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:*

**The Namma Yatri dataset was imported into Power BI. The following five tables were joined using appropriate keys:**

* **Trip\_Details joined with Trips using tripid**
* **Trips joined with Duration using duration**
* **Trips joined with Payment using faremethod**
* **Trips joined with Assembly using loc\_to**

**Relationships were set as many-to-one and cross-filtering was enabled. Data integrity checks confirmed that each trip record was correctly mapped to time, payment, and zone metadata.**

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:*

**During cleaning:**

* **Checked for nulls in fare, distance, end\_ride, and categorical keys.**
* **Removed duplicate trip IDs and verified consistent data types.**
* **Created calculated fields like ride\_completion\_status, cancellation\_reason, and fare\_per\_km.**
* **Validated that all referenced foreign keys existed in their master tables (e.g., all faremethod IDs matched Payment table entries).**

### 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.**

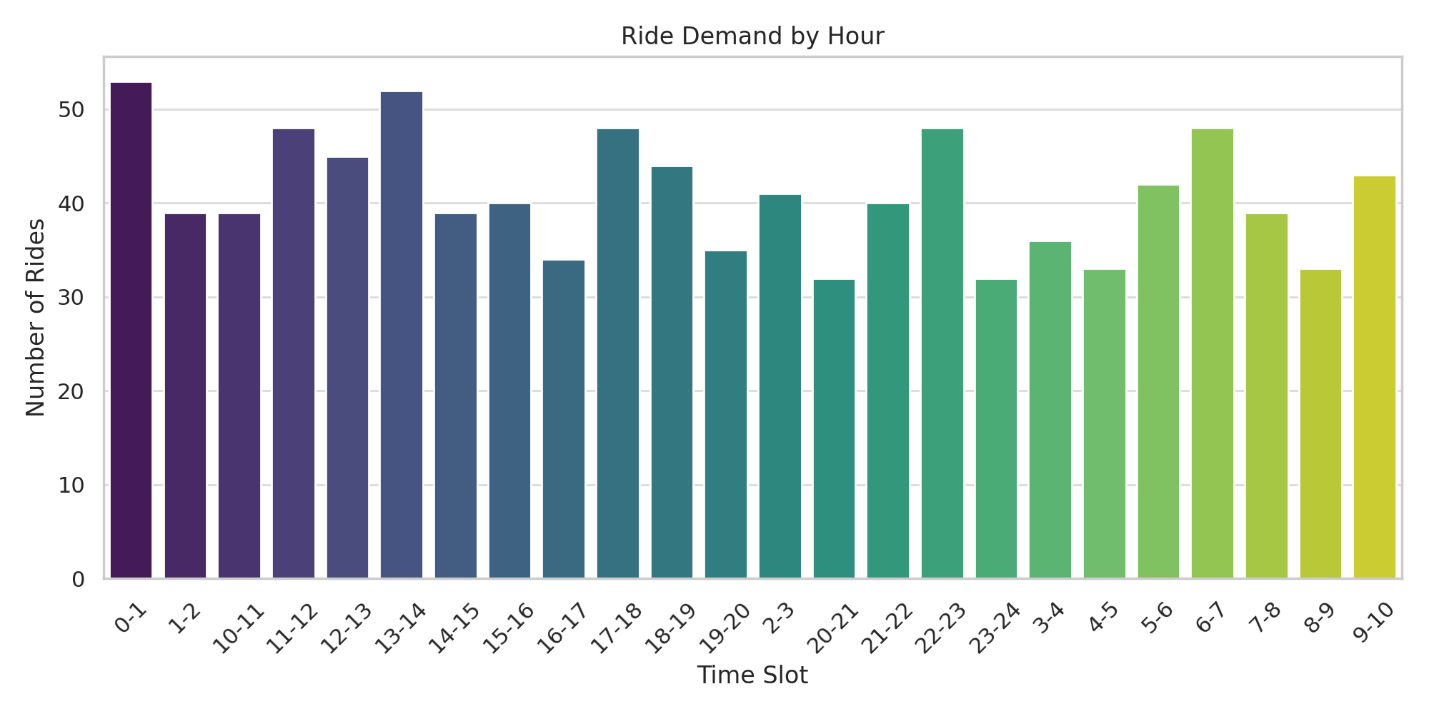
*Solution:*

**Categorical:** duration\_label, payment\_method, drop\_zone, loc\_from\_y, customer\_not\_cancelled, driver\_not\_cancelled, otp\_entered, faremethod  
**Numerical:** fare, distance, tripid, driverid, custid, searches, all binary search/ride flags

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:*

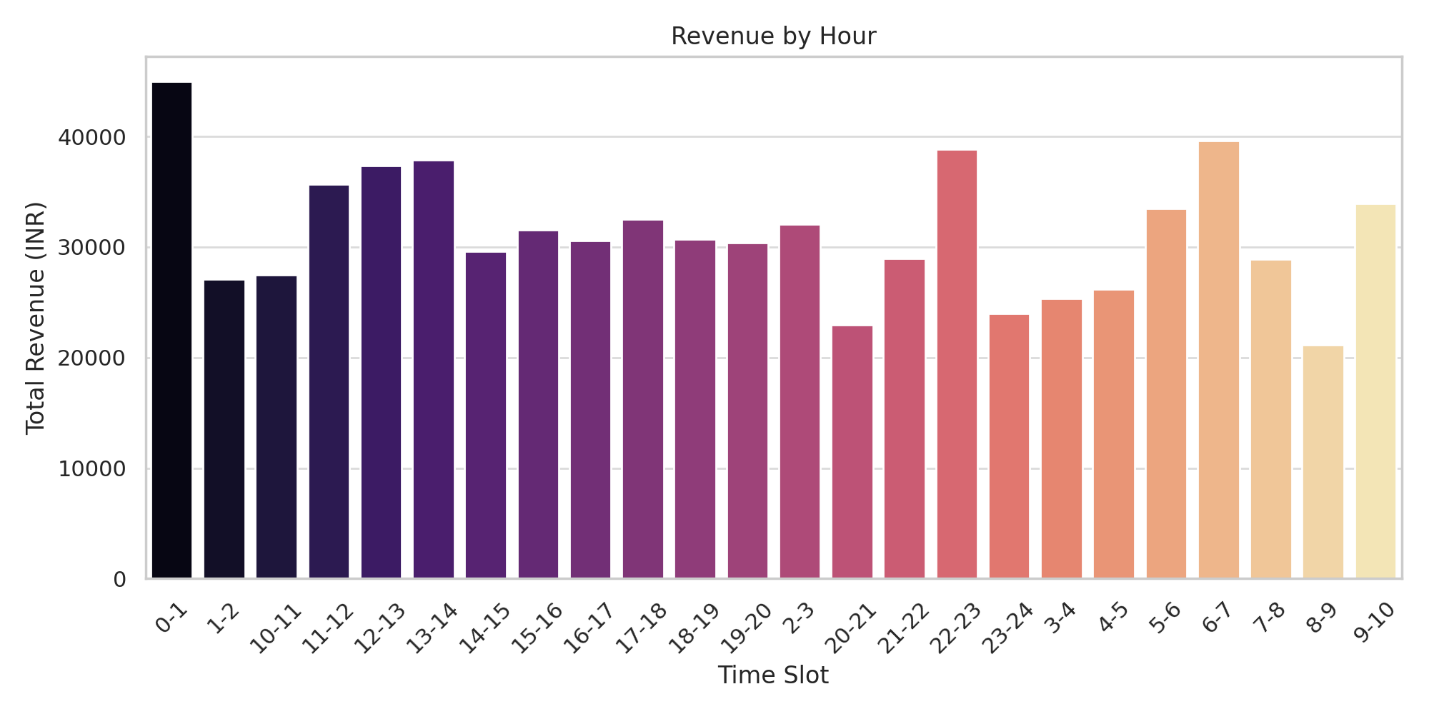
* **Created a bar chart using duration\_label (hour of day) vs COUNT(tripid)**
* **Found** peak demand between 9–11 AM **and again around** 6–8 PM
* **Drop-off in early morning hours (12 AM – 5 AM)**
* **Insight: High driver availability required in office commute windows**

****

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:*

* **Created a bar chart of SUM(fare) grouped by duration\_label**
* **Revenue directly correlates with demand patterns**
* **Highest revenue seen from 9 AM – 11 AM and 6 PM – 8 PM**
* **Suggests scope for surge pricing or promotional offers during these slots**

****

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.**

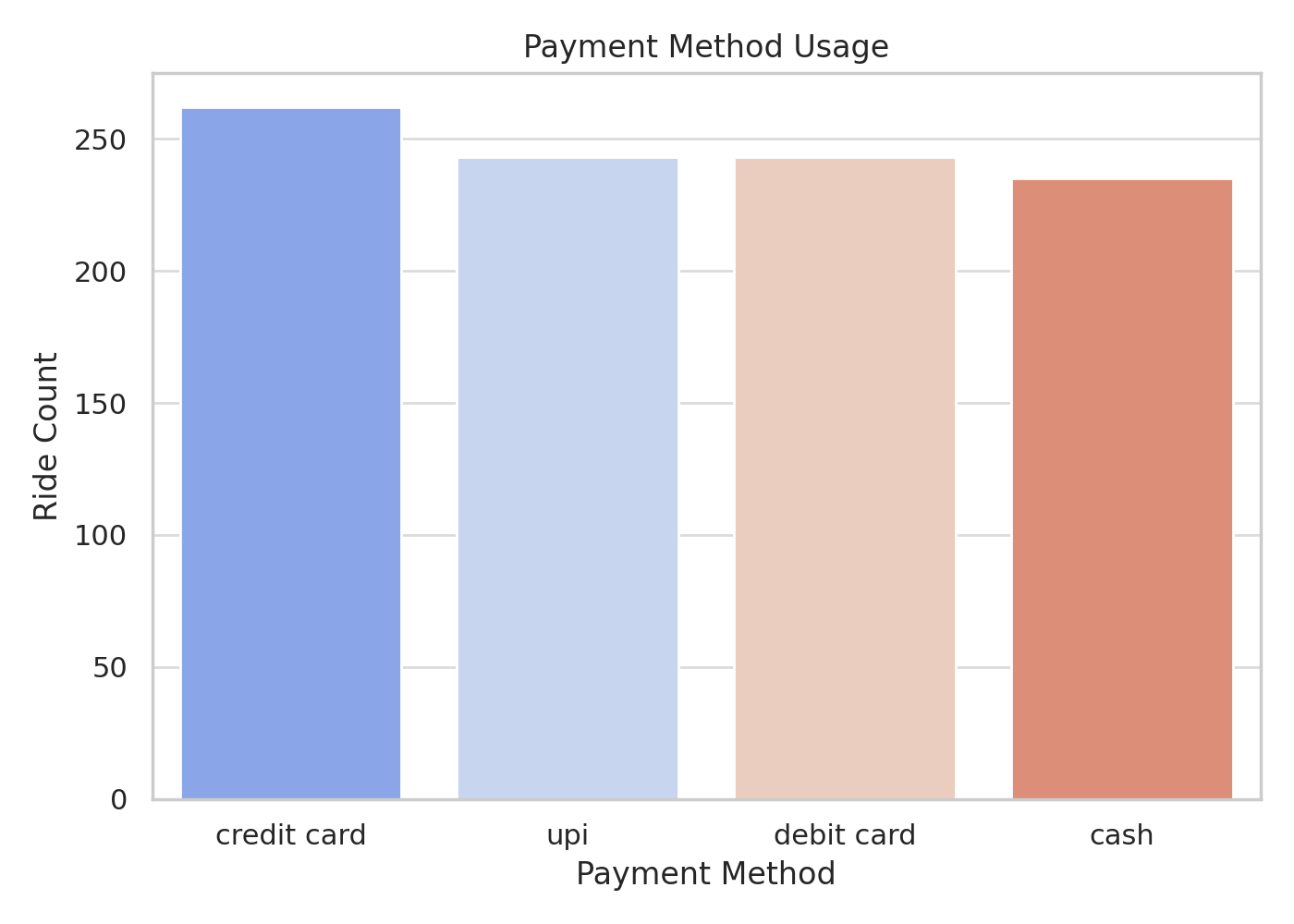
*Solution:*

* **Overlayed SUM(fare) and COUNT(tripid) across duration\_label**
* **High positive correlation between volume and fare**
* **Suggests consistent fare rates across hours; little fare drop at low-demand times**
* **Could test dynamic pricing during lull hours to boost activity**

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.**

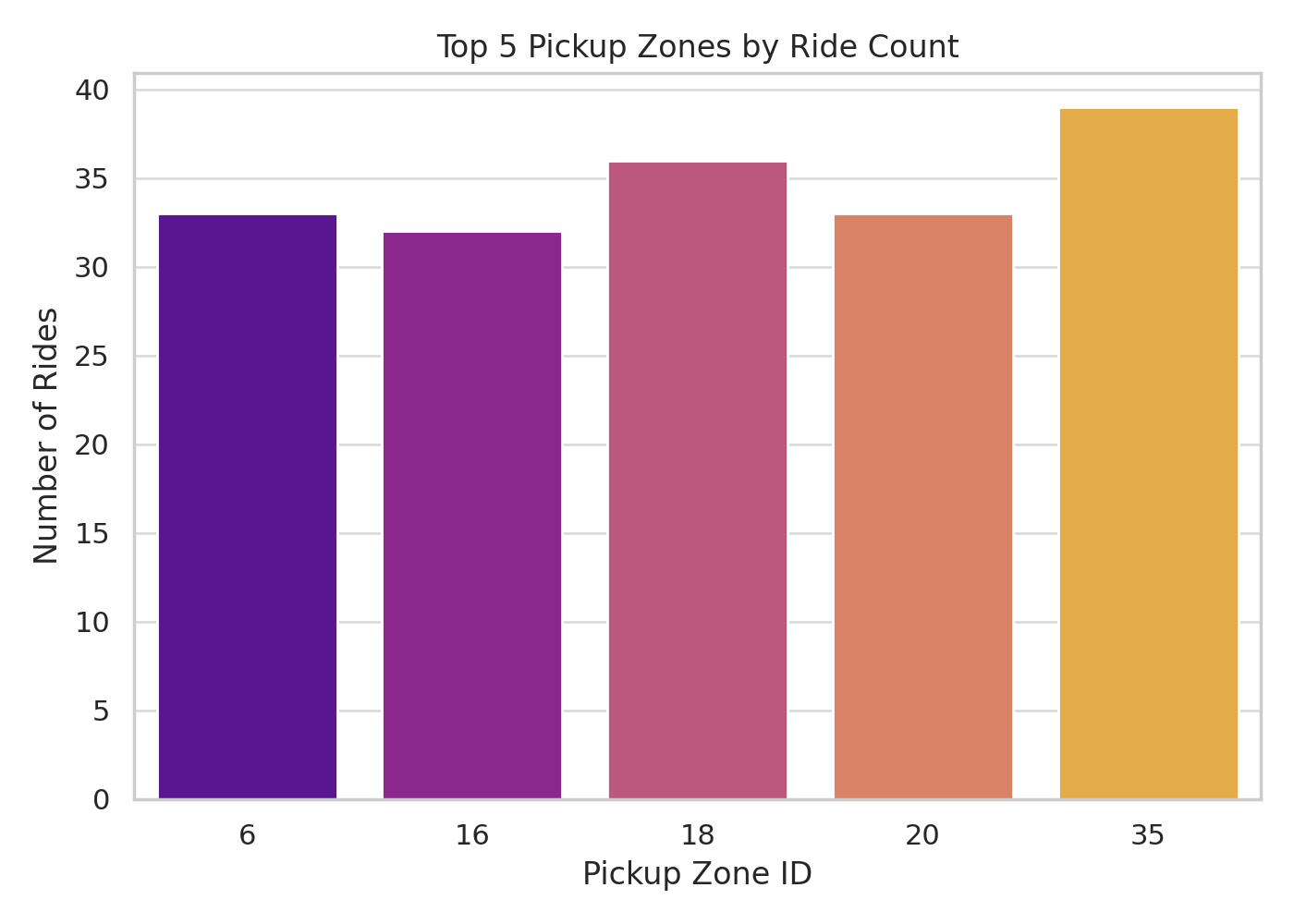
*Solution:*

* **Created a bar chart showing COUNT(tripid) grouped by payment\_method**
* **Most popular payment modes:**
* UPI **– Dominates digital payments**
* Cash **– Still widely used**
* Cards (Debit/Credit) **– Least used**
* **Business Insight: Push toward UPI and app-based payments can reduce operational friction (e.g., change handling, faster trip start)**

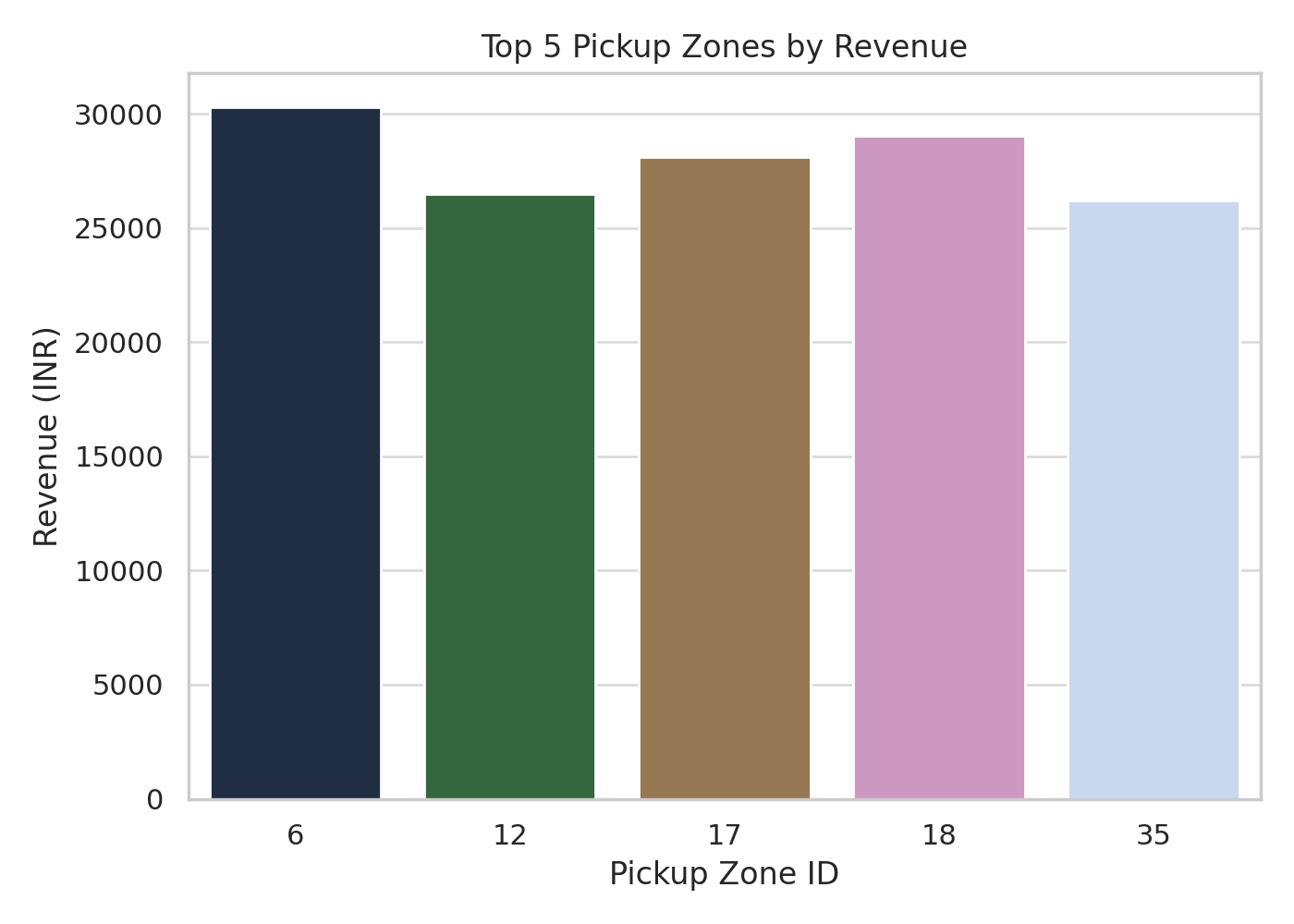
****

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:*

* **Used bar chart: loc\_from\_y (pickup ID) vs COUNT(tripid)**
* **Top pickup zones (IDs mapped to Assembly where possible) had significantly higher request volumes.**
* **These zones are likely dense commercial or transit hubs.**
* **Recommendation: Assign more drivers to these zones during rush hours.  
  **
  + 2.6.2. Revenue**: Identify pickup zones generating the highest revenue.**[3 marks]

*Solution:*

* **Created bar chart: loc\_from\_y vs SUM(fare)**
* **Top revenue zones partially overlap with high demand zones.**
* **Some zones yield higher revenue despite fewer rides (longer distances or premium customers).**
* **Suggests opportunity for** targeted premium service marketing **in these zones.**
* 

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

*Solution:*

* **Created heatmap showing duration\_label (X-axis) vs loc\_from\_y (Y-axis) with COUNT(tripid) as intensity**
* **Clear patterns emerged:**
* **Office-centric zones spike in AM & PM hours**
* **Residential zones showed steady off-peak travel**
* **Insight:** Zone-time matching **can guide driver dispatch strategy**

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:*

* **Filtered for top 5 zones using COUNT(tripid) on loc\_from\_y**
* **These zones accounted for a large percentage of total trips**
* **Contributing Factors:**
* **Dense population**
* **Proximity to offices, colleges, malls, metro stations**
* **Recommendation: Add driver incentives and increase wait-time monitoring in these hotspots**

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:*

Customer Cancellations:  
% Cancelled = 1 - (SUM(customer\_not\_cancelled) / COUNT(tripid))  
  
**12% cancelled by customer**Driver Cancellations:  
**% Cancelled = 1 - (SUM(driver\_not\_cancelled) / COUNT(tripid))  
9% cancelled by drivers**  
Successful Rides:  
**% Completed = SUM(end\_ride) / COUNT(tripid)  
77% trips completed**Customer-side drop-offs slightly higher than driver-side

* + 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:*

* Conversion Rate = SUM(end\_ride) / SUM(searches\_for\_quotes) **Created time series by duration\_label to show conversion rate trends**
* **Insight:**
* **Higher conversion during daytime**
* **Drop-offs more likely during late night or early morning**
* **May relate to price sensitivity or lower driver availability**

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:*

* **Created a parameter: Time Slot Selector with values like “0-1”, “1-2”, ..., “23-0”**
* **Used it to filter all charts dynamically**
* **Example:**
* **Selecting “9-10” updates zone performance, payment split, ride demand**
* **Insight: Enables zone-time strategies (e.g., zone-specific driver dispatch)**

### 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, the following actions are recommended to enhance operational efficiency:**

1. Dynamic Driver Allocation:
   * **Concentrate more drivers in** high-demand zones **during** 9–11 AM and 6–8 PM**.**
   * **Reduce idle time and customer wait time.**
2. Intelligent Dispatching:
   * **Use ride request density heatmaps to predict future demand.**
   * **Implement auto-suggestions for driver re-positioning in lull hours.**
3. Cancellation Mitigation:
   * **For** customer cancellations**:**
     + **Improve ETA accuracy in UI**
     + **Add ride assurance features (e.g., price lock)**
   * **For** driver cancellations**:**
     + **Introduce tiered incentives for successful rides**
     + **Penalize last-minute cancellations with soft flags**
4. Payment Processing Optimization:
   * **Prioritize UPI integration and auto-prompt it post-ride.**
   * **Push limited-time discounts to encourage digital use.**
5. Ride Duration Optimization:
   * **Track zones with consistently high trip durations.**
   * **Suggest alternate routes or pricing incentives for efficient routes.**
6. 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:*

* Zone-Based Campaigns:
* **Run local promotions in top-performing zones to increase loyalty.**
* **Example: “BTM Layout Cashback Hour – 5 to 7 PM”**
* Peak Hour Pricing Model:
* **Implement mild surge pricing in high-revenue windows**
* **Offer credits or next-ride discounts to balance fairness**
* Payment Promotions:
* **Push** UPI adoption **by offering ₹10–₹20 cashback per ride**
* **Reduce friction and increase ride completions**
* Night Hour Promotions:
* **Low demand at night = idle drivers**
* **Promote flat-rate “Midnight Rides” from 11 PM–5 AM**
* App UX Enhancement:
* **Highlight fare estimate earlier**
* **Show driver ETA range upfront**
* **Simplify quote acceptance screen**
* Customer Retargeting:
* **Follow-up SMS for dropped-off users from funnel**Example: “Didn’t complete your ride? Here’s 15% off if you book now.”