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:

Okay, to begin, imported the Namma Yatri dataset into Power BI.

The five tables were joined using the following relationships:

From: table (column) ↑	Relationship	To: table (column)	Status	
☐ Trips (duration)	＊—◀—1	Duration (id)	Active	⋯
☐ Trips (faremethod)	＊—◀—1	Payment (id)	Active	⋯
☐ Trips (loc_from)	＊—◀—1	Assembly (ID)	Active	⋯
☐ Trips (tripid)	1—◀▶—1	Trip_Details (tripid)	Active	⋯

(Toolbar above table: + New relationship · ⚡ Autodetect · ✏ Edit · 🗑 Delete · ≡ Filter ⌄)

From the above list:

- ***Trips->Duration*** will help us understand trip hours.
- ***Trips->Payment*** will help us figure out the payment modes used.
- ***Trips->Assembly*** will help us understand pickup zones.
- ***Trips->Trip_Details*** will help us in checking the trip details.

Used the visualisations on Model View to verify relationships between the tables. There are no inactive joins and no many-to-many relationships either, which means we are good to go for now.

You can check the Model View screenshot below.
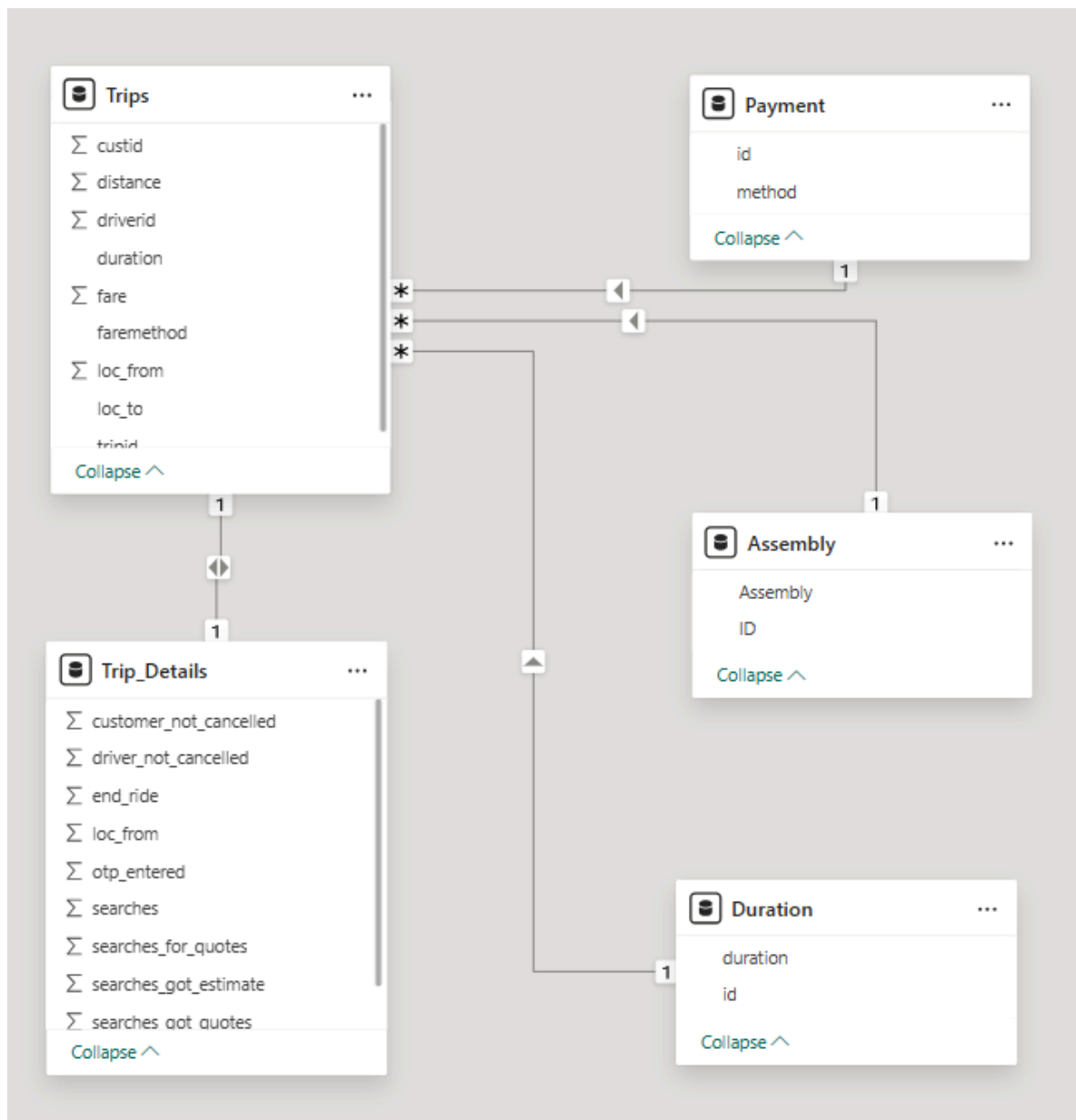
Looks clean to me!

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:

So, let's break down all the data cleaning and calculated fields steps:

1. Started with checking for duplicate columns in all datasets. This was easily done from the Table View and by going through column names. There weren't many, so no fuss.

2. Then, moved to checking for null/blank values. This was also easily done from the Table View by using the drop-down button on each column. We did come across some blank values, such as in Duration (duration) and Assembly (Assembly), but those were taken care of during the visualisation steps, using the Filters on the Report View.

3. Then, I had to change the datatypes of several columns. Primarily, columns that carried ID-related details were converted to text, which was the bulk of the operations in this step. A few, like distance or fare, were changed to decimals.

4. Now, coming to Calculated Fields.

Created a new measure for Total Revenue in Trips - This will help us in revenue-related insights, such as time period, payment method, etc.

Total Revenue = SUM(Trips[fare])

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:

First, a look at the Categorical Variables:

From Trips Table:

- tripid
- driverid
- custid
- faremethod
- loc_from
- loc_to

From Trip_Details Table:

- tripid
- loc_from

From Duration Table:

- duration (the original range)
- id

From Payment Table:

- Method
- id

From Assembly Table:

- Assembly
- ID

Now, a look at numerical variables:

From Trips Table:

- fare (Decimal)
- distance (Decimal)
- Total Revenue (measure)

From Trip_Details Table:

- searches (Whole Number)
- searches_got_estimate (Whole Number)
- searches _got_quotes (Whole Number)
- searches_for_quotes (Whole Number)
- customer_not_cancelled (Binary)
- driver_not_cancelled (Binary)
- otp_entered (Binary)
- end_ride (Binary)

From Duration Table:

- Duration Midpoint (calculated)

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 parameter I chose is the Sum of Searches from the trip_details table to act as a proxy for Ride Demand. The logic is simple: more searches = more demand.

When we plot it against the Duration (duration) column using a clustered bar chart, it shows the trends against different periods.

See below:

Ride Demand Over Time
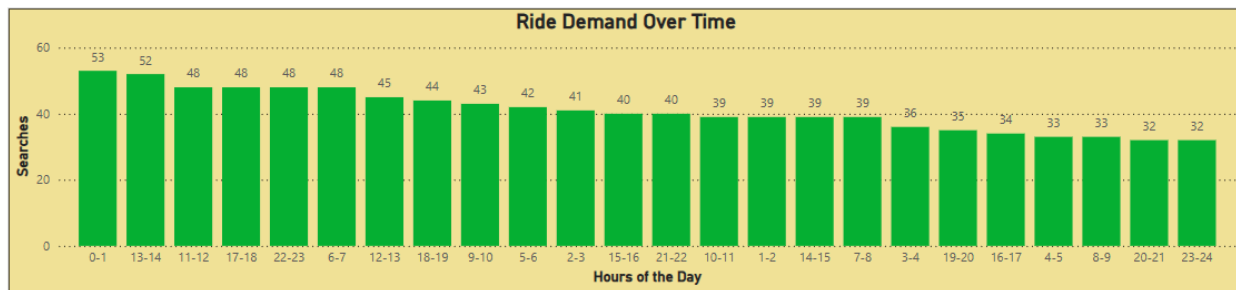
As we can see from above, the period of 12 am - 1 am (0-1) sees the highest number of searches, with a count of 53.

The period of 1 pm - 2 pm (13-14) ranks second, with a count of 52.

Lastly, the third rank is co-occupied by 11 am - 12 pm (11-12), 5 pm - 6 pm (17-18), 10 pm - 11 pm (22-23), and 6 am - 7 am (6-7), with counts of 48 each.

In terms of low demand, the 11 pm - 12 am (23-24) sees the least ride demand, at 32.

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:

For this, we started by creating a calculated measure for Proportion of Revenue.

This is the DAX formula:

Proportion of Revenue =

DIVIDE(

    SUM(Trips[fare]),

    CALCULATE(SUM(Trips[fare]), ALL(Duration)),

    0

)

Then, plotted a Donut Chart against Duration and this was the result:

Proportion of Total Revenue over Different Time Periods
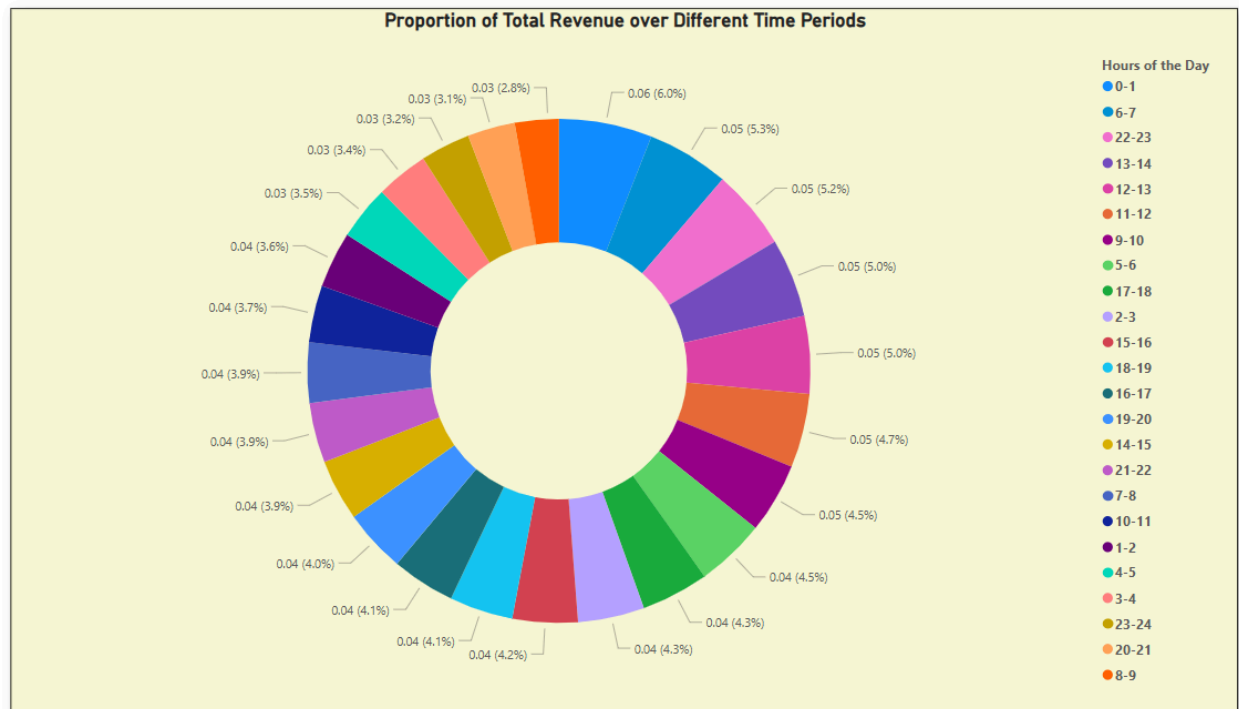
As we can see, the period of 12 am - 1 am (0-1) saw the lion's share of total revenue, at 6%. That makes sense, due to the late hours' demand as well as surge prices for drivers working in late-night shifts.

This is followed by 6 am - 7 am (6-7) at 5.3%, which again makes sense as that is the time when parents might take children to school and some professionals may also commute to the office for an early morning shift.

A close third rank is held by 10 am - 11 am (22-23) at 5.2%, which makes sense for late-night commuters as well as the beginning of surge pricing.

On the other end of the spectrum, we have the period of 8 am - 9 am (8-9), when the total revenue is the lowest, at 2.8%.

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:

For this, we had already created a calculated measure for Total Revenue, as mentioned earlier.

This is the DAX formula used: Total Revenue = SUM(Trips[fare])

Plotting a clustered bar chart, here's what we found:

Revenue Over Trip Hours
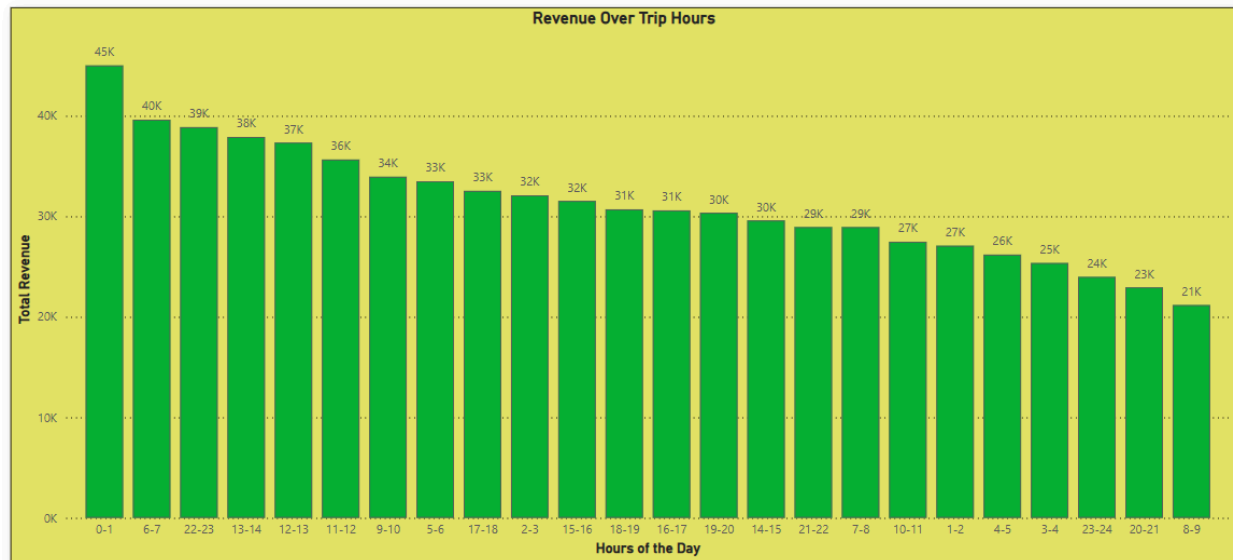
As we can see, the 12 am - 1 am (0-1) period again holds the lion's share of revenue. The observations are exactly the same as the previous question.

The hour between 12 am - 1 am (0–1) accounts for the highest share of total revenue at 6%. This aligns with expectations, given the higher demand during late-night hours and the likely impact of surge pricing for drivers operating at night.

Coming in next is the 6 am to 7 am (6–7) slot, contributing 5.3% of the total revenue, possibly driven by early morning commutes, school drop-offs, or professionals heading to early shifts.

Interestingly, the 10 pm to 11 pm (22–23) window also ranks high, with 5.2% revenue share, likely due to late-night travel and the start of nightly fare surges.

At the other end, the 8 am to 9 am (8–9) period sees the lowest revenue contribution at just 2.8%, which may indicate either lower trip fares or fewer rides during that hour.

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:

For this, the fare method was plotted against the payment ID in a stacked bar chart. The result was pretty straightforward:
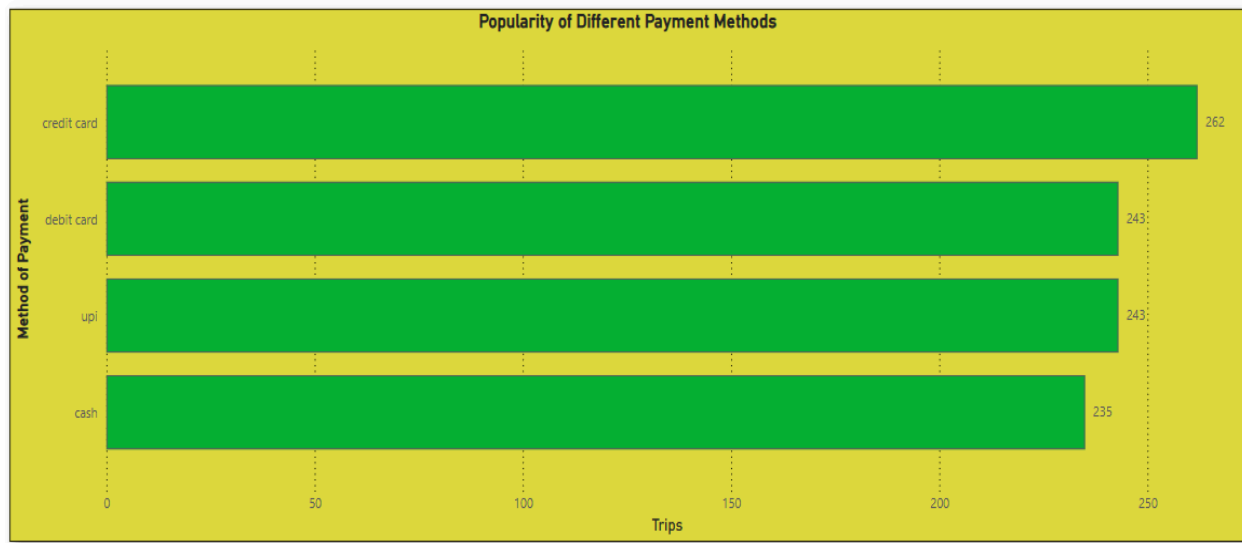
As we can see, credit cards appear to be the most favourite mode of payment for passengers, with a total share of 262 trips paid by CCs.

Debit cards and UPI jointly follow second, at 243 trips each.

Lastly, we have cash payments, which was used in 235 trips.

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:

For this, I plotted the Assembly data (denoted by pickup zone) against the sum of searches (denoted by Trips) using a stacked bar chart, and the result was pretty clear:
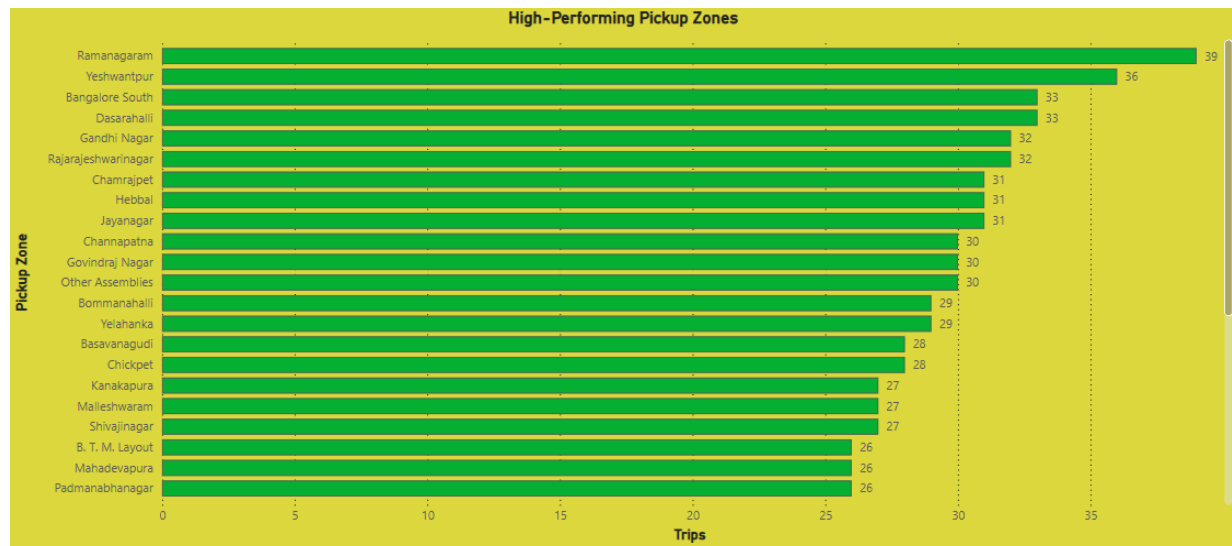
With 39 trip searches, Ramnagaram is the most popular pickup zone in terms of requests. Yeshwantpur occupies the #2 rank, with 36 trip requests. The third spot is shared by Bangalore South and Dasarahalli, with 33 trip requests each.
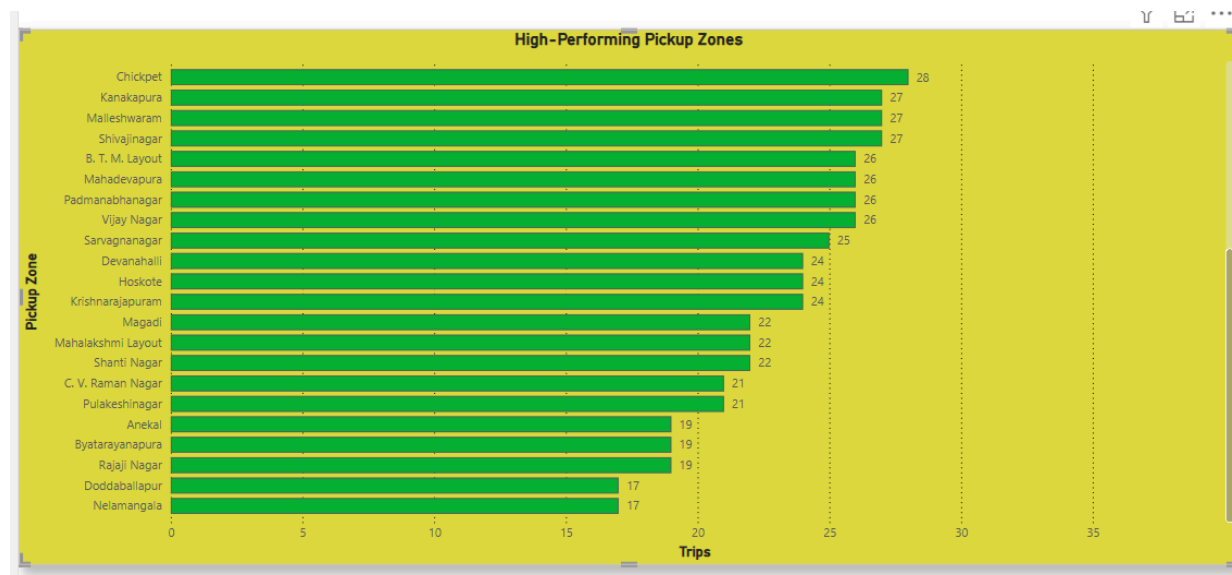

Curious to see the non-popular spots, I checked the other side of the spectrum and saw that Nelmangala and Doddaballapur appear to be the least popular pickup spots, with 17 trip requests each.

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

Solution:

For this, simply plotted a stacked bar chart of Assembly data (denoted by pickup
zone) against the Total Revenue measure (denoted by Trips), and the result was
pretty straightforward:
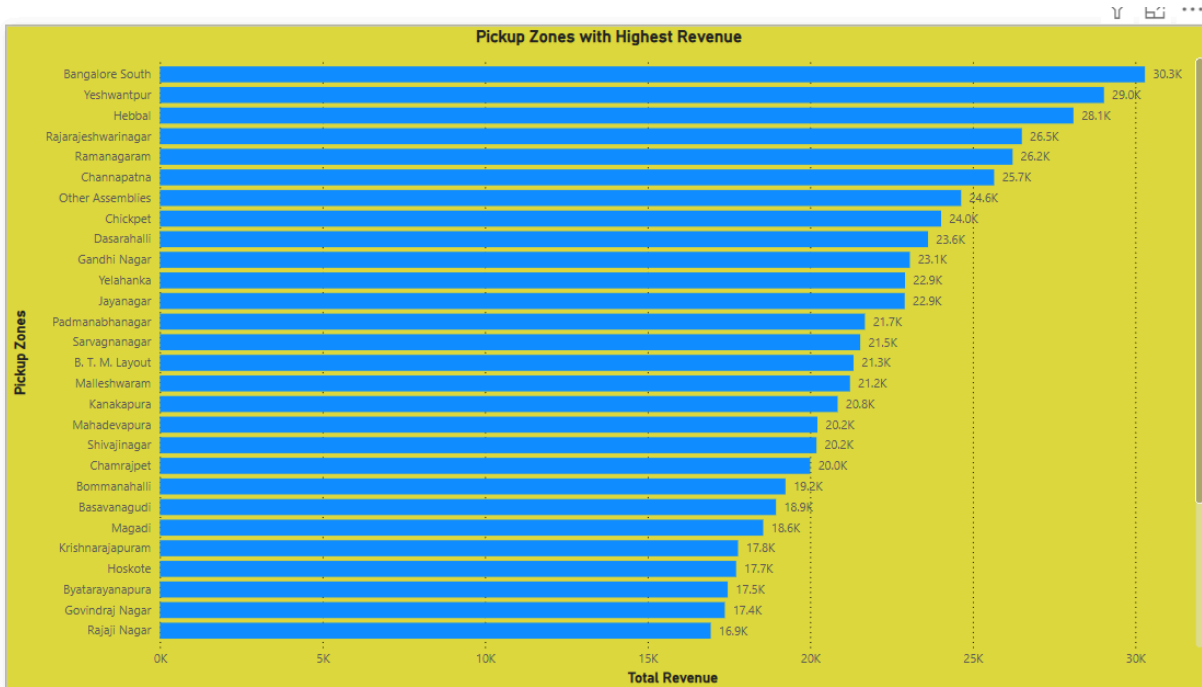
As we can see, Bangalore South has the lion's share of revenue among pickup
zones, with a total of ~Rs 30,300. Yeshwantpur and Hebbal make up the remainder
of the top 3 ranks.

On the other end of the spectrum lay Nelamangala, with Rs 11,700 of revenue
generated *(see below)*.

Pickup Zones with Highest Revenue
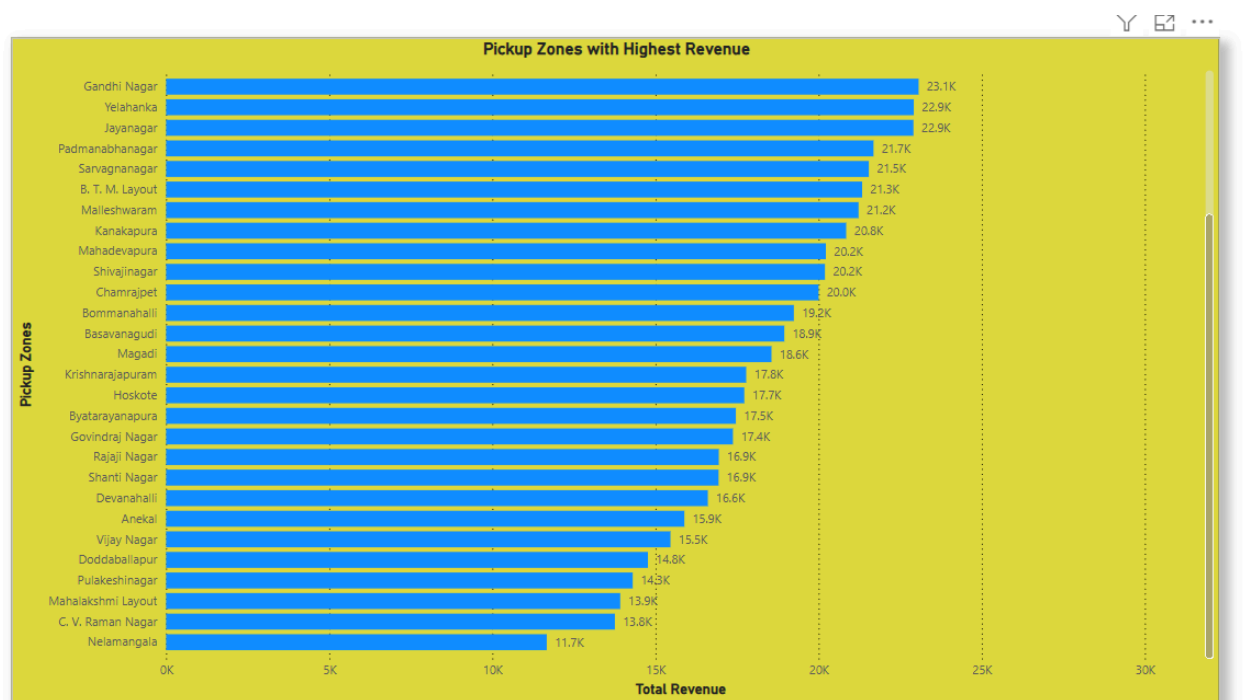
2.7. Analyse Ride Time Periods Across Zones [4 Marks]

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

Solution:

For this, I created a stacked column chart for Duration and tripID. This way, we can see which zones saw the most pickup at any given hour of the day.

Added Assembly in the legends to help understand which colour marks which zone.

Ride Time Periods Across Zones
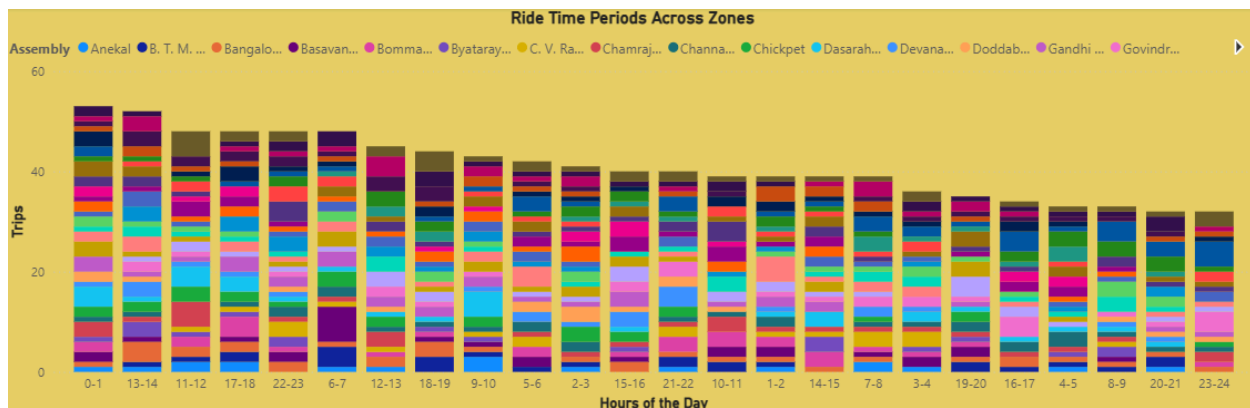
From what we can see, the 12 am - 1 am (0-1) time period saw the most trip

pickups, with Dasarahalli leading with 4 trips.

On the other end, we have the 11 pm - 12 am (23-24) period, which saw the least number of trips. Ramanagaram leads that slot with the most trips, with a count of 5.

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:

For this, started by creating a Calculate Measure for Competed Trips.

This is the DAX formula:

Completed Trips =

CALCULATE(

   COUNTROWS(Trip_Details),

   Trip_Details[end_ride] = 1

)

Then, created a stacked bar chart, plotting Completed Trips against loc_from, which marks the Zones. Lastly, finished by adding Assembly to the Legends, so that we can see the location names.

On the y-axis, we can only see the code names for locations, but a simple hovering on the bars showed us all the details we need.

This is what we got: *(chart on next page)*
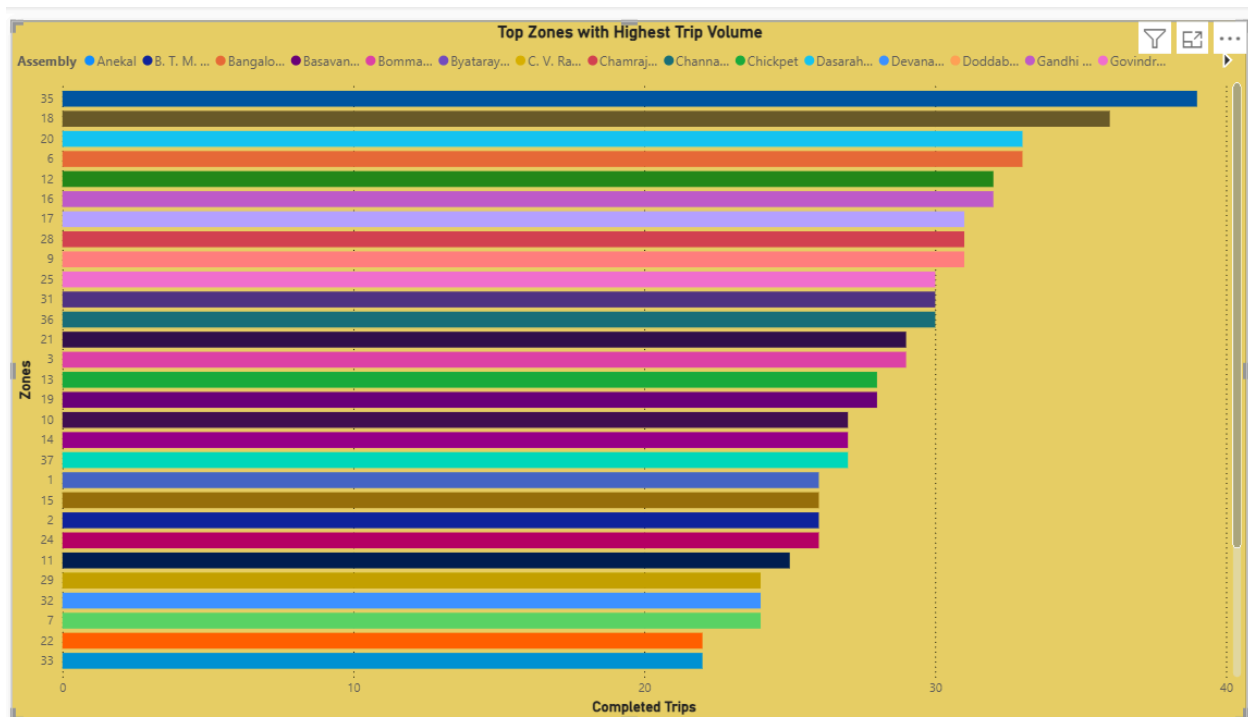
Ramanagaram takes the crown with 39 completed trips — no surprise, given its busy highway links and steady commuter flow. Yeshwantpur isn't far behind at 36, thanks to its railway station and buzzing markets.

Dasarahalli and Bangalore South share the third spot with 33 trips each. One's an industrial-residential mix, the other's a tech and education hub.

Bringing up the rear is Doddaballapur with just 17 trips. Fewer people, fewer rides — and perhaps less visibility for the app out there.

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:

For this, created 4 new calculated measures: Driver Cancellation %, Driver Success %, Customer Cancellation %, and Customer Success %.

Here are the DAX formulae:

Driver Cancellation % =

DIVIDE(

```
    CALCULATE(COUNTROWS(Trip_Details), Trip_Details[driver_not_cancelled] =
0),
    COUNTROWS(Trip_Details)
) * 100


Driver Success % =
DIVIDE(
    CALCULATE(COUNTROWS(Trip_Details), Trip_Details[driver_not_cancelled] =
1),
    COUNTROWS(Trip_Details)
) * 100


Customer Cancellation % =
DIVIDE(
    CALCULATE(COUNTROWS(Trip_Details), Trip_Details[customer_not_cancelled]
= 0),
    COUNTROWS(Trip_Details)
) * 100


Customer Success % =
DIVIDE(
    CALCULATE(COUNTROWS(Trip_Details), Trip_Details[customer_not_cancelled]
= 1),
    COUNTROWS(Trip_Details)
) * 100
```

Created cards for the same, here are the results:

Customer Cancellation

Customer Cancellation %

48.17

Customer Success Rate

Customer Success %

51.83

Driver Cancellation

Driver Cancellation %

47.25

Driver Success Rate

Driver Success %

52.75
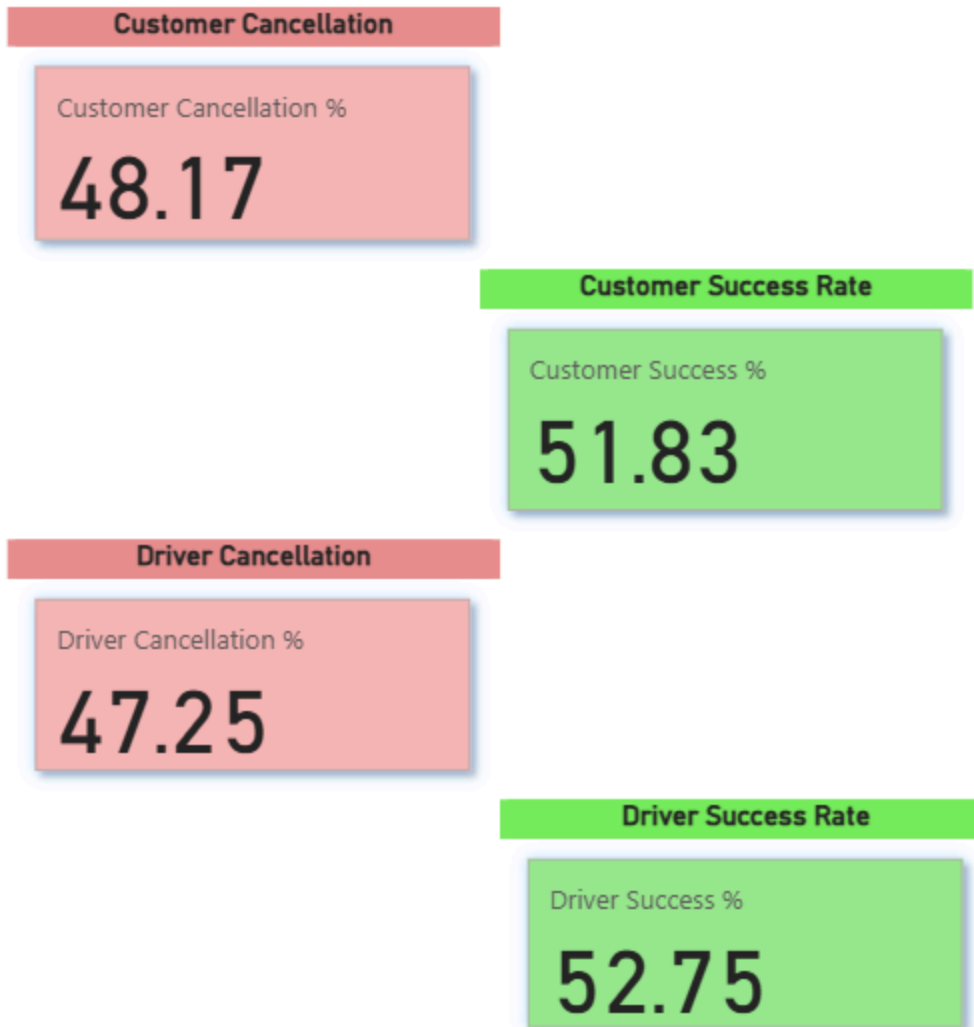
As we can see from above, although there is no chasm of difference, the success rates are a little better than the cancellation rates.

- **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:

For this, started with creating a DAX measure using quote searches and end rides,

which, if both are 1, marks a successful search and ride:

% Completion After Quote Search =

DIVIDE(

   CALCULATE(

      COUNTROWS(Trip_Details),

      Trip_Details[searches_for_quotes] = 1,

      Trip_Details[end_ride] = 1

   ),

   CALCULATE(

      COUNTROWS(Trip_Details),

      Trip_Details[searches_for_quotes] = 1

)

)

Then, plotted a line graph between Duration hours and the searches_for_quotes data. This is what we got:
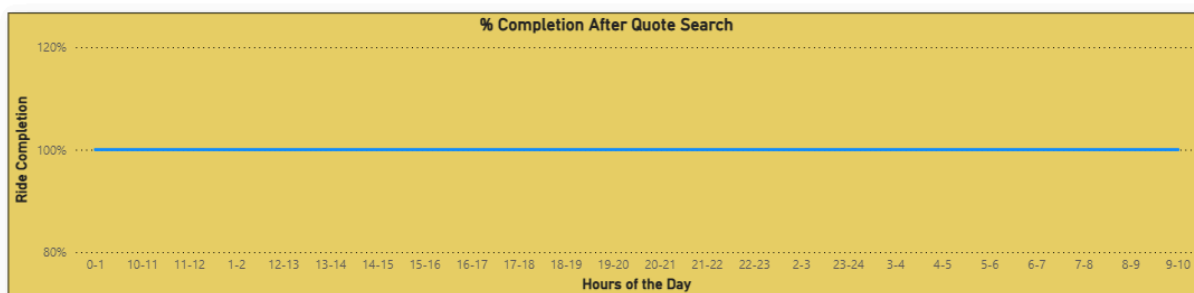
As we can see, all searches for quotes by customers ended in successful trips, across all hour periods.

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:

For this, decided to create a filter to check the total revenue earned by zones, to

check which time periods are the most lucrative for each pickup zone.
For this, created a Zone Filter table, added a slicer.
And also, added this DAX measure for Revenue by Zone and Duration:

```
Revenue by Zone and Duration =
CALCULATE(
    SUM(Trips[fare]),
    FILTER(
        ALL(Assembly[Assembly]),
        Assembly[Assembly] IN VALUES('Zone Filter'[Assembly])
    )
)
```

Then, proceeded with plotting a clustered column chart for Duration and the new measure created above.

Added a text box for clear instructions.

Lastly, as an added quick reference, added a card to show the total revenue for the selected zone.

This gives a complete understanding of the total revenue from each zone as well as the most popular time zone, and the most lucrative one.

Here's the result:

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 our analysis, here are the top recommendations to boost operational efficiency for Namma Yatri.

1. Allocate More Drivers During High-Demand Hours

- **Recommendation:** Increase driver availability in the below-mentioned time slots to reduce wait time and cancellations.
- **Reason:** 12 AM–1 AM (0-1) and 1 PM–2 PM (13-14) are peak search hours.
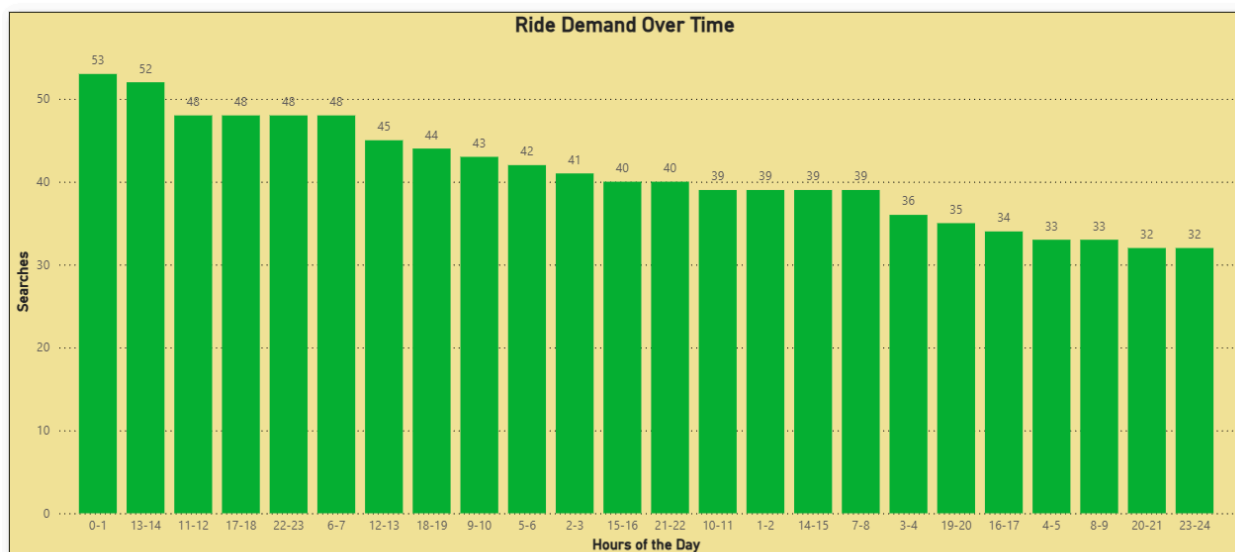
2. Focus on High-Revenue Time Slots

- **Recommendation:** Offer surge bonuses and optimise driver deployment during below-mentioned windows.
- **Reason:** 12 AM–1 AM, 6–7 AM, and 10–11 PM drive the most revenue.

Proportion of Total Revenue over Different Time Periods

Revenue Over Trip Hours
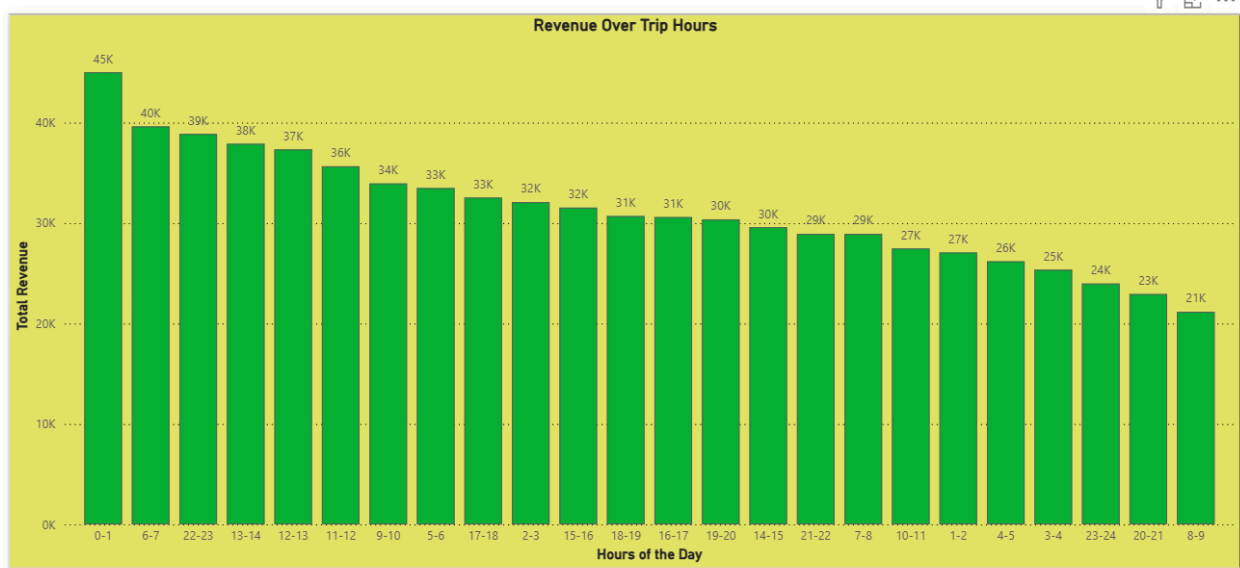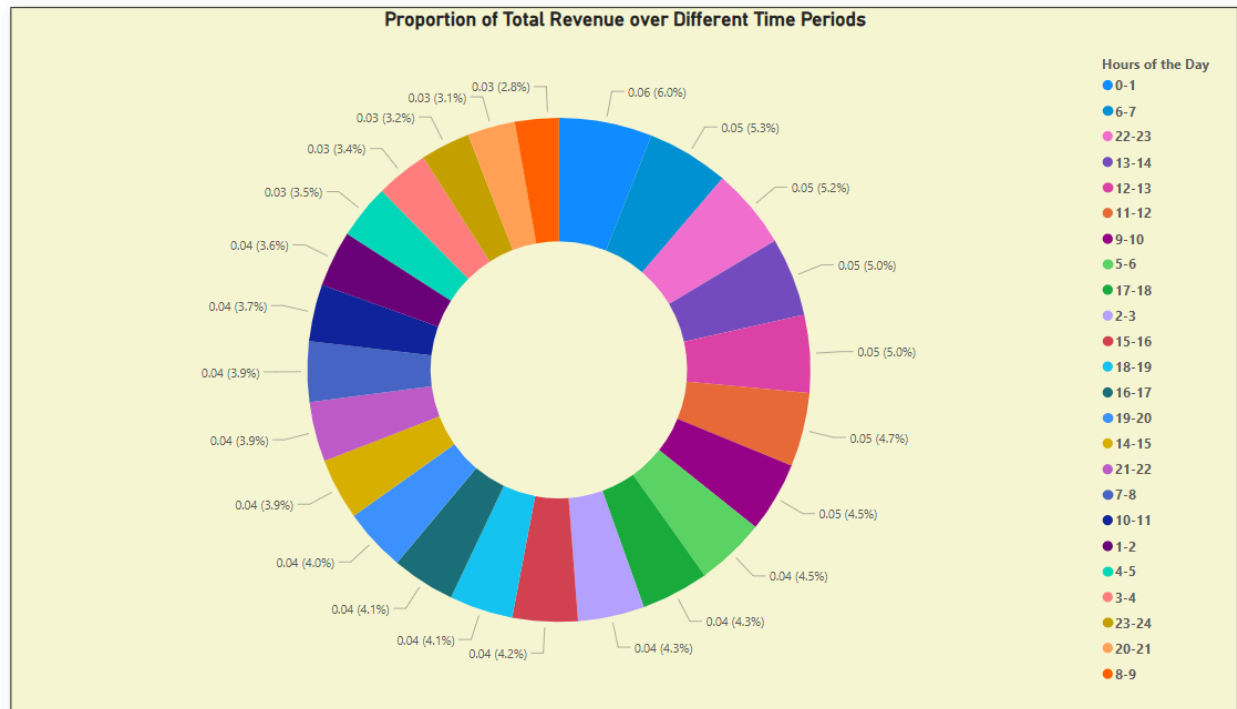
3. Reduce Cancellations via Root Cause Analysis

- **Recommendation:** Identify frequent causes and address them through app improvements or incentives for completed rides.
- **Reason:** There is a noticeable % of both driver and customer cancellations.

Customer Cancellation

Customer Cancellation %

48.17

Customer Success Rate

Customer Success %

51.83

Driver Cancellation

Driver Cancellation %

47.25

Driver Success Rate

Driver Success %

52.75

4. Use Zone-Specific Revenue Analysis

- **Recommendation:** Dynamically adjust driver distribution based on historical zone-wise profitability.
- **Aid:** The Revenue by Zone and Duration chart with slicer allows filtering revenue by zone and time.
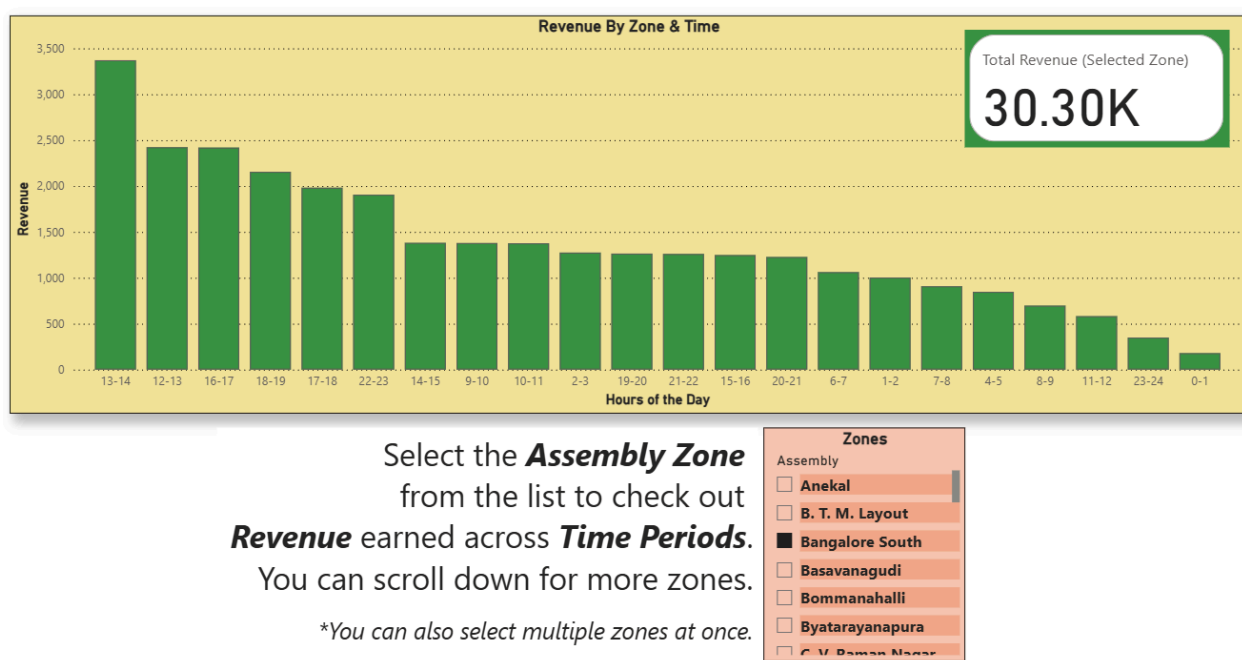
3.2.   Marketing and Operational Strategy Improvements   [10 Marks]

- 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:


Based on our analysis, here are actionable strategies:

1. Promote Low-Revenue Time Slots

● **Recommendation:** Launch time-based discounts or loyalty rewards to increase ridership during the below-mentioned hours.
● **Reason:** 8–9 AM was identified in charts (2.3, 2.4) as the lowest revenue slot.
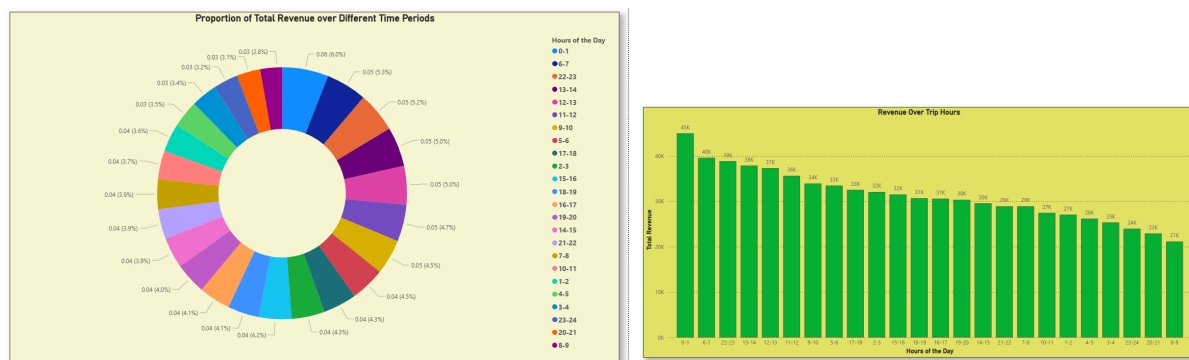
2. Hyperlocal Marketing in Underperforming Zones

● **Recommendation:** Run targeted promotions or awareness campaigns in the below-mentioned regions.
● **Reason:** From the pickup zone bar chart (2.6.1) and completed trips chart (2.8), Doddaballapur and Nelamangala underperform.
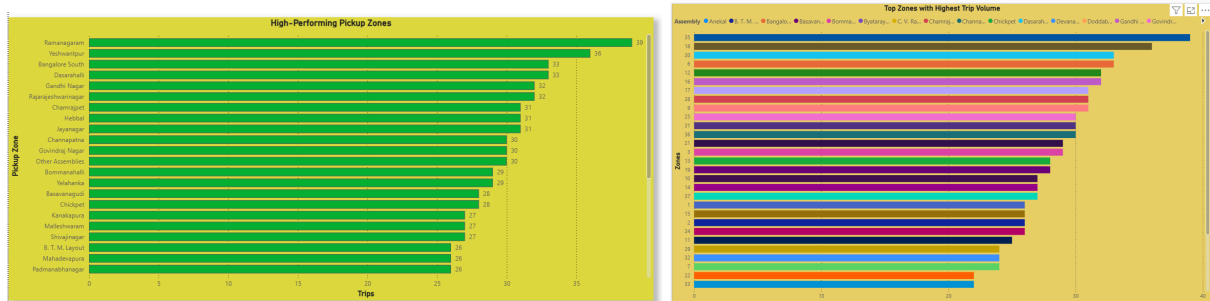
3. Promote Card and UPI Payments

- **Recommendation:** Incentivise digital payments through cashback or app-exclusive discounts.
- **Reason:** The stacked bar chart (2.5) shows that credit cards and UPI are the top choices.
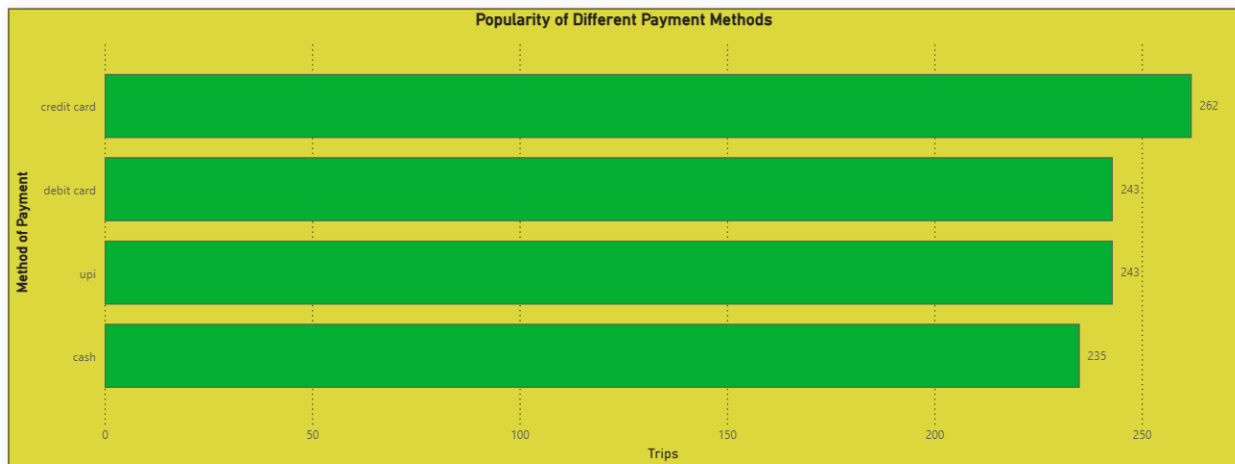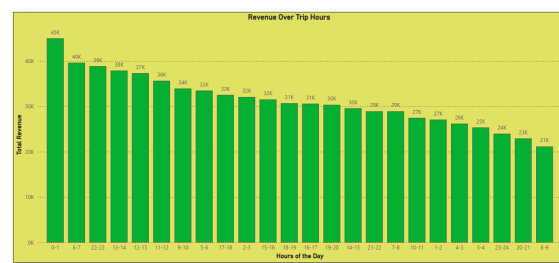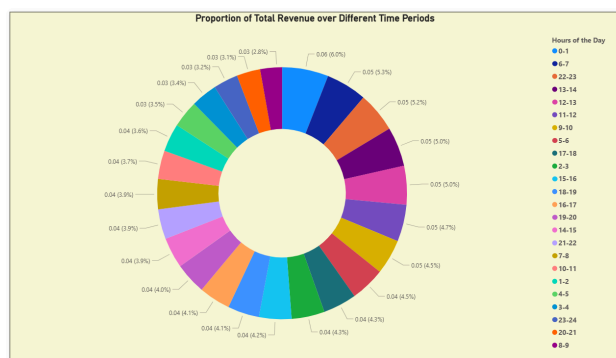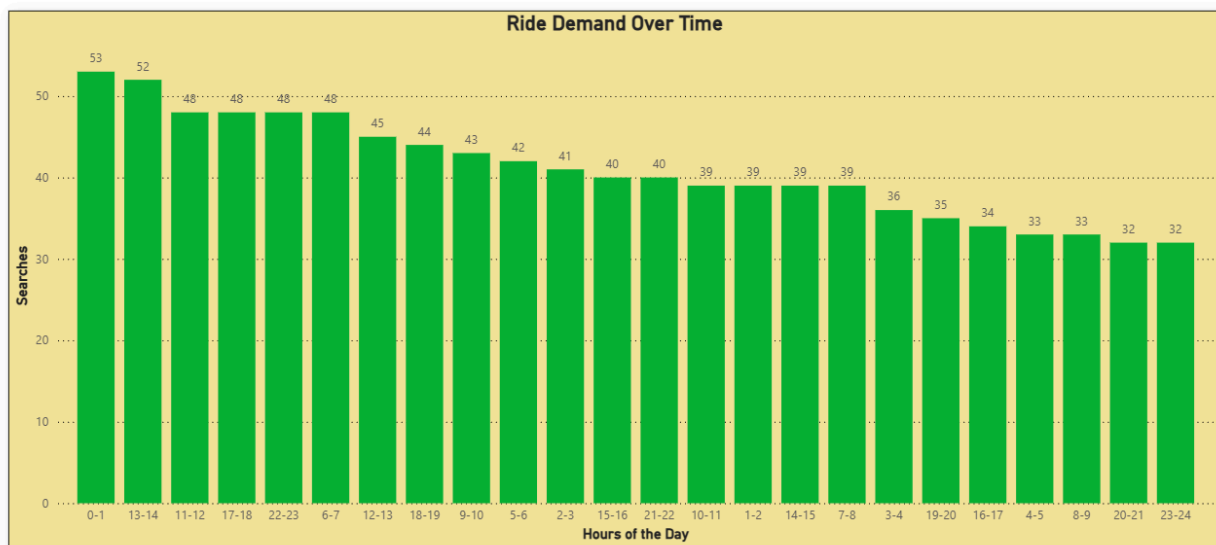

4. Driver Incentives in High-Demand Zones

- **Recommendation:** Offer bonuses for completing rides in these zones, especially during peak hours.
- **Reason:** Based on the Revenue by Zone chart (2.6.2), Bangalore South and Yeshwantpur lead in revenue.
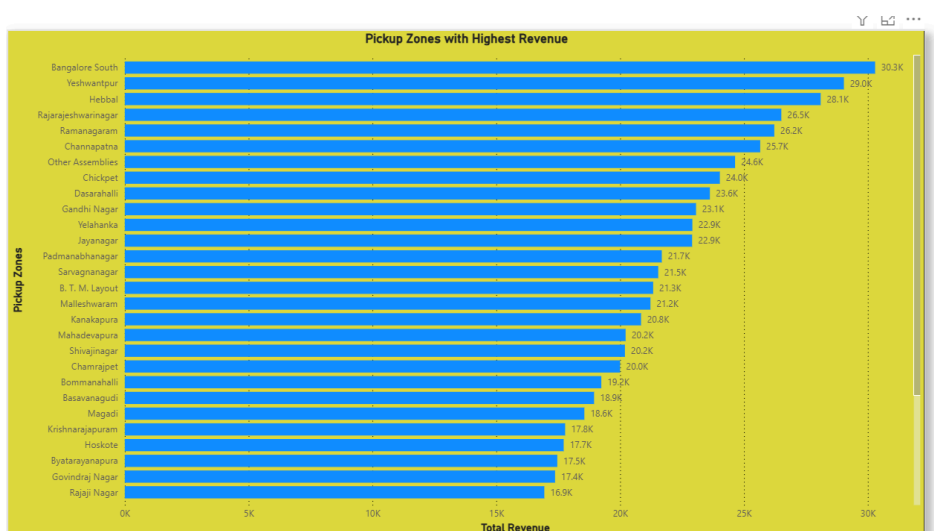
5. Advertise Safety and Reliability in Night Hours

- **Recommendation:** Position Namma Yatri as a safe night-time option through marketing with emphasis on OTP and successful completion rates.
- **Reason:** The clustered bar chart (2.2) and revenue charts (2.3, 2.4) show strong usage in late-night hours.

Ride Demand Over Time

Proportion of Total Revenue over Different Time Periods

Revenue Over Trip Hours