

Maximizing Revenue for NYC Yellow Taxi through Payment Type Data Analysis



NYC TAXI STUDY

Agenda

- Business Problem
- Research Question
- Data Overview
- Methodology
- Analysis and Findings
- Hypothesis Testing
- Regression Analysis
- Conclusion

Business Problem

In the fast-paced taxi booking sector, making the most of revenue is essential for long-term success and driver happiness.

In the competitive taxi booking industry, optimizing revenue is crucial for both sustained success and driver satisfaction. We aim to leverage data-driven approaches to enhance income streams for taxi drivers, prioritizing their financial well-being. Our study seeks to explore the potential influence of payment methods on fare pricing by examining the correlation between different payment types and the amount charged for fares.



Research Question

"What is the relationship between payment methods and fare pricing in the taxi booking industry, and how does this correlation impact taxi driver revenue and financial well-being?"

Data Overview

For this study, we employed the extensive dataset of NYC Taxi Trip records, employing data cleaning and feature engineering techniques to focus exclusively on the pertinent columns crucial for our analysis.

Relevant columns used for this reseach:

- passenger_count (1 to 5)
- payment_type (card or cash)
- fare_amount
- trip_distance (miles)
- duration (minutes)
-

[6]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	RateCodeID	store_and_fwd_flag	dropoff
--	----------	----------------------	-----------------------	-----------------	---------------	------------------	-----------------	------------	--------------------	---------

0	2	2015-01-15 19:05:39	2015-01-15 19:23:42	1	1.59	-73.993896	40.750111	1	N
1	1	2015-01-10 20:33:38	2015-01-10 20:53:28	1	3.30	-74.001648	40.724243	1	N
2	1	2015-01-10 20:33:38	2015-01-10 20:43:41	1	1.80	-73.963341	40.802788	1	N
3	1	2015-01-10 20:33:39	2015-01-10 20:35:31	1	0.50	-74.009087	40.713818	1	N
4	1	2015-01-10 20:33:39	2015-01-10 20:52:58	1	3.00	-73.971176	40.762428	1	N

[6]:

```
taxi_data.tail()
```

[6]:

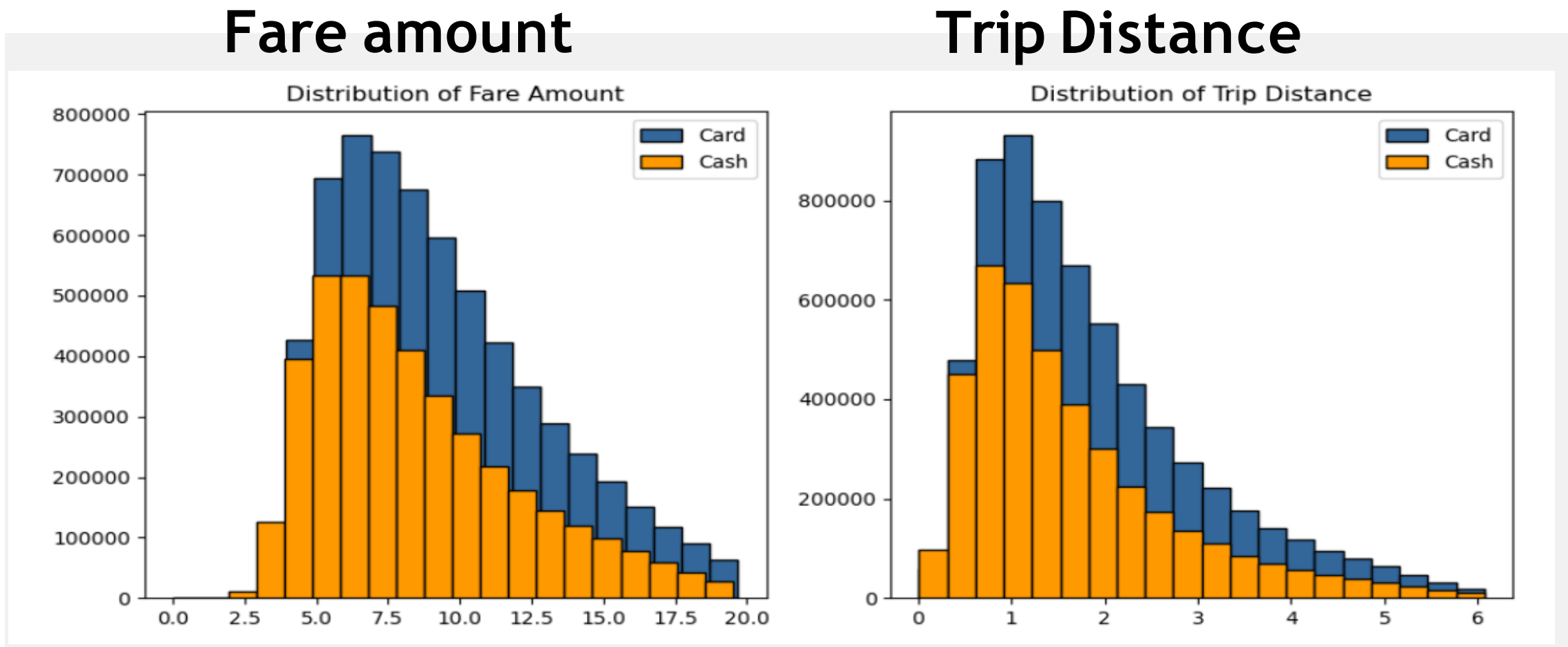
	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	RateCodeID	store_and_fwd_flag
12748981	1	2015-01-10 19:01:44	2015-01-10 19:05:40	2	1.0	-73.951988	40.786217	1	N
12748982	1	2015-01-10 19:01:44	2015-01-10 19:07:26	2	0.8	-73.982742	40.728184	1	N
12748983	1	2015-01-10 19:01:44	2015-01-10 19:15:01	1	3.4	-73.979324	40.749550	1	N
12748984	1	2015-01-10 19:01:44	2015-01-10 19:17:03	1	1.3	-73.999565	40.738483	1	N
12748985	1	2015-01-10 19:01:45	2015-01-10 19:07:33	1	0.7	-73.960350	40.766399	1	N

Methodology

Step	Description
Descriptive Analysis	Performed statistical analysis to summarize key aspects of the data, focusing on fare amounts and payment types.
Hypothesis Testing	Conducted a T-test to evaluate the relationship between payment type and fare amount, testing the hypothesis that different payment methods influence fare amounts.
Regression Analysis	Implemented regression Analysis to explore the relationship between trip duration (calculated from pickup and dropoff times) and fare amount.

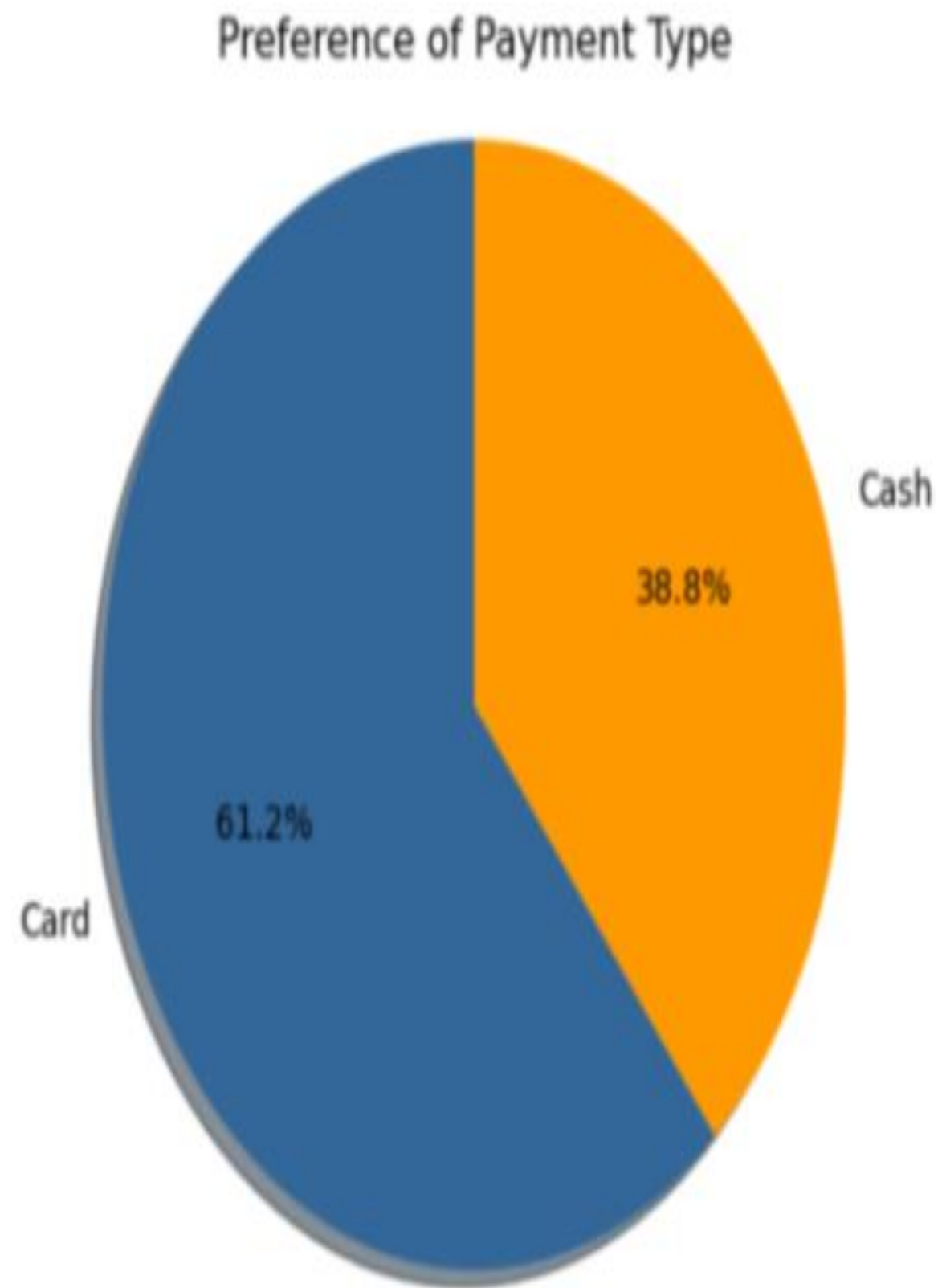
Journey Insights

- Customers who use cards for payment typically exhibit a slightly higher average trip distance and fare amount compared to those who opt for cash, suggesting a preference for card payments especially for longer trips with higher fares.
- Customers who opt to pay with cards tend to incur slightly higher average trip distances and fare amounts compared to those choosing cash payments. This trend implies a preference for card transactions, particularly for longer trips with higher fares.



	Payment Type	Mean	Standard Deviation
Fare amount	Card	9.15	3.67
	Cash	8.45	3.62
Trip Distance	Card	1.86	1.15
	Cash	1.45	1.09

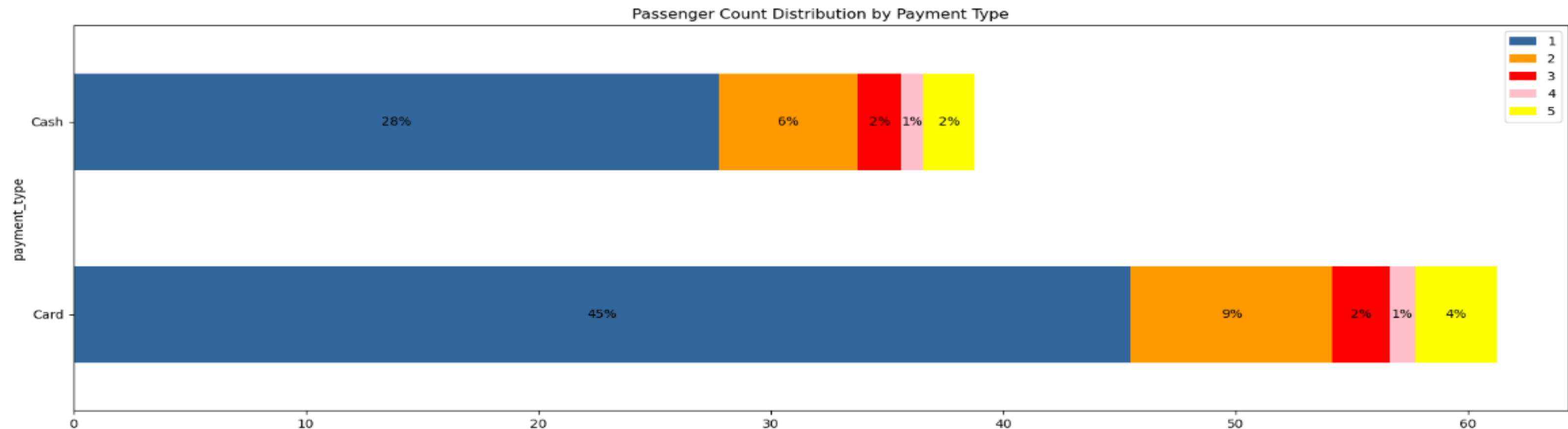
Preference of Payment Types



- This data highlights a notable trend favoring card-based transactions over cash, with card payments constituting a substantial majority of 61.2% compared to cash payments, which make up 38.8% of all transactions. This indicates a strong preference among consumers for the convenience, security, and potentially incentivized features offered by card payment methods.

Passenger Count Analysis

- Card payments reveal a predominant preference for single-passenger rides, constituting 45.02% of transactions.
- Similarly, cash transactions mirror this trend, comprising 20.04% of rides with solo passengers.
- The decline in transaction percentages as passenger count increases underscores the importance of analyzing payment methods and passenger count for valuable insights into customer behavior.
- Analyzing payment method and passenger count together yields key insights into customer preferences and behavior.



Hypothesis Testing

Null hypothesis: There is no difference in average fare between customers who use credit cards and customers who use cash.

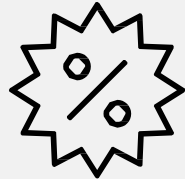
Alternative hypothesis: There is a difference in average fare between customers who use credit cards and customers who use cash

With a T-statistic of 304.51 and a P-value of less than 0.0, we reject the null hypothesis, **suggesting that there is indeed a significant difference in average fare between the two payment methods.**

Recommendations



The analysis revealed a significant association between payment methods and fare pricing. Customers paying with credit cards tend to have higher fares compared to those paying with cash.



By leveraging statistical analysis and hypothesis testing, valuable insights were extracted to aid taxi drivers in maximizing their earnings. Understanding the correlation between payment methods and fare pricing allows drivers to make informed decisions.



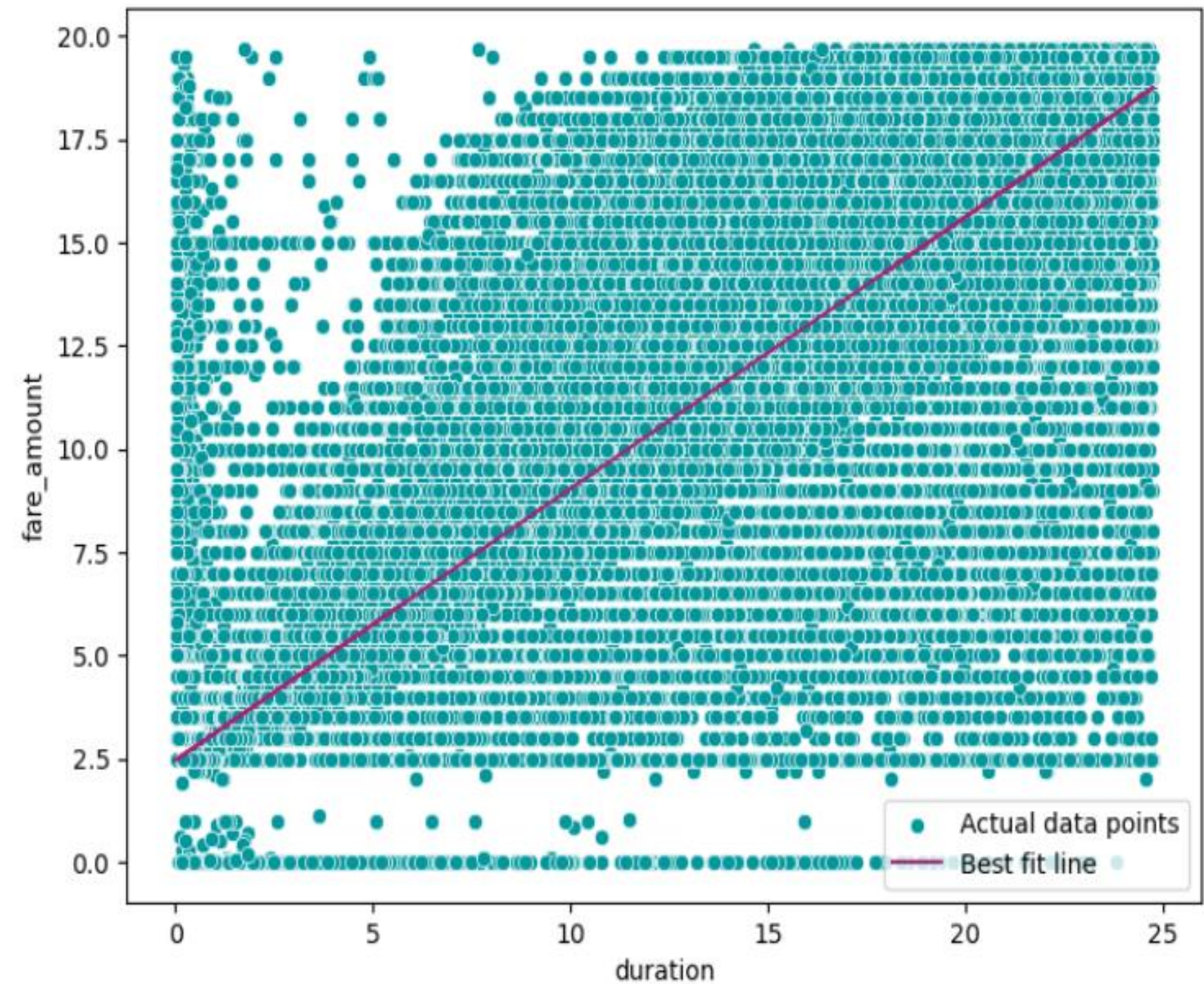
The project underscores the importance of ongoing data analysis and adaptation in the competitive taxi booking industry. As customer preferences and market dynamics evolve, taxi drivers and companies must continuously analyze data to identify trends, patterns, and opportunities for improvement.

Regression Analysis

This chart is a scatter plot depicting the relationship between the duration of taxi trips (in minutes) and the corresponding fare amount (in dollars). Each point on the plot represents an actual data point from the dataset, where the x-axis represents the duration of the trip and the y-axis represents the fare amount.

The data points are colored in a shade of teal. Additionally, a best-fit line is plotted on the graph, shown in a shade of pink ('#A62372'). This line represents the linear regression model's prediction of fare amount based on trip duration.

The line indicates the overall trend or relationship between trip duration and fare amount as captured by the model. By comparing the actual data points to the best fit line, we can assess how well the linear regression model fits the data and understand the direction and strength of the relationship between trip duration and fare amount.



Thank you!
