Waze PACE Strategy Document IV
Statistical Review

Introduction

PACE stands for Plan, Analyze, Construct and Execute. It is a framework that illustrates the foundation and structure for data analysis projects and each letter represents an actionable stage in a project. The stage "Plan" involves the definition of the project scope, the research of business data and the workflow development. The stage "Analyze" involves data scrubbing, data conversion and database formatting. The stage "Construct" involves building models and machine learning algorithms and selecting a modeling approach. The stage "Execute" involves the presentation of results to decision-makers, stakeholders and others in order to receive feedback. This framework is built upon an iterative cycle where each stage may reveal new insights, requiring the return to earlier stages. A PACE strategy document is used to record decisions and reflections at different stages of the data analytical process. It typically includes the definitions of roles and actions to ensure clarity and accountability.

Purpose

Waze's free navigation app makes it easier for drivers around the world to get to where they want to go. We want to analyze user data and develop a machine learning model that predicts user churn on the Waze app. Churn quantifies the number of users who have uninstalled the Waze app or stopped using the app. The project focuses on monthly user churn. Developing a churn prediction model will help prevent churn, improve user retention and grow Waze's business. An accurate model can also help identify specific factors that contribute to churn and answer questions such as "Who are the users most likely to churn?", "Why do users churn?" and "When do users churn?". For this stage of the project, we identify four main tasks that are presented in the following visual.

Task 1
Imports & data loading

Task 2
Data exploration

Task 3
Statistical test(s)

Task 4
Communicate insights with stakeholders

Considerations

PACE: Plan Stage

● What is the main purpose of this part of the project?

> The goal is to to reduce user churn and understand user behavior through data analysis.

● What specific question are we trying to answer with our data?

> The question we are trying to answer is whether there is a significant difference in the average number of drives between iPhone and Android users.

● What is the significance of random sampling?

> It ensures that the sample accurately represents the population, minimizing bias and allowing for generalizable conclusions.

● Provide an example of sampling bias that might occur if we do not use random sampling.

> If only users from a specific geographic area were sampled, it could skew results since driving patterns might vary by region.

PACE: Analyze & Construct Stages

● How do descriptive statistics help us understand data?

> They provide a summary of key metrics, making it easier to quickly compare different variables.

● How did calculating descriptive statistics help us make sense of our data?

> It allowed for an initial comparison between iPhone and Android users' average drives, highlighting potential differences for further testing.

- What is the difference between the null hypothesis and the alternative hypothesis in hypothesis testing?

> The null hypothesis is a statement that is assumed to be true unless there is convincing evidence to the contrary. The alternative hypothesis is a statement that contradicts the null hypothesis and is accepted as true only if there is convincing evidence of it. In our case, the null hypothesis assumes no difference between groups, while the alternative hypothesis assumes a difference exists.

- How did we establish our null hypothesis and alternative hypothesis?

> We defined them based on whether or not there is a difference in the average number of drives between iPhone and Android users.

- What does the hypothesis test reveal about our hypotheses and what conclusions can be drawn?

> The test showed no statistically significant difference between the average number of drives of iPhone and Android users, so the null hypothesis was not rejected.

PACE: Execute Stage

- What key findings did we uncover from our A/B test?

> Users of iPhone have a slightly higher average number of drives, but this difference is not statistically significant.

- What recommendations would we provide to the organization based on our findings?

> We should conduct further hypothesis tests on other factors influencing user behavior and consider temporary adjustments to marketing and the app interface to gather more data for understanding churn patterns.