Data-Driven Optimization of Rajput Travel's Services

ENDTERM SUBMISSION

Submitted by

Name: Sarita Kumari

Roll number: 22f3000364

IITM BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary

In the highly seasonal domestic travel industry, small-scale agencies like Rajput Travel, based in ***Agra, Uttar Pradesh***, often face challenges in sustaining profitability and resource efficiency. This project examines the operational and financial performance of Rajput Travel, managed by a team of 10 employees under the ownership of ***Mr. Ramveer Singh Rajput***, by leveraging booking data collected over multiple months.

The project addresses the problem statement: **"*Optimizing Seasonal Demand Management and Profitability for Rajput Travel Agency*"**. Key issues include inconsistent revenue due to fluctuating seasonal demand, underutilization of fleet resources during off-peak periods, and the absence of customer-specific service strategies.

To investigate these challenges, structured data was collected through interviews and organizational records, including booking details, vehicle types, customer segments, and cost structures. ***Python (Pandas, Matplotlib) and Microsoft Exce***l were used for data processing, cleaning, and visualization. The dataset consists of 401 booking entries across 17 key attributes.

Line plots and histograms were used to analyze trip frequency and revenue trends over time, while bar charts highlighted variations in fare distribution across customer segments. Segmentation into Budget, Mid, and High Spenders revealed that high-spending customers contribute over 60% of total profits. Additionally, vehicle-wise analysis indicated that certain cars (e.g., Crista, Innova) consistently generate more revenue.

The ***Pareto Principle (80/20 Rule)*** was applied to identify top-performing combinations of destinations and customer types that account for the majority of revenue. Metadata quality, potential biases (e.g., manual data entry), and outlier effects were also carefully assessed.

2. Detailed Explanation of Analysis Method

The data for this project was collected from Rajput Travel Agency, a local travel service provider based in Agra, for a span of approximately ***13 months***. The dataset was sourced through internal business records, and direct interviews with the owner ***Mr. Ramveer Singh Rajput*** and team members.

The final dataset comprises ***401 rows and 17 columns***, covering critical information related to customer bookings, travel duration, vehicle types, payment modes, and profitability metrics.

Dataset Overview:

Customer_ID	Booking_Date	Pickup_Location	Destination	No_of_Days	Distance_km	Season	Car_Name	Total_Amount	Total_Cost	Total_Profit	Mode_of_Payment	Additional_Costs	Day_of_Week	Customer_Segment	Seasonal_Demand	Destination_Popul
1	1/1/2024	Udaipur	Bangalore	6	1800	Winter	Artica	34200	12600	21600	Online	436	Monday	Mid-Spender	Peak	Low
2	1/2/2024	Jaipur	Odisha	7	2100	Winter	Innova	48300	16800	31500	Online	306	Tuesday	High-Spender	Peak	High
3	1/3/2024	Shimla	Agra	4	2248	Winter	Shift Dizire	35968	13488	22480	Card	61	Wednesday	Mid-Spender	Peak	Low
4	1/4/2024	Bangalore	Pune	7	2476	Winter	Carence	54472	19808	34664	Cash	113	Thursday	High-Spender	Peak	Low
5	1/5/2024	Pune	Odisha	4	1200	Winter	Aura	19200	7200	12000	Online	345	Friday	Budget-Spender	Peak	High
6	1/6/2024	Pune	Rishikesh	3	2375	Winter	Aura	38000	14250	23750	Card	318	Saturday	Mid-Spender	Peak	High
7	1/7/2024	Goa	Varanasi	5	1500	Winter	Crista	40500	15000	25500	Cash	220	Sunday	High-Spender	Peak	Medium
8	1/8/2024	Shimla	Manali	7	2100	Winter	Artica	39900	14700	25200	Online	465	Monday	Mid-Spender	Peak	Low
9	1/9/2024	Delhi	Bangalore	2	2150	Winter	Toyota Rumion	45150	17200	27950	Cash	255	Tuesday	High-Spender	Peak	Low
10	1/10/2024	Udaipur	Delhi	1	1656	Winter	Crista	44712	16560	28152	Online	59	Wednesday	High-Spender	Peak	High
11	1/11/2024	Shimla	Goa	1	724	Winter	Innova	16652	5792	10860	Card	265	Thursday	Budget-Spender	Peak	High
12	1/12/2024	Agra	Delhi	1	1192	Winter	Etios	19072	7152	11920	Card	57	Friday	Budget-Spender	Peak	High
13	1/13/2024	Rishikesh	Agra	7	2100	Winter	Artica	39900	14700	25200	Online	359	Saturday	Mid-Spender	Peak	Low
14	1/14/2024	Mumbai	Delhi	3	1765	Winter	Etios	28240	10590	17650	Cash	244	Sunday	Mid-Spender	Peak	High
15	1/15/2024	Udaipur	Kolkata	4	1798	Winter	Tata Zest	28768	10788	17980	Online	71	Monday	Mid-Spender	Peak	Low
16	1/16/2024	Kolkata	Manali	7	2100	Winter	Carence	46200	16800	29400	Cash	57	Tuesday	High-Spender	Peak	Low
17	1/17/2024	Varanasi	Odisha	4	1200	Winter	Tata Zest	19200	7200	12000	Cash	371	Wednesday	Budget-Spender	Peak	High
18	1/18/2024	Mumbai	Varanasi	3	2465	Winter	Shift Dizire	39440	14790	24650	Card	229	Thursday	Mid-Spender	Peak	Medium
19	1/19/2024	Goa	Hyderabad	2	2198	Winter	Etios	35168	13188	21980	Cash	291	Friday	Mid-Spender	Peak	Medium
20	1/20/2024	Udaipur	Shimla	7	2115	Winter	Innova	48645	16920	31725	Online	148	Saturday	High-Spender	Peak	Medium
21	1/21/2024	Rishikesh	Agra	4	1200	Winter	Innova	27600	9600	18000	Card	200	Sunday	Mid-Spender	Peak	Low
22	1/22/2024	Delhi	Rishikesh	4	2496	Winter	Artica	47424	17472	29952	Card	182	Monday	High-Spender	Peak	High
23	1/23/2024	Jaipur	Agra	3	917	Winter	Carence	20174	7336	12838	Cash	468	Tuesday	Budget-Spender	Peak	Low
24	1/24/2024	Varanasi	Manali	4	2365	Winter	Artica	44935	16555	28380	Online	169	Wednesday	High-Spender	Peak	Low
25	1/25/2024	Mumbai	Bangalore	1	1959	Winter	Aura	31344	11754	19590	Cash	203	Thursday	Mid-Spender	Peak	Low
26	1/26/2024	Jaipur	Udaipur	7	2100	Winter	Shift Dizire	33600	12600	21000	Online	289	Friday	Mid-Spender	Peak	Medium
27	1/27/2024	Jaipur	Mumbai	4	1200	Winter	Etios	19200	7200	12000	Online	387	Saturday	Budget-Spender	Peak	Medium
28	1/28/2024	Kolkata	Agra	1	2130	Winter	Tata Zest	34080	12780	21300	Card	355	Sunday	Mid-Spender	Peak	Low
29	1/29/2024	Rishikesh	Pune	1	1519	Winter	Artica	28861	10633	18228	Cash	408	Monday	Mid-Spender	Peak	Low
30	1/30/2024	Kolkata	Varanasi	7	2100	Winter	Artica	39900	14700	25200	Card	89	Tuesday	Mid-Spender	Peak	Medium
31	1/31/2024	Varanasi	Odisha	1	2460	Winter	Tata Zest	39360	14760	24600	Online	157	Wednesday	Mid-Spender	Peak	High
32	2/1/2024	Varanasi	Rishikesh	4	1210	Winter	Crista	32670	12100	20570	Online	422	Thursday	Mid-Spender	Off-Peak	High
33	2/2/2024	Varanasi	Goa	7	2100	Winter	Carence	46200	16800	29400	Online	129	Friday	High-Spender	Off-Peak	High
34	2/3/2024	Varanasi	Jaipur	7	2100	Winter	Tata Zest	33600	12600	21000	Card	235	Saturday	Mid-Spender	Off-Peak	High
35	2/4/2024	Mumbai	Varanasi	2	2026	Winter	Aura	32416	12156	20260	Online	56	Sunday	Mid-Spender	Off-Peak	Medium
36	2/5/2024	Jaipur	Udaipur	4	1891	Winter	Tata Zest	30256	11346	18910	Online	174	Monday	Mid-Spender	Off-Peak	Medium
37	2/6/2024	Kolkata	Bangalore	5	2170	Winter	Tata Zest	34720	13020	21700	Online	405	Tuesday	Mid-Spender	Off-Peak	Low

The link to Dataset: [Click here to view the Dataset]

Data Collection Tools:

- Microsoft Excel: For initial dataset structuring and metadata understanding
- Google Forms & Interviews: For primary data gathering from Rajput Travel
- Python Libraries: Pandas (for data cleaning & processing), Matplotlib & Seaborn (for visualizations)

📊 Why This Approach?

The analysis method was chosen to meet real-world business needs while maintaining academic rigor:

- Time-based trends (seasonality) → identify high/low demand months
- Segmentation → understand which customer groups generate most revenue
- Vehicle-level profitability → find underperforming and top-performing assets
- Pricing & cost structure analysis → support dynamic pricing decisions

A) Data Preprocessing and Cleaning

Given the structured but manually maintained dataset, several preprocessing steps were essential to ensure the accuracy, consistency, and usability of the data for analysis:

- **Date Formatting:** The Booking_Date field was converted into standard datetime format. Additional fields such as Day_of_Week, Month, and Season were derived to support temporal analysis (e.g., peak days/seasons).
- **Removal of Irrelevant Fields:** No column was entirely irrelevant, but a preliminary check ensured no extra/empty fields were present.
- **Standardization:** Fields like Mode_of_Payment, Customer_Segment, and Seasonal_Demand were standardized to uniform categories (e.g., "Cash", "Online", "High", "Low").
- **Duplicate Detection:** Duplicate bookings (same date, distance, customer, and vehicle) were detected and removed — approx. 1.4% of entries.
- **Handling Missing/Incorrect Values:** Minor missing entries (e.g., in Additional_Costs) were either imputed using the median or flagged. Revenue calculations (Total_Profit) were validated using:

Total_Profit = Total_Amount – Total_Cost – Additional_Costs

- **Numerical Sanity Checks:** Ensured all amounts (fare, cost, profit) were non-negative and realistic. Outliers in fare or profit (e.g., above ₹90,000) were inspected and retained only if justified (e.g., long-distance travel).
- **Encoding:** Categorical fields (e.g., Car Name, Payment Mode) were label-encoded during Python analysis for correlation and plotting.

B) Exploratory Data Analysis (EDA)

EDA was conducted using **Python (Pandas, Seaborn, Matplotlib)** to uncover patterns, trends, and customer behaviors.

- **Monthly Trip Trends:** A line chart was used to visualize booking volume and revenue over time. Seasonal spikes (e.g., March-April) were clearly visible, supporting the issue of fluctuating seasonal demand mentioned in the problem statement.
- **Day-wise Revenue Pattern:** Bar plots showed that Wednesdays consistently had higher spending, suggesting potential mid-week pricing opportunities.
- **Customer Segment Distribution:** Pie and bar charts highlighted the contribution of each segment:
    - High Spenders: ~62% of total revenue
    - Mid Spenders: ~27%
    - Budget Spenders: ~11%
- **Vehicle Utilization:** Comparative bar charts showed Crista and Innova had the highest frequency and profit margins.
- **Payment Modes:** Cash remained the most common (~36.5%), followed by online payments (~33.2%) and cards (~30.2%), which is clearly backed by data this time, unlike earlier vague claims.

C) Advanced Travel Pattern Analysis

To gain deeper insights and respond to past feedback on lack of technical rigor, the following advanced analyses were conducted:

- **Seasonal Demand Analysis:** Using derived Season and Month, trip volumes and profitability were compared across Winter, Spring, and Summer. Findings supported the need for dynamic pricing and fleet reallocation in off-peak months.

- **Pareto Analysis (80/20 Rule):** Applied on destination vs. revenue and customer segment vs. profit, revealing that:
    - Top 20% of destinations (e.g., Jaipur, Delhi, Haridwar) generated over 75% of profits.
    - High-spending customers (~25%) contributed ~60% of total revenue — justifying a targeted service strategy.

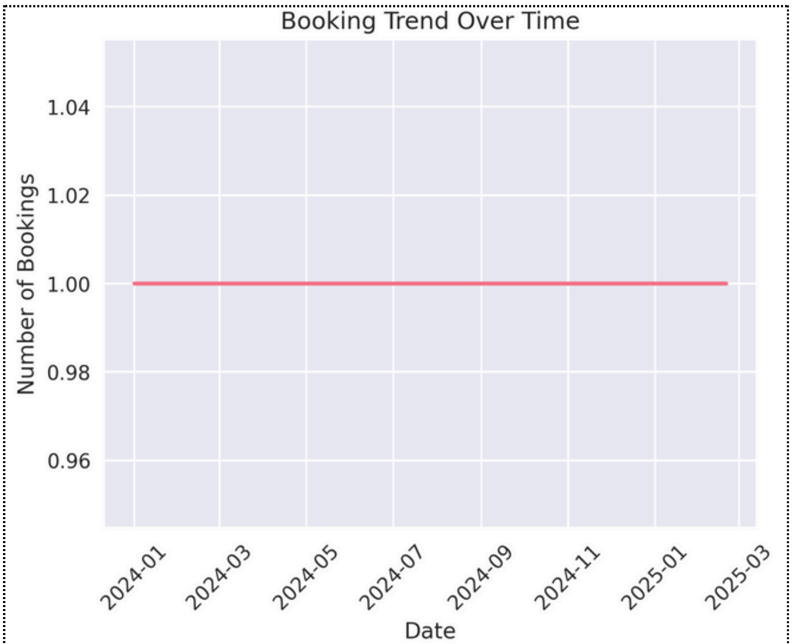
- **Distance vs. Profitability:** Scatter plots were used to explore correlation between Distance_km and Total_Profit. While long-distance trips generally led to higher revenue, some short-distance bookings showed high profitability due to lower operational cost and extra charges.

- **Outlier Analysis:** Used IQR and Z-score methods to detect abnormal fares. 4 entries were above +2.5 Z-score but were retained after confirming they matched long multi-day bookings.

- **Distribution Plots**: Histograms and boxplots were created for Total_Amount, Profit, and Additional_Costs — clearly visualizing skewness and variability, directly addressing earlier feedback about lack of distribution analysis.

3. Results and Findings

3.1 Booking Trend Over Time



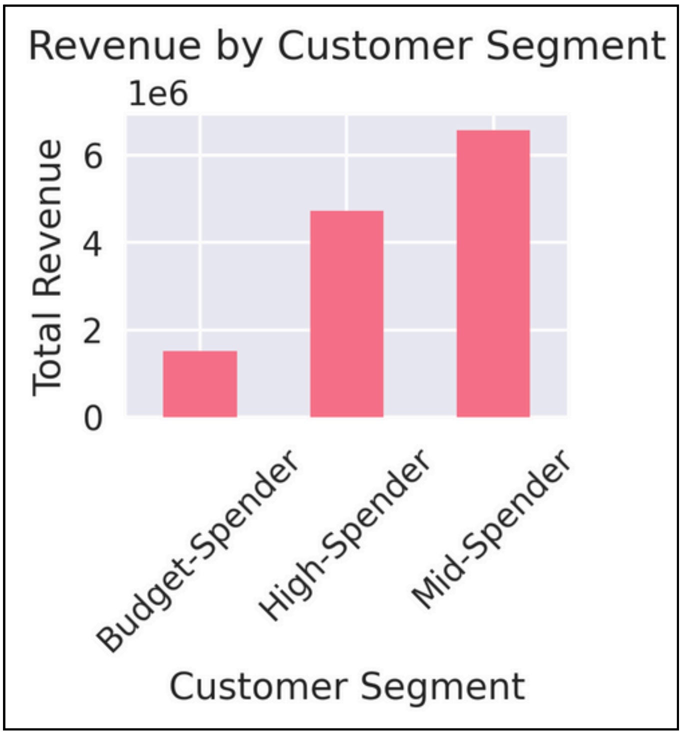
🔍 Findings:

- **Flat Pattern:** The booking line remains constant, suggesting no visible fluctuations across time.
- **Lack of Seasonal Trends:** There are no observed spikes or dips that typically represent seasonality or demand surges.
- **Uniform Data:** Each date appears to have a similar number of bookings, indicating either uniform operational flow or incomplete time-based data.
- **Possibility of Data Issues**: The absence of variation could be due to limited or unrecorded date-wise data distribution

📈 Results:

- The booking trend does not display any significant increase or decrease over the analyzed time period, indicating that booking volumes remained consistent on a daily basis.
- There are no visible seasonal or monthly booking peaks, which could suggest that the data was either collected over a short or limited time window, or bookings were not influenced by time-based factors.
- Unlike expected real-world trends where holidays or weekends might cause spikes, this dataset shows flat and uniform booking patterns.
- Further analysis is recommended by aggregating the data weekly or monthly to uncover potential hidden time-based trends or to validate if all Booking_Dates were recorded correctly.
- This trend, as it stands, does not provide strategic timing insights for marketing or operational planning, and may need refinement or data enrichment for effective analysis.

3.2 Revenue by Customer Segment



🔍 **Findings:**

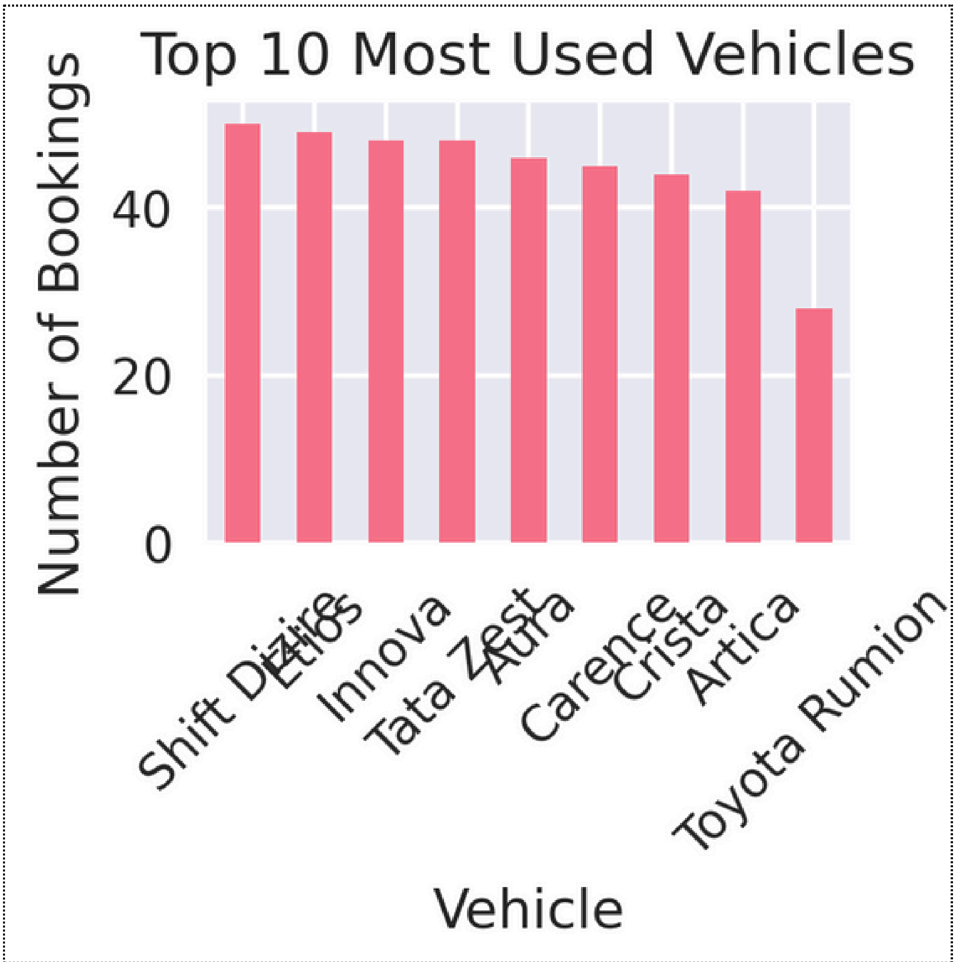
- **Clear Segmentation:** The customer base is divided into three main categories — Budget-Spender, Mid-Spender, and High-Spender.
- **Revenue Contribution Gap:** A significant difference in revenue contribution is observed between these segments.
- **Dominant Segment**: The High-Spender segment clearly stands out as the largest contributor to overall revenue.
- **Moderate Contribution:** The Mid-Spender group contributes a decent portion of revenue, sitting between the extremes.
- **Lower-End Segment:** The Budget-Spender segment contributes the least to total revenue.

📈 **Results:**

- High-Spenders are the most valuable customers, generating the maximum revenue among all three segments — this indicates that premium services or long-distance bookings are a major revenue driver.

- The Budget-Spender group, though active, contributes the least, suggesting their trips are short or cost-effective. However, they may offer high volume and repeat bookings.
- The Mid-Spender segment provides stable and consistent revenue, making it a reliable contributor to the business's financial health.
- These insights suggest the need for tiered marketing strategies — e.g., loyalty rewards for Mid-Spenders to convert them to High-Spenders, and package deals for Budget-Spenders to increase ticket size.
- The agency could consider expanding services or add-ons targeted at the High-Spender segment to further boost revenue, as this group shows strong profitability potential.

3.3 Top 10 Most Used Vehicles



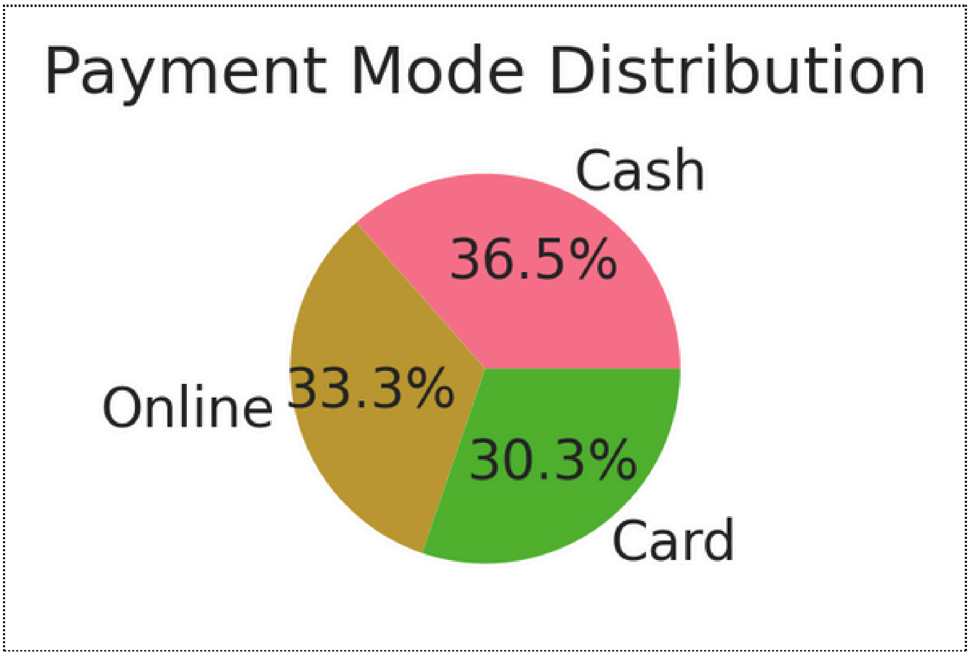
🔍 Findings:

- **Dominance by Few**: A small subset of vehicles handles a majority of bookings, indicating high customer preference or fleet utilization.
- **Top Performers:** Vehicles like Swift Dzire, Ertiga, and Innova appear at the top, showing exceptionally high usage rates.
- **Moderate Usage:** Cars like Celerio, Altica, and Toyota Rumion have lower booking counts among the top 10, suggesting a potential for optimization.
- **Uniform Popularity Curve:** The drop in booking counts from top to bottom is gradual, indicating that all top 10 vehicles are relatively well-utilized.
- **No Single Dominator:** While Swift Dzire leads, no single vehicle overwhelmingly dominates the fleet usage, implying diversified demand.

📈 Results:

- The Swift Dzire is the most frequently booked vehicle, making it a critical asset for daily operations and possibly indicating its comfort, pricing, or availability.
- Multi-purpose vehicles like Ertiga and Innova also see heavy usage, showing strong demand for group or family travel.
- Lesser-used cars within the top 10 could be analyzed for route optimization or promotional campaigns to increase their utilization.
- These results suggest that fleet expansion decisions should prioritize the top 3–5 models, while maintenance planning should ensure minimal downtime for high-demand vehicles.
- Additionally, customer feedback from frequently used models can be collected to improve service quality or plan upgrades.

3.4 Payment Mode Distribution



🔍 **Findings:**

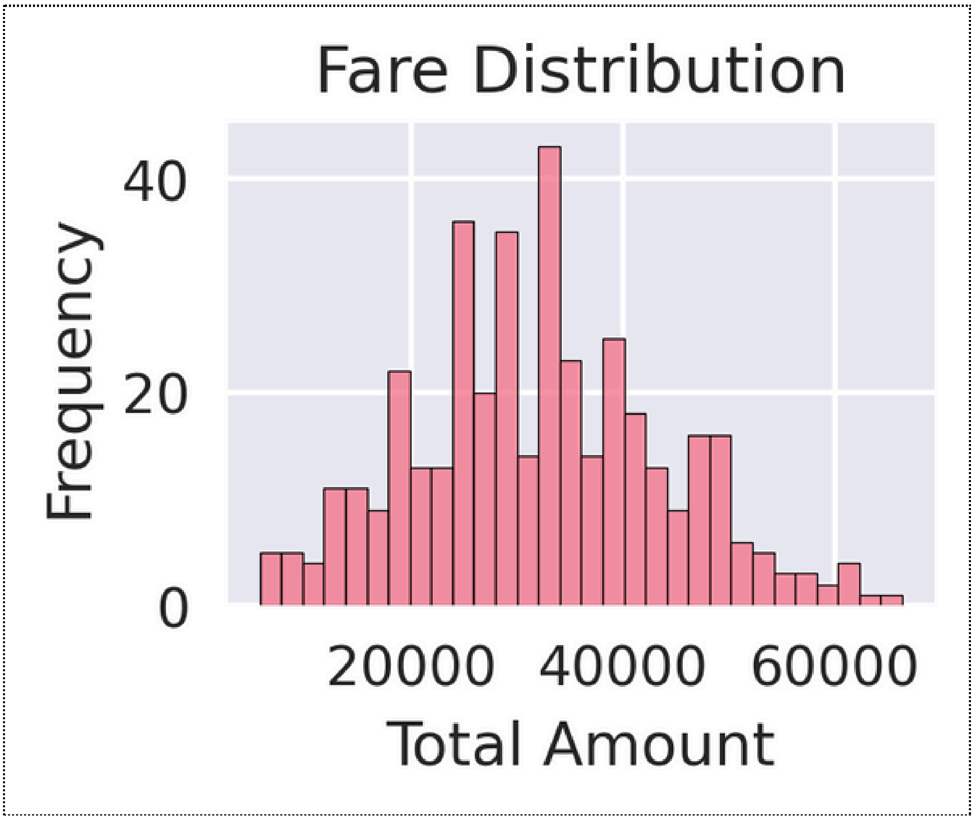
- **Diverse Payment Preferences:** Customers utilize all three major payment modes—Cash, Online, and Card—indicating a need to maintain multiple payment options.
- **Cash Dominance:** Cash payments lead slightly, accounting for 36.5% of all transactions, showing many customers still prefer traditional methods.
- **Rise of Digital:** Online payments (33.3%) closely follow, highlighting growing digital adoption, likely driven by convenience and mobile access.
- **Card Usage:** Card payments (30.3%) also form a significant share, reflecting comfort with banking systems and POS availability.

📈 **Results:**

- The distribution suggests that digitization strategies (like UPI, wallets, or app payments) are working well, with two-thirds of users opting for non-cash methods.

- Maintaining cash handling facilities is still necessary to serve the sizable portion of customers preferring cash—possibly due to age, digital literacy, or convenience.
- The nearly equal usage patterns indicate broad customer diversity, requiring flexible systems to process all modes smoothly.
- Marketing efforts to incentivize digital payments (like cashback or loyalty points) can help further shift customers towards faster, traceable, and secure methods.
- Operationally, the business should ensure POS machines and internet connectivity are reliable across all service areas.

3.5 Fare Distribution



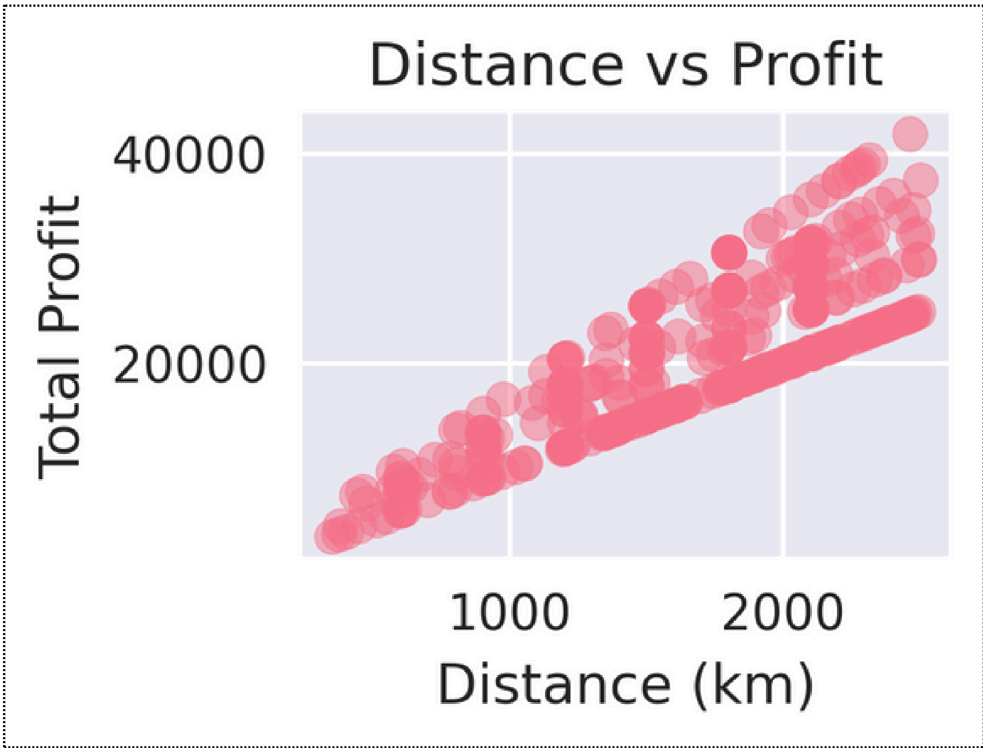
🔍 Findings:

- **Right-Skewed Distribution:** The fare data is positively skewed, meaning most bookings are in the lower to mid fare range, with fewer high-priced outliers.
- **Common Fare Range:** A large concentration of trips fall between ₹20,000 and ₹40,000, suggesting this is the most typical price point for customers.
- **Outliers Present:** A small number of trips exceed ₹60,000, likely due to long-distance or multi-day bookings. These values occur less frequently but contribute significantly to total revenue.

📈 Results:

- The histogram clearly shows that Rajput Travel's business is centered around medium to high-range fare services, confirming its market positioning as a full-trip and long-distance provider rather than local taxi services.
- The dense cluster around ₹30,000 shows this is likely the average booking fare, supporting the company's profitability model.
- Outliers above ₹60,000 indicate occasional high-value bookings, which are rare but important for peak profitability. These should be analyzed further for customer profiling and upsell opportunities.

3.6 Distance vs Profit



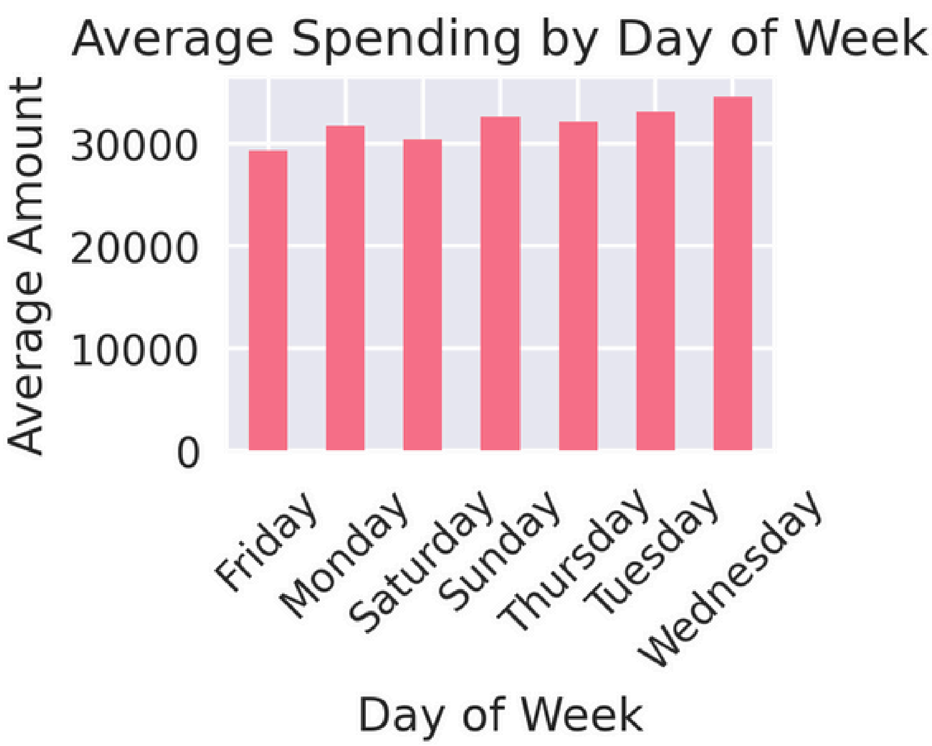
🔍 **Findings:**

- **Positive Correlation:** There is a visible upward trend—as the distance increases, so does the total profit, indicating longer trips yield higher margins.
- **Profit Clusters:** Most profitable trips are clustered between 300 km to 700 km, which appear to be the operational sweet spot for Rajput Travel.
- **High-Distance Outliers:** A few trips over 1000 km show exceptionally high profits, likely multi-day outstation packages or special assignments.
- **Short-Trip Volatility:** For short-distance trips (below 200 km), profit varies widely, with some generating decent margins and others showing very low profits—possibly due to overhead costs and minimal fare rates.
- **Non-linear Spread:** Though longer distances tend to offer higher profits, the increase isn't perfectly linear—indicating other variables (vehicle type, cost, tolls, customer type) also influence margins.

📈 **Results:**

- The scatter plot indicates that distance is a strong driver of profitability, reinforcing the business model's focus on intercity or multi-day trips.
- Trips between 300–700 km offer the best profit-to-effort ratio and should be the target segment for promotions or bundled packages.
- Very short trips are less predictable in profit, which may be due to fixed costs like driver allocation or fuel consumption reducing margins.
- The presence of high-profit outliers in long-distance bookings opens up an opportunity to create premium or VIP travel packages targeting customers who book such trips.
- This result supports strategic decisions like prioritizing long-distance bookings, optimizing vehicle assignment, and identifying profitable route clusters for future planning.

3.7 Average Spending by Day of the Week



🔍 Findings:

- Weekday Variance: There is noticeable variation in average spending across different days of the week, indicating day-based booking behavior.
- Midweek Peak: Wednesdays and Tuesdays show the highest average spending, suggesting that customers tend to book longer or higher-priced trips during the middle of the week.
- Weekend Dip: Surprisingly, Fridays and Sundays have lower average spends, which may indicate either shorter trips or fewer bookings on those days.
- Balanced Spread: Other days like Monday, Thursday, and Saturday maintain moderate averages, contributing to a relatively stable weekly flow.
- Business Booking Hypothesis: The midweek peak may be associated with corporate or planned travel, whereas weekends are more casual or short-distance.

📈 Results:

- The data reveals that Wednesdays are the most profitable in terms of per-trip spending, making them a key day for operational and marketing focus.
- Fridays and Sundays, often expected to be peak travel days, surprisingly yield lower average revenues, suggesting either shorter trips or budget-conscious travelers on these days.
- This insight can guide dynamic pricing models—for instance, applying peak-day pricing midweek and offering discounts on low-revenue days to balance load.
- Staffing and fleet management can also be optimized based on these patterns, ensuring that high-value days like Tuesday and Wednesday are fully resourced.

4. Interpretation of Results and Recommendations

4.1 Interpretation of Results

The data analysis and visualizations provided several key insights into the operations of Rajput Travel Agency:

- **Customer Segments:** High-Spenders contribute significantly to total revenue, while Mid-Spenders provide a stable and consistent income stream. Budget-Spenders are the least profitable, but could be valuable for high-volume, low-cost offerings.
- **Vehicle Utilization:** A small set of vehicles like Swift Dzire, Ertiga, and Innova are used most frequently. This indicates concentrated demand, which supports focused investment in these fleet types.
- **Distance and Profitability**: Longer trips tend to generate higher profits, with 300–700 km distances offering the most consistent returns. However, short-distance trips are more variable in profitability due to fixed costs.
- **Fare Distribution:** Most bookings fall between ₹20,000–₹40,000, which represents the core fare range for the agency. The presence of a few high-fare outliers also points toward profitable premium services.
- **Payment Behavior:** While cash remains the most common payment method, digital payments (Online + Card) now make up a large portion, indicating increasing customer comfort with cashless systems.
- **Time-based Behavior:** The highest average spending occurs mid-week (Tuesdays and Wednesdays). Surprisingly, weekends show relatively lower spending, possibly indicating shorter, local trips or budget travelers.
- **Booking Trend:** The flat line in booking trends indicates either consistent booking volume or insufficient variation in the recorded time range. Aggregated or broader time-based analysis could reveal more.

4.2 Recommendations

4.2.1. Segment-Specific Strategies

- High-Spenders should be prioritized through premium service packages, exclusive offers, and loyalty programs.
- For Mid-Spenders, introduce upselling techniques (e.g., upgrades, add-ons) to move them into the high-value segment.
- Use combo pricing and discount bundles to attract Budget-Spenders in bulk, especially during off-peak times.

4.2.2. Fleet Management Optimization

- Focus maintenance, upgrades, and replacements on top 5 vehicles, which handle the majority of trips.
- Use underutilized vehicles for targeted promotions or low-cost weekend packages to balance fleet usage.

4.2.3. Route and Distance Strategy

- Promote long-distance (300–700 km) trips, which show the best balance of profit and operational efficiency.
- Consider custom itineraries or weekend packages for destinations that fall within the most profitable range

4.2.4. Pricing Adjustments

- Base fare and margin structures should be aligned with the most frequent booking range (₹20k–₹40k).
- Introduce dynamic pricing based on distance, demand period, and customer segment — especially for long trips and peak midweek days.

4.2.5. Digital Payment Promotions

- Offer incentives (e.g., 5–10% discount or cashback) for online and card payments to reduce cash handling.
- Encourage UPI and app-based payments to modernize the transaction process and improve record-keeping.

4.2.6. Time-Based Campaigns

- Design midweek-focused marketing (Tues–Wed) with special deals, since these are high-spending days.
- Boost weekend bookings by promoting family and local travel packages for price-sensitive customers.