Statistical Analysis of New York City TLC Data

Summary Report – Project 4
Automatidata – NYC TLC 2017

Project Overview

In this part of the project, the data needs to be statistically described and compared. Meaningful insights and statistical analysis testing is performed for future modeling.

Key Insights

The Problem: The primary objective of this phase was to determine if a relationship exists between payment type and fare amount.

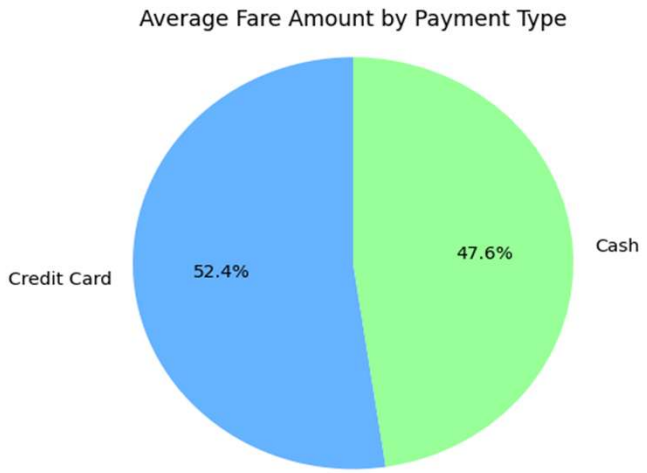
Proposed solution: Statistical analysis through a two-sample t-test would reveal if there is a significant difference between payment methods.

Keys to success

- **A significance level of 5% was utilized to conduct a two-sample t-test, providing a mathematically sound basis for rejecting the null hypothesis.**
- **Computing descriptive statistics like mean, standard deviation, and quartile distributions allowed for an initial understanding of data patterns before moving to inferential testing.**

Details

The pie chart was created with Pyplot to show the differences between cash and credit card payments (on average). This relationship should further be studied to allow for insights into achieving better profit margins.



Alt Text: New York City TLC data pie chart showing mean cash payments against credit card payments.

Next Steps

- Investigate external factors beyond payment choice, such as trip distance or modern payment habits, that may influence why credit card users pay more.
- Perform future modeling to quantify how much a shift toward credit card payments would increase overall industry profitability.