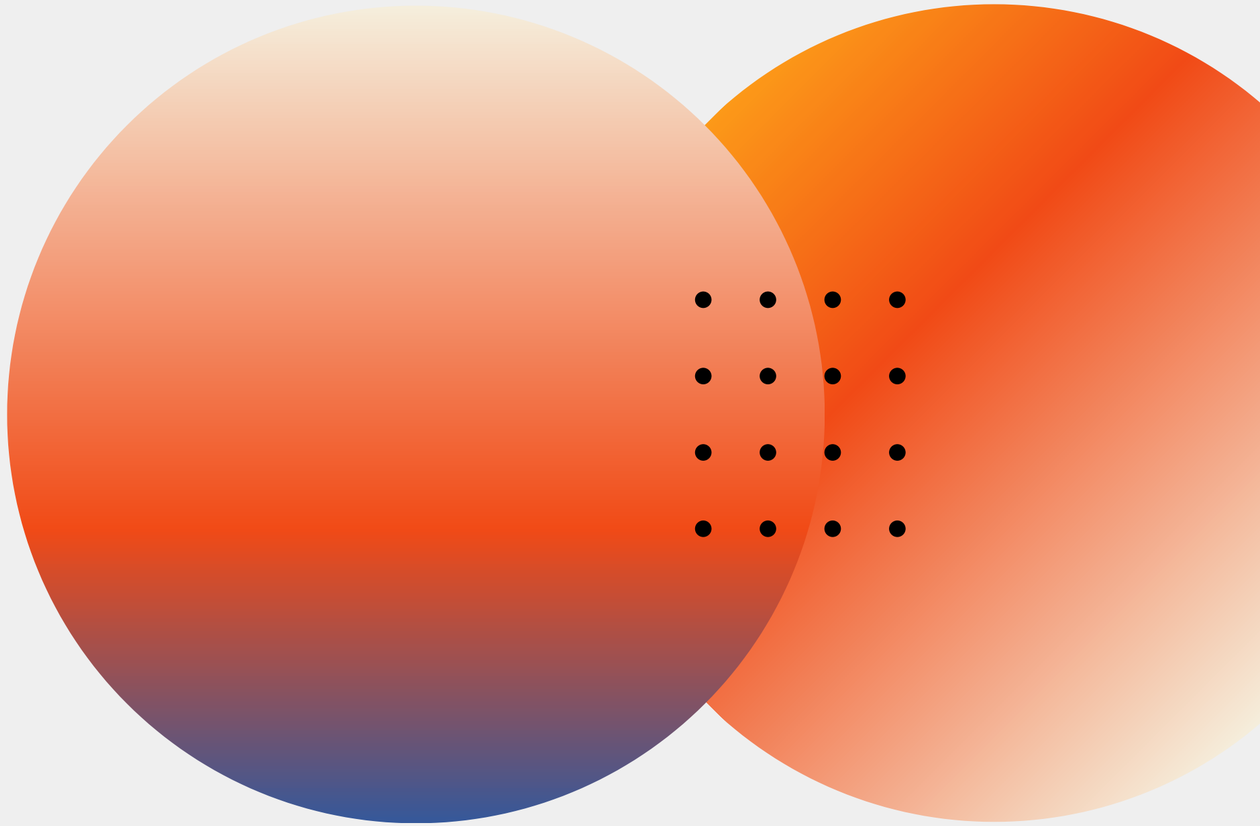
Let's Start

Statistical Analysis

Maximizing Revenue for cab Drivers through Analyzing Data using Python



Agenda

Objectives

Problem Statement

Research Question

Overview of Data

What Methodology Used?

Journey Insights

Testing of Hypothesis

Result: Regression Analysis

Recommendations

Key Finding

Problem Objectives

This project's goal is to run an A/B test to examine the relationship between the total fare and method of payment.
In this, we would use Python programming language, hypothesis testing, and Descriptive Statistics to extract useful information that can help cab drivers generate more cash.
In this particular case, we want to find out if there is big difference in the for those who pay with credit cards versus those who pay with cash.

Increase Financial Awareness

Attract New Customers by Schemes

Increase Customer Base

Increase Revenue for Drivers

Problem Statement

In the fast-paced booking sector, making the most of revenue is essential for long-term success and driver happiness. Our goal is to use data-driven insights to maximize revenue streams for taxi drivers in order to meet this need. Our research aims to determine whether payment methods have an impact on fare pricing by focusing on the relationship between payment type and fare amount.

Research Question

Is there a relationship between total fare amount and payment type and can we nudge customers towards payment methods that generate higher revenue for drivers without negatively impacting customer experience?

Overview of Data

For this analysis. we utilized the comprehensive dataset of NYC trip records, which is collect from Kaggle.com, records, used data cleaning and feature engineering procedures to concentrate solely on the relevant fields essential for our analysis wherein we get statistically significant result.

Fields used for Analysis:

- passenger_count (from one to five)
- payment_type i.e., Cash & Card
- fare_amount
- trip_distance
- duration (to be taken in mins

	passenger_count	payment_type	fare_amount	trip_distance	duration
0	1	Card	6.0	1.20	4.800000
1	1	Card	7.0	1.20	7.416667
2	1	Card	6.0	0.60	6.183333
3	1	Card	5.5	0.80	4.850000
5	1	Cash	2.5	0.03	0.883333

Methodology Used

Descriptive Statistics

Perform statistical Analysis to summarize key aspects of the data, focusing on fare amounts and payment types.

Exploratory Data Analysis

Removed unnecessary fields, null values, and negative values

Hypothesis Testing

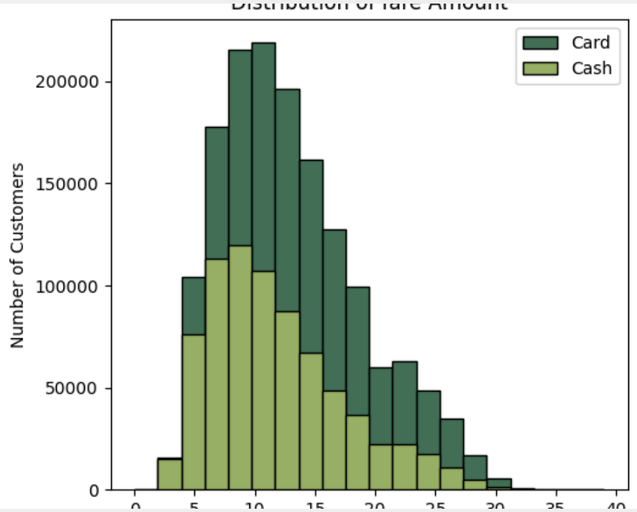
take a t-test to evaluate the relation between payment type and fare amount, testing the hypothesis that different payment methods influence fare amounts.

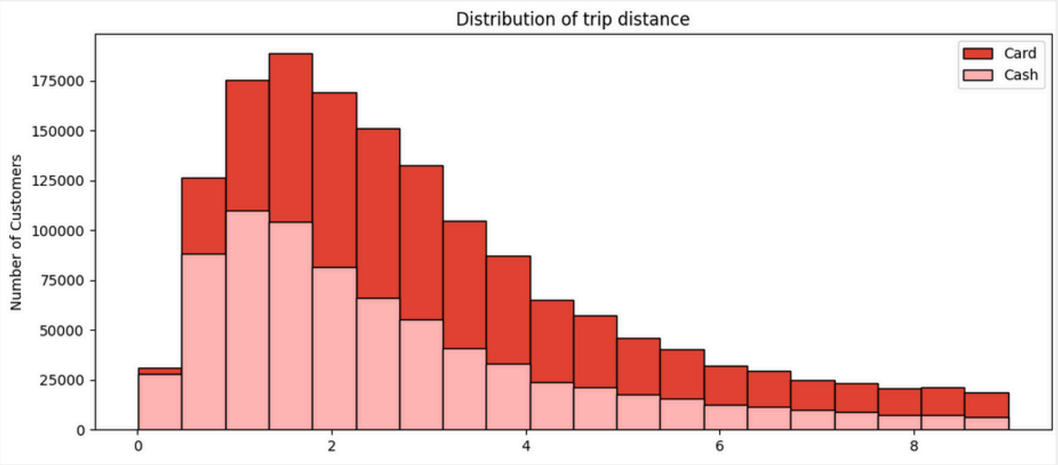
Regression Analysis

Implemented linear regression to explore the relationship between trip duration and fare amount

Journey Insights

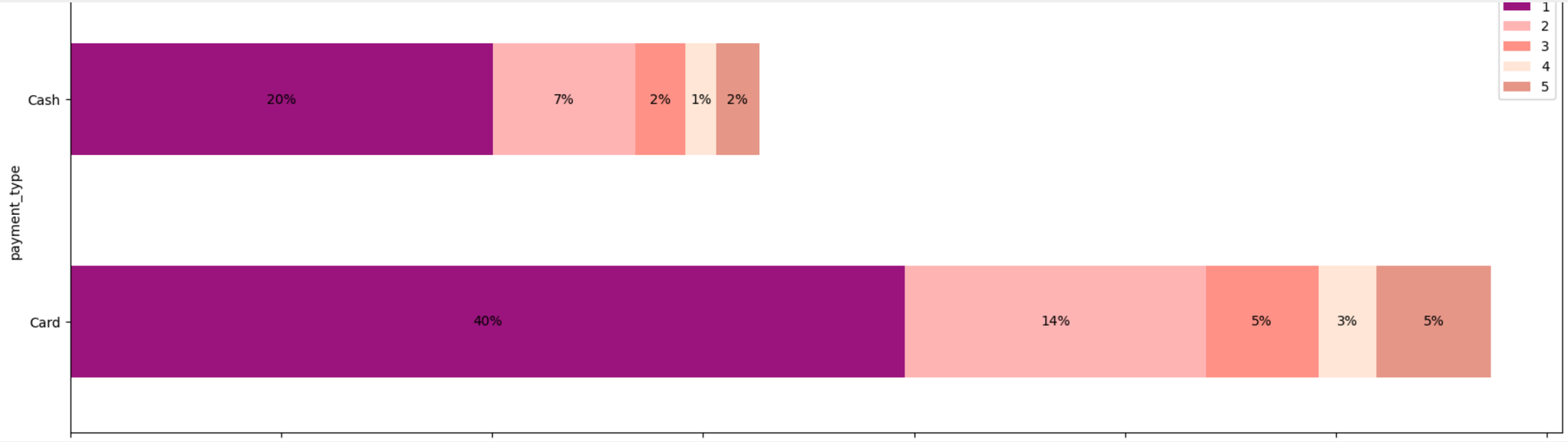
- Customers paying with cards tend to have a slightly higher average trip distance and fare amount compared to those paying with cash
- Indicates that customers prefers to pay more with cards instead of cash this is occasionally when fare amount and the trip of the distance is long.





Passenger Count Analysis

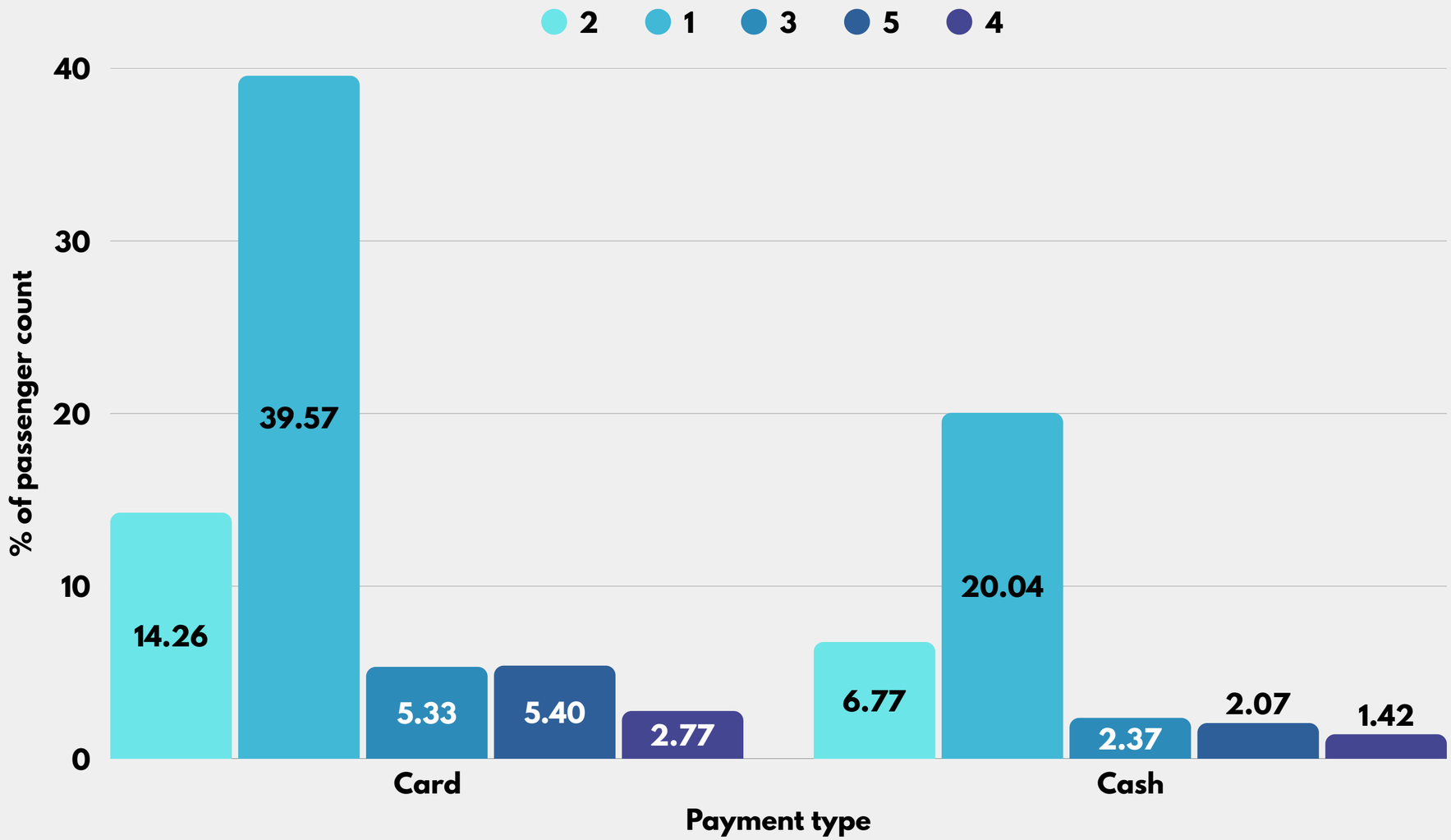
	payment_type	1	2	3	4	5
0	Card	39.57	14.26	5.33	2.77	5.40
1	Cash	20.04	6.77	2.37	1.42	2.07

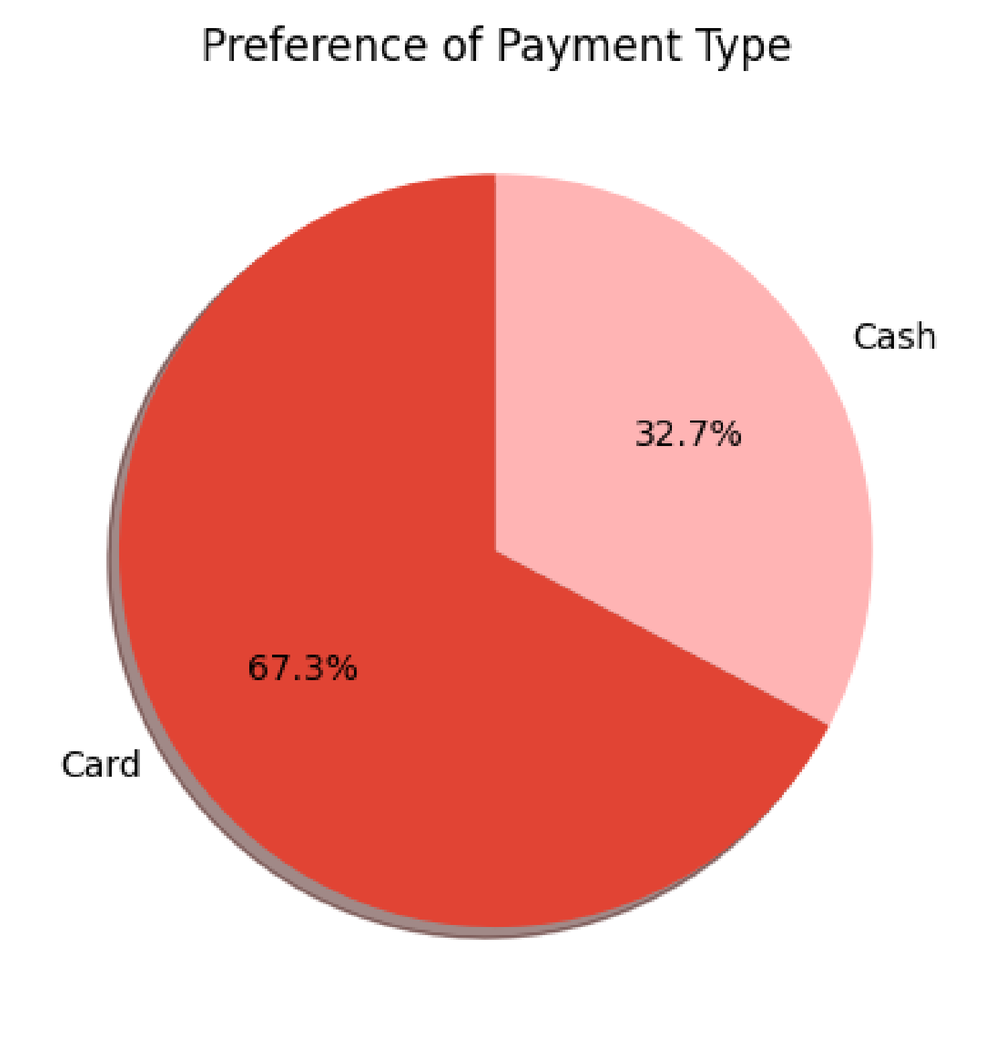


Passenger Count Analysis

- Among card payments, rides with a single passenger - passenger_count = 1, comprise the largest proportion, constituting 40.08% of all card transactions.
- Similarly, cash payments are predominantly associated with single-passenger rides, this all constitutes 20.04% of all cash transactions.
- Noticeable things, decrease in the % of transactions as the passenger count increases, indicating that larger groups are less likely to use cabs or they might opt for another payment option.
- In which These insights emphasize the importance of considering both payment_type and passenger_count.

Journey Insights





Preference of Payment Types

- The proportion of customers paying through cards are significantly higher than those paying with cash, with card payments accounting for 67.2% of all transactions compared to cash payments at 32.5%.
- The proportion of customers paying through cards are significantly higher than those paying with cash, with card payments accounting for 67.2% of all transactions compared to cash payments at 32.5%.

Hypothesis Testing

- **Null Hypothesis:** There is no difference in average fare between customers who use credits cards and customers who use cash.
- **Alternate Hypothesis:** There is difference in average fare between customers who use credits cards and customers who use cash.

Since, we got the value of t-Statistics of 165.5 and the p-value is less than the level of significance.
Herein , we reject the null hypothesis, suggesting that there is indeed a significant difference in average fare between the two payments methods.

Regression Analysis

Interpreting Regression Output

- Goodness-of-Fit Statistics:

- **R**-squared **(0.926) About 92.6%** of the variability in duration is explained by your two predictors—**excellent fit.**
- **Adj. R**-squared **(0.926)** Adjusted for number of predictors; very close to R^2, which is good.
- **F-statistic (1.433e+07)** Tests whether the model provides a better fit than a model with no predictors. This huge value means your model is statistically significant.
- **Prob (F-statistic) (0.00)** Confirms the overall model is significant.

Regression Analysis

Interpreting Regression Output

Regression Coefficients:

- **Intercept (const) (-3.9284):** When both fare_amount and trip_distance are zero, predicted duration is -3.93 minutes — which doesn’t make practical sense, but often the intercept isn't meaningful on its own.
- **fare_amount (2.3907)** For each **1-unit increase in fare**, duration increases by ~**2.39 minutes**, holding distance constant. Suggests higher fares are associated with longer trips.
- **trip_distance (-4.0875):** Surprisingly negative — each additional unit of distance is associated with a ~4.09 minute decrease in duration, all else being equal. This might seem counterintuitive and could hint at multicollinearity or differing fare structures for short vs. long trips.

Key Insights

- Encourage customers to pay with credit cards to capitalize on the potential for generating more revenue for cab drivers
- Implement strategies such as offering incentives, gifts, journey discounts, if some travelling simultaneously for credit card transaction.
- Ensure and assure that the payment option did he choose, would be convenient and seamless transaction, facility offered by the cabs should not delayed.

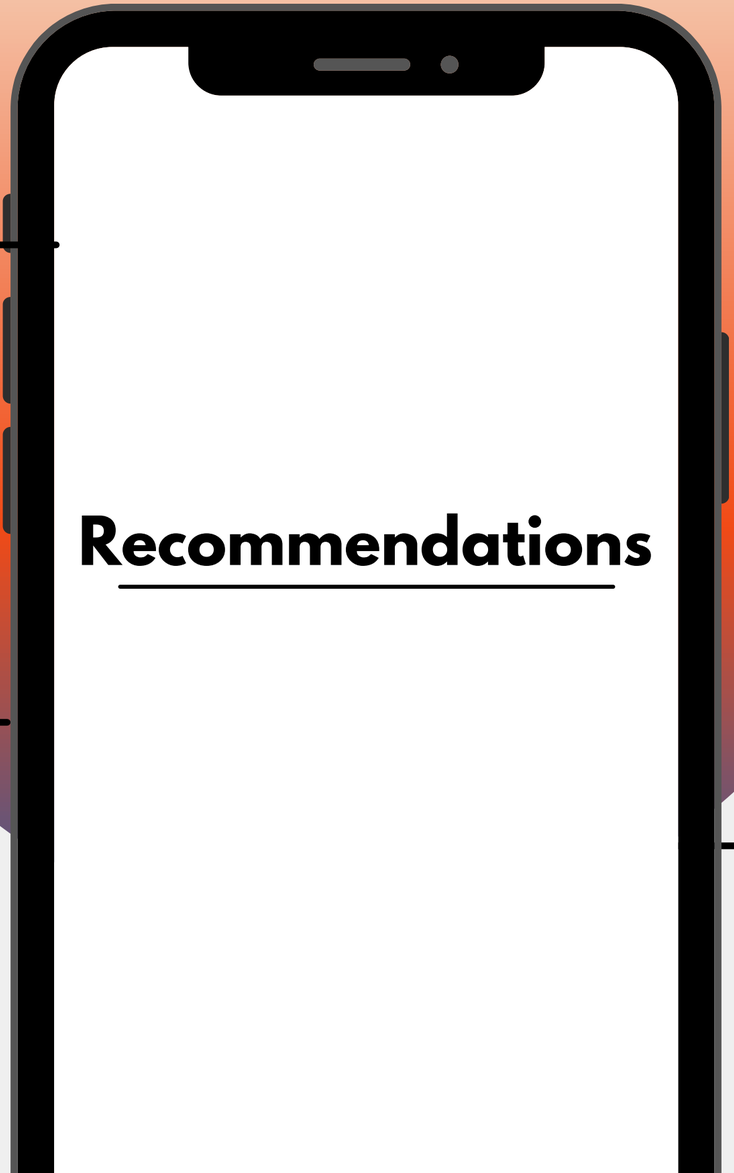
Recommendations

Ask for Customer Experience

Provide Credit Card Offers

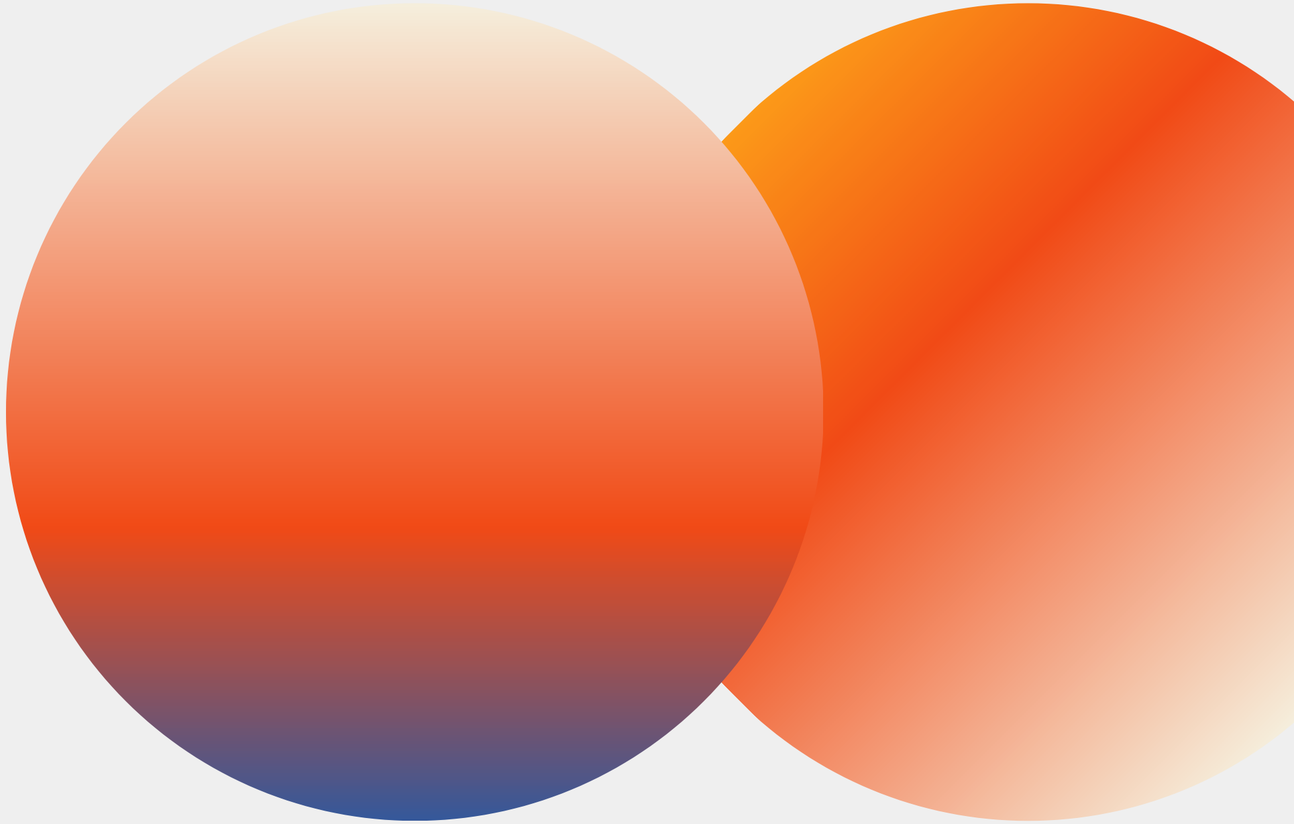
Ensure Security and safety of Cab ride

May Offer Discount as per passenger Count



Thank You

Do you have any questions?



End