Automated A/B Testing and Measurement Framework

INFORMS. DATA MINING

Tredence — Connect the Dots

Harshul Garg, Manya Dadarya, Phalguni Vatsa, Usama Ather, Matthew A. Lanham

Purdue University, Krannert School of Management

garg152@purdue.edu; mdadarya@purdue.edu; pvatsa@purdue.edu; uather@purdue.edu; lanhamm@purdue.edu

ABSTRACT

To create comparative solutions for the legacy brick and mortar business for measuring the effectiveness of their store innovation efforts an 'Accelerator' was developed that data scientists and statical analysts can use to analyze a broad set of use cases, reducing the time taken to design and run A/B Tests.

INTRODUCTION

Big companies like Amazon, Google and Facebook perform up to 10,000 A/B test experiments every year. These tests can take up a major chunk of time and investment from the companies end due to various operational challenges mentioned in the graph below. The identification of similar test and control groups is essential to reduce possible time frame of designing an experiment without compromising on quality the companies prefer automation. Our tool allows businesses to create multiple versions of an experiment and test them against target audience.

Source : Harvard Business Review

10,000 A/B experiments performed by Google, Amazon, Facebook and Booking.com yearly

1,000 A/B tests ran by Microsoft per month

20% Google and Bing's controlled experiments that yield positive results

Figure 1 : A/B Test Relevance

Source : Statista

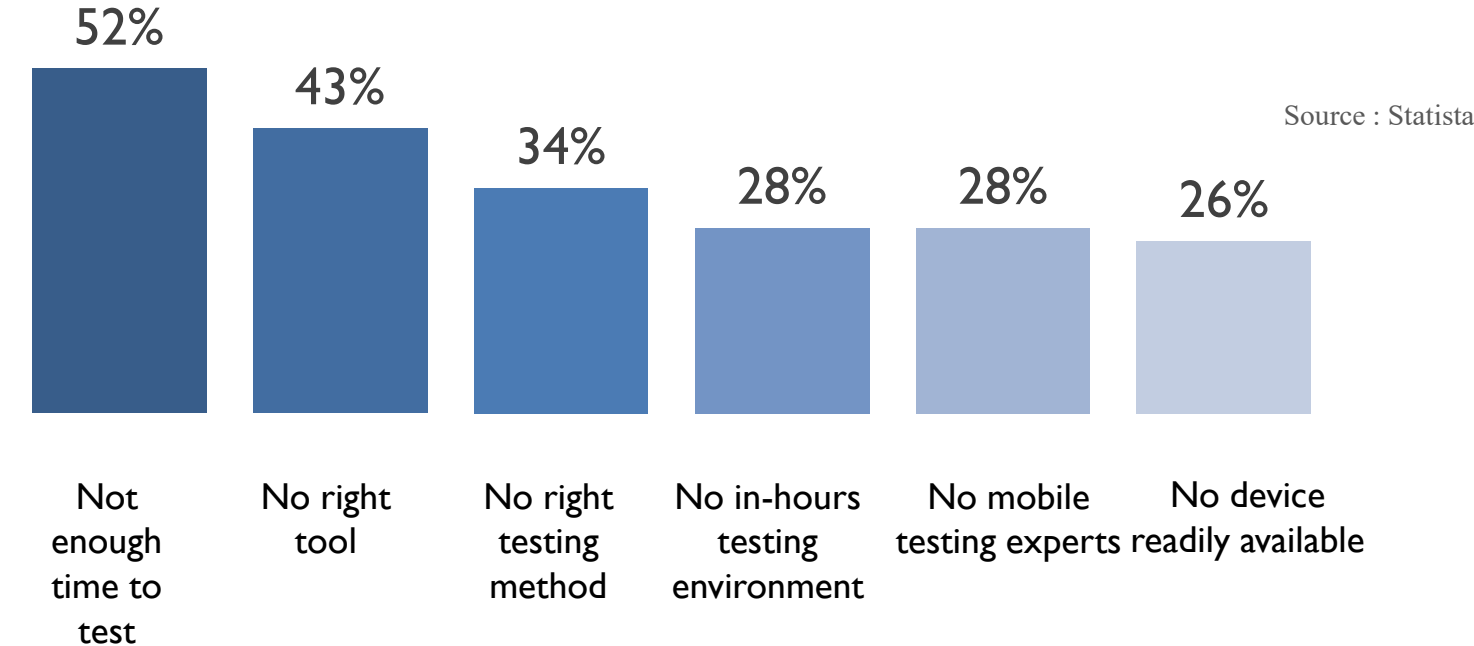| 52%	43%	34%	28%	28%	26%
Not enough time to test	No right tool	No right testing method	No in-hours testing environment	No mobile testing experts readily available	No device

Figure 2 : A/B Testing Challenges

LITERATURE REVIEW

There are not many literature references to automate A/B testing. Studies suggest an initial framework to automate the process for a wide range of industries and how to generalize it while considering diverse covariates presented with every problem.

1 Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead
By Maximilian Kasy, 2016
Proposes three optimal procedures instead of randomization- Bayes optimal estimation, Difference in means and Discrete optimization

2 A Complexity-Invariant Distance Measure for Time Series
By Gustavo E.A.P.A. Batista, Xiaoyue Wang and Eamonn J. Keogh, 2011
Studies importance of Complexity-Invariant Distance (CID) measure for time series over other popular classification algorithms

3 Towards Automated A/B Testing
By Giordano Tamburrelli and Alessandro Margara, 2014
Addresses the problems with traditional A-B testing methods which are time-consuming, error-prone and costly manual activity.

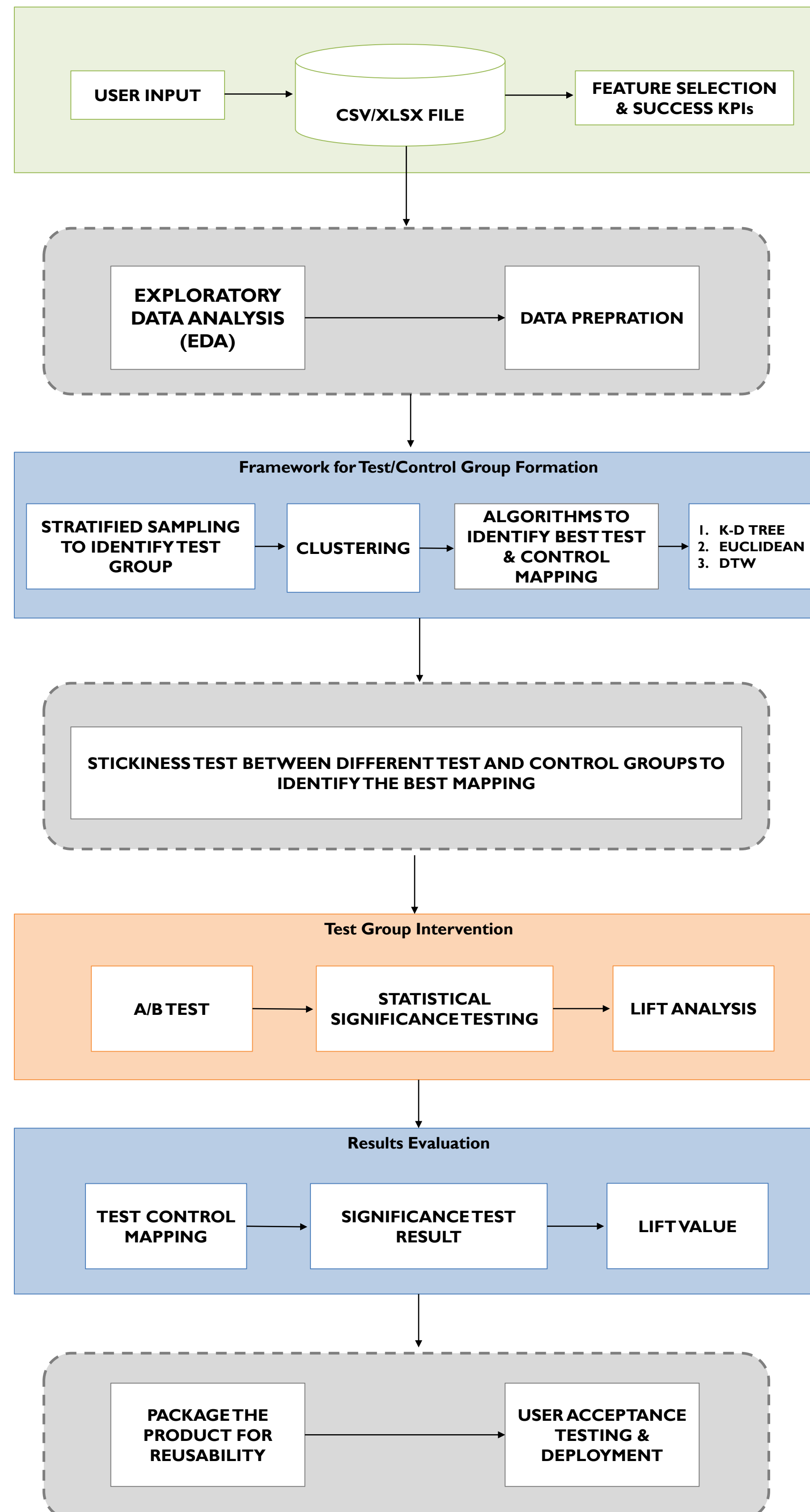Figure 3 : Literature Review

METHODOLOGY

USER INPUT → CSV/XLSX FILE → FEATURE SELECTION & SUCCESS KPIs

EXPLORATORY DATA ANALYSIS (EDA) → DATA PREPRATION

Framework for Test/Control Group Formation

STRATIFIED SAMPLING TO IDENTIFY TEST GROUP → CLUSTERING → ALGORITHMS TO IDENTIFY BEST TEST & CONTROL MAPPING →
1. K-D TREE
2. EUCLIDEAN
3. DTW

STICKINESS TEST BETWEEN DIFFERENT TEST AND CONTROL GROUPS TO IDENTIFY THE BEST MAPPING

Test Group Intervention

A/B TEST → STATISTICAL SIGNIFICANCE TESTING → LIFT ANALYSIS

Results Evaluation

TEST CONTROL MAPPING → SIGNIFICANCE TEST RESULT → LIFT VALUE

PACKAGE THE PRODUCT FOR REUSABILITY → USER ACCEPTANCE TESTING & DEPLOYMENT

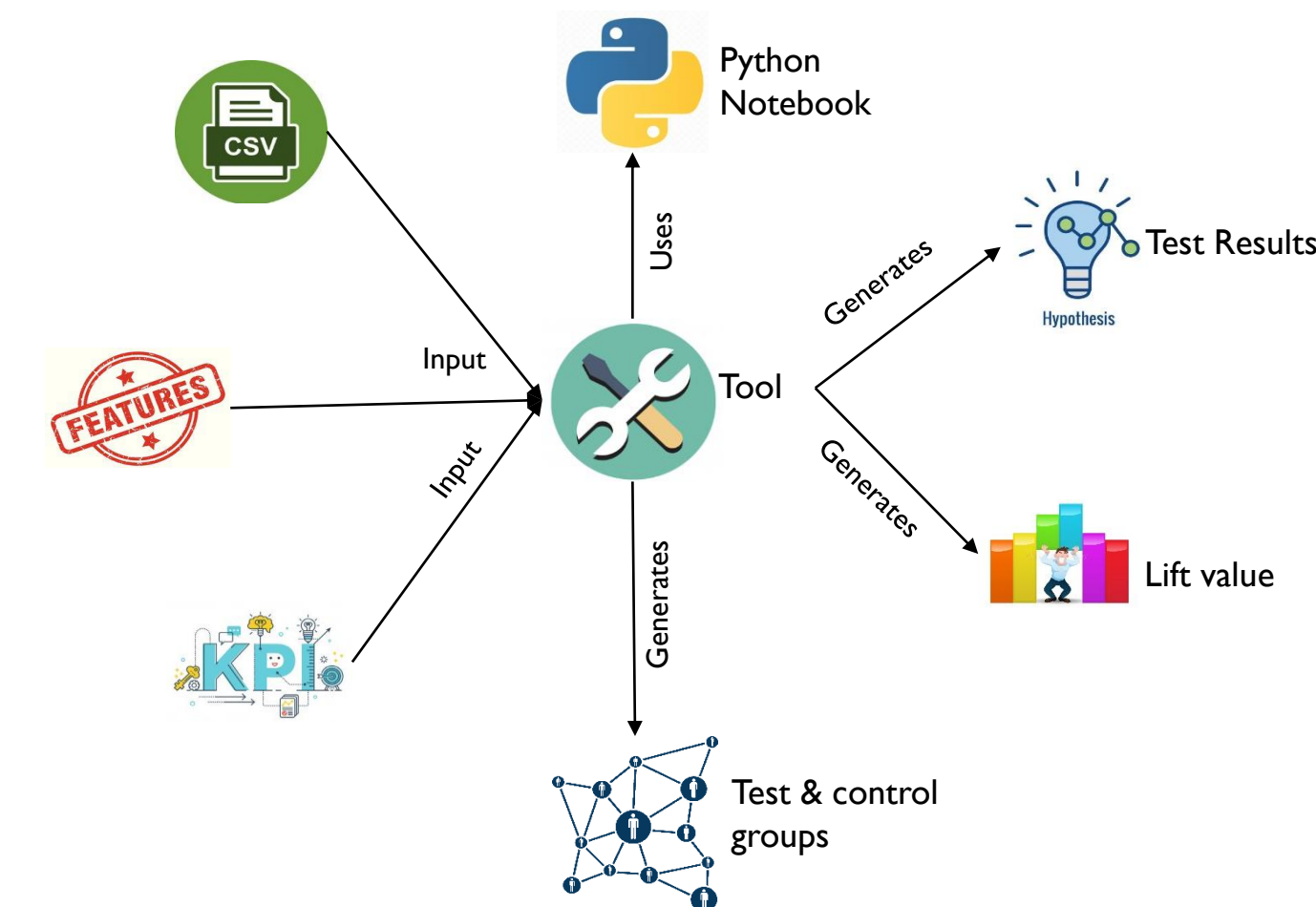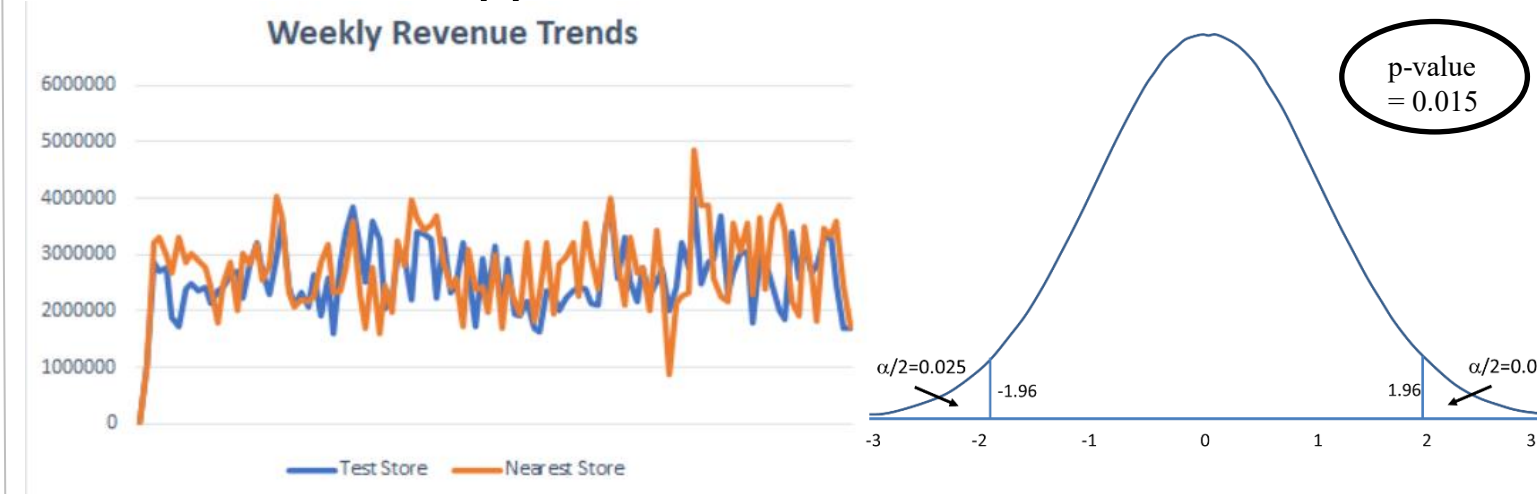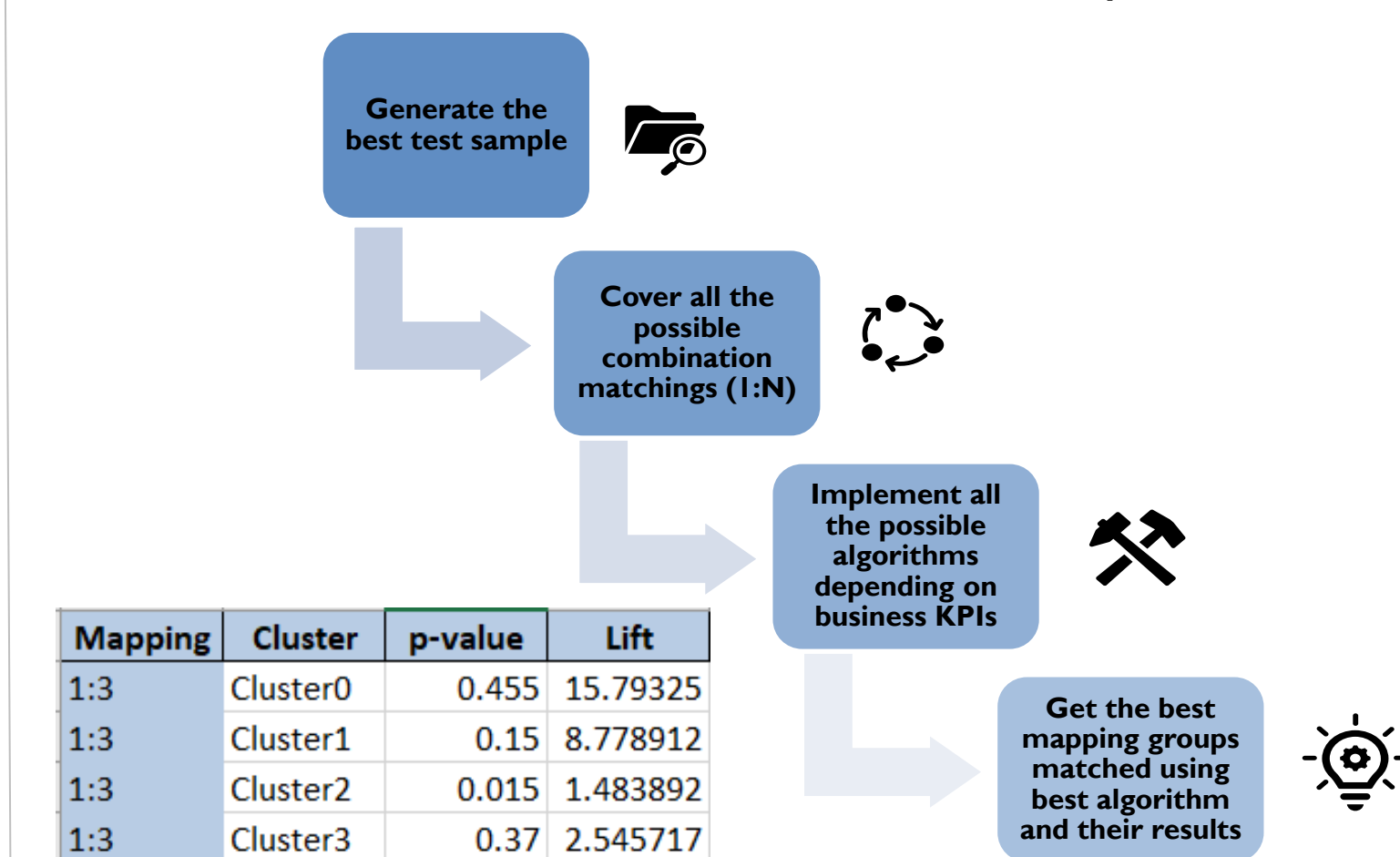Figure 4 : Framework

PRODUCT

Figure 5 : Product

STATISTICAL RESULTS

Different algorithms are incorporated to similar patterns, depending on the varied use of data; To find the neighboring stores we do different significance tests to find the best mapped stores

Weekly Revenue Trends

p-value = 0.015

$\alpha/2=0.025$ -1.96   1.96 $\alpha/2=0.025$

Accelerator results would help

Generate the best test sample → Cover all the possible combination matchings (1:N) → Implement all the possible algorithms depending on business KPIs → Get the best mapping groups matched using best algorithm and their results

Mapping	Cluster	p-value	Lift
1:3	Cluster0	0.455	15.79325
1:3	Cluster1	0.15	8.778912
1:3	Cluster2	0.015	1.483892
1:3	Cluster3	0.37	2.545717

Measure geolocation to identify maximum potential

Measure shopper behavior towards varied events

EXPECTED BUSINESS IMPACT

Traditional A/B Testing Approach: Data Entry → Manual Data Processing → Algorithms Testing & Evaluation → A/B Testing & Results

Automated A/B Testing Accelerator: Data Entry → A/B Testing Accelerator → Results
- 60% quicker
- 30% more accurate
- 60-120K per year savings
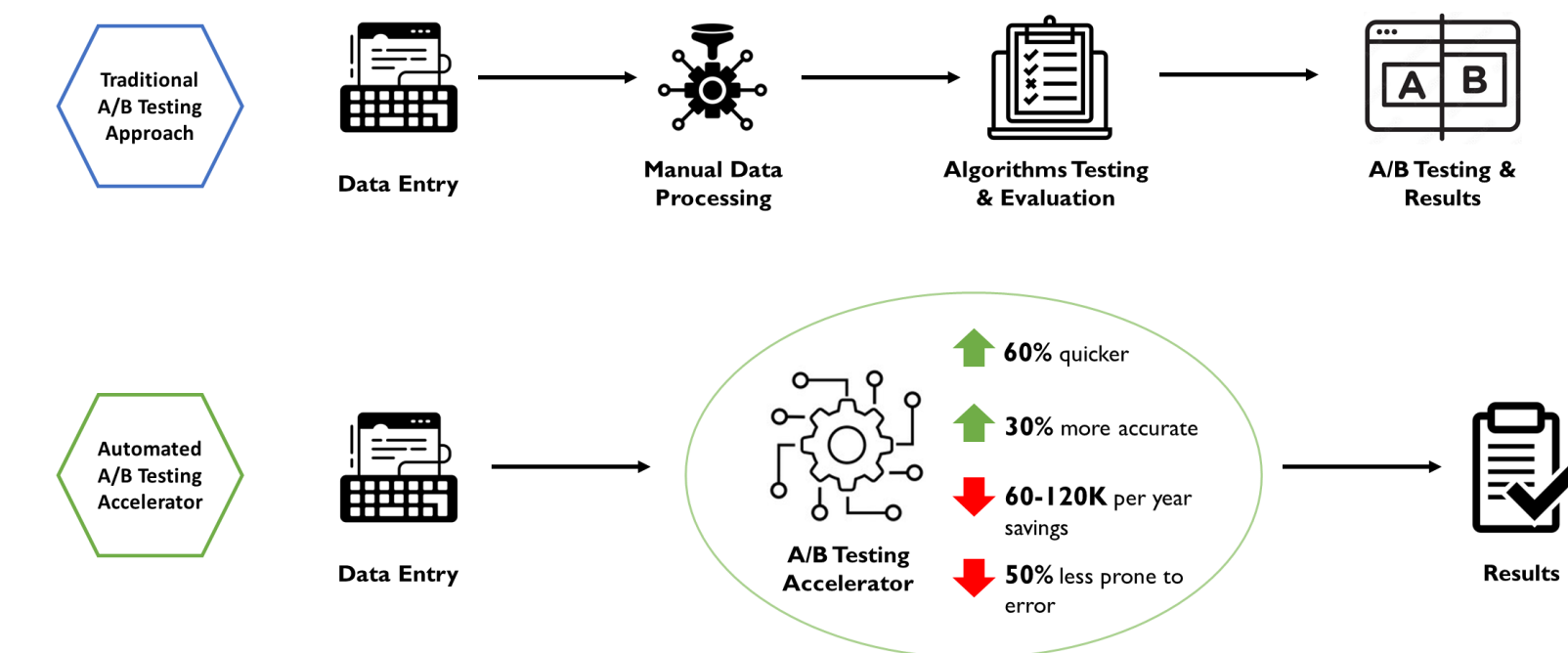- 50% less prone to error

Figure 6 : Business Impact

CONCLUSION

- Rapid Test Designing
- Automate Data Processing
- Design Multiple Tests in Parallel
- Elimination of Human Error
- Testing all Possible Algorithms

- Monetary Benefit
- Reduced Labor Cost
- Quality Assurance

FUTURE SCOPE

- Ingestion Data Pipeline
- Error Handling and User Alerts
- Compliance as per industry standards
- BI Dashboard

ACKNOWLEDGEMENT