Automated A / B Testing and Measurement Framework

Achintya Acharya, Zainab Aljaroudi, Padma Dwivedi, Akshay Jayan, Amit Zutshi, Matthew A. Lanham
acharya9@purdue.edu; zaljarou@purdue.edu; pdwivedi@purdue.edu; jayan@purdue.edu; zutshia@purdue.edu; lanhamm@purdue.edu

PURDUE UNIVERSITY | Krannert School of Management

ABSTRACT

A parametrized accelerator to significantly reduce time to market for data scientists while performing A/B and A/A testing under varying circumstances using Python programming.

INTRODUCTION

Given the need to understand whether an intervention succeeded in achieving the stated goals and whether the same can be ascertained with statistical significance, A/B tests need to be run on pre- and post-event data set(s) to frame the conclusions and decide on similar or dissimilar interventions in future for businesses that do not necessarily operate in the online space. This calls for the need for an accelerator that can cluster stores to determine appropriate test and control stores, perform forecasting for scenarios where intervention is global, and measure lift/drop numbers for the KPIs across the test vs. control and forecast vs. actual data sets.

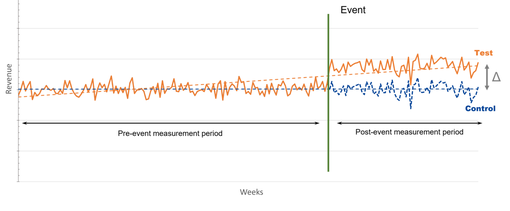

Fig 1. Representative Test Output

STUDY DESIGN

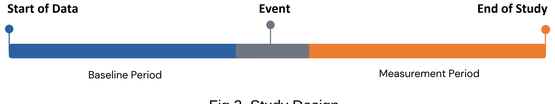

Fig 2. Study Design

- The data in the baseline period is used for control selection for local use cases and for cross-validation based model selection for global use cases.
- In the measurement period, the impact of intervention between two sets of data is measured using t-tests.

Test Types:

- Local Case (A/B Testing)
  - Examples: Store remodeling, local promotions
- Global Case (Forecast based A/B Testing)
  - Examples: Mass advertisement, adverse weather impact

METHODOLOGY

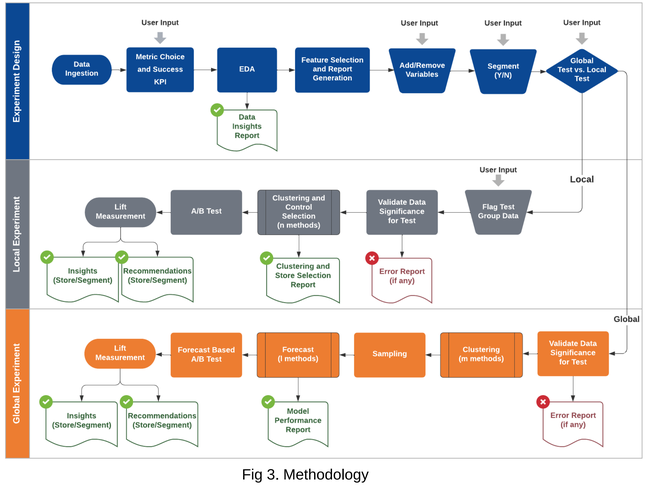

Fig 3. Methodology

STATISTICAL RESULTS

The clustering output from the tool showcases the store clusters, highlighted by difference within multiple attributes. Cluster #1 has Gas Station stores which have highest fuel revenue. Similarly, cluster #2 is High Performance stores and cluster #3 is Rural Super stores.

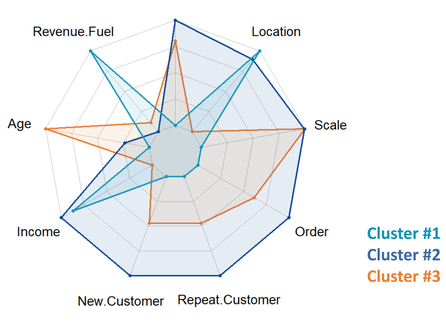

Fig 4. Accelerator Cluster Output

The optimum control stores have been selected for each test store based on proximity. Euclidean distance measurements between each test store and all control stores within the same cluster was done for the same.

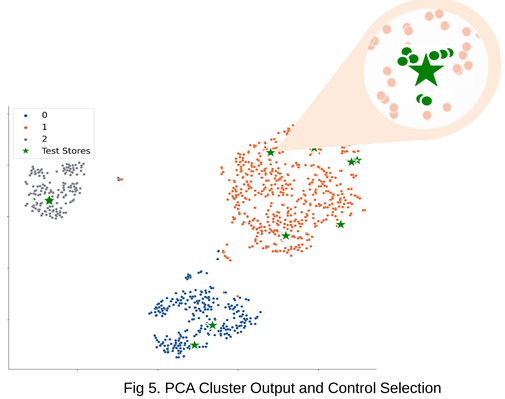

Fig 5. PCA Cluster Output and Control Selection

In the figure below, comparing actual vs. forecasted values in the measurement period shows us that the event leads to actual values being typically higher than the forecasts. This implies the positive impact of the event on the KPI i.e., Average Order Value (AOV).

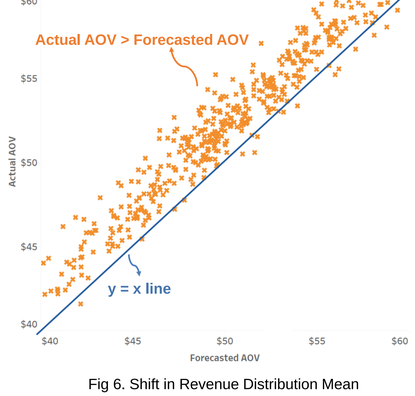

Fig 6. Shift in Revenue Distribution Mean

BUSINESS USE CASES

- An intuitive user interface driven tool will help data scientists as well as business users to perform A/B testing seamlessly
- Caters to all possible traditional scenarios: impact of adverse weather, introduction of discounts/promotions, store remodeling, advertisement efficacy et al.
- Delivers means to perform quick intervention impact analysis to brick-and-mortar organizations
- Provides detailed reports regarding EDA, model performance, and final impact to non-IT users

EXPECTED IMPACT

Reduce time and material costs

Minimize human error

Scalable reporting framework

Enable rapid iterations

Modular framework allows easy enhancements

Minimal onboarding requirement

CONCLUSIONS

- Organizations that run a multitude of A/B tests can gain immense value from automated frameworks
- Experiments would need to be designed keeping overarching global and local design methodology in mind
- Choosing right KPIs to test would be critical to success
- Anticipating business scenarios is difficult; parameterizing is the key
- Accelerator maintenance through incorporation of additional clustering and forecasting methods, as situations may demand, would be needed

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew A. Lanham and our industry partner for this opportunity. Their guidance and support throughout this project has been profoundly rewarding.