

Automatidata 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

Automatidata is consulting for the New York City Taxi and Limousine Commission (TLC). New York City TLC is an agency responsible for licensing and regulating New York City's taxi cabs and for-hire vehicles. The agency has partnered with Automatidata to develop a regression model that helps estimate taxi fares before the ride, based on data that TLC has gathered. For this stage of the project, we identify four main tasks that are presented in the following visual.





Considerations



PACE: Plan Stage

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

The goal is to explore the relationship between payment methods and total fare amounts in an effort to identify ways to increase revenue for taxi drivers.

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

The question we are trying to answer is whether there is a difference in fare amounts between customers who use credit cards and those who use cash.

- What is the significance of random sampling?

Random sampling ensures that any differences between groups can be attributed to the payment method rather than other factors, allowing for valid causal conclusions.

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

If more frequent travelers prefer credit cards, non-random sampling could falsely attribute their higher spending to the payment method rather than their travel frequency.



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 a direct comparison of average fare amounts between credit card and cash users.

- 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?

Based on the goal to test if there is a difference in fare amounts between payment methods, we set up the null hypothesis as no difference and the alternative as a difference.

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

The hypothesis test showed a significant difference in fare amounts, leading to the rejection of the null hypothesis and suggesting that credit card users tend to pay higher fares.



PACE: Execute Stage

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

Credit card users pay higher average fare amounts compared to cash users, indicating a potential revenue boost through encouraging credit card payments.

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

Credit card payments should be encouraged through signage in cabs and instructing drivers to promote this payment method verbally to passengers.