

Statistical Review and A/B Testing for New York City TLC Project

Executive summary report
Commission Prepared by **Automatidata**

Overview

The aim of this project is to predict taxi cab fares prior to each trip. At this point, the project is aimed at finding ways to generate more revenue for New York City taxi cab drivers. This project paper will examine the relationship between total fares and the type of payment.

Problem

Cab drivers receive a variety of tips. While examining the relationship between total fees and payment type, the project seeks to discover whether customers who pay by credit card pay higher fares than customers who pay cash.

Solution

The Automatidata team ran an A/B test to analyze the relationship between credit card payment and total fare amount. The key business insight is that encouraging customers to pay with credit cards will likely generate more revenue for taxi drivers.

Details

Steps conducted in the A/B test

1. Collected sample data from an experiment in which customers are randomly selected and divided into two groups:
 - a. Customers who are required to pay with credit card.
 - b. Customers who are required to pay with cash. This enables us to draw causal conclusions about how payment method affects fare amount.
2. Perform descriptive statistics to understand the average value of fares by each available payment method.
3. Conduct a two-sample t-test in order to determine if there is any significant difference in the average value of fares divided by payment method. The payment methods sampled were cash and credit card.

A/ B test results

During the test we found that there is a significant difference in the average value of the rates between cash and credit card customers. It was found that customers who use credit cards have higher fare amounts than those who pay with cash.

Next Steps

The Automatidata data team recommends that the New York City TLC encourages customers to pay with credit cards, and create strategies to promote credit card payments. For example, the New York City TLC can install signs that read "Credit card payments are preferred" in their cabs, and implement a protocol that requires cab drivers to verbally inform customers that credit card payments are preferred.

Tik Tok | Executive Summary: Statistical A/B Test

Executive summary report
A Tik Tok Claim Classification Project

Overview

The Tik Tok data team seeks to develop a machine learning model to assist in the classification of claims for user submissions. In this part of the project, the data team will conduct a hypothesis test to analyze the relationship between `verified_status` and `video_view_count`.

Problem

In previous results, the Tik Tok data team found a difference in the statuses of users who are claim and opinion. This part of the project seeks to understand if this difference is due to the number of views.

Solution

The Tik Tok data team will conduct an A/B test to understand the relationships between these two variables and to gain a better understanding of the differences in behavior towards unverified and verified users.

Details


The TikTok data team considered the relationship between `verified_status` and `video_view_count`.

One approach conducted was to examine the mean values of `video_view_count` for each group of `verified_status` in the sample data. The findings showed that unverified accounts have a mean of 265,663 views vs. 91,439 views for verified accounts

The second approach was a two-sample hypothesis test. Aligned with preliminary findings from the mean values, this statistical analysis shows that any observed difference in the sample data is due to an actual difference in the corresponding population means.

Next Steps

This model on the `verified_status` variable can provide important information about user behaviors in the verified and unverified groups. The results of this model can provide context that will help in the construction of the claim classification model.



Waze | User Churn Project: Executive Summary

Two-sample Hypothesis Test Executive Summary

Overview

The Waze data team is currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn on the Waze app. As part of the effort to improve retention, Waze wants to learn more about users' behavior.

Problem

The Waze data team seeks to make inferences about the population from the data obtained to understand more about users and their behavior. Among the variables that seek to provide more information on behavior are the use of the application and the platform used.

Solution

Develop a two-sample hypothesis test to analyze and determine whether there is a statistically significant difference between mean number of rides and device type

Details

In a first approach, the average use of the app by users while driving was calculated. The calculations indicate that iPhone users use the app more, with 68 uses on average, while Android users use 66 uses on average.

Despite this difference, the results obtained in the t-tests of both samples indicate that there is no significant difference in app usage between the two groups.

This shows that the experience and use of the app is the same for users of both platforms.

Next Steps

The data team recommends conducting further t-tests and regression models on other variables to understand user behavior and then using these results in building the model that predicts user churn.

