Executive Summary of Statistical Analysis for New York TLC Data

Commission Prepared by Automatidata

Project Overview

The NYC Taxi & Limousine Commission has consulted with Automatidata to build a regression model to predict taxi cab fares. The Automatidata team performed statistical analysis to determine statistically significant relationship between payment type and fare amount paid to maximize earnings from taxicab drivers.

Key Insights

EDA:

- Negative fare amounts and 0 trip_distance values were removed for a clean data set.

A/B Test

- Only credit card and cash payments were considered for this analysis. Other payment methods were disregarded.

Hypothesis Testing

- Samples from credit card users and cash users were randomly selected with replacement to equally represent both user types.

Details

A/B Test
A two-sample t-test was conducted to determine if there was a statistically significant relationship between the variables payment_type and fare_amount.

Null hypothesis: there is no difference in the average fare amount paid between credit card users and cash users.

Alternative hypothesis: there is a difference in the average fare amount between credit card users and cash users.

Using a significance level of 5%, the p-value found was 1.21e-6%, and the null hypothesis was rejected.

Next Steps

- Encourage taxicab users to accept credit card payments to maximize earnings.