ONLINE VEHICLE BOOKING MARKET

BY

Nishant Kumar

INTRODUCTION:

The online vehicle booking market has witnessed significant growth over the past decade, driven by rapid urbanization, increasing smartphone penetration, and evolving customer preferences towards convenience and flexibility. With the rise of digital platforms and mobile applications, consumers today prefer the ease of booking vehicles on-demand for a variety of purposes, including daily commutes, business travel, tourism, and logistics.

India, being one of the fastest-growing markets in terms of internet users and digital payments, has seen a surge in app-based vehicle rental and ride-hailing services. Established players like Ola and Uber have already captured substantial market share in metropolitan areas, while newer startups and niche services are targeting emerging segments such as EV rentals, subscription models, and micro-mobility solutions.

The market is not limited to cars; there is a growing demand for two-wheelers, electric vehicles, and specialized services catering to different customer needs across Tier-1 and Tier-2 cities. Factors like rising fuel prices, increasing awareness of sustainability, and changing work patterns (remote work, flexible schedules) have further fueled this shift towards flexible mobility solutions.

This report aims to analyze the current landscape of the online vehicle booking market, identify potential customer segments through data-driven analysis, and provide strategic recommendations for businesses seeking early market advantage in this competitive and evolving sector.

A. <u>FERMI ESTIMATION</u>

Objective:

Estimate the potential market scope for a vehicle booking startup by logically breaking down the Indian vehicle booking market into sub-categories.

1. Market Segmentation by Vehicle Type

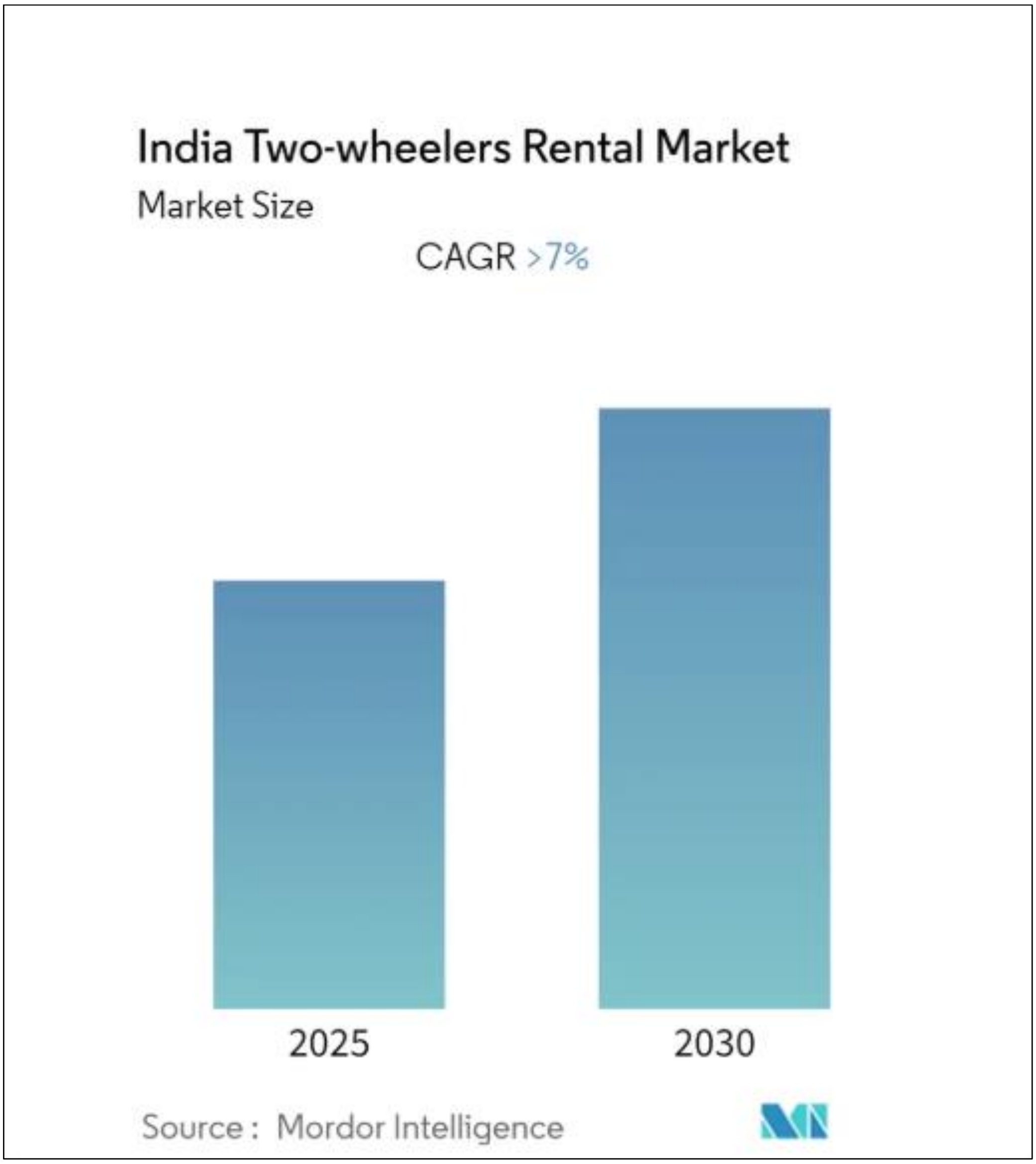
- **Cars (Rental, Chauffeur, Self-Drive)**

The India car rental market size was valued at **USD 2,742 Million** in 2024. Estimated to grow around **USD 9,793M** by 2033, exhibiting a CAGR (Compound Annual Growth rate) of 14.5%.

With the advantage of door-to-door services and customized packages, car rentals have emerged as a preferred choice for large segments of travellers. To cater to all customers and offer easy accessibility, these websites provide a variety of types of vehicles from economy options to luxury and SUVs-at very competitive prices.

Advanced technologies include GPS tracking, real-time vehicle availability, and contactless payment systems, which enhance the convenience and efficiency of car rentals. Increased business travel also contributes to the growing demand for short-term car rentals, as corporate professionals are often traveling and prefer renting cars for their comfort and flexibility. More urbanized, high requirements of personalized on-demand kind of services in the car rental market are expected to grow manifold, coupled with a significant transformation in consumer patterns and technological aspects, thereby, boosting the India car rental demand.

- **Scooters / Bikes (Rental, Short Commutes)**



The India Two-wheelers Rental Market is expected to register a CAGR of greater than 7% during the 2024-2033. The booming tourism, rapid urbanization, increasing interest of people in adventure tourism, rapid digitalization, international and national migration, micro-mobility, traffic congestion, and the cost associated with owning and maintenance of two-wheelers have been some of the factors that are expected to fuel the demand for two-wheeler rental market during the forecast period. Scooters are commonly used for short-term rentals, especially for micro-mobility within urban city limits.

In urban areas, the new app-based startups are strengthening their presence by providing app-based dockless scooter rental/sharing services as the urban located player caters to the working professionals and students for their commuting needs.

2. Market Segmentation by Customer Type

- **Business Travellers**

  Corporate employees consume a section of customer share of rental vehicle companies. Since, Business traveller are to be travelling on a regular basis, they are one of the key targets for a company in rental vehicle market.

  They often require vehicles to travel for meetings, conferences, temporary work assignments. Unlike leisure travellers, they prioritize efficiency, reliability, and premium service.

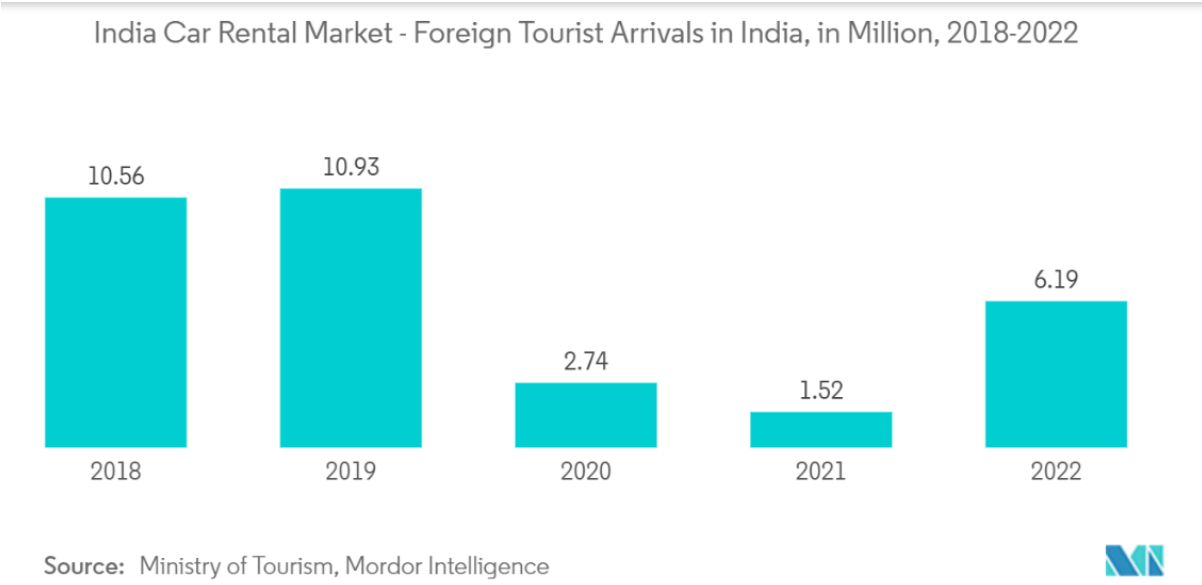
  Preferred vehicles: Business professionals tend to favor mid-size sedans, luxury cars, and fuel-efficient models that project professionalism and offer comfort during long drives.

  Key Locations: Airport, Railway Stations, Business Hubs are primary locations for business travellers, where they might need the rental vehicle.

  Service Expectations: Fast check-in/check-out, premium services, corporate rental agreement, GPS navigation, loyalty perks.

  Whether a new start company or a well-established company, it's essential for the companies to establish partnerships with businesses and offer corporate rental programs can build a steady stream of repeat customers in this segment. Large corporations often seek long-term rental agreements for their traveling employees, making it essential for rental companies to offer customized corporate packages.

- **Tourists and Vacationers**



Leisure travellers represent another core market, particularly in destinations with high tourism activity. Rental vehicles provide flexibility and convenience for exploring new locations, making them an attractive option for tourists who prefer self-guided travel.

Peak Demand Periods: Holidays, summer months, and travel seasons see an influx of rental requests, especially in regions with popular tourist attractions, beaches, or national parks.

Common Rental Preferences: Families often opt for SUVs or minivans with extra luggage space, while couples and solo travellers may choose economy cars for affordability or convertibles for an enhanced travel experience. Adventure seekers may look for off-road-capable vehicles.

Booking Behaviour: Many tourists book rentals in advance through online platforms, travel agencies, or vacation package deals. Others may book on arrival, especially in high-traffic locations like airports or train stations.

Rental companies in tourist hotspots could offer flexible booking options, multilingual customer support, and tailored packages, such as discounts for extended rentals or bundled deals with hotels and attractions.

- ## Rideshare and Delivery Drivers

With the rise of gig economy jobs, many drivers working for rideshare services (such as Uber and Lyft) and delivery platforms (such as DoorDash, Uber Eats, and Instacart) rely on rental cars. These customers often seek long-term rental solutions with affordable pricing and flexible terms.

Rental Preferences: Fuel-efficient, hybrid, or electric vehicles to reduce operational costs. Reliability and good mileage limits are key factors for these customers.

Challenges: High mileage usage, increased wear and tear on rental cars, and insurance considerations. Some rental companies may impose mileage restrictions or require special rental agreements for these drivers.

Opportunities for Rental Companies:

- Special rental programs designed for gig workers, offering flexible pricing and extended rental terms.
- Exclusive deals for rideshare drivers, including maintenance and insurance add-ons.
- Partnerships with rideshare companies to provide pre-approved rental programs.

Providing rental solutions tailored to this segment can create a consistent revenue stream, particularly in urban markets where rideshare and delivery work are prevalent.

- ## Event and Special Occasion Renters

Some customers rent cars for specific events such as weddings, proms, business functions, or weekend getaways. These rentals often include luxury, specialty, or high-performance vehicles.

Popular Choices:

- High-end sedans (Mercedes-Benz, BMW, Audi) for business events or VIP transport.
- Sports cars (Ferrari, Lamborghini) for thrill-seekers or once-in-a-lifetime experiences.
- Vintage or classic cars for themed events, film productions, or unique wedding transportation.
- Limousines for weddings, proms, or corporate clients.

Customer Priorities: Impeccable vehicle condition, premium customer service, and flexible rental terms. Customers in this category expect a flawless experience, often requiring delivery and pickup services.

Marketing Strategies:

- Partnerships with event planners, wedding venues, and upscale hotels.
- Targeted advertising through social media, wedding magazines, and lifestyle blogs.

- Offering exclusive rental packages that include chauffeurs, decoration services, and extended rental hours.

Niche and Emerging Markets

Beyond traditional customers, several niche and emerging markets offer new opportunities for car rental companies. As consumer behaviors shift and new mobility trends develop, rental businesses can expand their reach by targeting specialized customer groups with tailored services.

- ### Eco-Conscious Consumers

  With the growth in sustainability awareness, citizens are now preferring eco-friendly vehicle. Nearly 70% of consumers consider sustainability when making purchasing decisions. Cities worldwide are also implementing stricter emission regulations, pushing rental companies to use EV/Hybrid Vehicles.

  Preferred Vehicles: Hybrid models, fully electric cars (such as Tesla, Nissan Leaf, and Chevrolet Bolt), and fuel-efficient vehicles attract environmentally conscious renters.

  Infrastructure Considerations: To successfully cater to this market, rental locations should offer charging stations or partner with charging networks to ensure seamless access for EV users.

- ### Long-Term Renters and Expats

Not all renters need a car for just a few days. Expats, business consultants, seasonal workers, and digital nomads often require vehicles for extended periods but prefer renting over committing to a lease.

Key Demographics:

- International professionals relocating temporarily.
- Students studying abroad or on internship programs.
- Families transitioning between homes who need temporary transportation.

Advantages Over Leasing: Unlike traditional leases, long-term rentals provide greater flexibility, eliminating the need for down payments and long-term commitments.

Ideal Rental Offerings: Monthly rental programs with competitive pricing, inclusive maintenance, and flexible return options attract this segment.

To capture long-term renters, companies should offer discounts for extended bookings, bundled insurance packages, and easy renewal options that provide hassle-free mobility.

- **Subscription based car rentals**

A growing trend in the automotive industry is car subscription services, where customers pay a monthly fee for vehicle access instead of purchasing or leasing a car. These services blend the benefits of car ownership with the flexibility of rentals.

How It Works: Customers subscribe to a rental plan, which often includes insurance, maintenance, and roadside assistance in a single monthly payment.

Target Audience:

- Young professionals who prefer flexible mobility over ownership.
- Urban residents who occasionally need a car but don't want the responsibility of ownership.
- Frequent travellers who need reliable transportation without long-term commitments.

Competitive Advantage: Subscription-based rentals reduce customer churn by creating predictable, recurring revenue. Companies can differentiate themselves by offering tiered subscription plans based on vehicle types and mileage needs.

By embracing this model, car rental companies can tap into the shift away from traditional car ownership and position themselves at the forefront of mobility innovation.

2. Market Segmentation by Regional Analysis

<u>South India:</u>

Key Cities: Bengaluru, Chennai, Hyderabad

Demand Drivers:

- Tourism hotspots (Kerala, Tamil Nadu)
- IT Hubs with high disposable incomes
- Strong demand for chauffeur-driven vehicles for executives

Urbanization Impact: By 2036, 40% of India's population projected in urban areas (World Bank)

<u>North India:</u>

Key Cities: Delhi, Jaipur, Chandigarh

Demand Drivers:

- Tourism (heritage sites)
- Increased business travel
- Infrastructure growth boosting rentals

West & Central India

Key Cities: Mumbai, Pune, Ahmedabad, Indore, Bhopal

Demand Drivers:

- Mumbai-Pune corridor: premium services, corporate demand
- Tourism in Goa boosting short-term rentals
- Growing middle-class travel habits in central India

East India

Key Cities: Kolkata, Bhubaneswar, Ranchi, Patna

Demand Drivers:

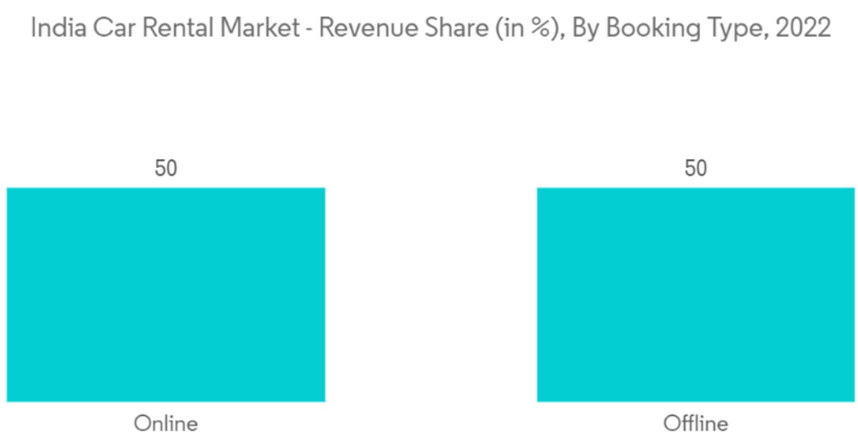
- Urban development on the rise
- Focus on tourism
- Market still emerging, good future potential

Competitive Landscape (Summary Points):

- Integration of **AI, ML, data analytics** for optimization
- Growing inclusion of **EV fleets** for sustainability
- Rise of **subscription-based models**
- Partnerships with **hotels, airlines, travel platforms**
- Aggressive **offers, loyalty programs** to capture demand
- Penetration into **Tier-2 / Tier-3 markets**
- Travel industry expected to hit **$125 Billion by FY27 (IBEF)**

Source:

3. Market Segmentation based on Booking Type



B. DATA COLLECTION

1. Behavioural Data Collection

```
behavioural_data_set.head()
```

	Date	Time	Booking_ID	Booking_Status	Customer_ID	Vehicle_Type	Pickup_Location	Drop_Location	V_TAT	C_TAT	...	Cancel
0	2024-07-26 14:00:00	14:00:00	CNR7153255142	Canceled by Driver	CID713523	Prime Sedan	Tumkur Road	RT Nagar	NaN	NaN	...	Pe
1	2024-07-25 22:20:00	22:20:00	CNR2940424040	Success	CID225428	Bike	Magadi Road	Varthur	203.0	30.0	...	
2	2024-07-30 19:59:00	19:59:00	CNR2982357879	Success	CID270156	Prime SUV	Sahakar Nagar	Varthur	238.0	130.0	...	
3	2024-07-22 3:15:00	3:15:00	CNR2395710036	Canceled by Customer	CID581320	eBike	HSR Layout	Vijayanagar	NaN	NaN	...	
4	2024-07-02 9:02:00	9:02:00	CNR1797421769	Success	CID939555	Mini	Rajajinagar	Chamarajpet	252.0	80.0	...	

What was Collected?

- The behavioural dataset was sourced from a publicly available GitHub repository related to **Ola Ride-Booking Data.** The dataset contains **103,024 ride records**, providing attributes that reflect customer behaviour during vehicle bookings. Relevant columns include:
- Booking Status (Success / Unsuccessful)
- Vehicle Type (Bike, Car, etc.)
- Booking Value (Fare Amount)
- Ride Distance
- Payment Method (Later cleaned to Offline / Online)

- Customer and Driver Ratings

How was it Collected?

- This dataset was sourced from the following GitHub repository: OLA-Data-Analysis-and-Visualization/Bookings-100000-Rows.xlsx
- It was identified through secondary research on platforms like GitHub and Kaggle during the data collection phase focused on behavioural aspects of the Indian vehicle booking market.

Why was this Dataset Chosen?

- The dataset represents **actual booking behaviours** (success rates, ride types, payment preferences) aligned with the study's **behavioural segmentation goal**.
- It provides relevant indicators to analyse **user preferences, booking trends, and customer habits**.
- The dataset includes variables that support segmentation tasks in customer behaviour, making it fit for clustering and behaviour-based analysis.

Summary for This Section:

Aspect	**Content**
What	Ride-booking records reflecting behaviour
How	Sourced from GitHub repository
Why	Matches behavioural segmentation goals; provides essential variables

2. Demographic Data Collection

For the demographic segmentation, due to the lack of a single comprehensive public dataset with detailed age, income, and occupation profiles tied to vehicle booking behaviour, a **synthetic dataset** was generated to simulate customer patterns typically observed in the Indian urban mobility market.

The dataset was generated using **random sampling logic based on real-world distributions**, incorporating features such as:

- **Age** (18–60 years): Common user age group for digital bookings

- **Gender**: Male, Female, and Other
- **Occupation**: Student, Working Professional, Self-employed, Unemployed
- **Monthly Income**: ₹8,000 to ₹1,00,000 — based on Indian income trends
- **Preferred Vehicle Type**: Bike, Car, Auto, Scooter
- **Payment Mode**: Online / Offline
- **Booking Frequency**: 1 to 15 rides per month (estimated)

The data mimics what one would expect from **user surveys, booking logs, and mobility reports** gathered by companies in the online transportation domain. This dataset provides a solid foundation for demographic-based segmentation analysis.

Source Note:
This synthetic dataset was generated manually for research purposes to support segmentation modelling due to the unavailability of publicly consolidated Indian mobility demographic data.

C. Data Pre-Processing

1. Behavioural Segment Data Pre-Processing

Dataset Used:

Bookings-100000-Rows.xlsx (cleaned and saved as CSV)

Preprocessing Steps:

- **Column Selection:**
  Selected only relevant behavioral features: Booking_Status, Vehicle_Type, Booking_Value, Ride_Distance, Payment_Method, Driver_Ratings, Customer_Rating.

```python
columns_to_drop = [
    'Date', 'Time', 'Booking_ID', 'Customer_ID', 'Pickup_Location',
    'Drop_Location', 'V_TAT', 'C_TAT','Canceled_Rides_by_Customer', 'Canceled_Rides_by_Driver',
    'Incomplete_Rides', 'Incomplete_Rides_Reason', 'Vehicle Images', 'Unnamed: 20'
]
cleaned_behavioural_data = behavioural_data_set.drop(columns = columns_to_drop)
```

```python
cleaned_behavioural_data.head(10)
```

	Booking_Status	Vehicle_Type	Booking_Value	Payment_Method	Ride_Distance	Driver_Ratings	Customer_Rating
0	Canceled by Driver	Prime Sedan	444	NaN	0	NaN	NaN
1	Success	Bike	158	Cash	13	4.1	4.0
2	Success	Prime SUV	386	UPI	40	4.2	4.8
3	Canceled by Customer	eBike	384	NaN	0	NaN	NaN
4	Success	Mini	822	Credit Card	45	4.0	3.0
5	Success	Mini	173	UPI	41	3.4	4.6
6	Success	Bike	140	Cash	49	3.2	4.5
7	Canceled by Driver	Prime Plus	344	NaN	0	NaN	NaN
8	Driver Not Found	Mini	839	NaN	0	NaN	NaN
9	Canceled by Driver	Auto	893	NaN	0	NaN	NaN

Fig: 10 columns in behavioural data after column selection

- **Handling Missing Values:**

```python
[13]: cleaned_behavioural_data.isnull().sum()
```

```
[13]: Booking_Status     0
      Vehicle_Type       0
      Booking_Value      0
      Payment_Method     0
      Ride_Distance      0
      Driver_Ratings     0
      Customer_Rating    0
      dtype: int64
```

```python
[14]: cleaned_behavioural_data.head(10)
```

[14]:	Booking_Status	Vehicle_Type	Booking_Value	Payment_Method	Ride_Distance	Driver_Ratings	Customer_Rating
0	Canceled by Driver	Prime Sedan	444	Unknown	0	4.0	4.0
1	Success	Bike	158	Cash	13	4.1	4.0
2	Success	Prime SUV	386	UPI	40	4.2	4.8
3	Canceled by Customer	eBike	384	Unknown	0	4.0	4.0
4	Success	Mini	822	Credit Card	45	4.0	3.0
5	Success	Mini	173	UPI	41	3.4	4.6
6	Success	Bike	140	Cash	49	3.2	4.5
7	Canceled by Driver	Prime Plus	344	Unknown	0	4.0	4.0

  - Removed rows with missing Payment_Method, Driver_Ratings, and Customer_Rating (approx. 39% of rows).
  - This step was crucial to maintain data integrity for segmentation.

- **Feature Engineering:**

```python
cleaned_behavioural_data.head(10)
```

	Booking_Status	Vehicle_Type	Booking_Value	Ride_Distance	Driver_Ratings	Customer_Rating	Offline_Payment
0	0	Prime Sedan	444	0	4.0	4.0	0
1	1	Bike	158	13	4.1	4.0	1
2	1	Prime SUV	386	40	4.2	4.8	0
3	0	eBike	384	0	4.0	4.0	0
4	1	Mini	822	45	4.0	3.0	0
5	1	Mini	173	41	3.4	4.6	0
6	1	Bike	140	49	3.2	4.5	1
7	0	Prime Plus	344	0	4.0	4.0	0
8	0	Mini	839	0	4.0	4.0	0
9	0	Auto	893	0	4.0	4.0	0

- Transformed **Payment_Method** into a binary column **Offline_Payment** (0 for Online, 1 for Offline).
- Mapped **Booking_Status: 1 = Successful, 0 = Unsuccessful** (for classification/clustering).

- **Data Type Validation:**

```python
[29]: cleaned_behavioural_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103024 entries, 0 to 103023
Data columns (total 7 columns):
 #   Column           Non-Null Count   Dtype  
---  ------           --------------   -----  
 0   Booking_Status   103024 non-null  int64  
 1   Vehicle_Type     103024 non-null  object 
 2   Booking_Value    103024 non-null  int64  
 3   Ride_Distance    103024 non-null  int64  
 4   Driver_Ratings   103024 non-null  float64
 5   Customer_Rating  103024 non-null  float64
 6   Offline_Payment  103024 non-null  int64  
dtypes: float64(2), int64(4), object(1)
memory usage: 5.5+ MB
```

Ensured all numerical and categorical features were of correct data type for further analysis.

2. Demographical Segment Data Pre-Processing

```
demographic_dataset.head(10)
```

	Age	Gender	Occupation	Monthly_Income
0	56	Female	Working Professional	54915
1	46	Female	Retired	56963
2	32	Male	Student	63718
3	60	Male	Self-Employed	30611
4	25	Male	Working Professional	30760
5	38	Female	Working Professional	51544
6	56	Female	Self-Employed	46567
7	36	Male	Student	40301
8	40	Other	Student	67084
9	28	Female	Student	42106

The demographic dataset comprises 100,000 customer records with the following attributes:

- Age
- Gender
- Occupation
- Monthly_Income

Initial Inspection and Data Types

- All four columns were verified for correct data types:
    - Age and Monthly_Income → Numeric (int64)
    - Gender and Occupation → Categorical (object)
- No type correction or conversion was required.

Missing Value Check

```python
demographic_dataset.isnull().sum()
```

```
Age              0
Gender           0
Occupation       0
Monthly_Income   0
dtype: int64
```

A thorough inspection revealed **no missing values** in any column. This ensured we could proceed without imputation or row dropping.

Data Cleaning Decisions

- The dataset was already clean and well-structured.
- **Categorical Columns (Gender, Occupation)** were **retained in their original form** rather than encoded. This improves human interpretability during EDA and segment profiling.
- **No columns were dropped**.

Summary Statistics

	Age	Monthly_Income
count	100000.000000	100000.000000
mean	41.022480	50040.552880
std	13.526548	14964.114137
min	18.000000	8000.000000
25%	29.000000	39940.750000
50%	41.000000	50052.500000
75%	53.000000	60114.250000
max	64.000000	115427.000000

- **Age**:
  - Mean: ~41 years
  - Range: 18 – 64 years
- **Monthly Income**:
  - Mean: ₹50,000

- Range: ₹8,000 – ₹1,15,000

These ranges are consistent with a diverse adult working and retired population, making the dataset ideal for demographic segmentation.

D. Exploratory Data Analysis (EDA)

1. Behavioural Segment EDA

- **Distribution of Booking Status**:

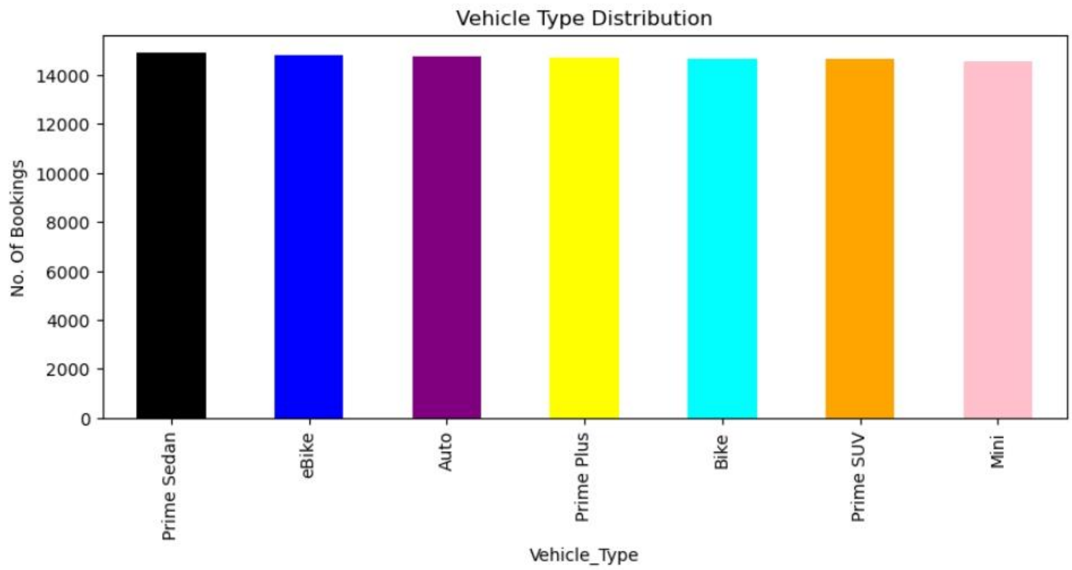
Observation:

- Out of 100,000+ bookings, approximately 65**% are successful**.
- The remaining 35% were either **cancelled** or **incomplete**.

Interpretation:

- Indicates a reasonably efficient system.
- However, **35% failure rate** presents an opportunity to study and improve service reliability, vehicle availability, or payment methods.

- **Popular Vehicle Types**:



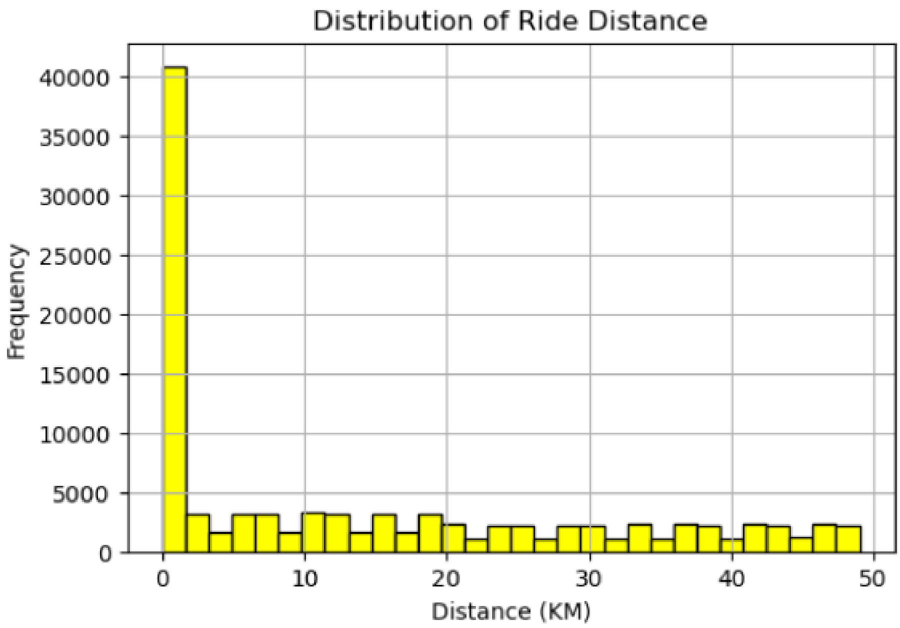
Observation:

- The booking distribution across all vehicle types is **surprisingly uniform**.
- Categories such as **Prime Sedan**, **eBike**, **Auto**, **Prime Plus**, **Bike**, **Prime SUV**, and **Mini** all have **similar booking volumes (~14,000–15,000)**.

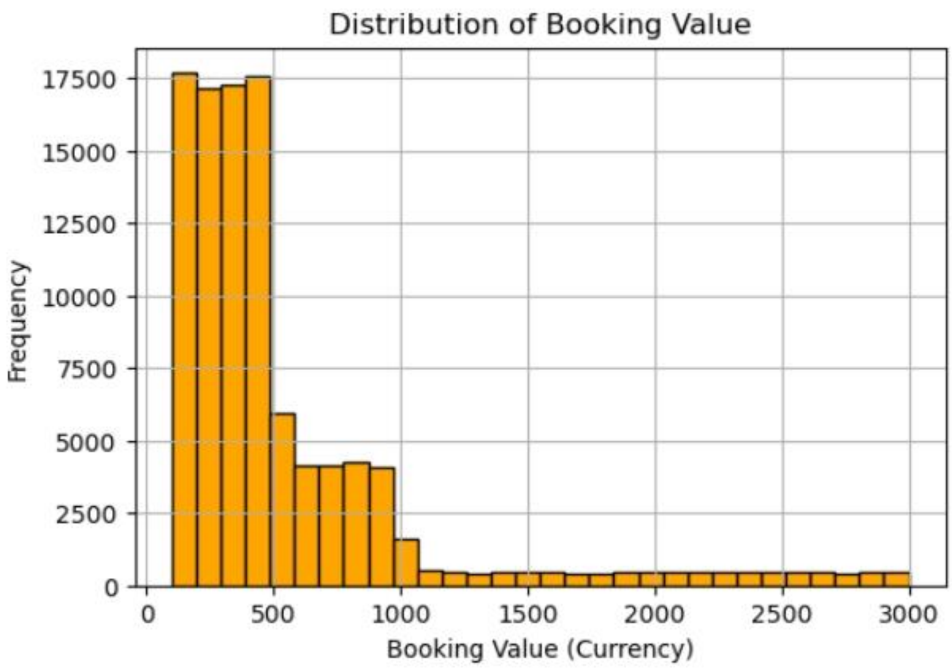
Interpretation:

- The data indicates a **diversified demand** — customers are utilizing **all types of vehicles** almost equally.
- This might reflect:
  - A **balanced supply** strategy from the platform.
  - Users' **diverse travel needs** — short commutes (bike/eBike), group travel (SUV), budget travel (Auto/Mini), and executive needs (Prime).
- This can be leveraged for:
  - Offering **customized plans or pricing tiers** per vehicle type.
  - Strategizing **location-based vehicle stocking** based on further geographic analysis.
  - Running **vehicle-type-specific promotions** for better targeting (e.g., discounts on SUVs for weekend bookings).

- **Ride Distance Distribution:**



- **Booking Value:**



Observation:

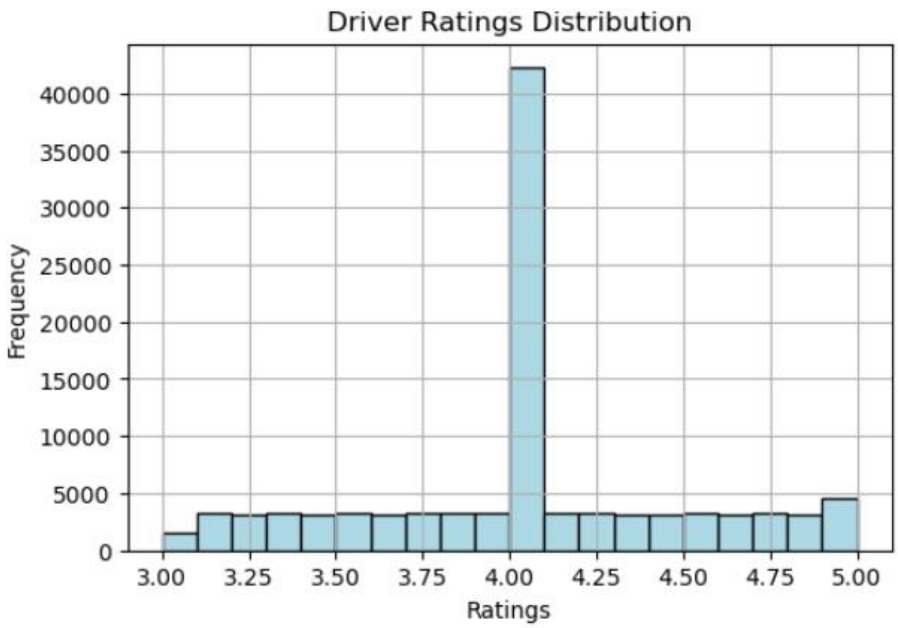
- Most bookings lie between ₹100 to ₹500.
- A small number of luxury or long-distance rides cost up to ₹2,000+.

Interpretation:

- Majority of users are **price-sensitive**, preferring economical rides.

- Helps define price tiers or dynamic pricing strategies to appeal to both budget and premium segments.

- **Ratings Analysis**



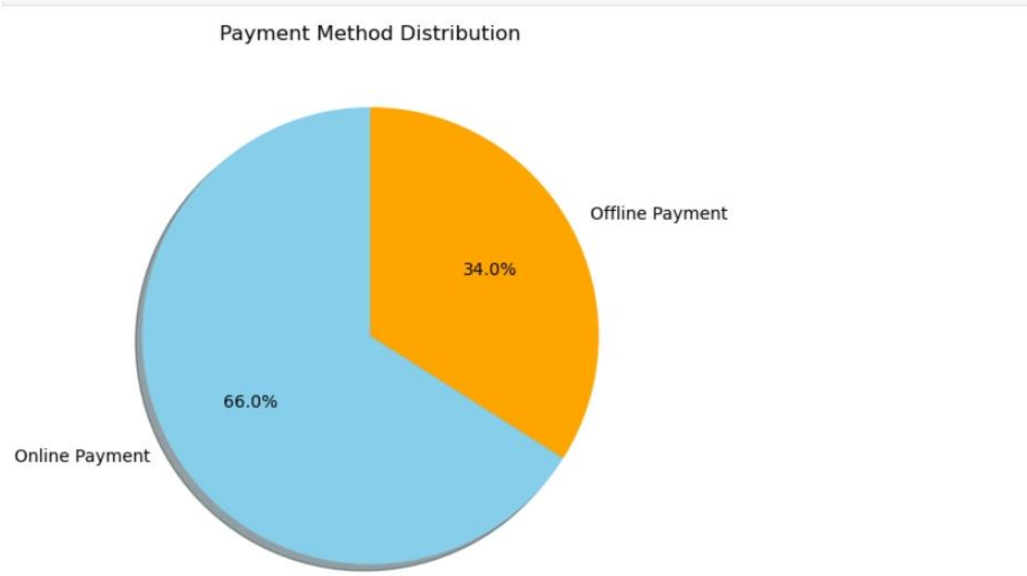
Observation:

- Driver ratings mostly range between 3.5 to 5.
- Customer ratings are similar, slightly skewed toward higher scores.

Interpretation:

- Positive customer experience is key.
- A small percentage of low-rated drivers or users could help in **creating quality-based segmentation** or **flagging churn risks**.

- **Offline vs. Online Payment Behaviour**



Observation:

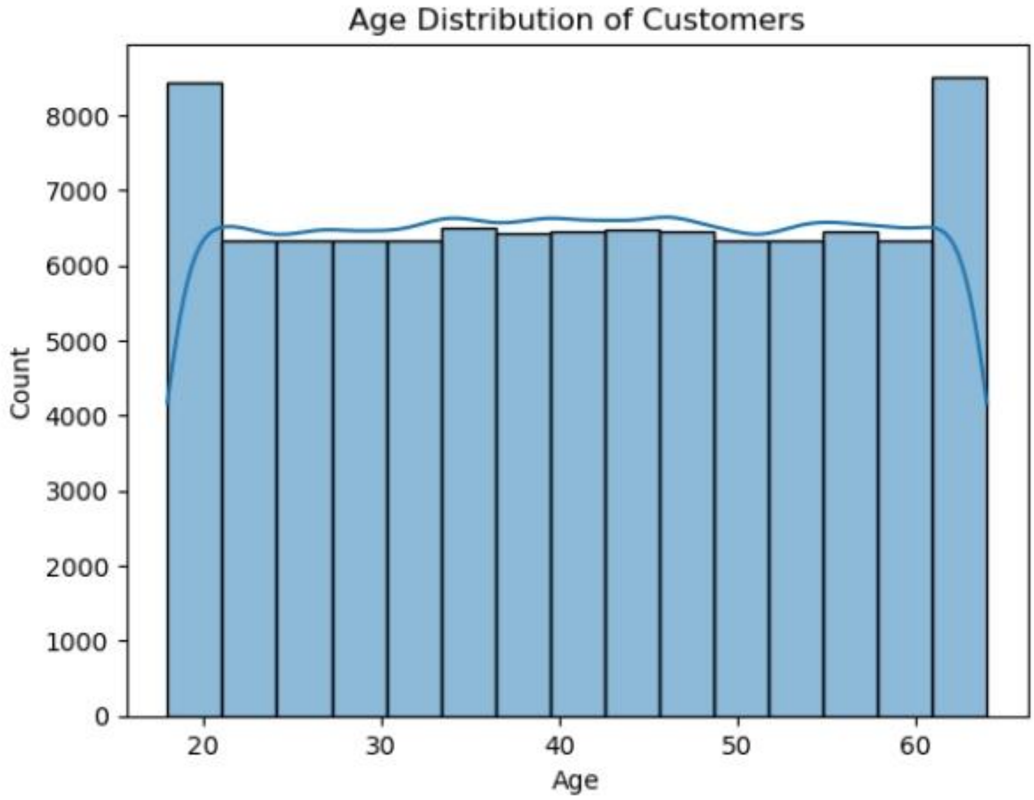
- Bookings paid **online** tend to succeed more than **offline**.
- Offline bookings may fail due to payment delays, cancellations, or availability mismatches.

Interpretation:

- Focus on **promoting online methods** (wallets, UPI) through discounts or convenience features.
- Indicates room for adoption improvement, especially in Tier 2/3 markets.

2. Demographic Segment EDA

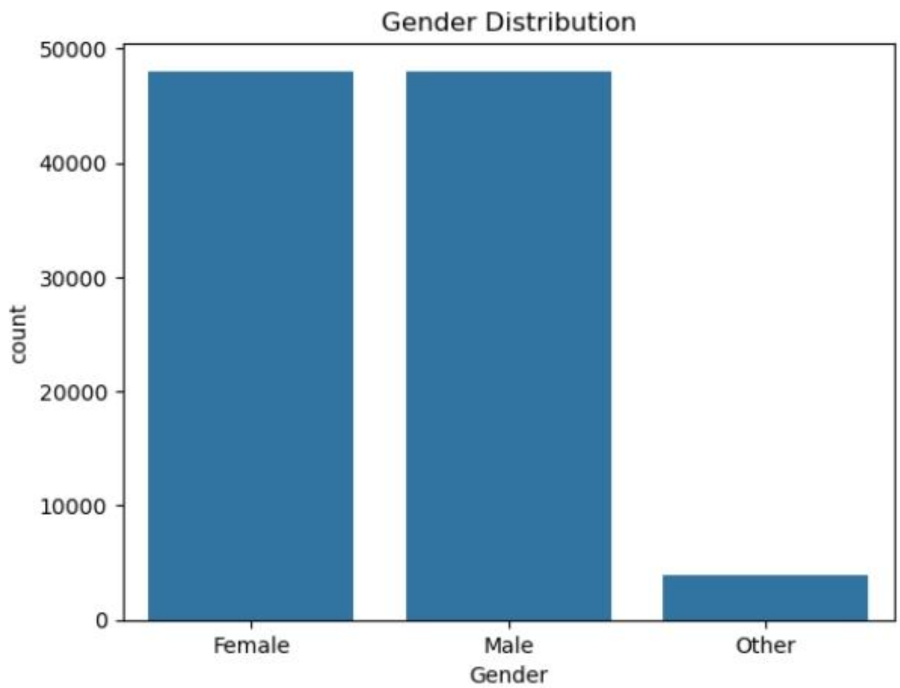
Insight: Age Distribution of Customers



The histogram shows that:

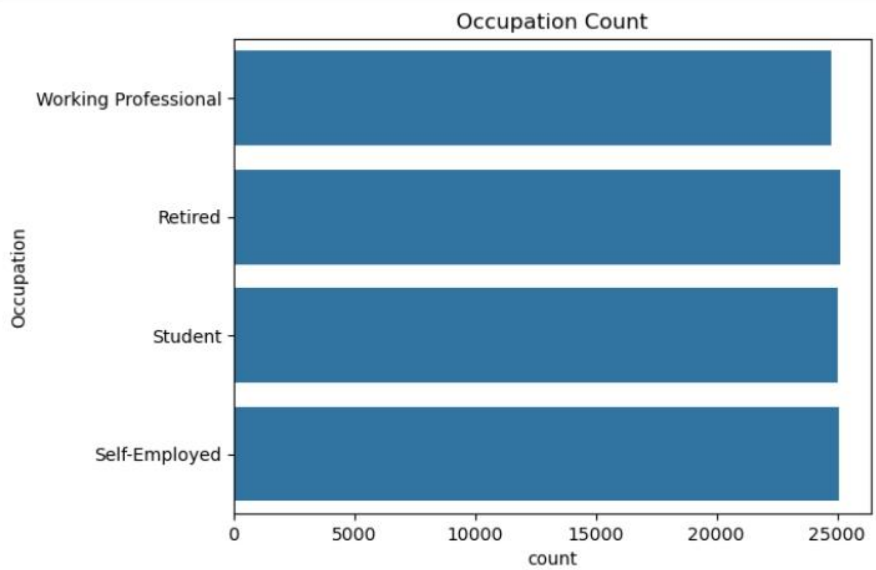
- There are **notable spikes** in customer count at the **youngest (around age 20)** and **oldest (above 60)** ends of the spectrum.
- The **age group between 25 and 60** maintains a **fairly uniform distribution**, indicating a consistent user base across working-age adults.
- This suggests two strong user segments:
  - 🎓 **Young users (~20)** – possibly college students or early-career individuals preferring cost-effective or flexible travel.
  - 🧓 **Older adults (60+)** – likely retirees or senior citizens using rentals for convenience.

Insight: Gender Distribution of Customers



- The chart shows a **balanced usage** of the rental service among **male and female** customers, each contributing nearly equally to the customer base.
- The **“Other” gender category** has noticeably fewer users — which is expected, as it reflects both the **smaller population size** and possibly **lower representation** in digital ride booking behaviour.

Insight: Occupation Distribution of Customers

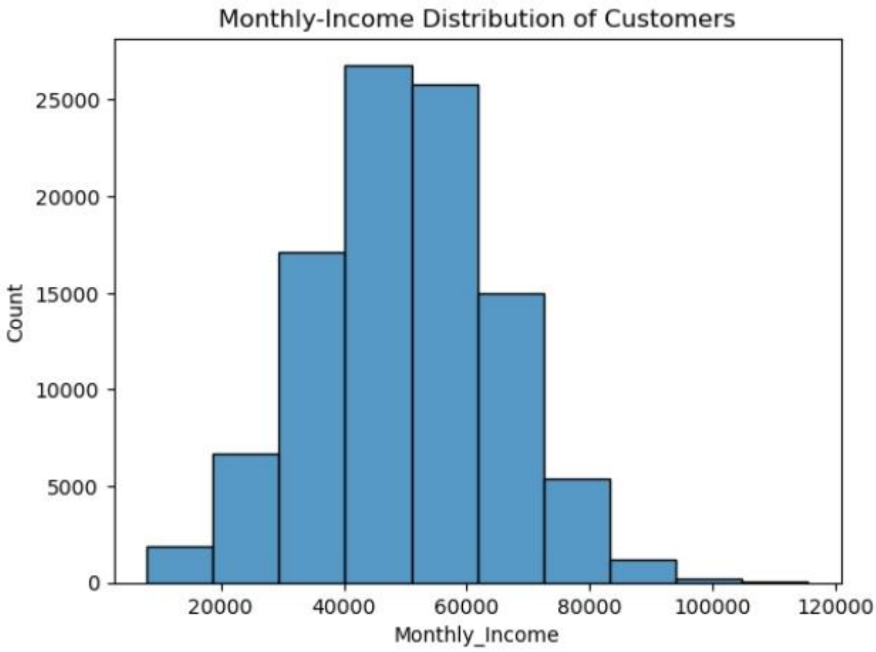


- The rental service appears to be **evenly used** across all major occupational segments:
  - **Working Professionals**

- Students
- Self-Employed
- Retired Individuals

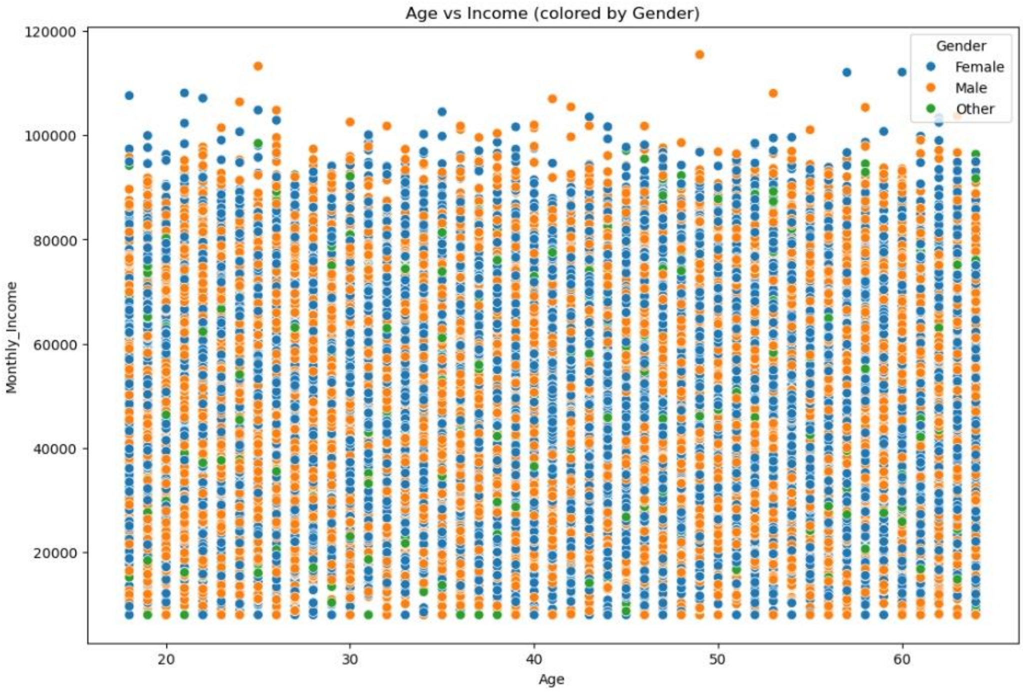
- Each group contributes approximately equally, with **no segment dominating** the user base.

Insight: Monthly Income Distribution of Customers



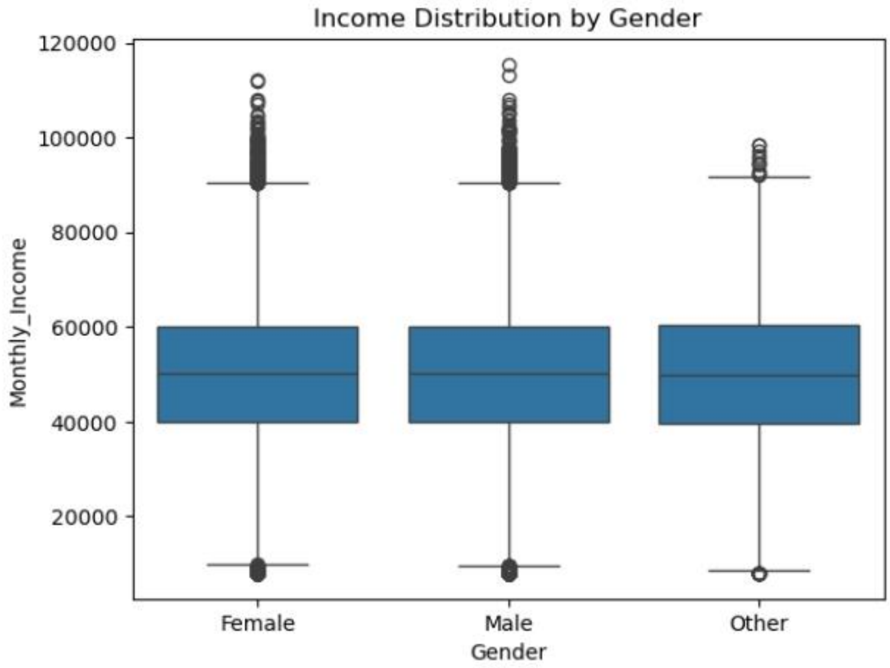
- The income distribution follows a **normal (bell-shaped) curve**, peaking around ₹45,000–₹60,000.
- Most users earn between **₹30,000 and ₹70,000** per month.
- Very few customers fall in either extreme — **low-income (< ₹20,000)** or **high-income (> ₹80,000)** brackets.

Insight: Age vs Income (by Gender)



- The scatter plot shows **income spread across different age groups**, with color coding for gender:
  - 🔵 **Female**
  - 🟠 **Male**
  - 🟢 **Other**
- **No strong upward or downward trend**: Income levels are quite scattered across all age ranges, suggesting:
  - High earners exist across both younger (20s) and older (50s–60s) segments.
  - There's **no clear correlation** between age and income in this sample.
- **Gender-wise spread appears balanced** — both male and female users have a similar spread in income brackets.

Interpretation of "Income Distribution by Gender"



Central Tendency

- The **median income** (thick horizontal line in each box) is **similar across all genders**, indicating fair consistency in central income levels.

Spread of Income

- The **interquartile range (IQR)** (box height) is comparable, showing that most individuals across all genders earn within a similar range (≈ ₹40,000 to ₹60,000).

Outliers

- All genders show **numerous outliers on the higher end** (dots above the whiskers), indicating the presence of high earners in all gender groups.
- Few outliers exist on the lower side as well (those below ₹15,000), especially among "Other" and "Male" categories.

Symmetry

- The distribution is **symmetric**, suggesting no heavy skew toward extreme low or high incomes.

E. <u>Segment Extraction (Clustering Analysis)</u>

1. Behavioural Segment Extraction

	Booking_Status	Vehicle_Type	Booking_Value	Ride_Distance	Driver_Ratings	Customer_Rating	Offline_Payment
0	0	Prime Sedan	444	0	4.0	4.0	0
1	1	Bike	158	13	4.1	4.0	1
2	1	Prime SUV	386	40	4.2	4.8	0
3	0	eBike	384	0	4.0	4.0	0
4	1	Mini	822	45	4.0	3.0	0

To uncover customer segments based on behavioral characteristics, **K-Means Clustering** was applied to the dataset. To enhance the accuracy and interpretability of clustering, the dataset was divided into three meaningful behavioral sub-segments:

a. Booking Behaviour

Features used: Booking_Value, Ride_Distance

```
booking_cluster_behavioural_data.head()
```

	Ride_Distance	Booking_Value	Cluster
0	0	444	2
1	13	158	2
2	40	386	0
3	0	384	2
4	45	822	0

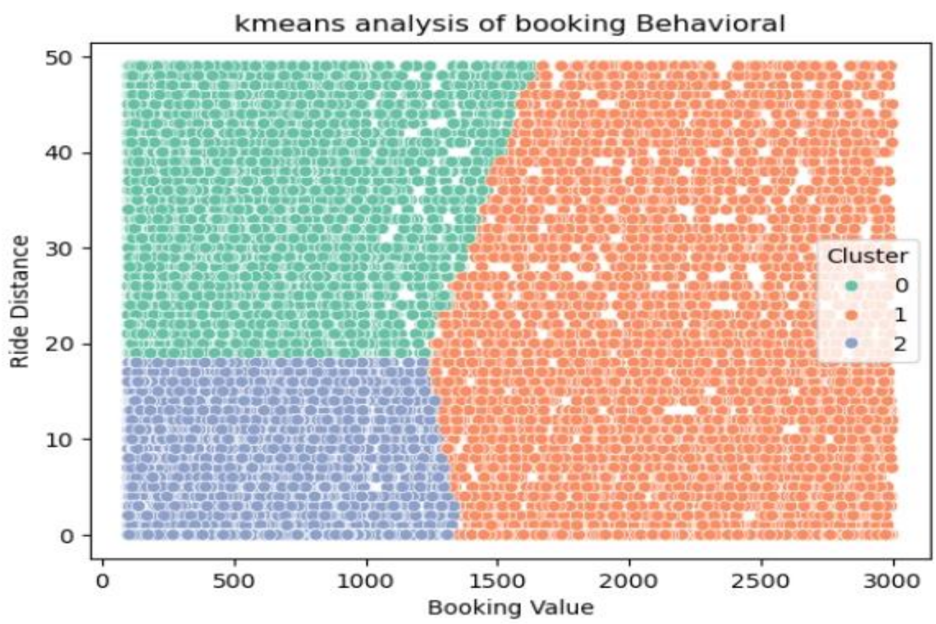
Features Used:

- **Booking_Value**
- **Ride_Distance**

Number of Clusters: 3 (identified via the Elbow Method)

Cluster Analysis:

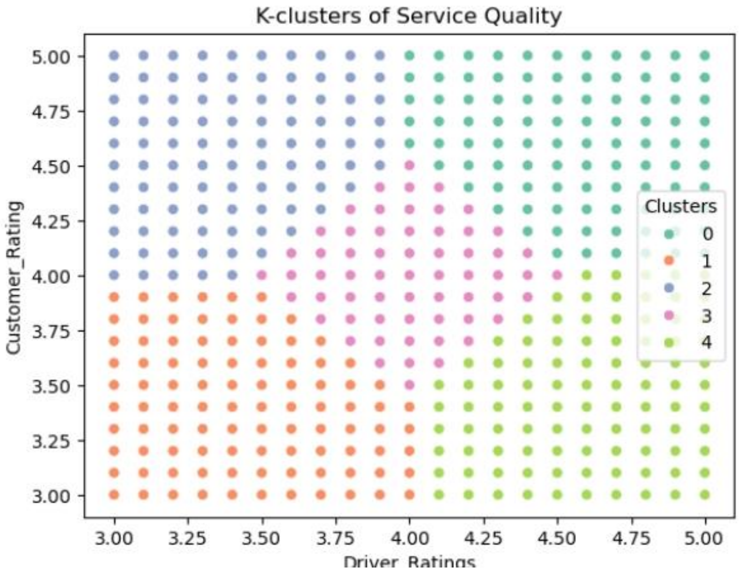
- **Cluster 0**: Represents customers who typically travel longer distances (20–50 km) but at a **lower booking value** (₹0–1500). These may include budget-conscious long-distance travellers.
- **Cluster 1**: Includes **short-distance riders** (0–20 km) who also spend **less (₹0–1500)**. Possibly local commuters or casual users.
- **Cluster 2**: These customers book rides across **all distances (0–50 km)** but have a **higher booking value (₹1500–3000)**. Likely to be frequent or premium service users.
    - This clustering helps identify **price-sensitive**, **casual**, and **premium customer groups** based on how much they spend and how far they travel.



b. Service Quality Perception

Features Used:

- Driver_Ratings
- Customer_Rating



Number of Clusters: 5

Cluster Analysis:
The clustering plot reveals how ratings from drivers and customers combine to form unique service-based segments. Some clusters may represent:

- Highly satisfied customers with high mutual ratings.
- Groups where **driver ratings are higher than customer ratings**, possibly indicating issues in customer service perception.
- Clusters where both ratings are moderate or low, pointing to **quality gaps** in service delivery.

This segmentation allows businesses to identify **satisfaction zones**, detect **potential conflict areas**, and take steps to improve driver-customer alignment.

2. Demographic Segment Extraction

To uncover hidden customer segments based on demographic characteristics, **K-Means Clustering** was employed.

a. Feature Selection

We selected key demographic attributes for clustering:

- Age
- Monthly Income

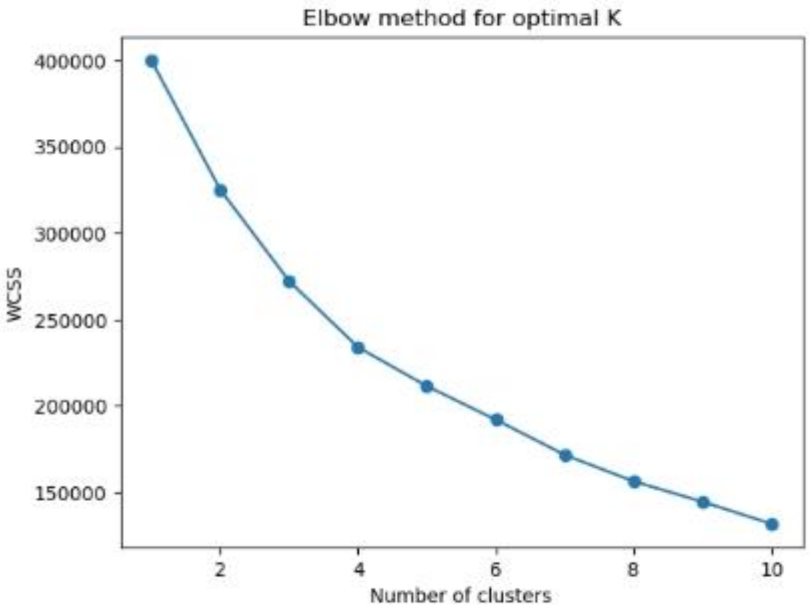
- Gender (encoded)
- Occupation (encoded)

These features were normalized categorical/continuous values contributing to customer differentiation.

b. Preprocessing for Clustering

- Gender and Occupation were label-encoded using LabelEncoder() from sklearn.preprocessing to convert categorical variables into numeric format.
- A feature matrix was constructed containing: [['Age', 'Monthly_Income', 'Gender', 'Occupation']]

c. Optimal Number of Clusters

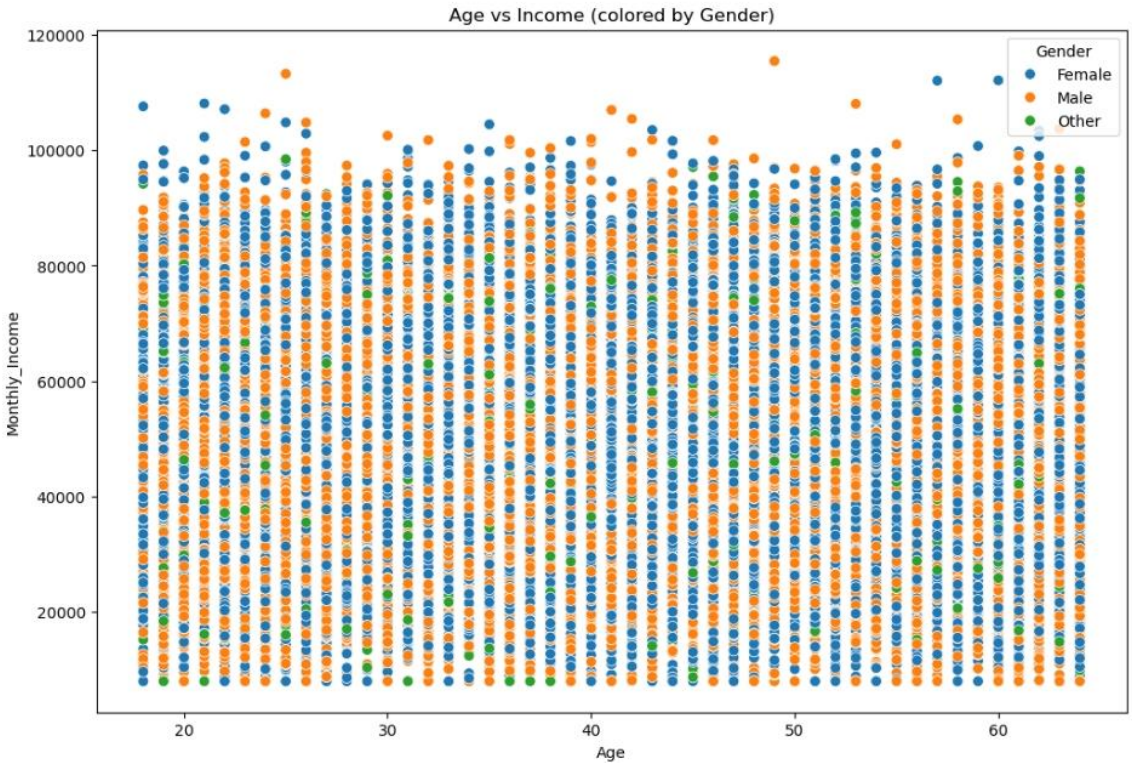


To determine the ideal number of clusters (**K**), the **Elbow Method** was applied. The Within-Cluster Sum of Squares (WCSS) was plotted for K ranging from 1 to 10.

The elbow was observed at **K = 4**, indicating four distinct customer groups.

d. Clustering Execution

K-Means was applied with **n_clusters = 4**. Customers were grouped into 4 clusters based on demographic similarity.



e. Cluster Centroids (Mean Attributes)

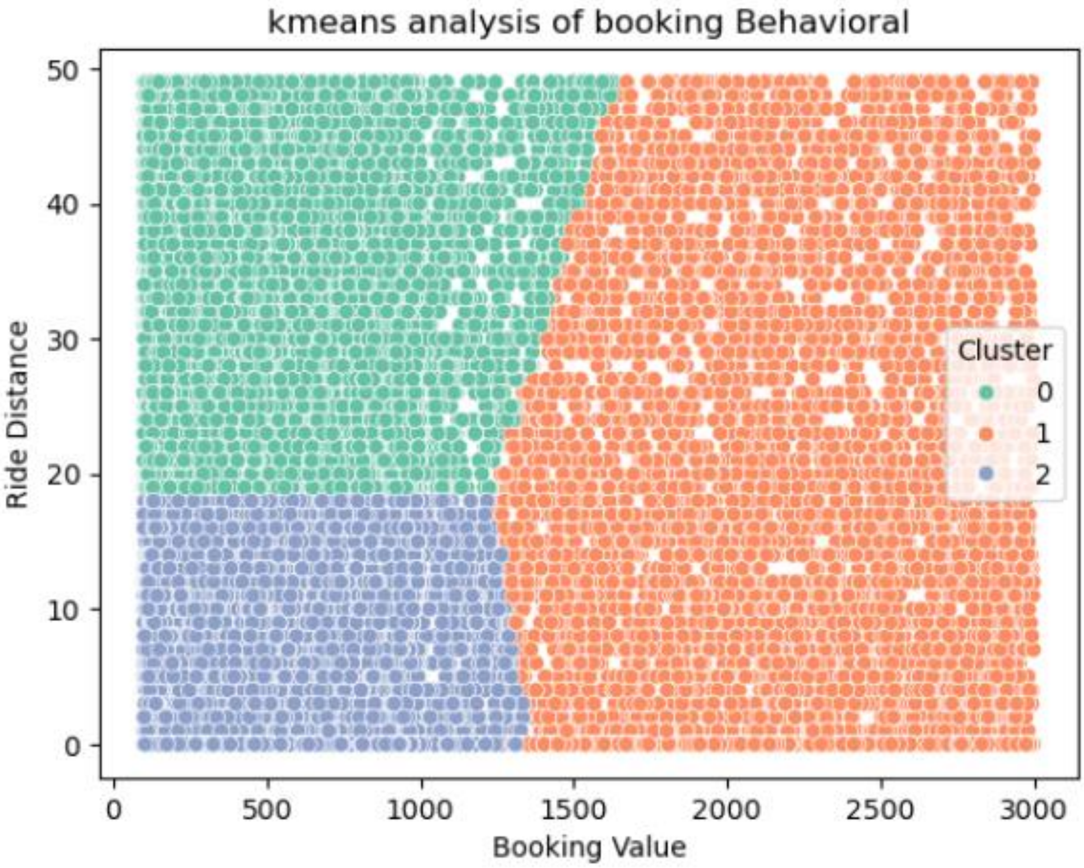
Cluster	Age	Monthly_Income
0	41.047218	27686.075928
1	40.972933	57188.351967
2	41.179940	73039.315793
3	40.986886	43592.845145

F. Profiling the Segments (Insights)

1. Behavioural Segment Profiling

To uncover actionable insights from customer behaviour patterns, we performed **K-Means clustering** on three distinct dimensions of user behaviour: **Booking Behaviour**, **Service Quality**, and **Payment Behaviour**. Each was analysed independently to isolate core behavioural traits and form well-defined segments for strategic targeting.

1.1 Booking Behaviour Segments



Features Used: Booking_Value, Ride_Distance
Optimal Number of Clusters: 3 (as determined by the Elbow Method)

Cluster 0: Budget Long-Distance Travellers

- **Ride Distance**: 20–50 units
- **Booking Value**: ₹0–1500
- **Insight**:
  These customers **travel longer distances** while keeping their spending low. Likely to be **cost-conscious** or **value-oriented**, possibly using shared or economy rides.
- Prefer budget-friendly options
- Might use the service for **commuting**, long regular routes, or intercity travel

Cluster 1: High-Spending Power Users

- **Ride Distance**: 0–50 units
- **Booking Value**: ₹1500–3000

- **Insight**:
  These users are **willing to spend more** regardless of ride length. They span **all distances**, but choose premium options.
  Likely include:
  - **Business professionals**, tourists, or high-income individuals
  - Prioritize **convenience, speed, or vehicle quality**
  - Potential for **upselling premium services**

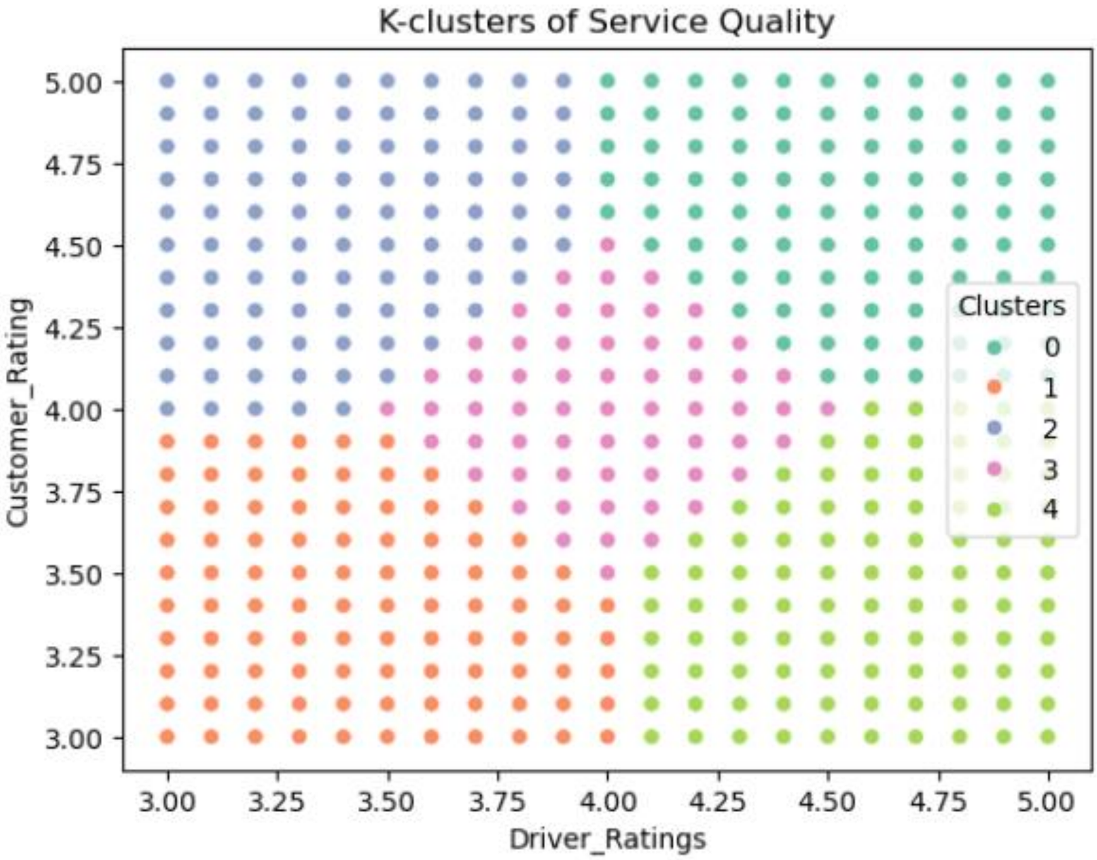
Cluster 2: Premium Short-Distance Riders

- **Ride Distance**: 0–20 units
- **Booking Value**: ₹0–1500
- **Insight**:
  These users travel **short distances**, but still spend more per kilometer. This suggests a **preference for quality, safety, or brand loyalty** even for small trips.
  - Possibly elderly, female, or student riders
  - Could reflect **city-center users**, airport drops, or high-demand zones
  - Opportunity for targeted **micro-ride premium plans**

Strategic Value:

- Each cluster represents a **distinct customer mindset**, and this segmentation can guide:
- Pricing models
- Promotional offers
- Route optimization
- Loyalty programs

1.2 Service Quality Segments



Features Used: Driver_Ratings, Customer_Rating
Optimal Number of Clusters: 5

Cluster 0:

- High customer satisfaction with **ratings between 4.5 and 5** from both customers and drivers.
- Indicates **strong mutual appreciation** and a highly satisfactory service experience.
- Marketing Insight: Promote testimonials and loyalty campaigns within this happy user group.

Cluster 1:

- Low ratings on both ends (mostly **3.0–3.8**), reflecting dissatisfaction or service quality issues.
- This may include new users, occasional riders, or customers in underperforming service areas.
- Marketing Insight: Focus on service improvement, driver retraining, or targeted feedback collection.

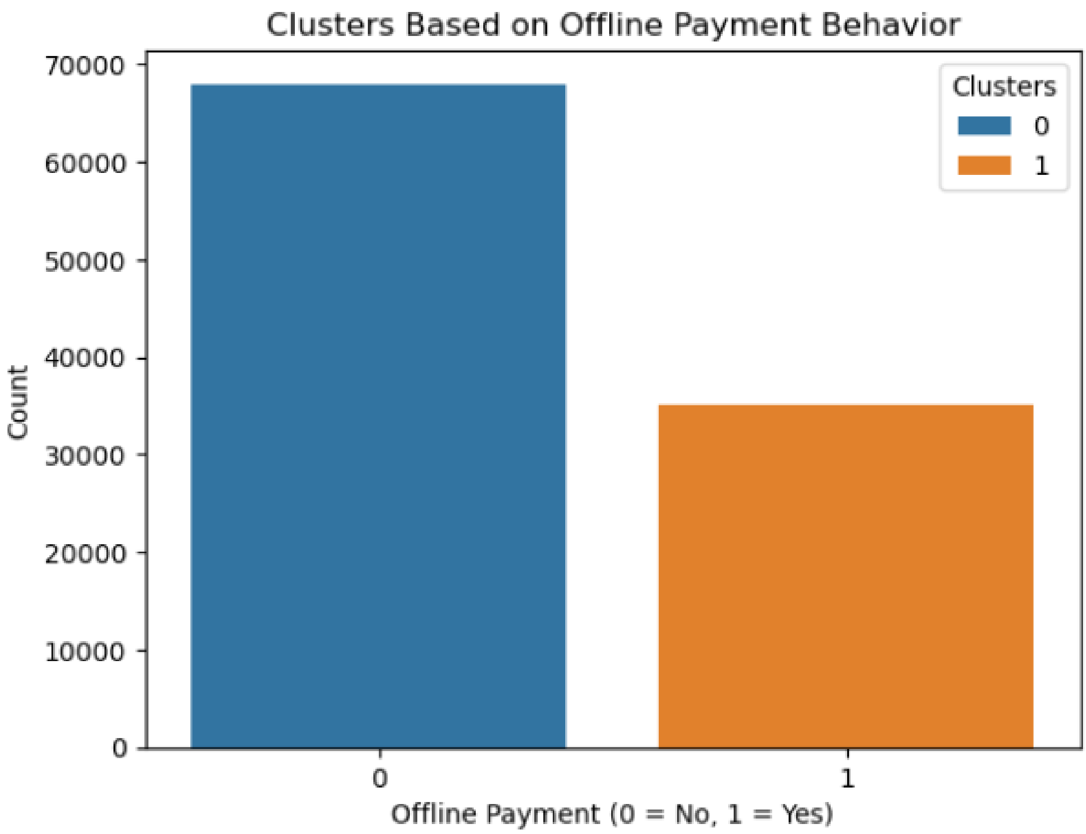
Cluster 2 & 3:

- Average ratings (~4.0–4.3), representing **neutral to moderately satisfied** experiences.
- These groups may not be loyal yet, but can be swayed toward retention with minor enhancements.
- Marketing Insight: Run follow-up campaigns and post-ride discounts to nudge loyalty.

Cluster 4:

- **Imbalanced rating patterns** — either high customer ratings with lower driver ratings, or vice versa.
- May indicate **discrepancies in expectations or communication gaps**.
- Marketing Insight: Improve two-way rating awareness and address mismatches in rider-driver dynamics.

1.3 Payment Behaviour Segments



Feature Used: Offline_Payment
Optimal Number of Clusters: 2

Cluster 0 (Blue):

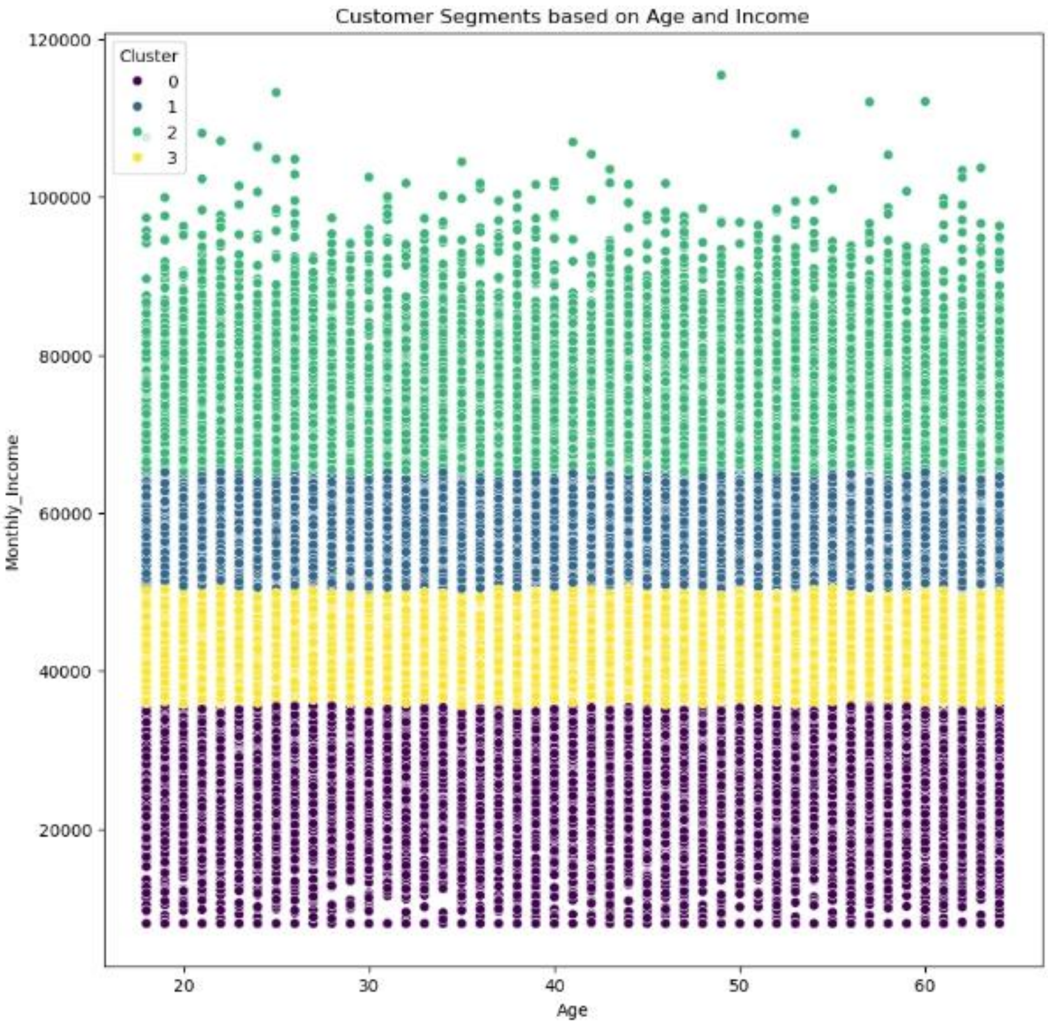
- Represents customers who predominantly **use online or digital payment methods**.

- This segment is the **largest share of the base**, signalling strong digital adoption and trust in cashless systems.
- Marketing Insight: Promote app-based offers, cashback, and UPI-integrated campaigns.

Cluster 1 (Orange):

- Comprises users who **prefer offline payments** such as cash.
- This may be attributed to **older age groups, lower digital literacy**, or regions with limited digital infrastructure.
- Marketing Insight: Offer incentives to transition to online payments or ensure robust support for offline convenience.

2. Demographic Segmentation – Cluster Profiles



Cluster 0: Budget-Conscious Middle-Aged Group

- **Average Age**: ~41 years
- **Monthly Income**: ~₹27,600

- **Likely Profile**:
  This segment includes **middle-aged customers** with **lower income levels**.
  They might represent:
    - Entry-level employees
    - Skilled workers or support staff
    - Possibly from semi-urban or suburban areas
- **Behavioural Implication**:
  Prefer **economical ride options**, discounts, or subscription-based value plans.

Cluster 1: Working Professionals (Mid-Income Tier)

- **Average Age**: ~41 years
- **Monthly Income**: ~₹57,100
- **Likely Profile**:
  These are likely **mid-level professionals or young executives**.
  Balanced in income and lifestyle preferences.
- **Behavioral Implication**:
  Open to **moderate pricing**, value-for-money services, occasional use of premium offerings.

Cluster 2: Affluent Commuters / High Earners

- **Average Age**: ~41 years
- **Monthly Income**: ~₹73,000
- **Likely Profile**:
  This segment likely consists of:
    - Senior professionals, business owners, or tech workers
    - Urban, possibly from metro cities
- **Behavioral Implication**:
  More **price-insensitive**, focused on comfort, brand, and speed.
  Strong candidates for **premium, loyalty, or business ride plans**.

Cluster 3: Moderately Earning Adults

- **Average Age**: ~41 years
- **Monthly Income**: ~₹43,500

- **Likely Profile**:
  Positioned between Clusters 0 and 1.
  Could include:
  - Early-career professionals
  - Freelancers or private sector employees
- **Behavioural Implication**:
  Value seekers, open to both budget and premium services depending on context.

Observations:

- **Age is nearly uniform** (~41 years) → segmentation is driven more by **income and encoded categorical values (Gender & Occupation)**.
- This segmentation allows targeted **marketing by economic tier**, not age bracket.

Cluster	Age	Monthly_Income
0	41.047218	27686.075928
1	40.972933	57188.351967
2	41.179940	73039.315793
3	40.986886	43592.845145

G. Selection of Target Segment

To decide which segment(s) to target, we evaluate them based on business potential, profitability, and strategic fit.

1. Recommended Target Segments

1.1 Behavioural Segment: Cluster 1 — High Expenditure, All Distance Travellers

- Booking Value: ₹1500–3000 (high)
- Ride Distance: 0–50 km (short and long)
- Insight:

These customers are willing to pay more regardless of distance, suggesting:

  - Strong spending power
  - Use rides frequently for both short and long distances
- Why target?
  - High revenue potential
  - Ideal for premium or loyalty-based plans
  - Opportunity to upsell and cross-sell

1.2 Demographic Segment: Cluster 2 — Affluent Commuters / High Earners

- Monthly Income: ~₹73,000
- Insight:

  Likely urban professionals or business-class customers with:

  - High disposable income
  - Preference for comfort, reliability, and convenience
- Why target?
  - Most price-insensitive
  - More likely to adopt new or premium services
  - Can become brand loyalists if experience is optimized

Strategic Reasoning

Overlap of Behavioural Cluster 1 + Demographic Cluster 2 gives us high-income users who spend more across all ride types

- This segment offers the best ROI, allowing the company to:
  - Increase revenue per user
  - Improve brand perception through tailored offerings
  - Introduce high-end features like priority bookings, executive rides, or memberships

Segments *not selected* for targeting (but useful for future strategies):

- Behavioural Cluster 0: Low spenders but long-distance travellers – may respond well to economical packages or ride passes
- Behavioural Cluster 2: High spenders but only on short distances – suitable for micro-mobility or subscription plans
- Demographic Cluster 0: Lower income – good for budget offerings but not priority for premium strategy

H. Marketing Strategy Recommendation

This step builds directly on the **target segment** you selected:

- **High-income individuals (Demographic Cluster 2)**
- **High spenders across short and long trips (Behavorial Cluster 1)**

These customers are your **most valuable and profitable group** — so your marketing strategy should focus on **value, exclusivity, and convenience**.

Marketing Strategy Recommendation for the Selected Segment

1. Positioning Strategy: Premium & Personalized

- Position your service as a **premium, reliable, and time-efficient** commuting solution.
- Messaging should highlight:
  - Comfort
  - Safety
  - Time savings
  - Executive-level treatment

2. Product/Service Customization

- **Priority Booking**: Skip the queue for peak-hour demand.
- **Executive Rides**: Offer better vehicles, trained drivers.
- **Loyalty Program**: Points-based rewards or subscription packages.
- **On-Ride Benefits**: Free WiFi, refreshments, or charging ports.

3. Communication & Promotion Channels

- **Email & App Notifications** with personalized ride suggestions.
- **LinkedIn Ads**: Professional targeting based on income bracket.
- **Influencer Campaigns**: Promote convenience and luxury for busy professionals.

4. Pricing Strategy

- **Value-based pricing** – they're ready to pay for comfort, so:
  - Emphasize benefits, not cost.

- Offer premium monthly plans.
- **Bundle Offers**: e.g., “10 executive rides/month at ₹X – save 15%”

5. Geographic & Time Targeting

- Target **urban metro zones** with high density of professionals.
- Promote during **weekday mornings and evenings** — prime commute times.

6. Customer Relationship Management (CRM)

- Use ride history & feedback to send:
  - Tailored deals
  - Thank-you messages
  - Automated issue resolution
- Assign **dedicated support chat** for VIP users.

7. Business Growth Levers

- **Referral Campaigns**: Reward premium users for inviting peers.
- **Cross-Selling**: Promote corporate ride plans or outstation services.

8. Retention Strategy

- Auto-renewing ride plans
- “Platinum” tier for long-term users
- Surprise upgrades or anniversary rewards

Git-hub Repository

Link: [FeynnLabs-Internship/Rental_vehicle_Analysis/Online_Vehicle_Segmentation at main · Ishan534/FeynnLabs-Internship]