



Hypothesis Testing for Customer Insights

This project focuses on leveraging statistical methods to derive data-driven insights from customer data, supporting informed business decisions.

Project Objectives



Hypothesis Testing

Perform rigorous hypothesis testing on customer-related data.



Validate Assumptions

Validate assumptions using robust statistical methods.



Derive Insights

Generate data-driven insights to support strategic business decisions.

Key Concepts Covered

Data Cleaning & Preparation

Ensuring data quality and readiness for analysis.

Exploratory Data Analysis (EDA)

Uncovering patterns and anomalies in datasets.

Hypothesis Formulation

Developing testable statements about data relationships.

Statistical Testing

Applying statistical methods to validate hypotheses.

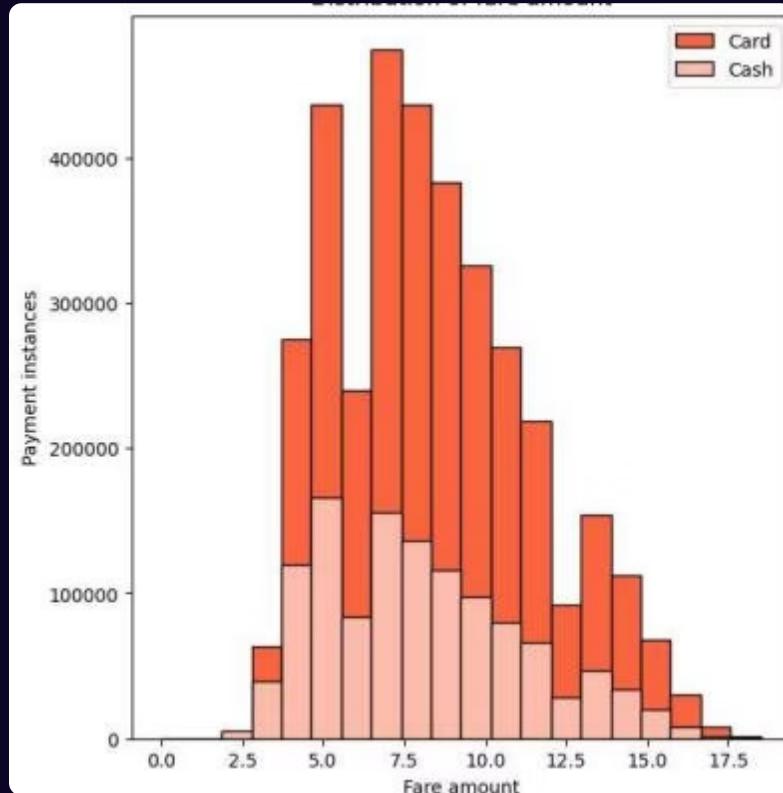
Data Visualization

Presenting complex data in an understandable visual format.

Insight Interpretation

Translating analytical findings into actionable business strategies.

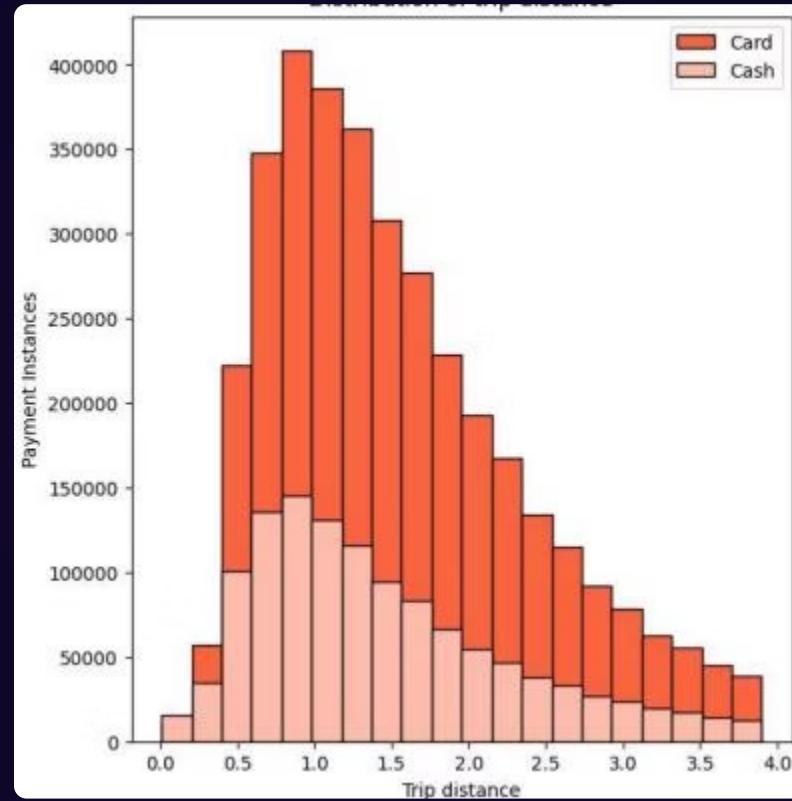
Fare Amount Distribution



The distribution of fare amounts is bimodal, with peaks around \$5.0 and \$7.5. Card payments are significantly higher than cash payments in the higher fare ranges, indicating a preference for card use on more expensive trips.

Trip Distance Distribution

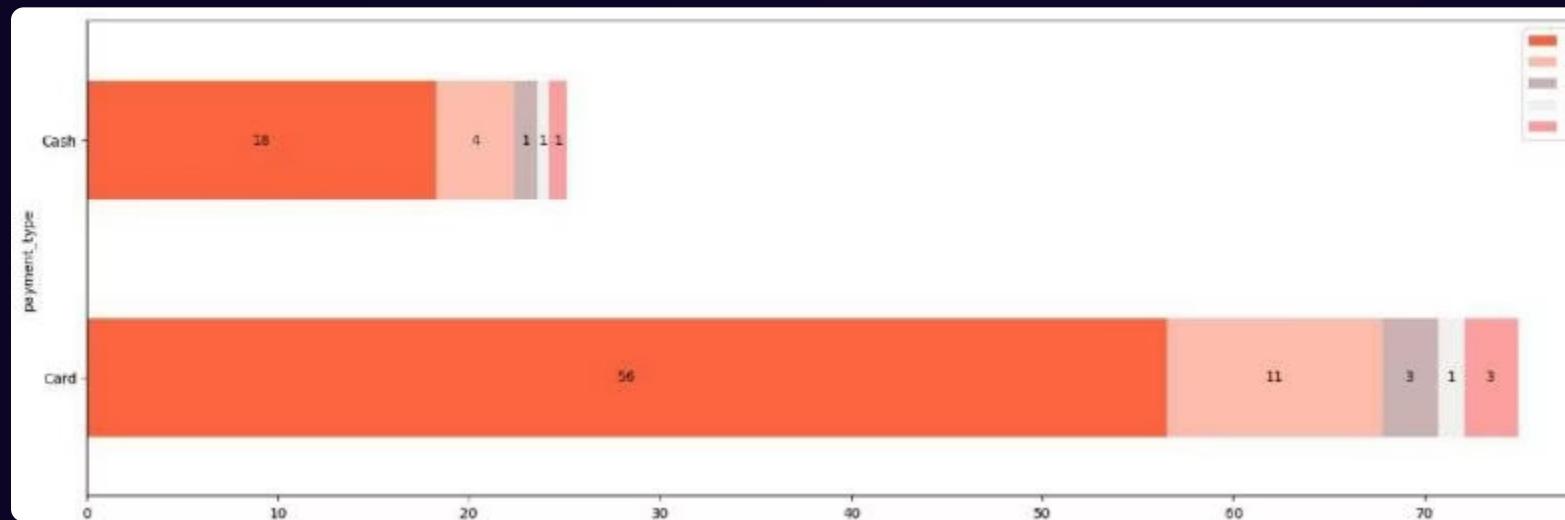
Trip distance distribution is highly skewed right, peaking between 1.0 to 1.5 miles. Similar to fare amounts, card payments dominate for longer trip distances, suggesting convenience for extended travel.



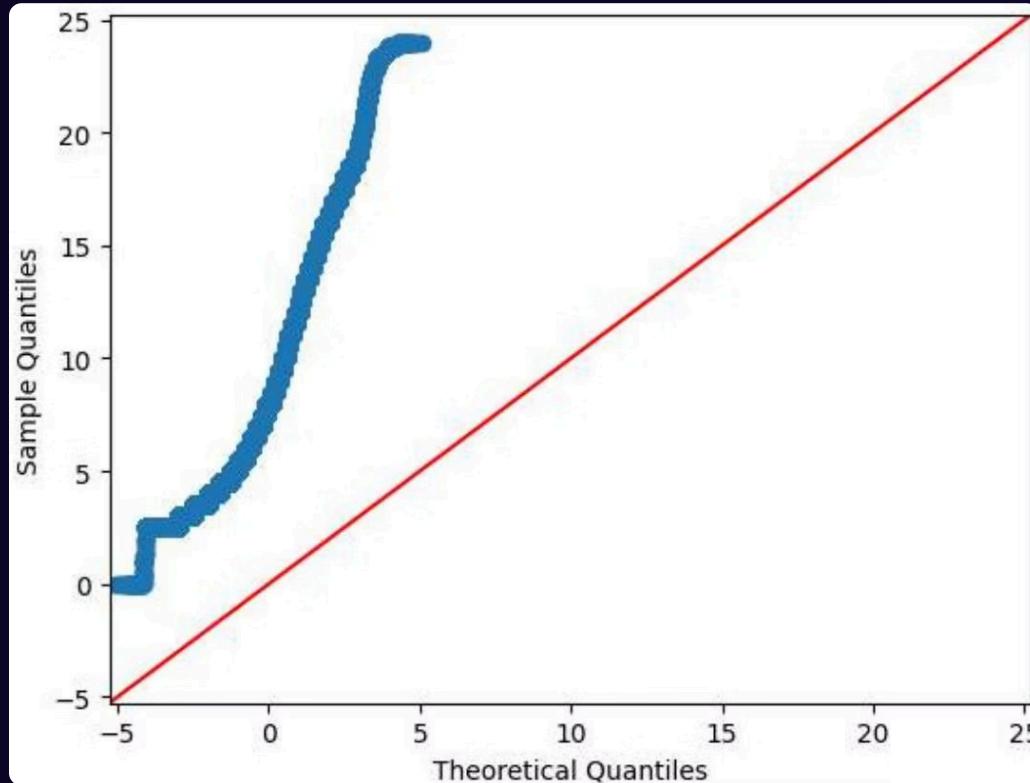
Payment Type Breakdown by Category

| | | | | | |
|------|----|----|---|---|---|
| Cash | 28 | 4 | 1 | 1 | 1 |
| Card | 56 | 11 | 3 | 1 | 3 |

This stacked bar chart illustrates the distribution of payment types across five distinct categories. Card payments consistently show higher values across all categories, reinforcing their overall dominance.



Q-Q Plot Analysis



The Q-Q plot generally follows the identity line, indicating a good fit between sample and theoretical quantiles. However, a noticeable deviation in the upper right quadrant suggests that sample quantiles are consistently higher, implying a heavier tail or positive skewness in the data compared to the theoretical distribution.

Implications for Business Decisions

1

Optimize Payment Options

Focus on enhancing card payment infrastructure due to its high usage, especially for higher fares and longer trips.

2

Targeted Promotions

Develop promotions for cash users in specific categories or for shorter distances where cash is still prevalent.

3

Data-Driven Strategy

Utilize these insights to refine pricing strategies and service offerings, aligning with customer payment behaviors.



Conclusion & Next Steps

Key Takeaways

Card payments dominate, especially for higher fares and longer distances. Data distributions reveal important patterns for customer behavior.

Future Analysis

Further investigate the deviations in Q-Q plots to understand underlying data characteristics and potential biases.

Actionable Insights

Implement targeted strategies based on payment type and trip characteristics to optimize operations and customer satisfaction.

