Empowering Supplier Strategies
Network Clustering for Smarter A/B Testing

Wayfair GST Analytics Ads Team: Stefan Poninghaus, Cameron Madge, Siddarth Kumar S
MIT Advisors: Prof. Mohammad Fazel Zarandi

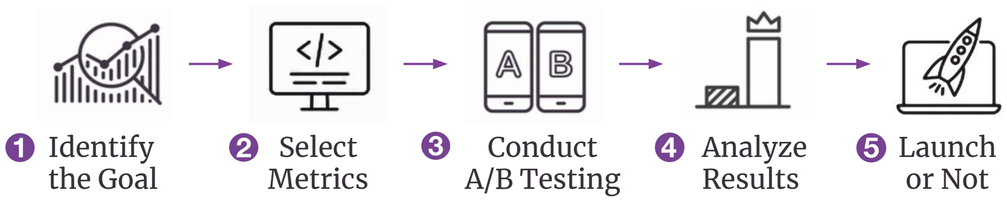

Crystal Li

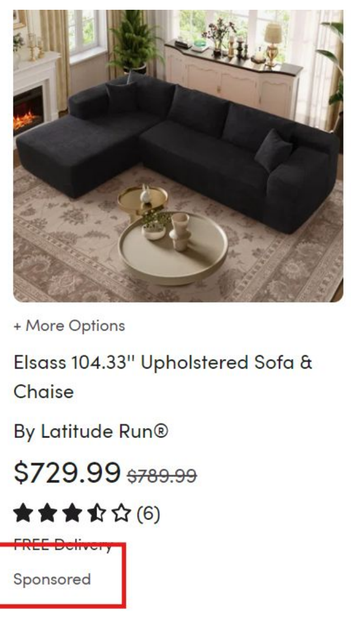

Tim Zhou

Problem Statement

Context: Wayfair is an e-commerce company that sells the world's largest online selection of home goods. We are focusing on suppliers who compete with each other in **Wayfair Sponsored Product (WSP) auctions** to win sponsored slots.

A/B Testing Framework





Problem

- Suppliers directly competing with each other are not truly independent, which violates underlying assumptions of independent samples within A/B testing
- Network effect: interactions and interventions within the treatment group can unintentionally influence the control group

Goal

- Build a **network clustering model** to identify independent groups of suppliers
- Use robust **experimentation** to understand and measure the impact of ML driven recommendation models

Data & Scope

Scope: Wayfair Sponsored Product | Worldwide | 2022 - 2024

Google Cloud Platform

Supplier Auction Data

SupplierID, AuctionID, Date, Auction Participants, BidAmount, Slots Fill Rate etc.

Supplier Feature Data

SupplierID, Marketing Category, Spend on Auctions, IsBranded, WSP Attribution, etc.

~ 2 Trillion
Rows of Auction Data on a Daily Basis

~ 3600
Suppliers Participating

Methodology

Graph Constructions

Purpose

- Build a network structure of suppliers to perform further analysis

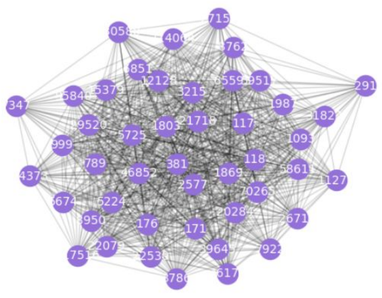
Co-occurrence Score

- Use **Jaccard Co-occurrence Score** (0-1) to measure how often two suppliers participate in the same auctions

$$\text{Jaccard Index}_{A,B} = \frac{|A \cap B|}{|A \cup B|}$$

Matrix & Graph Construction

- Adjacency matrix and distance matrix based on pairwise Co-occurrence Score
- Network Graph based on Co-occurrence Score as edge weight



Network Clustering

Purpose

- Cluster suppliers into groups based on their connections and competitions

01 Louvain Algorithm

Maximizes modularity in graphs (edges density inside communities with respect to edges outside)

02 Hierarchical Agglomerative Clustering

Bottom-up approach that starts with every data point as a single cluster, and merge the closest pair of clusters iteratively

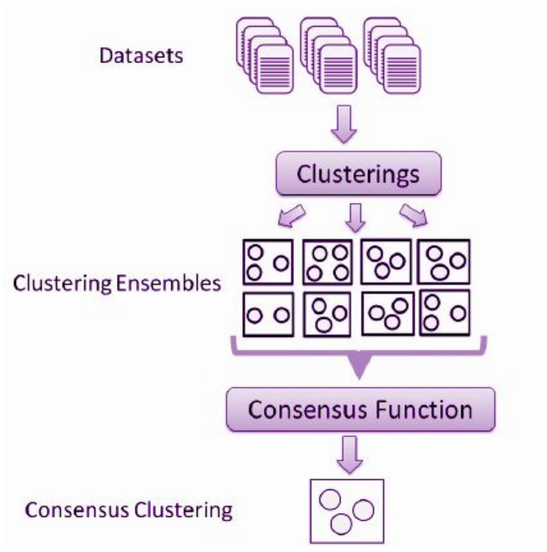
03 K-Medoids Algorithm

Partitions data by selecting data points as centers, minimizing the sum of dissimilarities

Consensus Clustering

Purpose

- **Ensemble learning** method to aggregate multiple clustering solutions to achieve a more reliable and stable clustering result



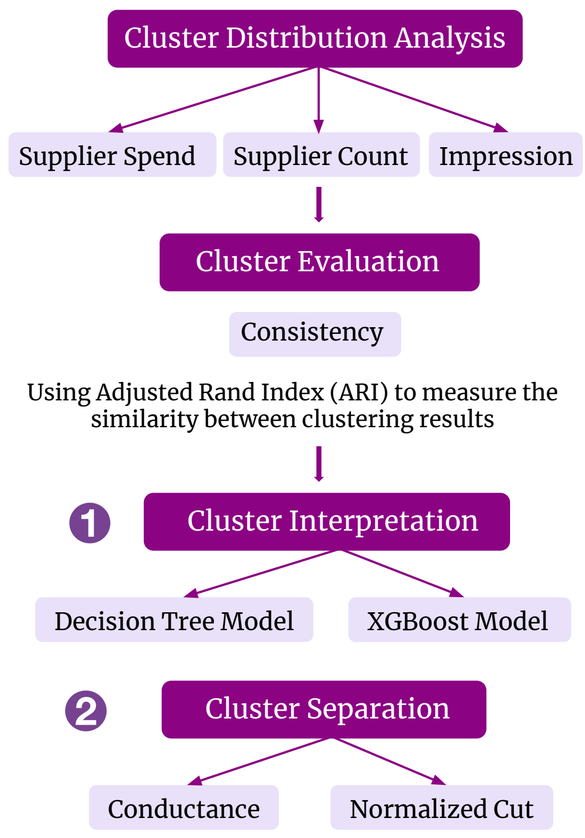
How does it work?

- Captures overlaps across clusters
- Determines the most cohesive solution
- Provides a final aggregated result

Cluster Analysis

Purpose

- Provide analysis, evaluation metrics, and interpretation for cross-functional teams and stakeholders



Results

We identified ideal clusters by considering the following metrics:

Consistency	Stability of clusters over time
Balance	Clusters should not have extreme size differences
Interpretability	Insights from the characteristics of each cluster

Identify cluster characteristics to provide context for product managers and other stakeholders.

Decision Tree			
Clustering Result Used	Number of Clusters	Testing Accuracy	F1-Score
Consensus Clustering	4	0.82	0.82
	6	0.81	0.80
HAC	6	0.72	0.72
Louvain	11	0.70	0.70

Different well-performed clustering results with various cluster numbers provide greater flexibility for further experimental implementations.

+ 15%
Interpretability than single clustering method

0.6
Cluster Similarity score (ARI)

3
Well-Performed Clustering Results

Impact

Alternative Experimentation Framework

Wayfair can apply treatments to suppliers and measure performances and outcomes at the customer level.

Flexibility to Support Multiple Use Cases

- **Provide Insights via Reporting**
- **Offer Recommendations**
- **Automate ML systems**

Next Steps

- Choose **cluster balancing criteria** based on what we are measuring
- Use **synthetic control methods** (e.g., CausalImpact from Google) to estimate the causal effect of an intervention on a time series; comparing treated groups with synthetically constructed control groups by weighting clusters