Assortment Planning Strategy

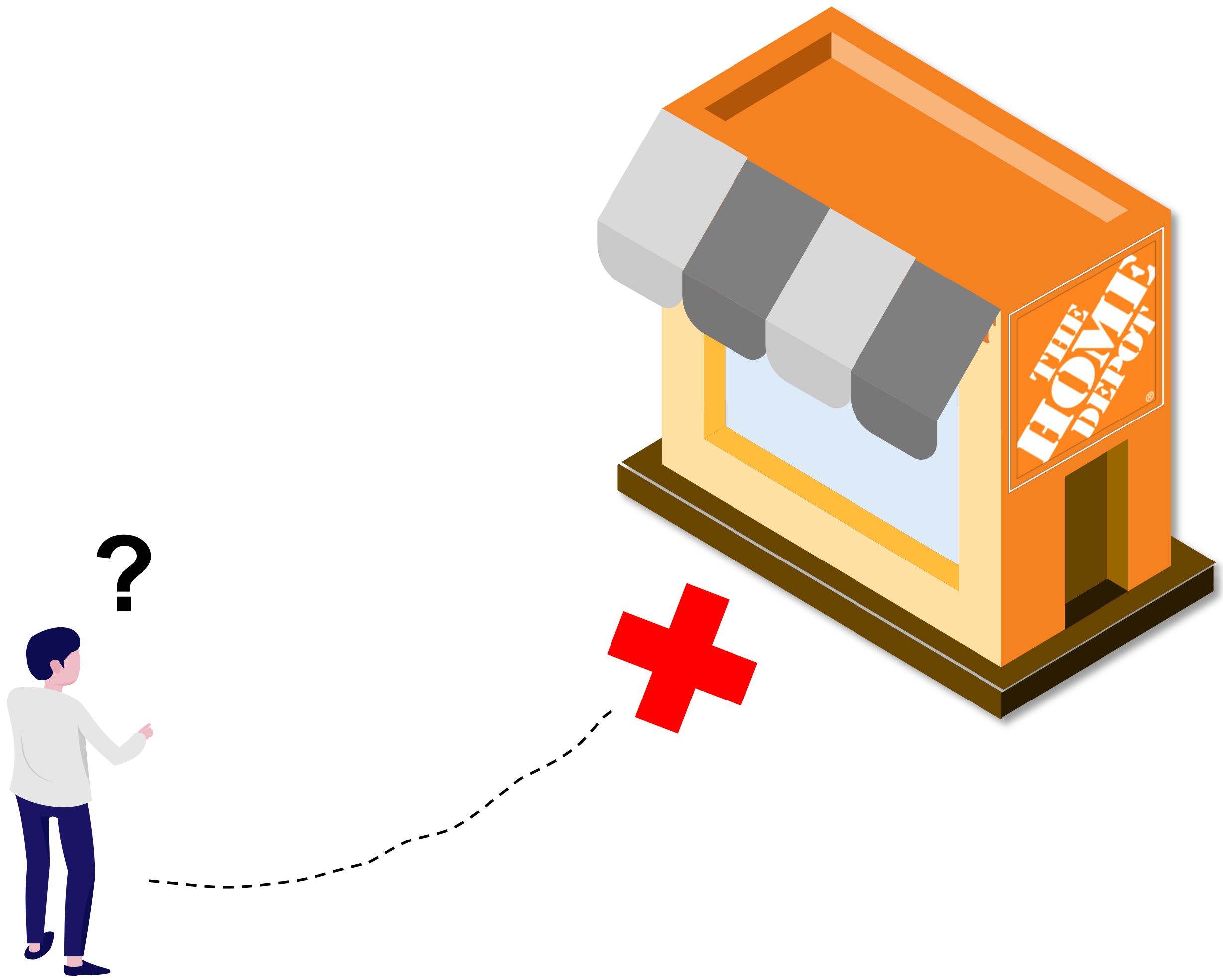
Final Presentation

Wednesday, November 14, 2018

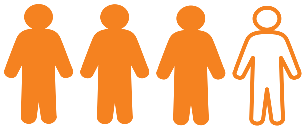
Team 1:
Hannah Chen, Joyce Kinsey, Sarah Lee, David Morrison
Matt Munns, Vince Nguyen, Suraj Sehgal, and Alina Wendt

Advisor: Dr. Alexander Shapiro
Client Contact: Stephanie Kalman

A Consumer's Dilemma



A Consumer's Dilemma



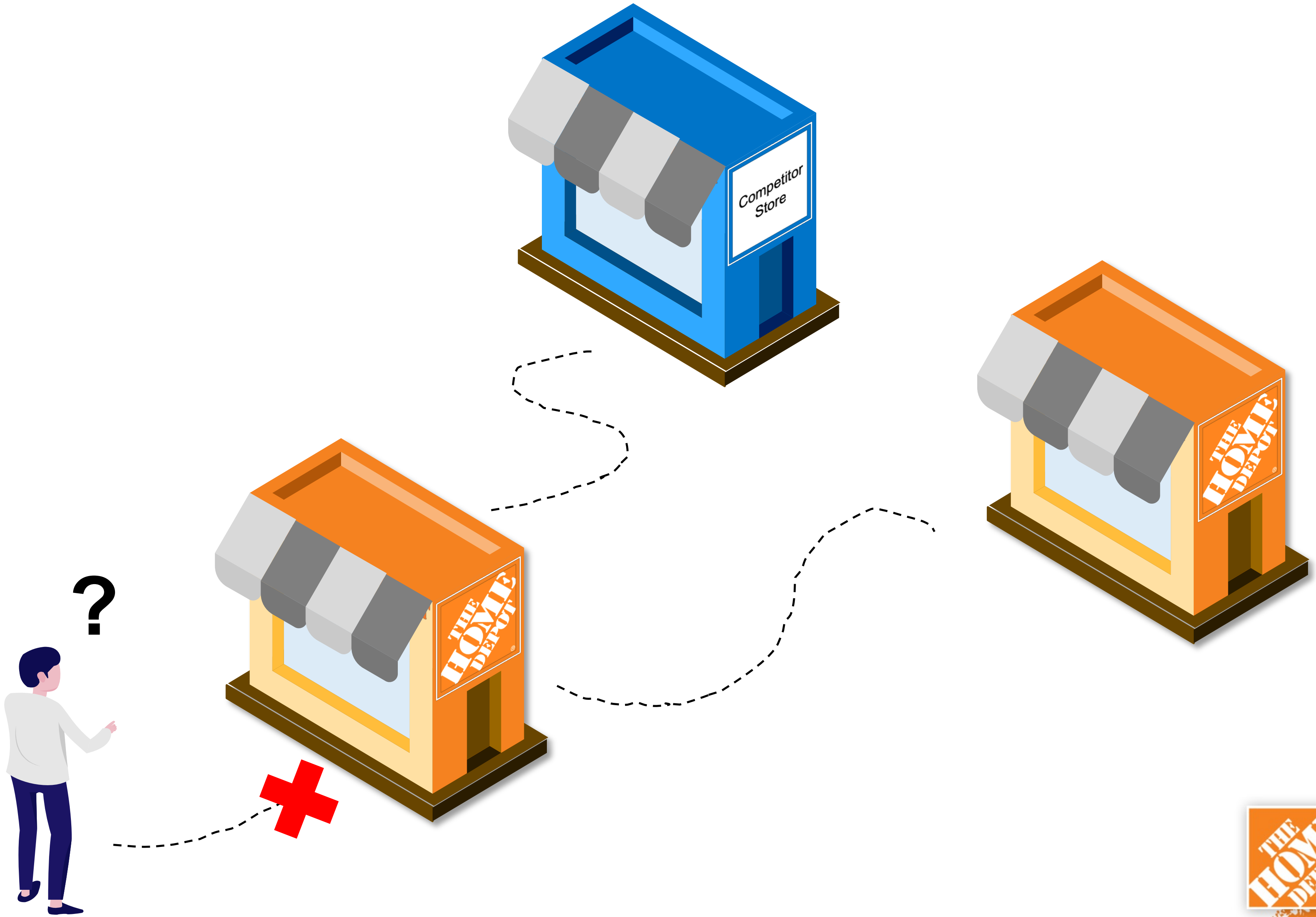
3 out of 4 consumers experience some **difficulty** while shopping for a product*.

14%

of consumers **could not find** the product they wanted.

Base: Consumers (n=101)
*Subject to specific category

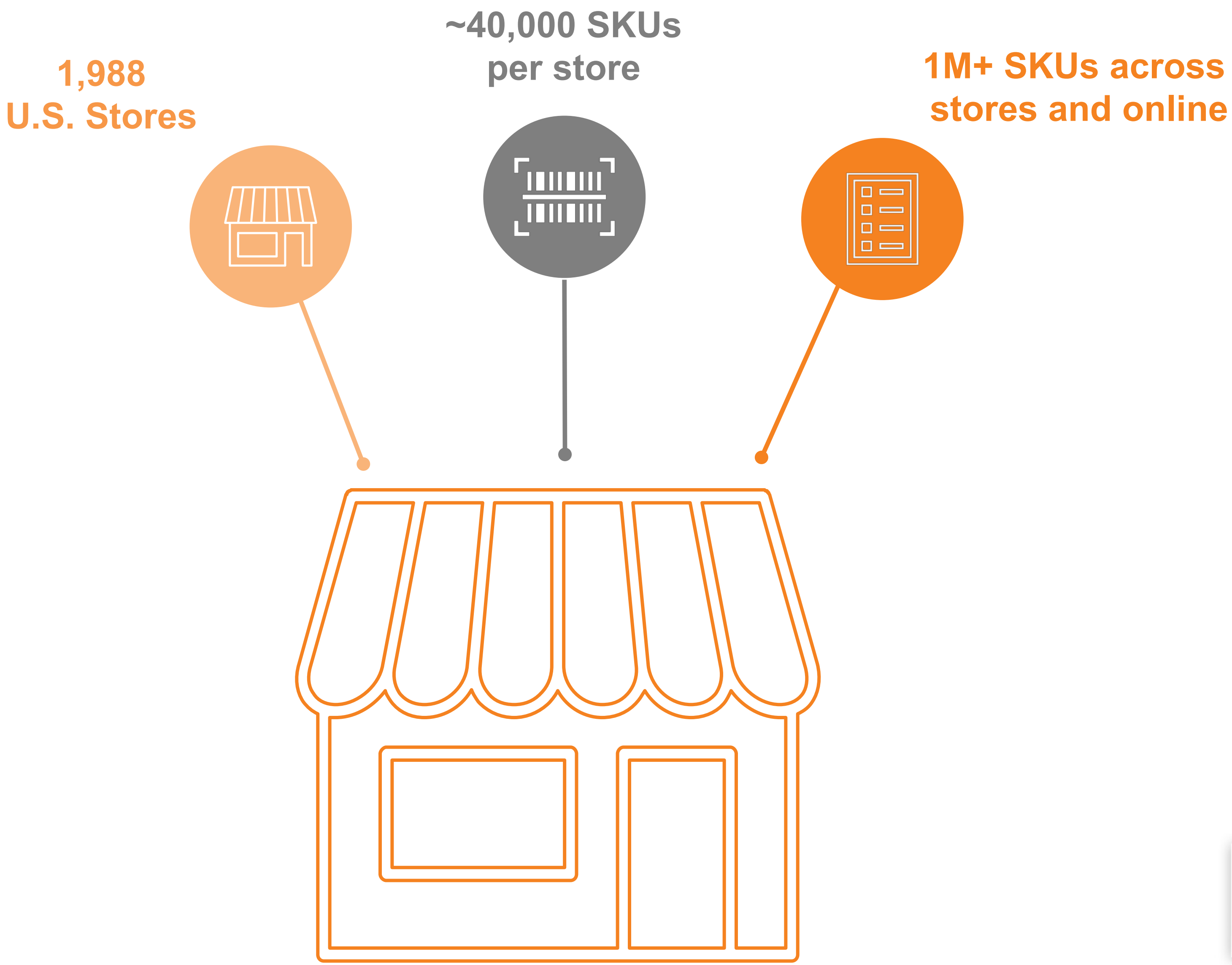
The Consumer's Dilemma



Agenda

- Client & Project Overview
- Initial Approach, Challenges, & Final Approach
- Triple Clustering Approach
- Deliverables & Impact

Client Overview



Product Line Review (PLR)

Purpose

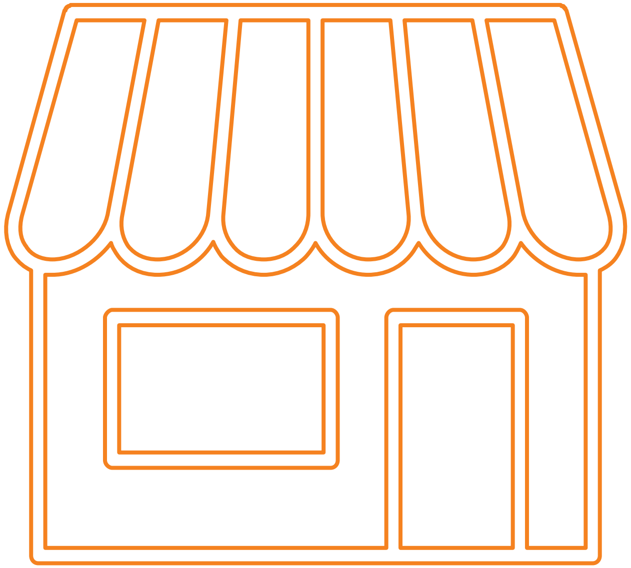
- Right product
- Right store
- Right price

Focus

Assortment Planning serves as the starting point for PLR decisions

Methods

- Store Clustering
- SKU Assignment

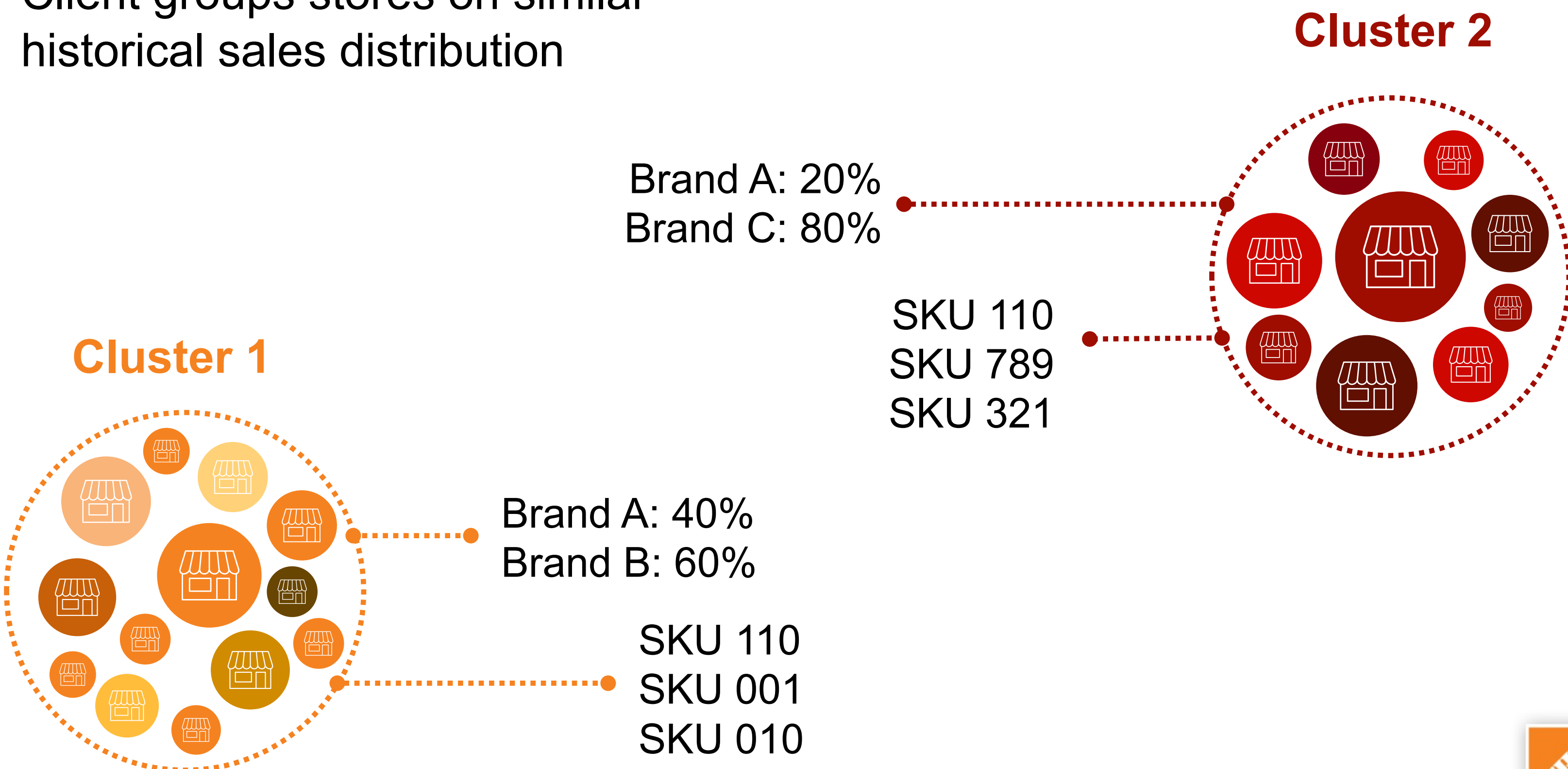


Clustering & SKU Assignment

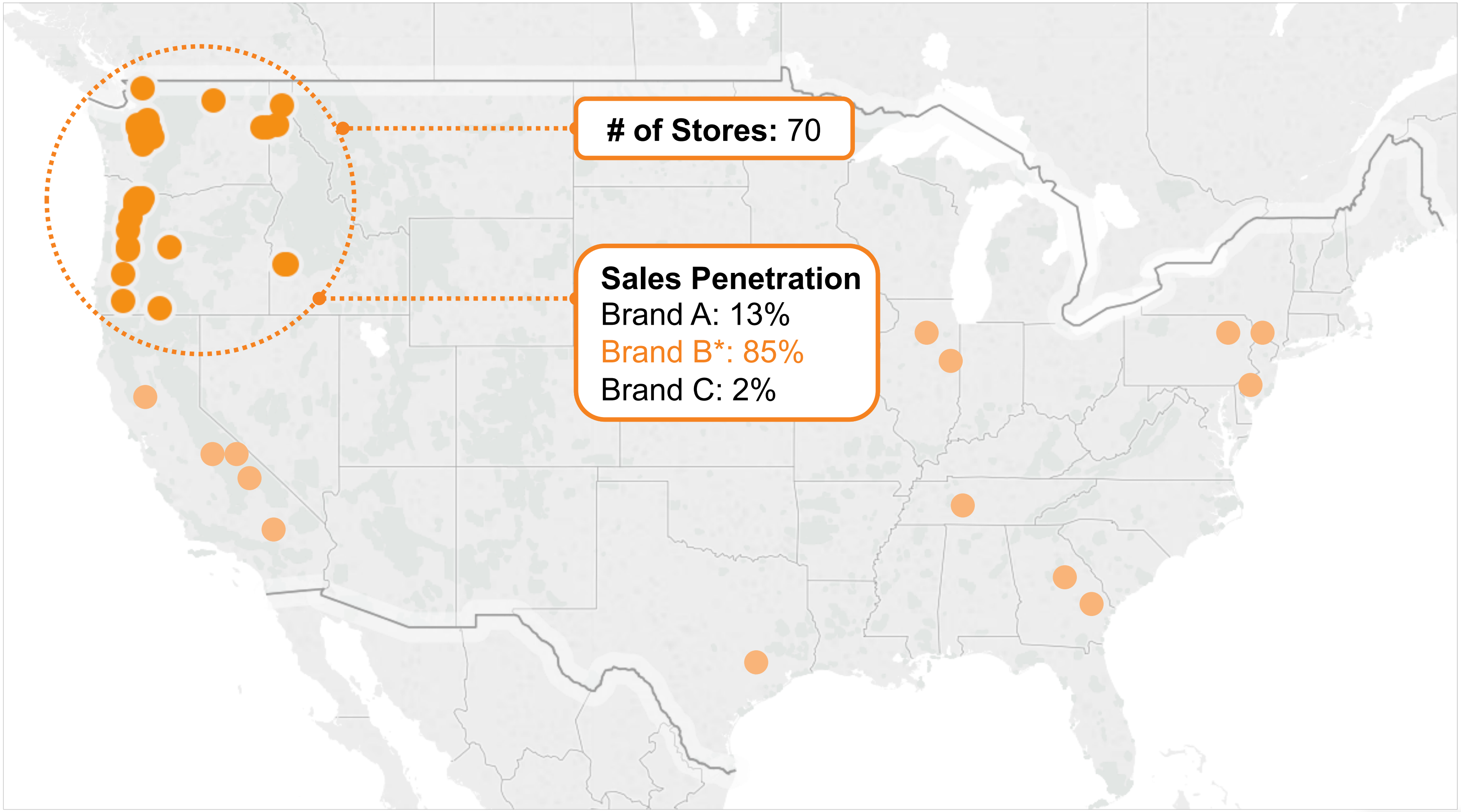
What is Clustering?
Grouping stores based on their similarities

Client groups stores on similar historical sales distribution

What is SKU Assignment?
Determining which products are placed into which clusters



Current Clustering and Limitations



Client's clustering approach groups stores based on similar historical sales
* Brand B was only piloted in the 70 stores.

Room for Opportunity

Context Recap

Product Line Review

- Assortment Planning
  - Store Clustering
  - SKU Assignment

Limitation

- Clusters based only on historical sales
  - Creates bias
  - Results in missed sales opportunities

Opportunity

- Leverage data on store traits
  - Produce more holistic clusters
  - Introduce products to new stores

Objective

Remove historical sales bias from store clustering and SKU assignment to **increase projected sales** from the PLR process.

Agenda

- Client & Project Overview
- Initial Approach, Challenges, & Final Approach
- Triple Clustering Approach
- Deliverables & Impact

Approach

Objective: Remove historical sales bias from store clustering and SKU assignment to **increase projected sales** from the PLR process.

1

How are stores similar?

- Identify significant store traits that drive sales

2

How do we group similar stores?

- Use store traits to overcome historical sales bias
- Research and implement new clustering algorithm

3

How do we assign SKUs to those groups?

- Predict demand of unsold SKUs
- Assign SKUs based on these predictions

Initial Approach and Concerns

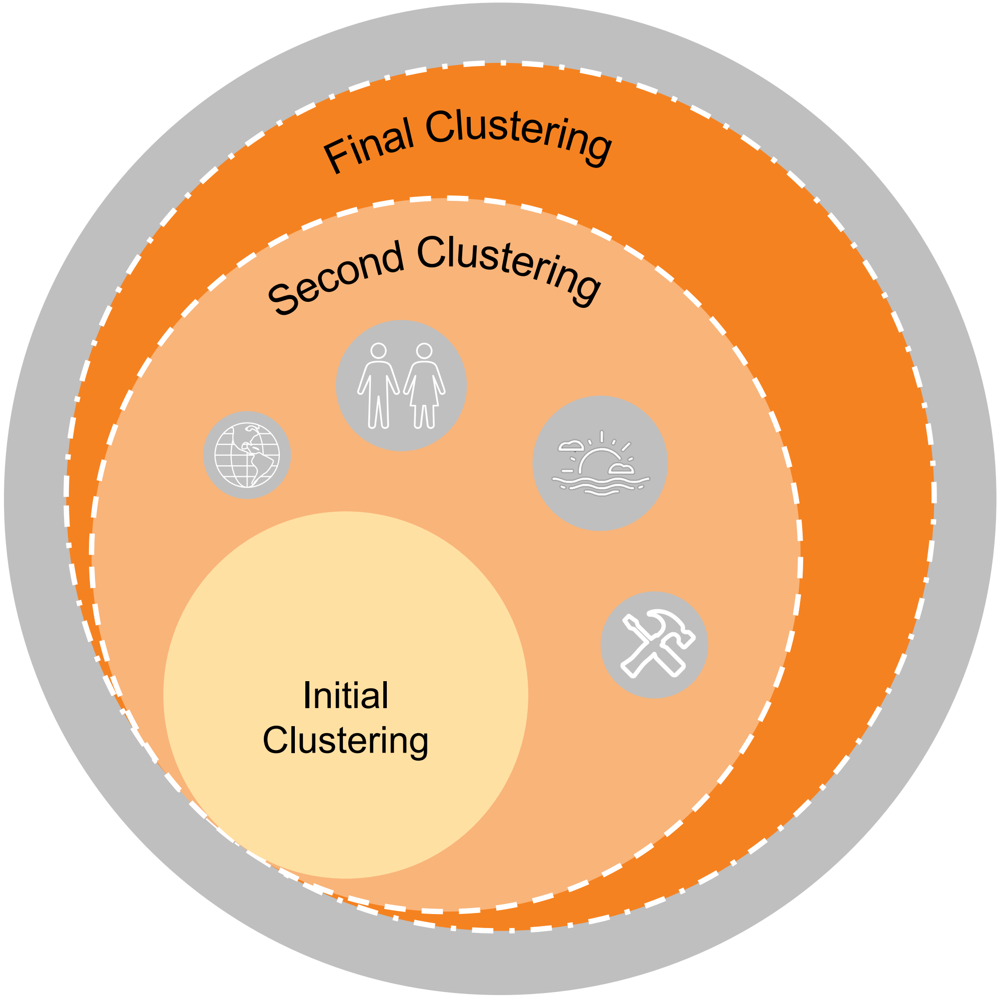
Approach	Concerns
1 Picked SKU with **largest variance** in sales and with greatest availability	**Inflated** the significance of one SKU's ability to predict an entire category
1 Performed **regression** on chosen SKU's sales across all stores to determine **significant store traits**	Regression is a **time intensive** method that not all managers are comfortable with
2 **Clustered** on significant store traits	**Store traits alone** were **not ideal** to use to predict similar stores
3 Selected **top 20 SKUs** for each cluster	**Did not consider** that some SKUs are grouped together

Concerns and Solutions

Concerns	Solutions
1 **Inflated** the significance of one SKU's ability to predict an entire category	Look at sales performance of SKUs that are **grouped together**
Regression is a **time intensive** method that not all managers are comfortable with	Evaluate store traits using **Chi-Square Test**
2 **Store traits alone** is **not ideal** to use to predict similar stores	**Cluster** on historical sales and store traits to predict sales
3 **Does not consider** that some SKUs are grouped together	Select **top-performing attribute groups**

Additional benefit:
category agnostic

Final Approach: Triple Clustering Overview



Initial Clustering

- Cluster on historical sales
- Current client approach

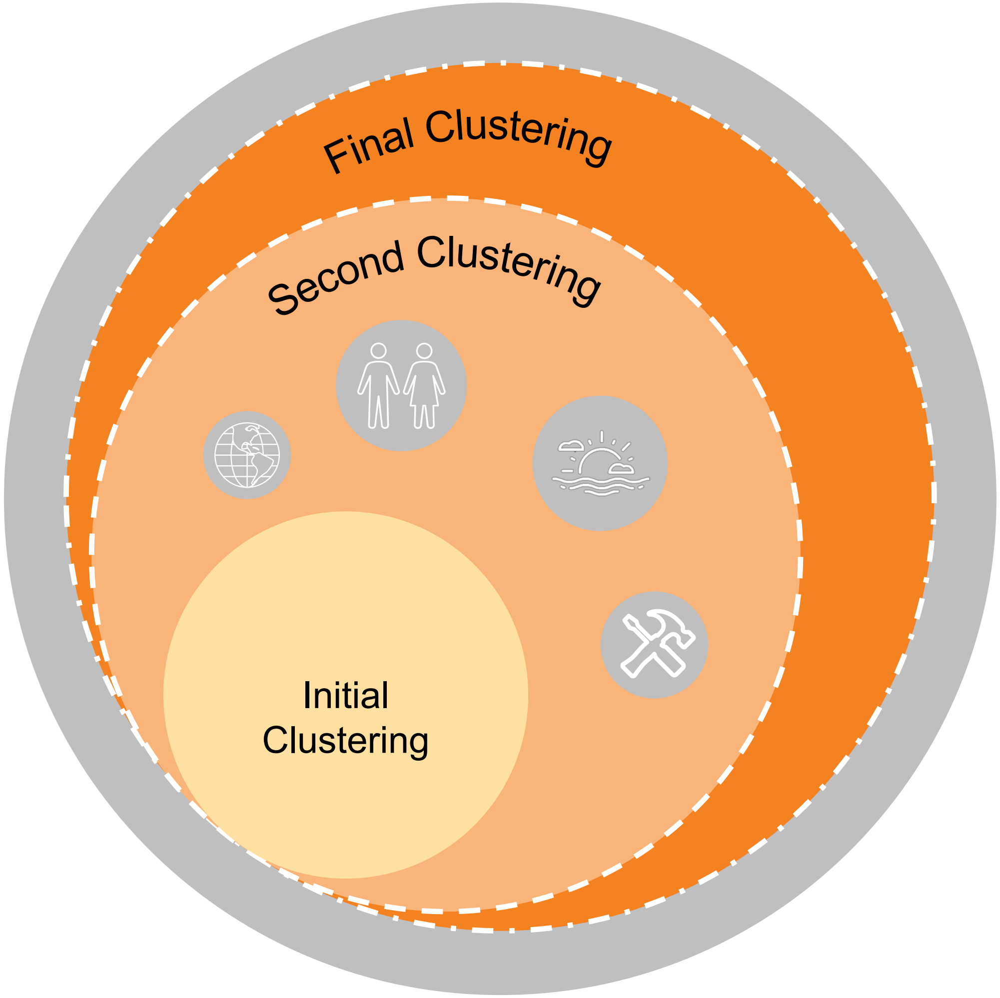
Second Clustering

- Identify significant store traits
- Re-cluster on those traits and historical sales

Final Clustering

- Calculate predicted sales for unsold SKUs
- Re-cluster on predicted and historical sales

Final Approach: Triple Clustering Overview



Our approach creates holistic clusters:

We consider both historical sales and store traits to **predict sales of unsold SKUs**

Incorporating these predictions **reduces historical sales bias**

Agenda

- Client & Project Overview
- Initial Approach, Challenges, & Final Approach
- Triple Clustering Approach
- Deliverables & Impact

Initial Clustering – Input Data

Sales File

SKU	Store	Sales	Units
101	1	$10	3
202	1	$15	4
101	2	$7	1
303	2	$8	1
123	2	$9	2
101	3	$11	3
...			

Attribute File

SKU	Brand	Color
101	A	White
202	B	Brown
303	A	Red
123	D	Yellow
505	E	Blue
321	C	Gray
…		

Create Attribute Groups

Initial Clustering – Create Attribute Groups

- For each store, APSW* sales are calculated for each attribute group (i.e. Brand-Color)

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	$20	$0	$80	$0	…
2	$78	$12	$0	$34	…
3	$41	$0	$145	$0	…

- According to the APSW sales, the sales distributions are calculated

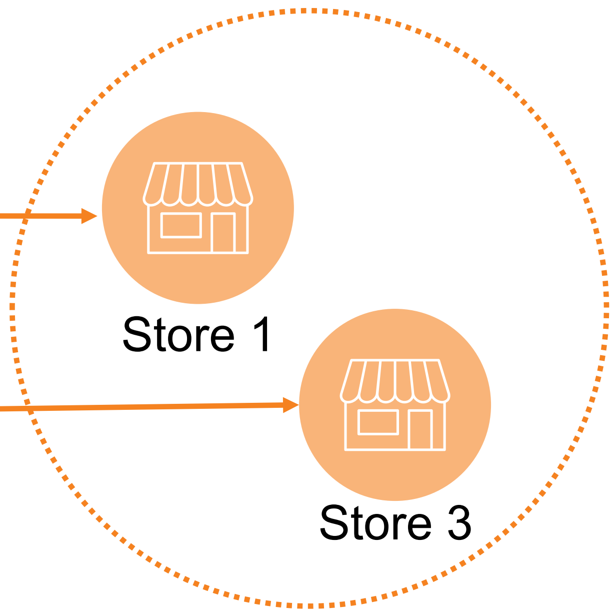
Store	A-White	A-Red	B-Brown	D-Yellow	…
1	20%	0%	80%	0%	…
2	63%	10%	0%	27%	…
3	22%	0%	78%	0%	…

* Average per store per week

Initial Clustering – K-Means on Attribute Groups

- Stores with similar sales distribution of the same attribute groups will be clustered together

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	20%	0%	80%	0%	…
2	63%	10%	0%	27%	…
3	22%	0%	78%	0%	…



Stores 1 and 3 have a similar sales distribution. Therefore, they will be clustered together.

Opportunity

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	20%	0%	80%	0%	…
2	63%	10%	0%	27%	…
3	22%	0%	78%	0%	…

- Attribute groups that were not previously sold are listed as 0%, implying that they did not sell well in a given store
- **Opportunity:** Remove historical bias by predicting sales for unsold SKUs

Expanding Client’s Methodology

Initial clustering on historical sales

Second clustering on significant store traits to predict sales

Final clustering on predicted and historical sales

Second Clustering – Choose Store Traits

- Assortment planner selects ~10 store traits to consider
- Select traits that historically **influence demand**

Lawn Mower	
Avg Precip	Med Age
Avg Temp	….

Hardware	
Urbanicity	Avg Precip
Med HH Inc*	….

* Median Household Income

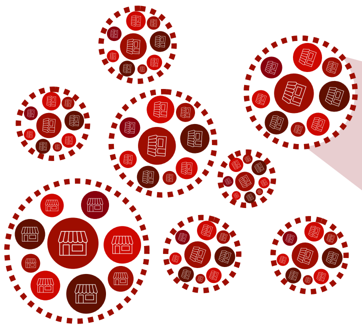
Second Clustering – Identify Significant Store Traits

Urbanicity Distribution in overall store data

Urban	Suburban	Rural
28%	50%	22%

Urban	Suburban	Rural
30%	50%	20%

Distribution for **Cluster 5**



Urban	Suburban	Rural
70%	15%	15%

Distribution for **Cluster 9**

Apply the **Chi-Square** test to identify if clusters' distributions **deviate** from the population's distribution.

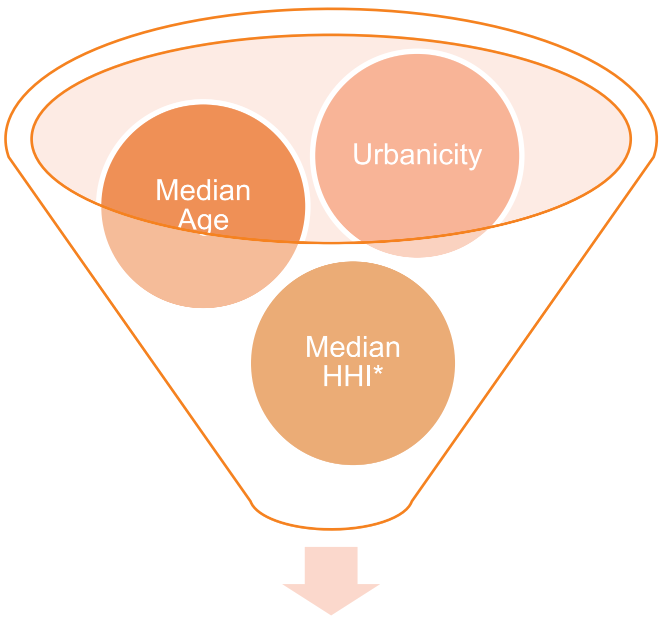
Conclusion: **Urbanicity** is a **significant store trait**.

Second Clustering – Identify Significant Store Traits

Households | Average Temp

Urbanicity | **Median HHI***

Median Age | # of Competitors



K-Prototype

* Median Household Income

Expanding Client's Methodology

Initial clustering on historical sales

Second clustering on significant store traits to predict sales

Final clustering on predicted and historical sales

Final Clustering – Predicting Sales

Sales %*	Store	A-White	A-Red	B-Brown	D-Yellow	…
0.02	1	$20	$0	$80	$0	…
0.10	2	$78	$12	$0	$34	…
0.05	3	$41	$0	$145	$0	…
0.07	4	$50	$15	$70	$21	…
0.03	5	$35	$7	$55	$30	…

Scale attribute group sales upwards - proportional to overall cluster sales:

Store 2: $\frac{\$12}{0.10}$ = $120 Store 4: $\frac{\$15}{0.07}$ = $215 Store 5: $\frac{\$7}{0.03}$ = $234

Average of scaled sales:

$$\frac{\$120 + \$215 + \$234}{3} = \$190$$

* % of total sales of cluster

Final Clustering – Predicting Sales

Sales %	Store	A-White	A-Red	B-Brown	D-Yellow	…
0.02	1	$20	$3.8	$80	$11	…
0.10	2	$78	$12	$244	$34	…
0.05	3	$41	$9.5	$145	$28	…
0.07	4	$50	$15	$70	$21	…
0.03	5	$35	$7	$55	$30	…

Average of scaled sales:

$$\frac{\$120 + \$215 + \$234}{3} = \$190$$

Sales scaled by sales percentage:

Store 1: $\$190 \times 0.02 = \$3.8$

Store 3: $\$190 \times 0.05 = \$9.5$

Final Clustering – Convert Sales to Distribution

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	$20	$3.8	$80	$11	…
2	$78	$12	$244	$34	…
3	$41	$9.5	$145	$28	…

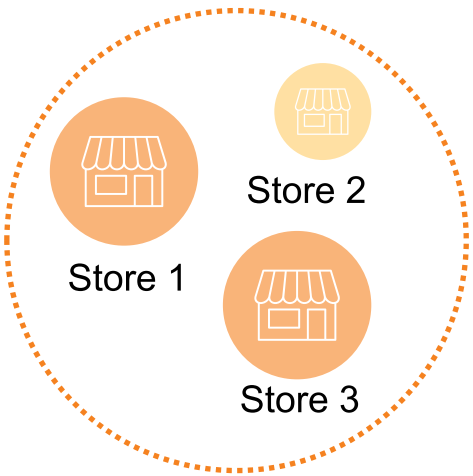
Dollar amount is converted to %, so that each store has an expected sales distribution per attribute-group

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	17%	9%	23%	19%	…
2	17%	10%	24%	18%	…
3	19%	8%	23%	20%	…

Final Clustering – K-Means on Predicted Sales Distribution

- Re-cluster stores using the predicted sales distribution

Store	A-White	A-Red	B-Brown	D-Yellow	…
1	17%	9%	23%	19%	…
2	17%	10%	24%	18%	…
3	19%	8%	23%	20%	…



Stores 1, 2, and 3 have **similar sales distribution**. Therefore, they will be **clustered together**.

Expanding Client's Methodology

Initial clustering on historical sales

Second clustering on significant store traits to predict sales

Final clustering on predicted and historical sales

Capturing the Opportunity

Agenda

- Client & Project Overview
- Initial Approach, Challenges, & Final Approach
- Triple Clustering Approach
- Deliverables & Impact

Deliverables

Final Cluster Result & SKU Assignment

to demonstrate the results of our new methodology

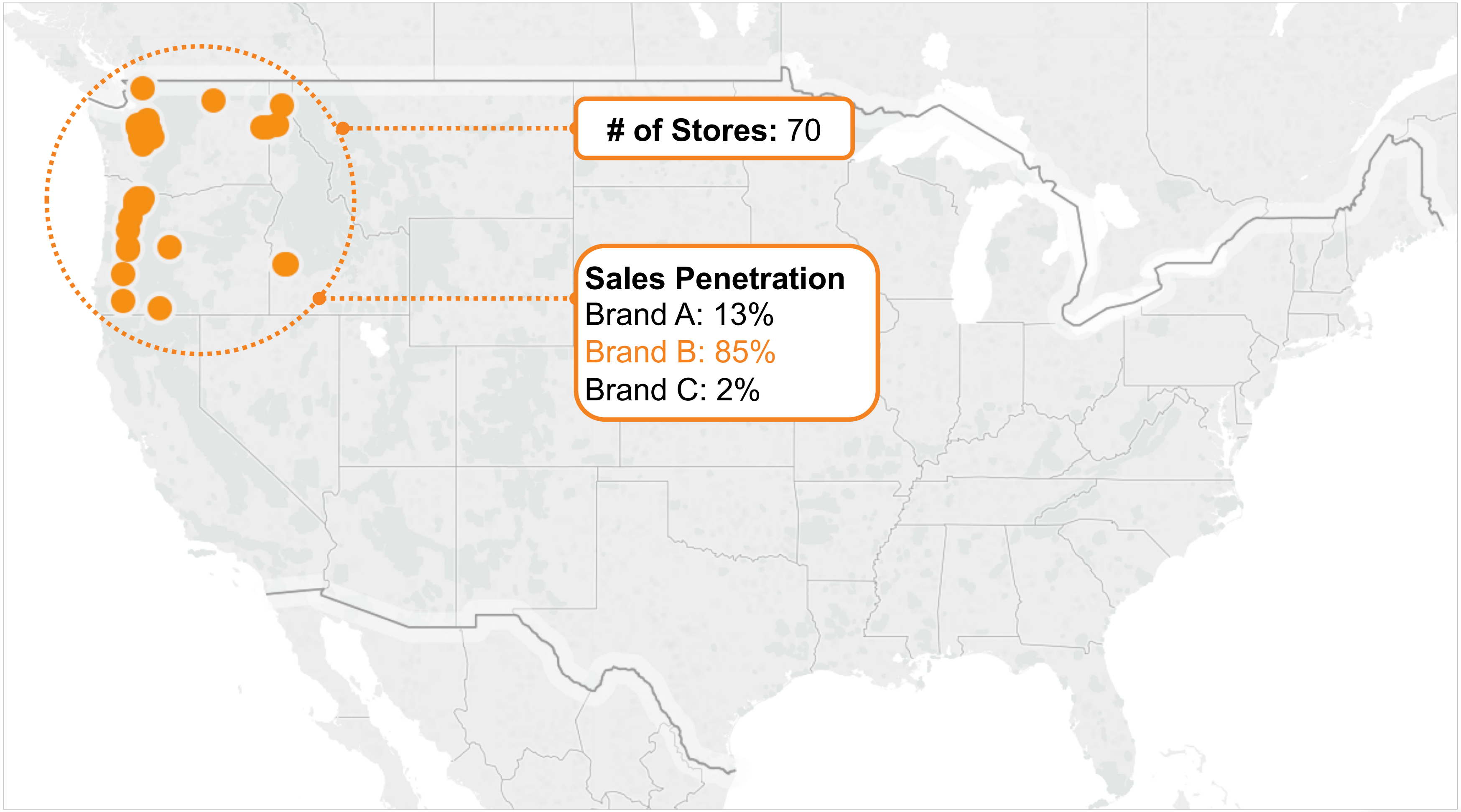
Documentation

to outline our methodology and assumptions

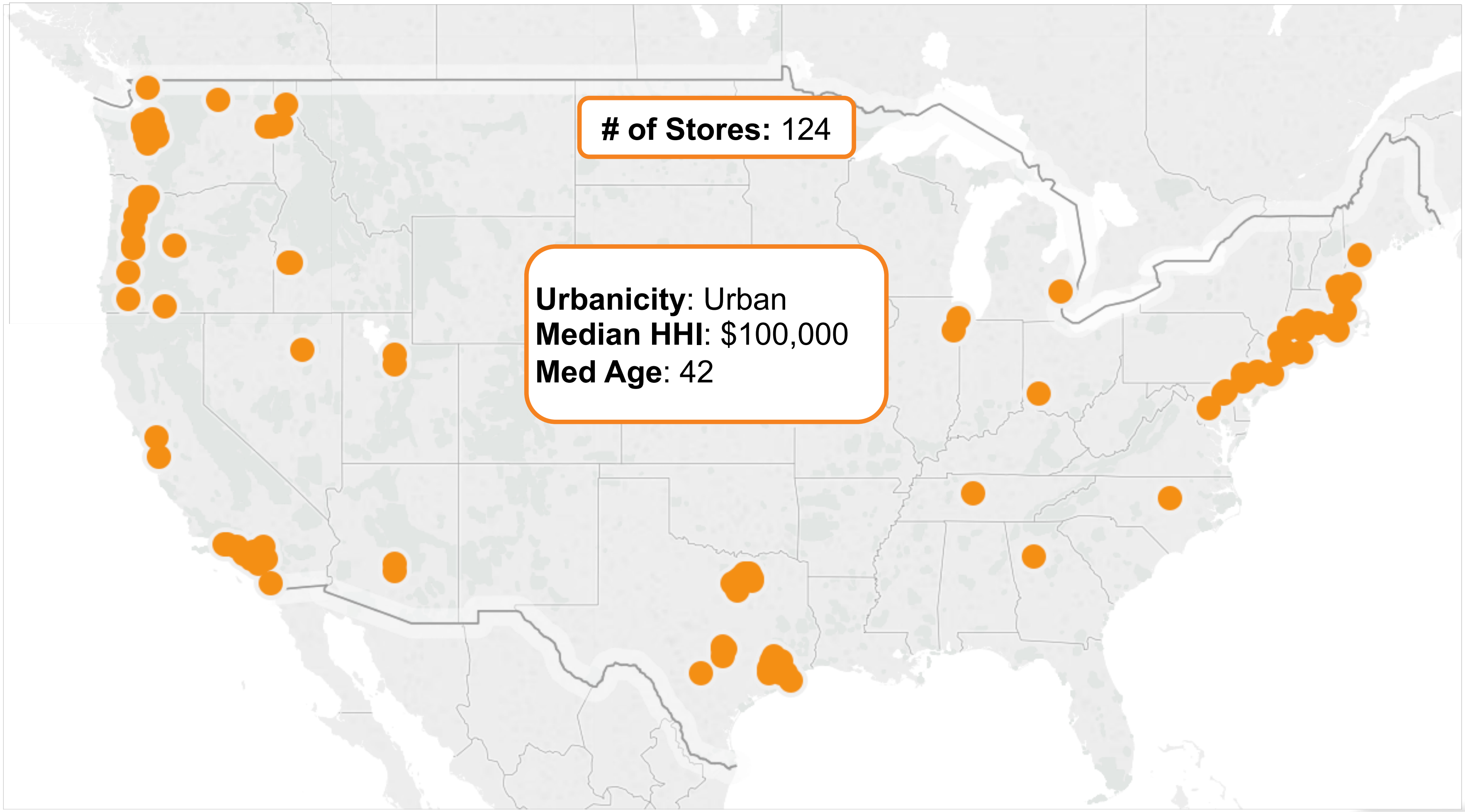
Python Script

to run additional clustering

Current Clustering



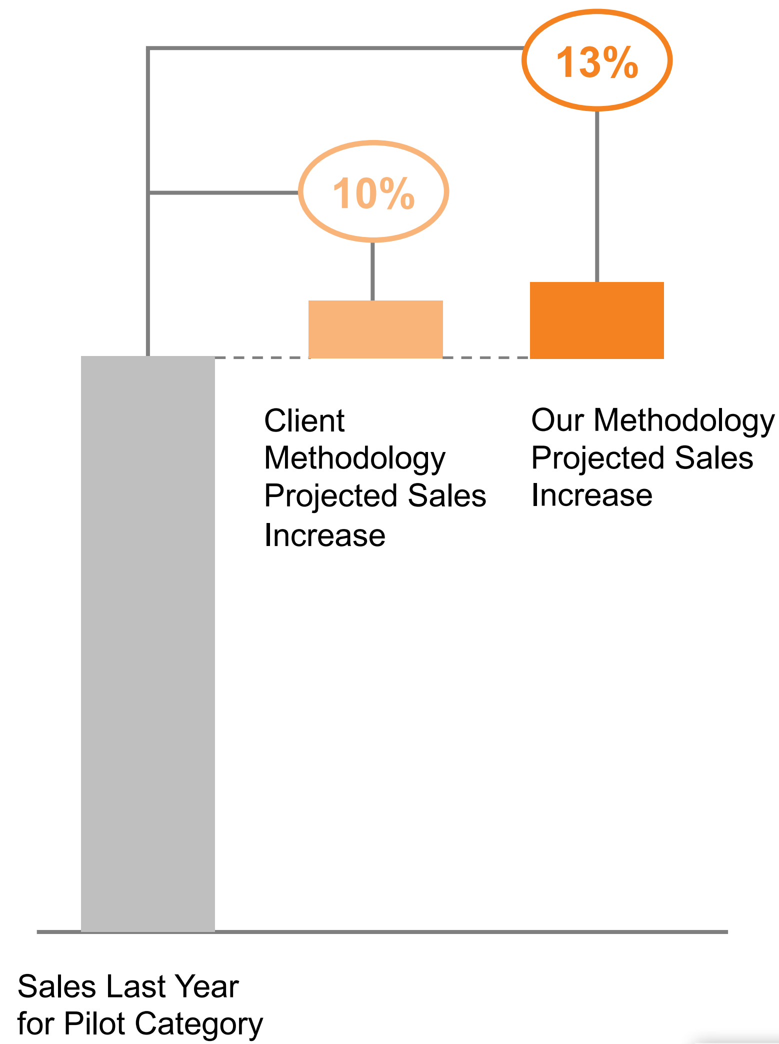
New Clustering



Impact on Pilot Category

$$\sum_{store}\sum_{SKU} \mathrm{APSW}_{store,SKU} * y$$

$$y = \begin{cases} 1 \text{ if SKU is expanded to store} \\ -1 \text{ if SKU is dropped from store} \end{cases}$$



* APSW: Average per store per week;
Historical APSW for SKU that are dropped
Predicted APSW for SKU that are expanded

Final Impact

3%

Additional increase in projected sales

Impact specific to the pilot category

Improved
Customer experience

Standardized
Cluster methodology

Incorporated
Additional data

Reduced
Historic sales bias

Q&A

Appendix

Chi-Square Test

- **Objective**: Determine if cluster distributions are *significantly different* from the expected distribution derived from the population

	Core	Value	Upscale
Population	50%	25%	25%
Actual	150	52	60
Expected	131	65.5	65.5

$$\chi^2 = \sum \frac{(\mathrm{O_i} - \mathrm{E_i})^2}{\mathrm{E_i}}$$

P-Value: 0.049787

Significance Level: 0.05

A **significant difference exists** between distributions

K-Prototypes Algorithm

Cost Function:

$$E = \sum_{l=1}^{k} \sum_{i=1}^{n} y_{il} d(X_i, Q_l)$$

Cost for a single store:

$$d(X_i, Q_l) = \sum_{j=1}^{m_r} (x_{ij}^r - q_{lj}^r)^2 + \gamma_l \sum_{j=1}^{m_c} \delta(x_{ij}^c, q_{lj}^c)$$