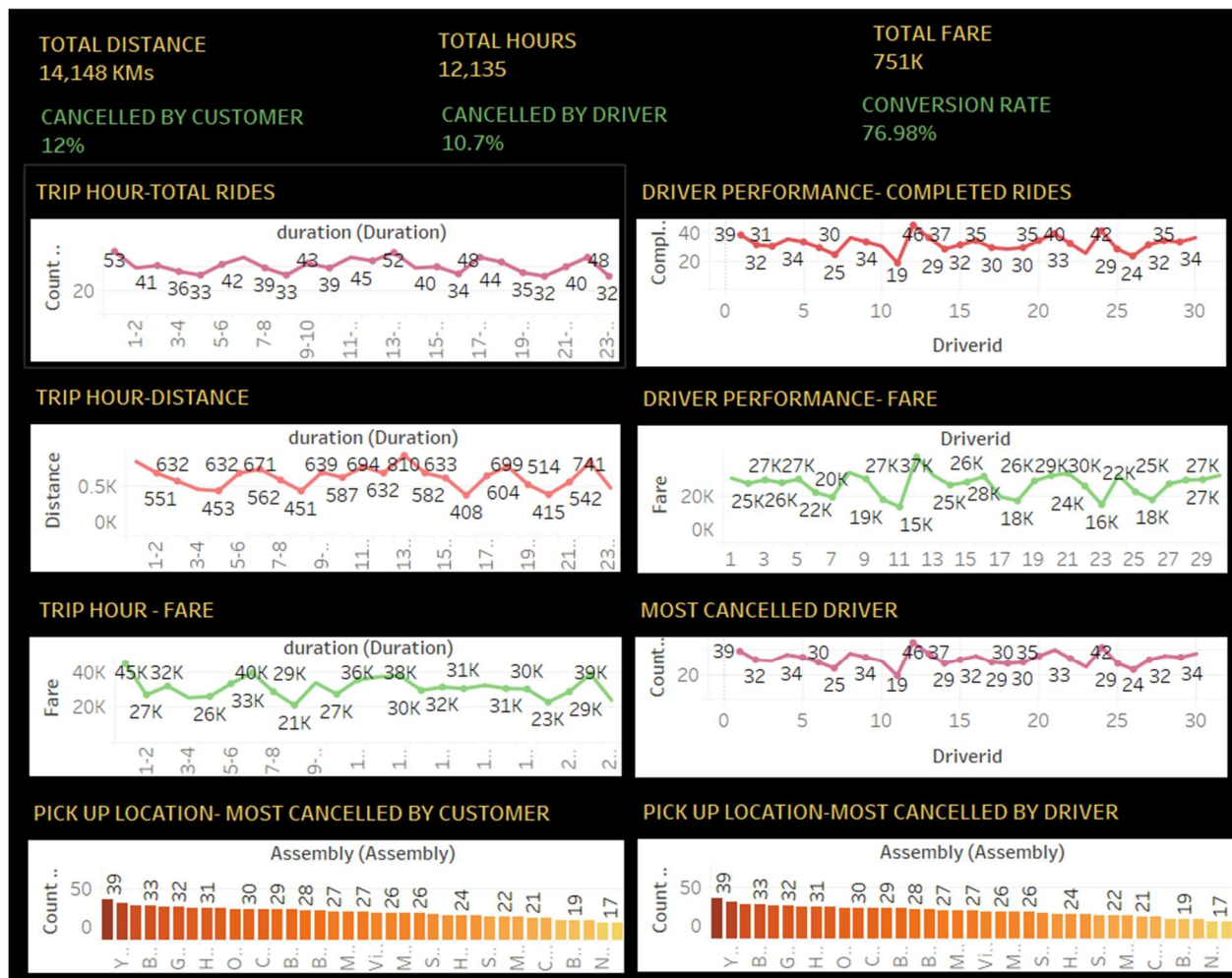
- Late-night trips have higher fare per km (longer rides).
- **Action:** Promote longer rides (e.g., airport rides) via discounts during off-peak

hours to improve utilization.

- **Supporting Dashboard:** Revenue per trip vs Duration

These recommendations are data-driven and directly implementable, helping Namma Yatri:

- Improve rider-driver match rates
- Lower operational idle time
- Reduce lost revenue due to cancellations
- Balance supply across time and zone



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. Zone-Based Promotional Campaigns

- High ride demand observed in **MG Road, Whitefield, Indiranagar**
- **Strategy:**
 - Launch promotions (first-ride discounts, cashback) targeted at users in these zones
 - Geo-targeted push notifications during peak times to increase retention
- **Supporting Dashboard:** Zone-wise trip volume and heatmap visualizations

2. UPI & Wallet-Based Incentives

- UPI and Wallet payments dominate, but some users still prefer cash.
- **Strategy:**
 - Offer ₹10–20 cashback on UPI rides to encourage digital payments
 - Partner with platforms like PhonePe, Paytm for co-marketing
- **Supporting Dashboard:** Payment Method Distribution bar chart

3. Driver Incentives for Underserved Time Slots

- Off-peak hours (12–6 AM) have lower ride supply despite demand in key zones (e.g., airport rides).
- **Strategy:**
 - Introduce surge bonuses or guaranteed fare slabs for drivers operating in low-supply windows
- **Supporting Dashboard:** Hour-wise trip demand + Fare per trip visual

4. Improvement of conversion of Quote to Complete Trip

- **Strategy:**

- 1.Run limited-time promo codes visible on the quote screen
- 2.Offer ₹20 off if booked within 30 seconds

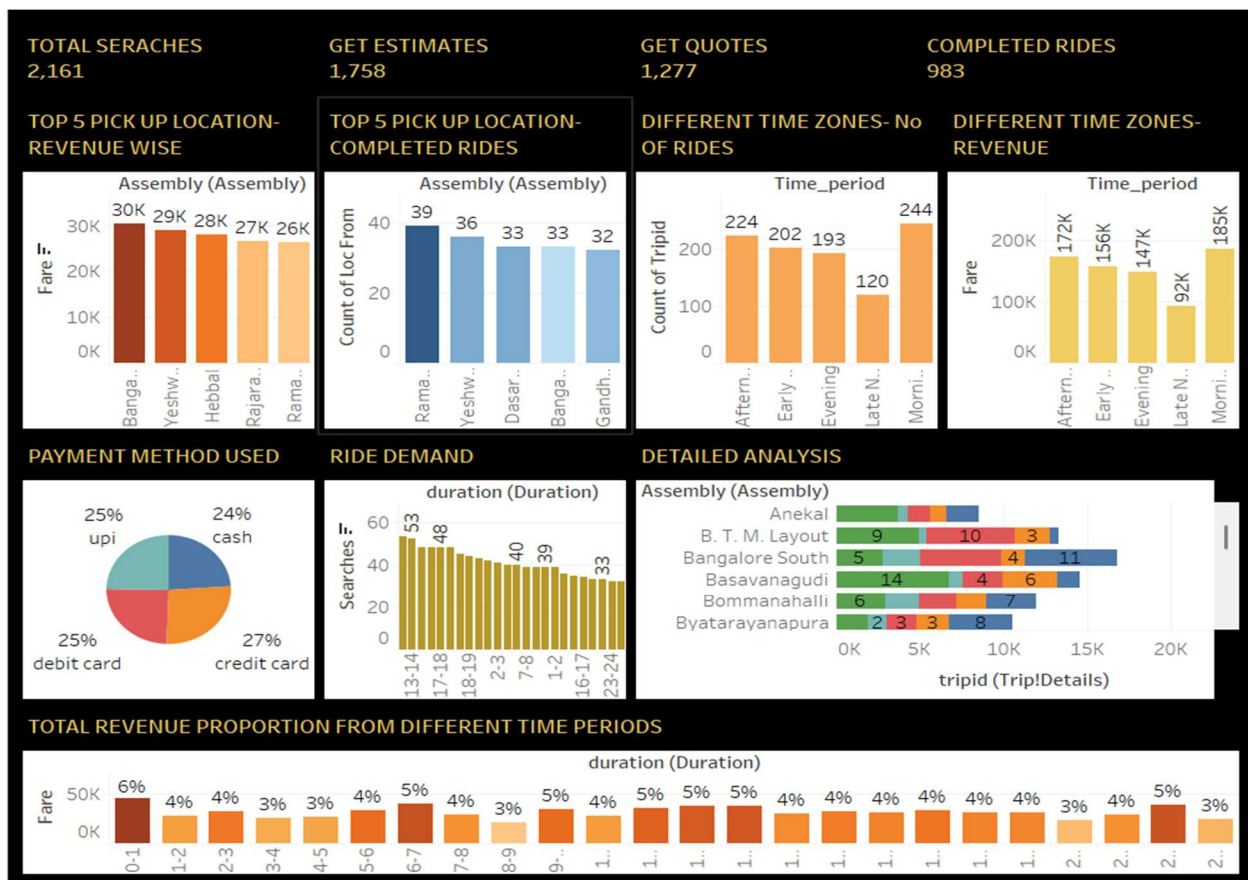
- **Supporting Dashboard:** Funnel visual (Search → Quote → Ride)

5. Seasonal & Festival Campaigns

- Use ride data to plan festival-period campaigns in high-volume zones.
- Push app-only offers to drive downloads and loyalty.

These combined **marketing + ops strategies** will help Namma Yatri:

- Boost digital adoption
- Improve quote conversion
- Increase off-peak and regional ride volumes
- Strengthen brand positioning in core neighborhoods.



Prepared By-

1. Shatrughan Patel
2. Shreya Aron
3. Surabhi Sharma