upGrad

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

DO NOT EDIT ANYTHING ON THIS PAGE

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

By - SHIVANI RAJESH MURDE, SHUBHRANGI SHARMA & SIDDHANT SINGH

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:

---

*The Namma Yatri dataset was imported into Power BI as an Excel file, with each relevant table(**Assembly, Duration, Payment, Trip Details, and Trips**)loaded and their data types explicitly defined to ensure consistency and integrity. The data type transformations were performed using Power Query, as shown below:*

- ***Assembly:** ID (Int64), Assembly (Text)*
- ***Duration:** id (Int64), duration (Text)*
- ***Payment:** id (Int64), method (Text)*
- ***Trip Details:** All key columns (e.g., tripid, loc_from, searches, etc.) set to Int64*
- ***Trips:** All key columns (e.g., tripid, faremethod, fare, loc_from, etc.) set to Int64*

After importing, we established and verified the relationships between all tables within the Power BI Model view. The schema ensures that each table is properly linked via their respective key columns (e.g., tripid, assembly ID, duration ID, payment ID), supporting a unified and accurate dataset for analysis. This alignment is visually confirmed in the attached schema diagram, which demonstrates the integrity of the model and the correct mapping of relationships between tables1.

*To further optimize the analytical workflow, we created two extended tables(**Merger1** and **Merger2**) using Power Query's table expansion functions.*

- ***Merger1*** *expands the Assembly table, flattening relevant fields for easier access and analysis.*

- ***Merger2*** *expands the Trips table, making all trip-related fields directly accessible for downstream analysis.*

Both Merger1 and Merger2 are fully integrated into the data model, with relationships defined to all relevant tables, ensuring seamless data alignment across the unified dataset.

This structured approach guarantees that the dataset is clean, relationships are robust, and the model is optimized for generating accurate and actionable insights in Power BI.

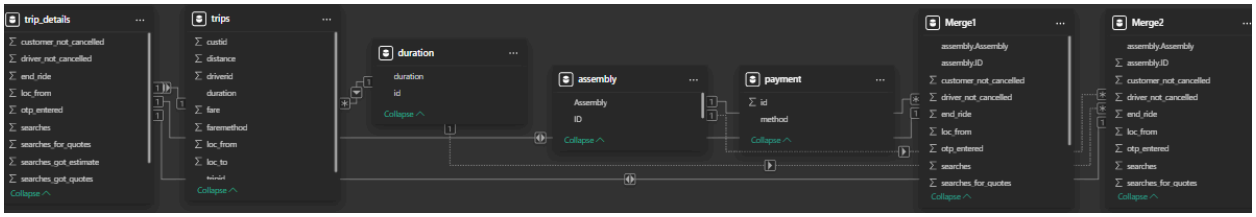


Fig1 - Schema Representing the Relationship between the Tables

---

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:

During the data preparation phase in Power BI, we identified and addressed several inconsistencies that could impact the quality of our analysis. Notably, Power BI initially set the ID fields in various tables to aggregate as sums by default. Recognizing that these identifiers are categorical and not meant for aggregation, we promptly updated their settings to “Do not summarize,” preserving their integrity for accurate relational mapping.

We conducted a thorough data cleaning process, which included:

- Removing irrelevant and duplicate records to eliminate redundancy.
- Filtering out erroneous entries, such as incomplete or logically inconsistent data points, to ensure only valid data was retained for analysis.

To enhance the relevance and accuracy of our insights, we created calculated fields tailored to our analytical needs. For example, we introduced a conversion rate metric, defined as:

$$\text{Conversion Rate} = \frac{\sum(\text{trip\_details[end\_ride]})}{\sum(\text{trip\_details[searches]})}$$

This calculated field provides a clear measure of the proportion of searches that resulted in completed rides, offering valuable insight into operational efficiency.

Through these steps, we ensured the dataset is clean, well-structured, and analytically robust, supporting precise and meaningful business insights in Power BI

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:

Answer:

The variables in the Namma Yatri dataset can be classified as follows:

Numerical Variables

These represent quantitative values that can be measured or counted, such as IDs, counts, fares, distances, and durations:

- **tripid**
- **fare**
- **distance**
- **duration**
- **loc_from**
- **loc_to**
- **driverid**
- **custid**
- **searches**
- **searches_got_estimate**
- **searches_for_quotes**
- **searches_got_quotes**

- **customer_not_cancelled**
- **driver_not_cancelled**
- **otp_entered**
- **end_ride**
- **id** (in Assembly, Duration, Payment tables)

Categorical Variables

These describe qualitative characteristics or categories, such as names, methods, and types:

- **Assembly** (zone or area name)
- **duration** (when used as a time slot label, e.g., "0-1", "1-2")
- **method** (payment method, e.g., "Cash", "UPI")
- **faremethod** (type of fare calculation or payment)

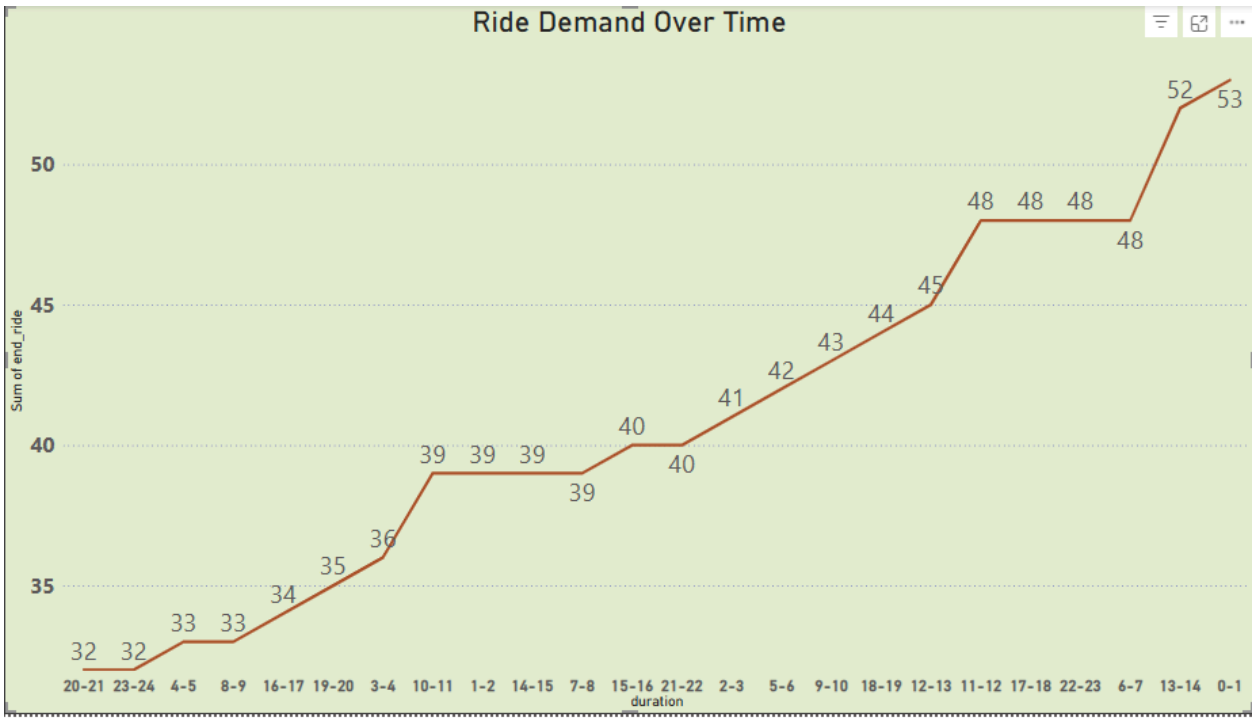
In summary:

- **Numerical variables** are all columns representing amounts, counts, or IDs.
- **Categorical variables** are those representing categories, labels, or descriptive groupings.

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:

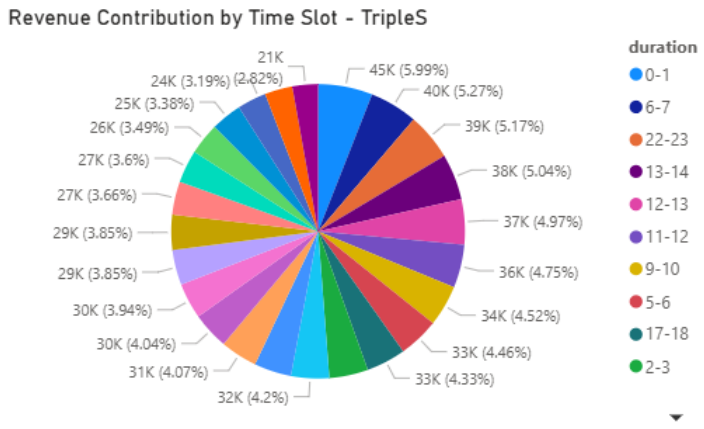


- **Gradual Increase During Early Hours:**Ride demand is relatively low and stable during the late night and early morning hours (e.g., 20-21, 23-24, 4-5) with completed rides in the low 30s.
- **Steady Growth Through the Morning:**From 8-9 onwards, there is a consistent upward trend in demand, indicating increased ride activity as the day progresses.
- **Significant Afternoon and Evening Growth:**Demand continues to rise through the afternoon, reaching the mid-40s by 12-13 and peaking at 48 rides during the 11-12, 12-13, and 22-23 slots.
- **Peak Demand Periods Identified:** The highest ride demand is observed during the 13-14 (1–2 PM) and 0-1 (12–1 AM) slots, with 52 and 53 rides respectively. These represent the clear peak periods in the dataset.
- **Parameter Justification:**The analysis uses the sum of end_ride as the demand parameter, which effectively captures the number of trips completed in each time slot and is a direct measure of actual 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 pie chart highlights that the **0–1 AM time slot** contributes the **highest proportion of revenue** at **5.99%**, with a total of ₹45,000. This suggests strong late-night demand, possibly from airport trips or night-shift workers.

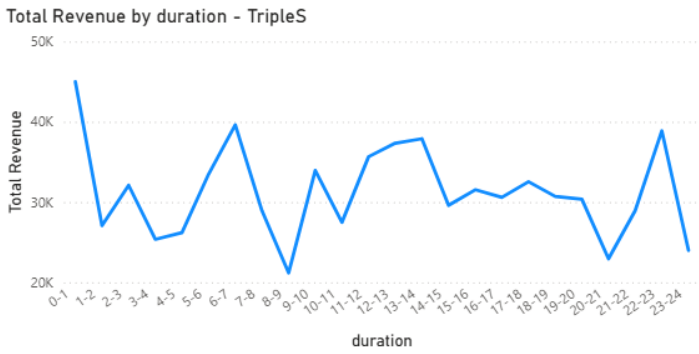
In contrast, the **8–9 PM slot** contributes just **2.82%** (₹21,000), indicating either **lower demand** or **shorter rides** during this period. This proportion analysis reveals that revenue is not evenly distributed across the day, and **late-night hours can outperform traditional peak hours** in revenue contribution.

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 line chart shows how revenue changes throughout the day based on trip hour.

- The **highest revenue** was generated between **12–1 AM (0–1)** with **₹45,000**, highlighting strong late-night demand, possibly due to **airport transfers**, **late-shift workers**, or **intercity travelers**.

- In contrast, the **8–9 PM (20–21)** time slot brought in only **₹21,000**, which is unexpectedly low for a typical evening rush hour. This suggests either **fewer trips** or **shorter ride distances** during that period.

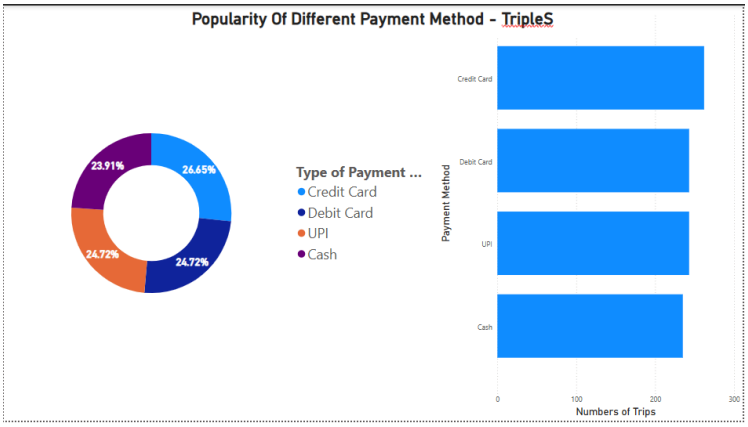
Pattern Observed:

- Revenue is **not uniformly distributed** across hours.
- Specific slots like **midnight** and **early morning** tend to contribute more.
- This insight can help in **driver scheduling** and **promotional planning** for high-revenue time slots.

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:



- Payment methods (Credit Card, Debit Card, UPI, Cash) are almost equally popular, each accounting for about 24–27% of rides.
- No single payment method dominates; usage is well balanced across all options.
- Ride frequency is similar across payment methods, indicating payment choice does not affect how often customers take rides.
- This balanced distribution suggests strong payment flexibility and customer preference diversity.
- Maintaining multiple payment options supports customer satisfaction and consistent ride volume.

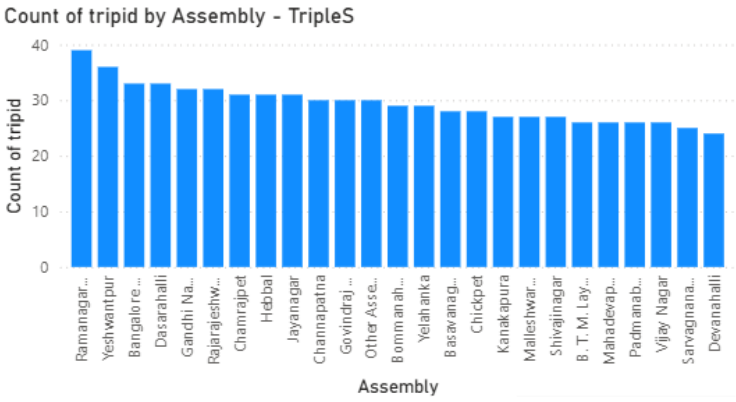
2.6. Identify High-Performing Zones [6 Marks]

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:

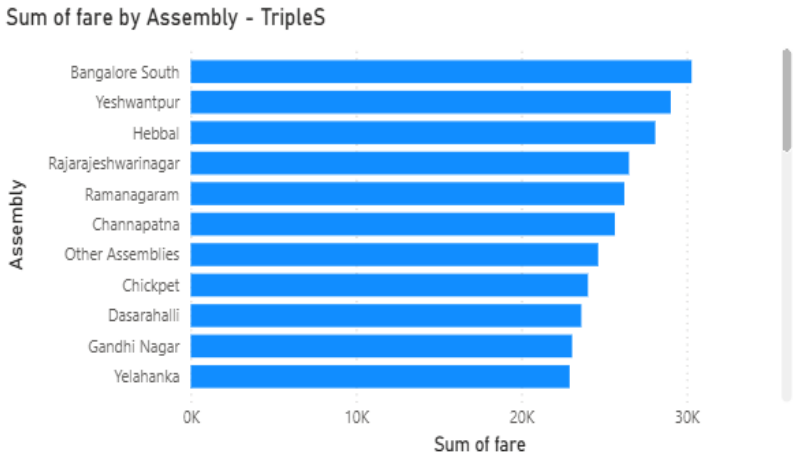


The analysis shows that **Ramangaram** and **Yeshwantpur** are the top-performing pickup zones, with 39 and 36 trip requests respectively.
These areas likely see high demand due to:

- Proximity to major transport hubs (e.g., highways or industrial areas)
- Limited local transport options, leading to higher app-based ride usage
- Possibly more long-distance or intercity trip bookings (like outstation rides)

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

Solution:



The analysis of revenue by pickup zone reveals that:

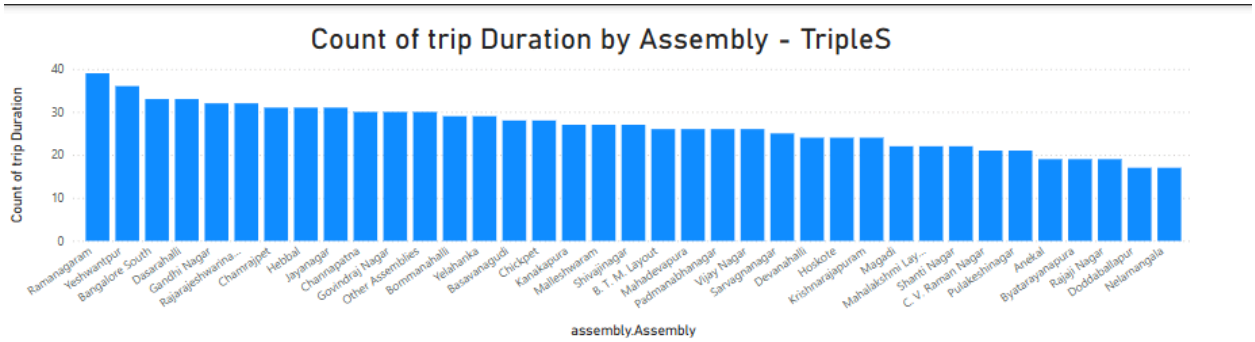
- **Bangalore South** is the highest revenue-generating zone with a total of **₹30,000**.
- **Yeshwanthpur** follows closely with **₹29,000**.
- On the lower end, **Neelamangala** generated the least revenue at only **₹11,000**.

These findings suggest that **Bangalore South and Yeshwanthpur** are strategic hubs with high ride volume or longer trip distances, leading to greater revenue. In contrast, **Neelamangala** may see fewer trip requests or primarily shorter, low-fare rides.

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

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

Solution:



assembly.Assembly		0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	Total
Mahadevapura	40	1		1	1	1	2	2	1	1	2		1	2	3	2				1			2	1	2	75
Ramanagaram	35	2	1	1	2	1	3	1	3	4	1					1		4	1	1	1	3	3	1	5	74
Gandhi Nagar	39	3	2	1			1	3			2	1	1	2	1	2	3		3	2	1	1		2	1	71
Other Assemblies	40	2		1			1		1	2		4	1	2	2	2	1	1		1	1	1	2	4	1	70
Yeshwantpur	33		1	1	2	1	2		1	1	1	1	5	2		1	2	1	2	4		1	2	2	3	69
Dasarahalli	33	4	1		2	1		1		1	5		4	1	1	3	1	1	3	1	1	1		1		66
Jayanagar	34	2	5		2		4	2	2		2				3	1		1	2	1		1	1	1	1	65
C. V. Raman Nagar	43		1		3		2	1	3	1	1		1	1		1			1				2	3		64
Rajarajeshwarinagar	32	1	2	2		3	1		1	3			1	3	1	1	2	2		2		3	1	2	1	64
Yelahanka	34	2	1	1	2	1	1	3		1	1	2			1		1	1	1	3	1	3	1	2		63
Hebbal	31		2	1		1	1	1	1			2	2	3	1	1	3	2	1	2	4	1		1	1	62
Chickpet	33	2		3				3	1		2		3	2	2		2		2		2		2	1	1	61
Kanakapura	34	1		1	2	1			2	3	1	3	1	3	2	1	1	1	1			1	1	1		61
Anekal	41	1	1	1	1	1		1	2		3		2	1	1				2			1	1			60
Malleshwaram	33	1	1	1		2	2			1		3	3		1	2	3	2	2		1		1	1		60
Basavanagudi	31	2	2		2		2	7	1	1	1	2			1		1		1	1	2			2		59
Govindraj Nagar	29		1		2				2	1	1		1	2	2	1	2	4		1	1	1	3	1	4	59
Total	**1178**	**53**	**39**	**41**	**36**	**33**	**42**	**48**	**39**	**33**	**43**	**39**	**48**	**45**	**52**	**39**	**40**	**34**	**48**	**44**	**35**	**32**	**40**	**48**	**32**	**2161**

Overall Distribution of Ride Durations:

- The heatmap shows that trips are distributed across all duration periods, indicating a mix of short, medium, and longer journeys ending in various zones.
- There isn't one single duration period that overwhelmingly dominates across all zones, suggesting a diverse range of travel needs.

High-Activity Dropoff Zones:

- **B. T. M. Layout** and **Yeshwantpur** appear to be among the most active dropoff zones overall, showing darker shades across a broader range of duration periods compared to others. This suggests they are frequent destinations for various trip lengths.
- Other zones like **Bangalore South** and **Bommanahalli** also show significant activity, but perhaps with more concentrated patterns.

Specific Zone-Duration Hotspots:

- **B. T. M. Layout**: Shows relatively high trip counts for very short durations (e.g., '1-2', '2-3' minutes) and also significant activity in medium-long ranges like '10-11' and '18-19' minutes.
- **Yeshwantpur**: Appears to be a consistent dropoff point across many duration periods, with notable activity in 11-12, 15-16, and 17-18 minutes.
- **Bommanahalli**: While it receives trips across durations, there's a particular concentration in the 13-14 minute range. Also shows consistent activity in the 5-6 and 21-22 minute periods.
- **Anekal**: Exhibits activity across the spectrum, with noticeable peaks in the 6-7 and 19-20 minute periods.
- **Bangalore South**: Shows a diverse pattern, with some distinct hotspots in 16-17 and 23-24 minute durations.

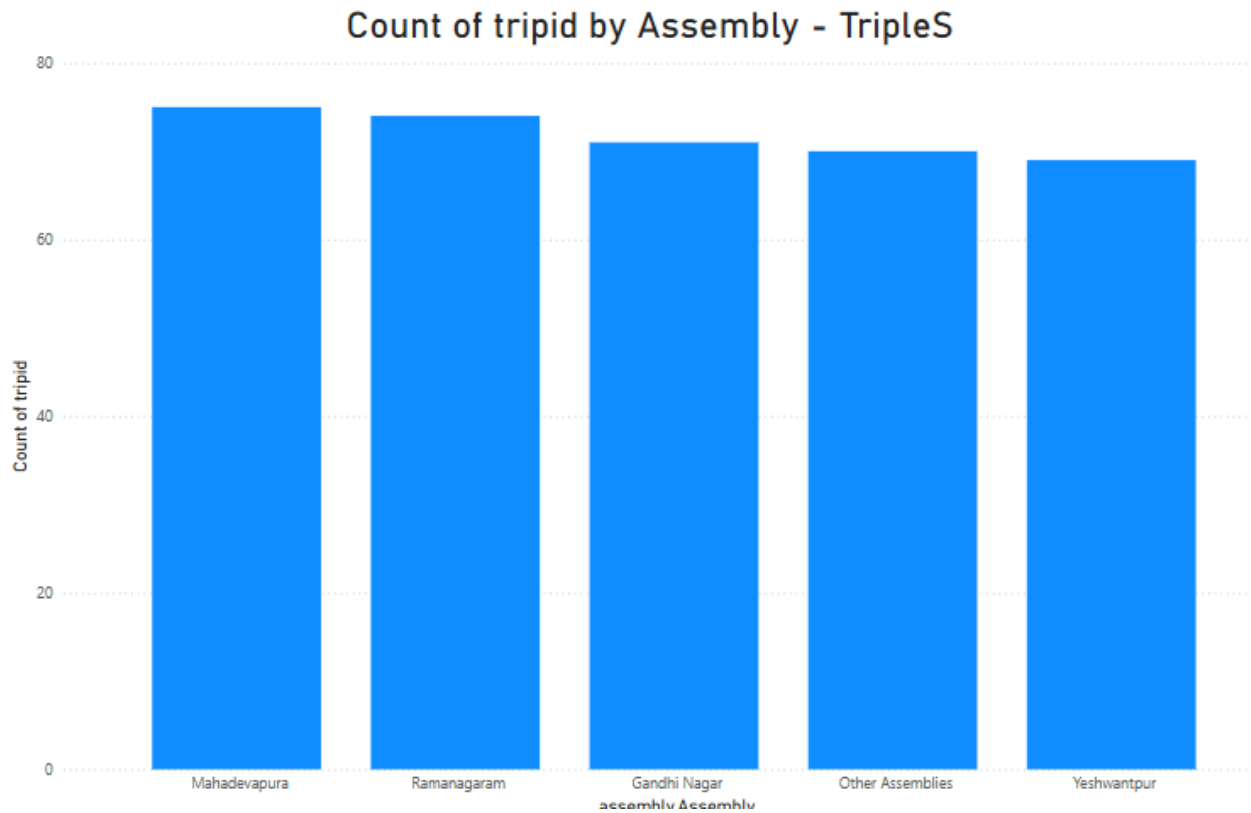
Least Common Dropoff Durations:

- For many zones, some duration periods, particularly those in the 20-21 to 23-24 minute range, show fewer trips (lighter cells), indicating these are less frequent travel times to those specific destinations. However, some zones like Bangalore South and Yeshwantpur still show a decent number of longer trips.

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 high trip volumes from these specific zones can be attributed to several factors, reflecting their unique characteristics within the urban and peri-urban landscape:

- **"Other Assemblies" (Collective Demand):** The fact that "Other Assemblies" collectively constitute the top pickup zone highlights the widespread adoption of Namma Yatri across a broad geographical spectrum. It suggests that demand isn't concentrated solely in a few major hubs but is significant even from smaller, aggregated areas, indicating strong penetration into diverse neighborhoods and local communities. This category's dominance implies that granular, localized demand, when combined, surpasses even major individual zones.
- **Yeshwantpur (Major Transport Hub):** As a prominent hub in Bengaluru,

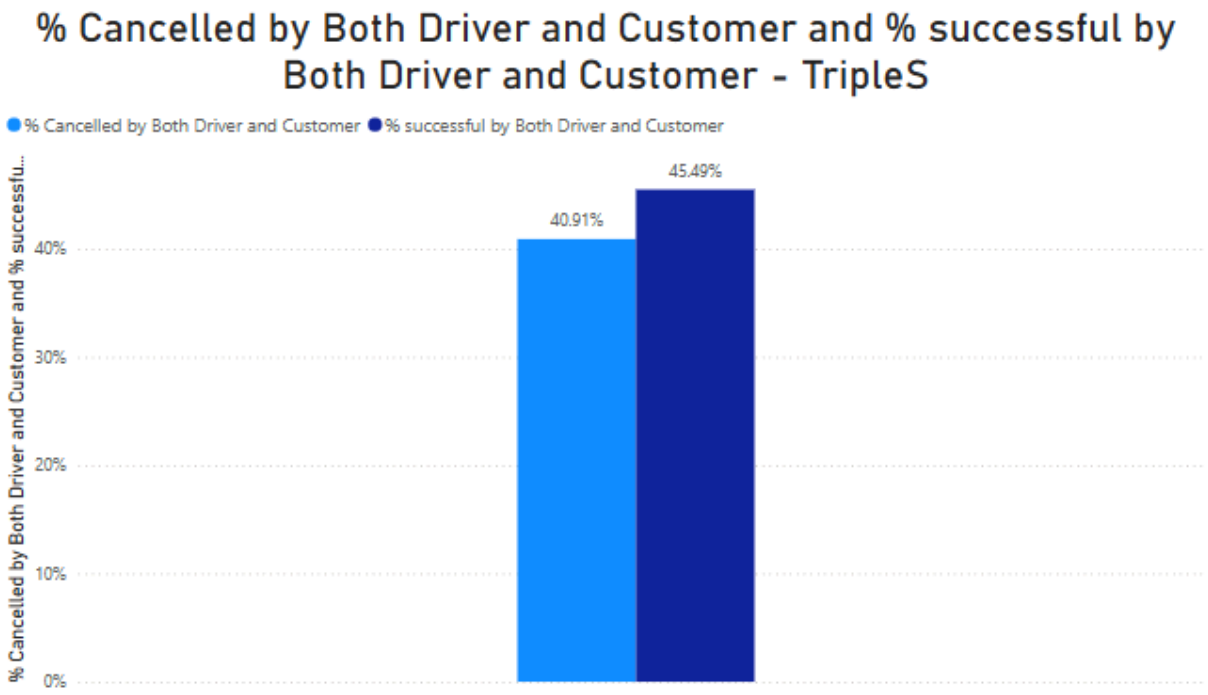
Yeshwantpur hosts a major railway station, a metro station, and a large bus terminal. This naturally generates immense trip volume from individuals starting their local journeys after arriving from other cities or using these inter-city transport modes. It's a key first-mile/last-mile connectivity point.

- **Mahadevpura (IT/Commercial Corridor):** Mahadevpura is a significant IT and commercial corridor in East Bengaluru, home to numerous tech parks and business establishments. The high volume of trips from here is likely driven by daily commutes of office workers during peak hours, as well as business-related travel throughout the day.
- **Gandhi Nagar (Central Commercial & Cultural Hub):** Located in the heart of Bengaluru, Gandhi Nagar is a bustling commercial area, known for its markets, theaters, and a mix of residential and business establishments. Its central location and diverse activities contribute to consistent trip generation for shopping, entertainment, and general city movement.
- **Ramanagara (Peri-Urban/Outskirts Connectivity):** Ramanagara, being further on the outskirts of Bengaluru, suggests a demand pattern related to journeys connecting peri-urban areas with the main city or other regional destinations. Trips from here might be longer commutes or serve specific local needs within its jurisdiction and surrounding areas, including industrial or agricultural activities.

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:



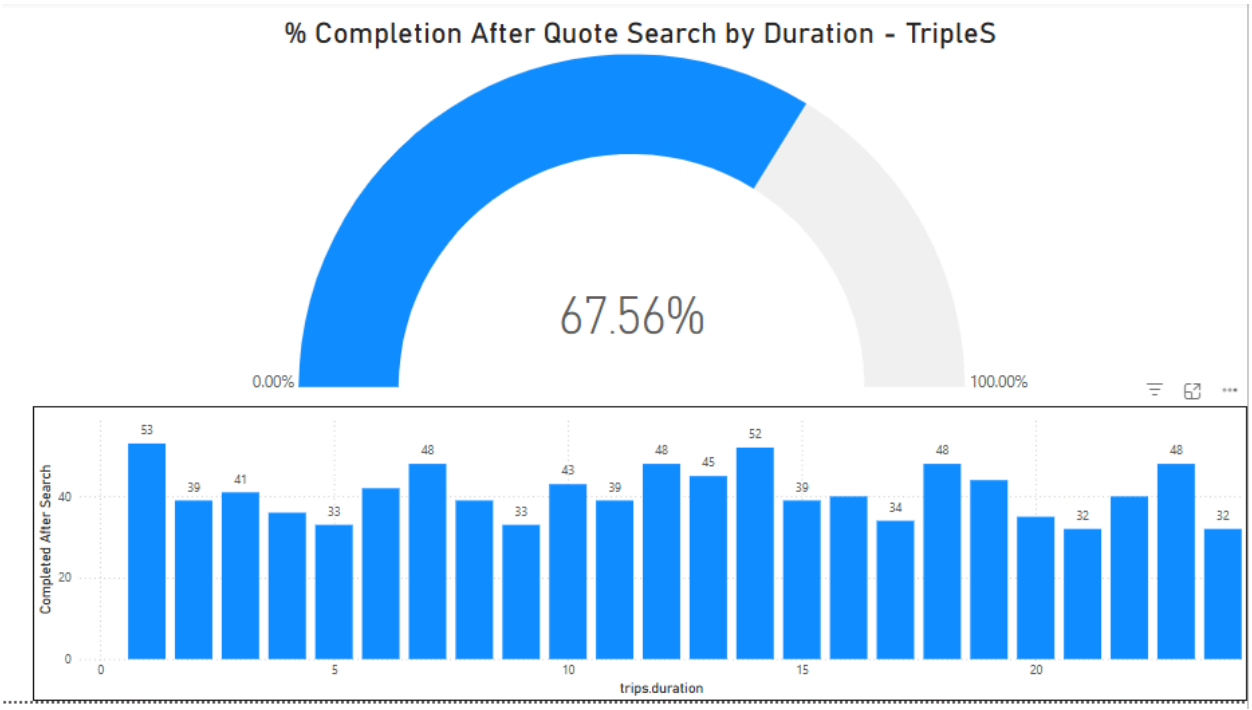
- Drivers and customers share an identical cancellation rate of 40.91%.
- Both groups also mirror each other with a 45.49% successful transaction rate.
- Successful trips exceed cancellations by just 4.58 percentage points for each

cohort.

- The remaining 13.6% of activity isn't captured by these two metrics, hinting at other outcomes (e.g., no-shows or pending).

- **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:

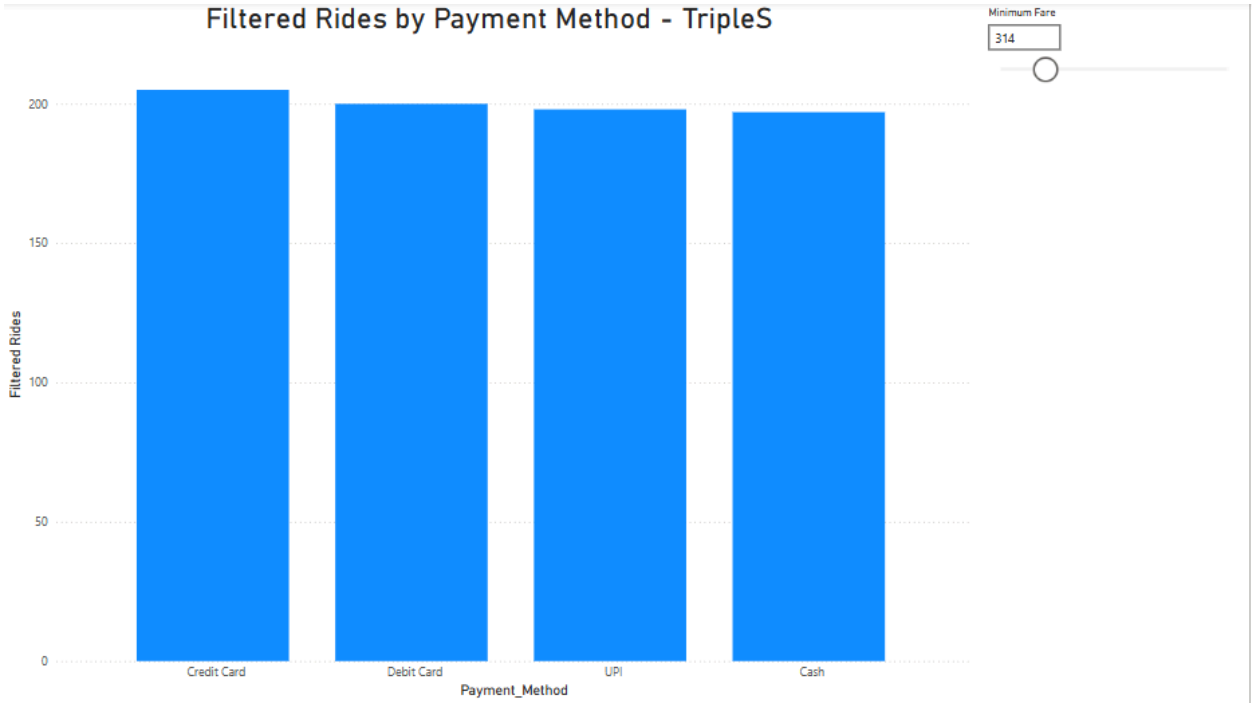


- The overall completion rate after searching for quotes stands at 67.56%.
- Completions peak for immediate trips (0 min lead time) with 53 successful bookings.
- As lead time increases, completions decline, falling to 38 by the 10 min mark and 32 at 20 min.
- A mid-range uptick occurs at 15 min (40 completions), indicating moderate wait times still convert well.

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:



- **Parameter Setup:**
  We created a parameter in Power BI (e.g., "Minimum Fare") using the Modeling

tab's parameter feature. This parameter allows interactive selection of a fare threshold.

- **Filter Application:**
  We used the parameter as a filter on the Trips data, so only rides with a fare equal to or above the selected threshold are displayed in the visuals. The parameter was added as a slicer, enabling dynamic adjustment.
- **Reason for Filter Choice:**
  Filtering by fare amount helps focus analysis on higher-value trips, revealing patterns among premium rides and customer segments.
- **Insights Drawn:**

  By adjusting the parameter, We observed how the distribution of rides, payment methods, and demand trends change for different fare levels. For example, as you can see the visual fares may correlate with the payment methods.

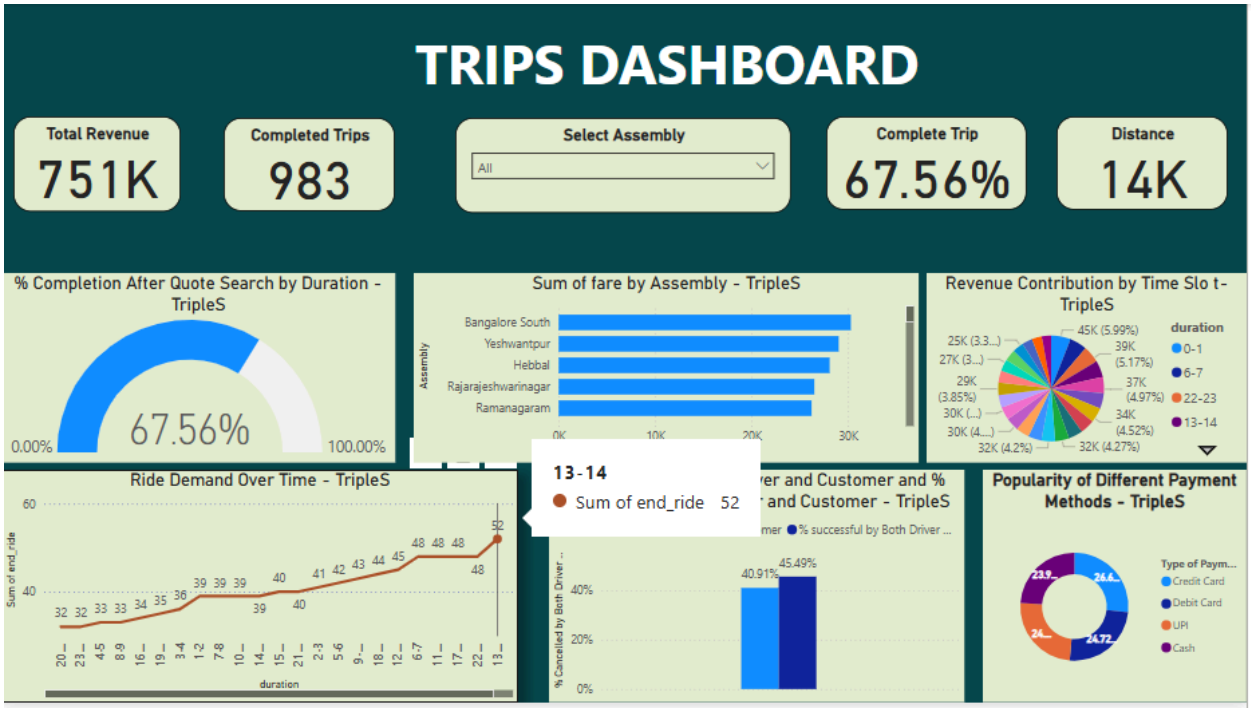
---

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. **Deploy Drivers Strategically:** Use demand data to position more drivers in high-demand zones and during peak hours to reduce wait times and missed rides.

2. **Reduce Cancellations:** Identify and address cancellation hotspots by improving communication, offering real-time updates, and incentivizing successful ride completions.

3. **Optimize Routes:** Implement smart routing based on real-time traffic to minimize ride durations and improve overall trip efficiency.

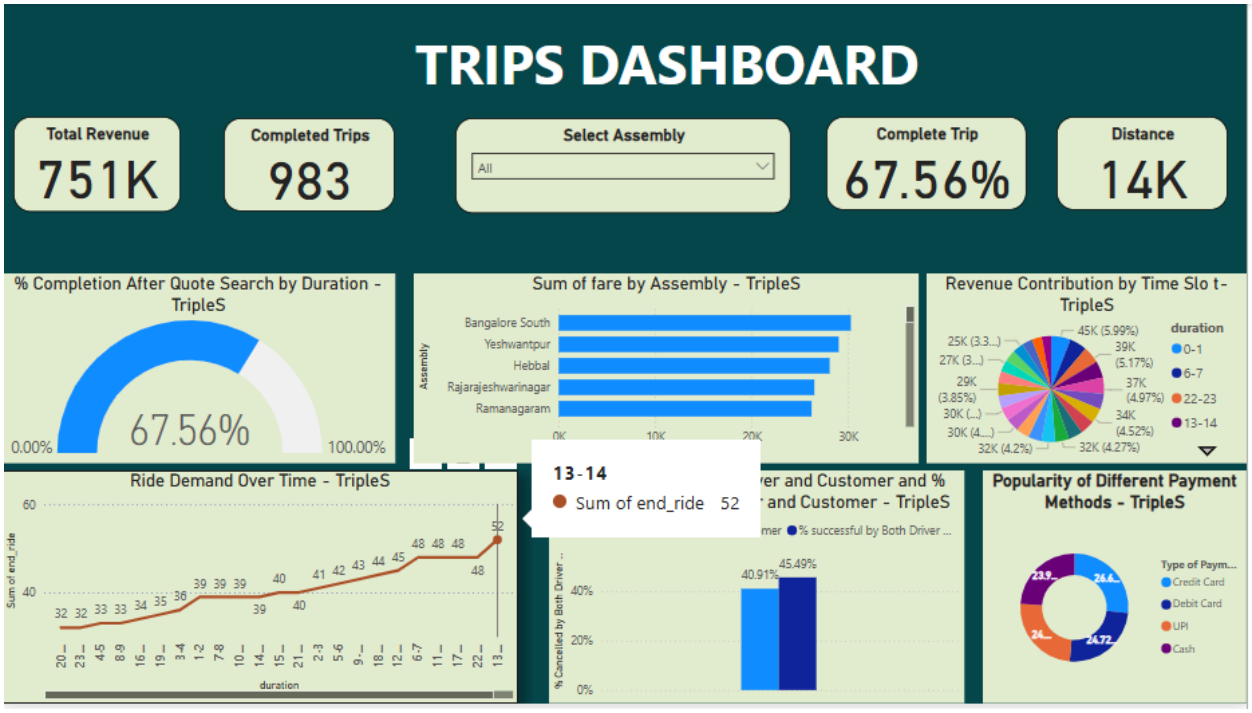
4. **Enhance Payment Experience:** Continue supporting multiple payment options and ensure smooth, hassle-free transactions for both riders and drivers.

5. **Leverage Real-Time Data:** Use interactive dashboards to monitor key metrics and make quick, data-driven decisions for resource allocation and service improvements.

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. **Region-Specific Promotions:** Launch targeted offers and discounts in high-potential zones and during local festivals to boost ridership and engage users at a community level.

2. **Referral and Loyalty Programs:** Incentivize both riders and drivers with referral bonuses, cashback, and loyalty rewards to encourage repeat usage and word-of-mouth growth.

3. **Strategic Partnerships:** Collaborate with local businesses, metro stations, and event organizers for last-mile connectivity and cross-promotional opportunities, expanding reach and convenience.

4. **Driver Incentives:** Offer performance-based incentives and flexible scheduling, especially during peak demand hours or in underserved regions, to improve service reliability and driver retention.

5. **Influencer and Social Media Marketing:** Leverage local influencers and social media campaigns to build brand awareness, share user stories, and attract younger, tech-savvy customers.