CAPSTONE: MARKET SEGMENTATION

Ian Gault

December 17th, 2023

Final Submission

EXECUTIVE SUMMARY

Background

Supermarket specializes in selling meats, fruit, sweets, wine, and gold products

Dataset of 2051 customers over a 2-year period of operations

Average response to campaigns: 7.5%

Goal

Cluster customers into groups

Allow for personalized, targeted marketing to improve campaign responses and return on investment (ROI)

Method

Use unsupervised learning techniques to uncover hidden components and clusters within the customer base

Find the most impactful features

Outcome

Parameters to inform campaign strategy for the following groups:

- Loyal Customers
- Untapped Opportunity Customers
- Restricted Customers

PROBLEM STATEMENT

Current State

- Have low campaign conversion rate
- Uniform distribution of frequency of purchasing and signing up for service
- Steady but stagnant growth

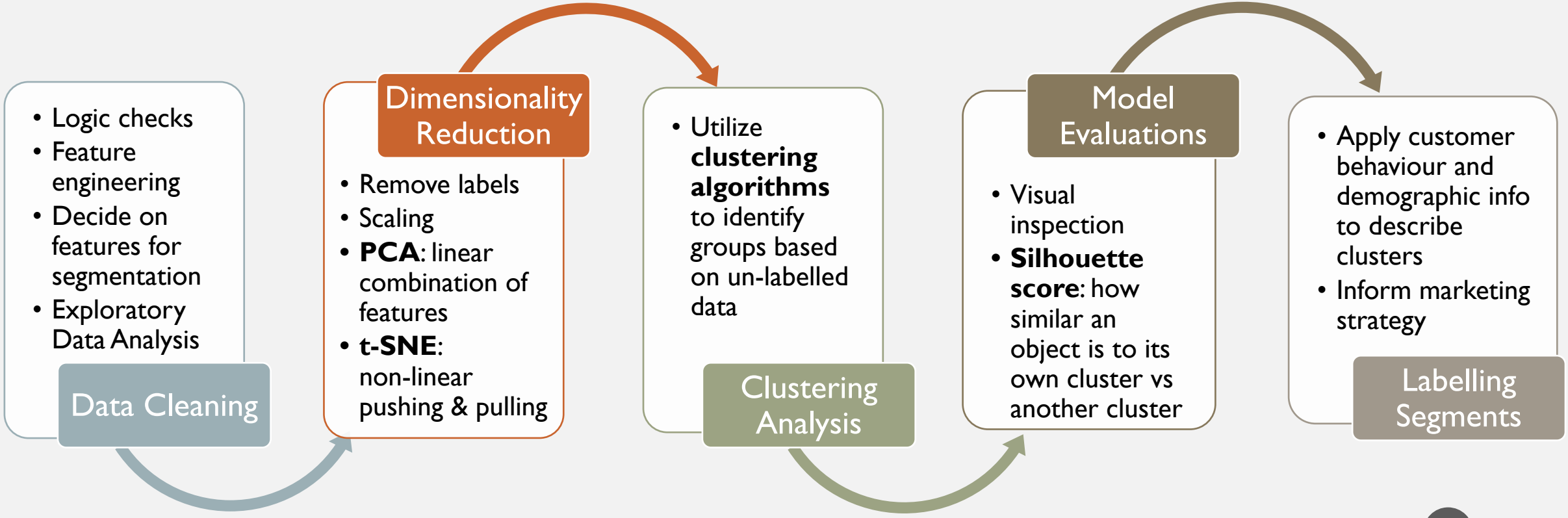
Gap

- Lack of targeted, segmented approach in marketing strategy
- Not meeting the customer where they are at or what their preference are or buying behaviour

Future State

- Improved campaign conversion rate
- Better ROI on marketing strategies
- More flexibility to make innovative marketing campaigns
  - Reinforce brand

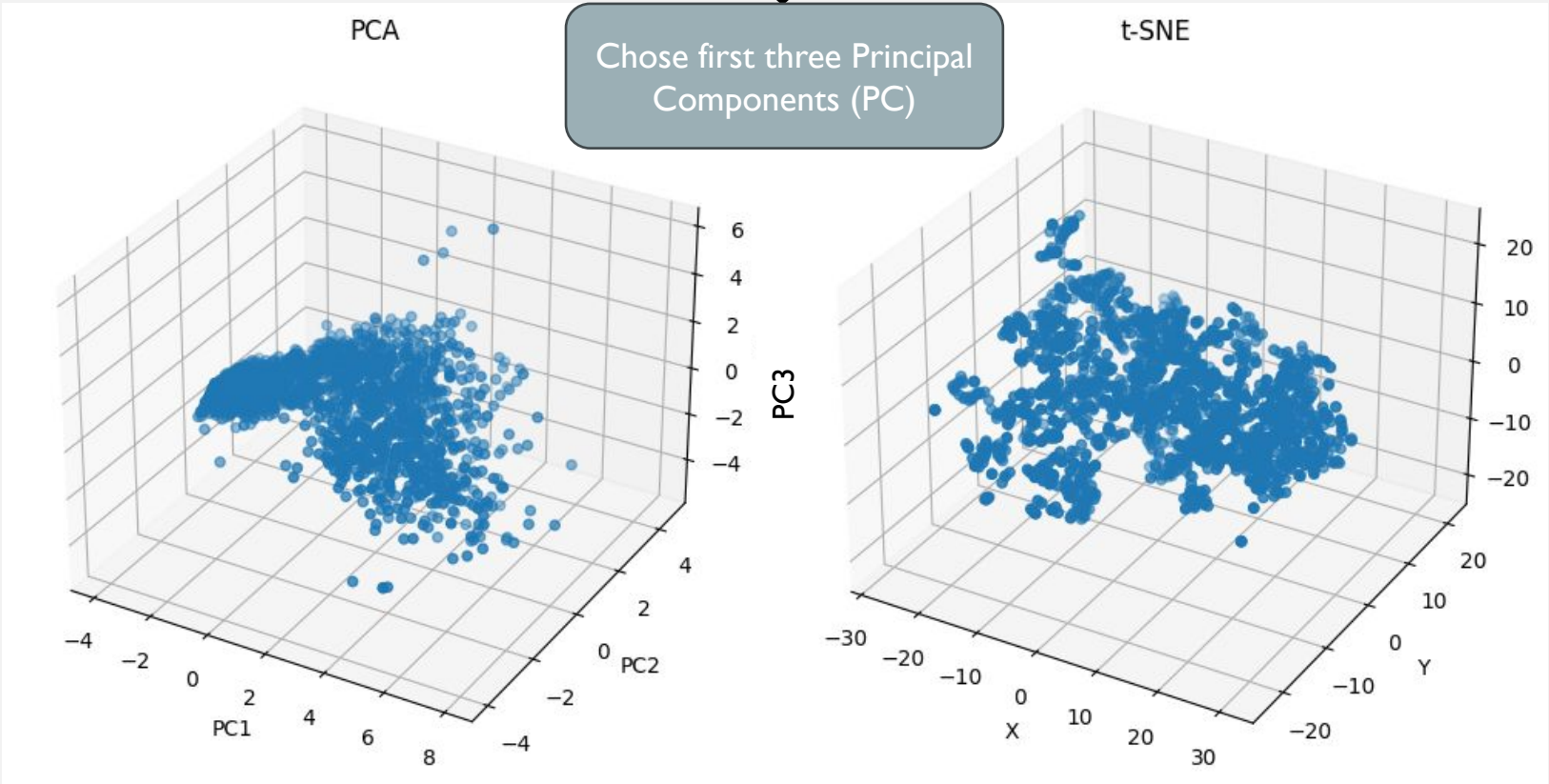
APPROACH


Data Cleaning
• Logic checks
• Feature engineering
• Decide on features for segmentation
• Exploratory Data Analysis
Dimensionality Reduction
• Remove labels
• Scaling
• PCA: linear combination of features
• t-SNE: non-linear pushing & pulling
Clustering Analysis
• Utilize clustering algorithms to identify groups based on un-labelled data
Model Evaluations
• Visual inspection
• Silhouette score: how similar an object is to its own cluster vs another cluster
Labelling Segments
• Apply customer behaviour and demographic info to describe clusters
• Inform marketing strategy

DIMENSIONALITY REDUCTION

29 variables containing distinct demographic and buying behaviour information for 2051 customers

Chose first three Principal Components (PC)



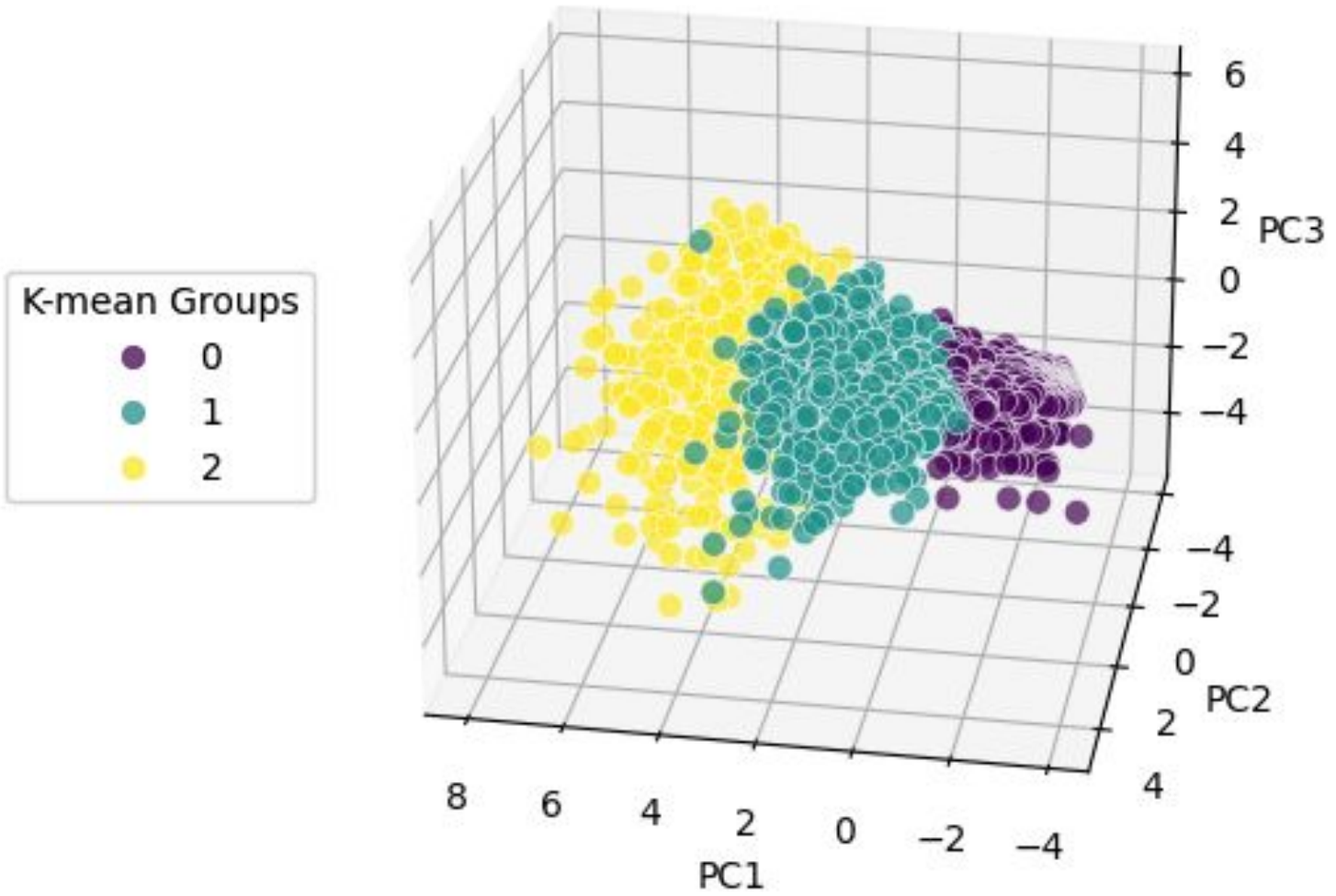
PCA more promising for use case

CHOOSING CLUSTERING MODEL

Clustering Algorithm	Methodology	Choice of K (Number of Segments)	Model Evaluation (Silhouette Score)
K Means	Centroid based on mean using Euclidean distance	3	**0.35** Highest Value
K Medoids	Centroid based on median using Euclidean distance	3	0.28
Hierarchical Clustering	Bottom-up agglomerative clustering	3	0.26
Density Based spatial Clustering	Detects concentrated areas of data points	5	0.14
Gaussian Mixture Model	Probability-based clustering using gaussian distributions	3	0.23

KEY FINDINGS & INSIGHTS

K-means Clusters Applied to PCA Dimensionality Reduction Outputs



LABELLING & DEFINING SEGMENTS

Segments		Demographics			Buying Behaviour			Engagement		
Segment Name	Segment (Sample Size)	Avg. Income ($)	Avg. Household Size	Avg. Age	Avg. Total # Items	Avg. Total Spent ($)	Avg. Spent per Item ($)	Avg. Accepted Campaign	Loyalty (days since joining)	Avg. Web Visits
Restricted Customer	0 (983)	36k	**2.9**	45	6	102	15	0.2	307 days	**6.4**
Untapped Opportunity	1 (567)	59k	2.7	**50**	18	734	41	0.4	**403 days**	**5.5**
Loyal Customers	2 (501)	**76k**	1.9	48	**20**	**1,454**	**77**	**1.0**	391 days	3.0

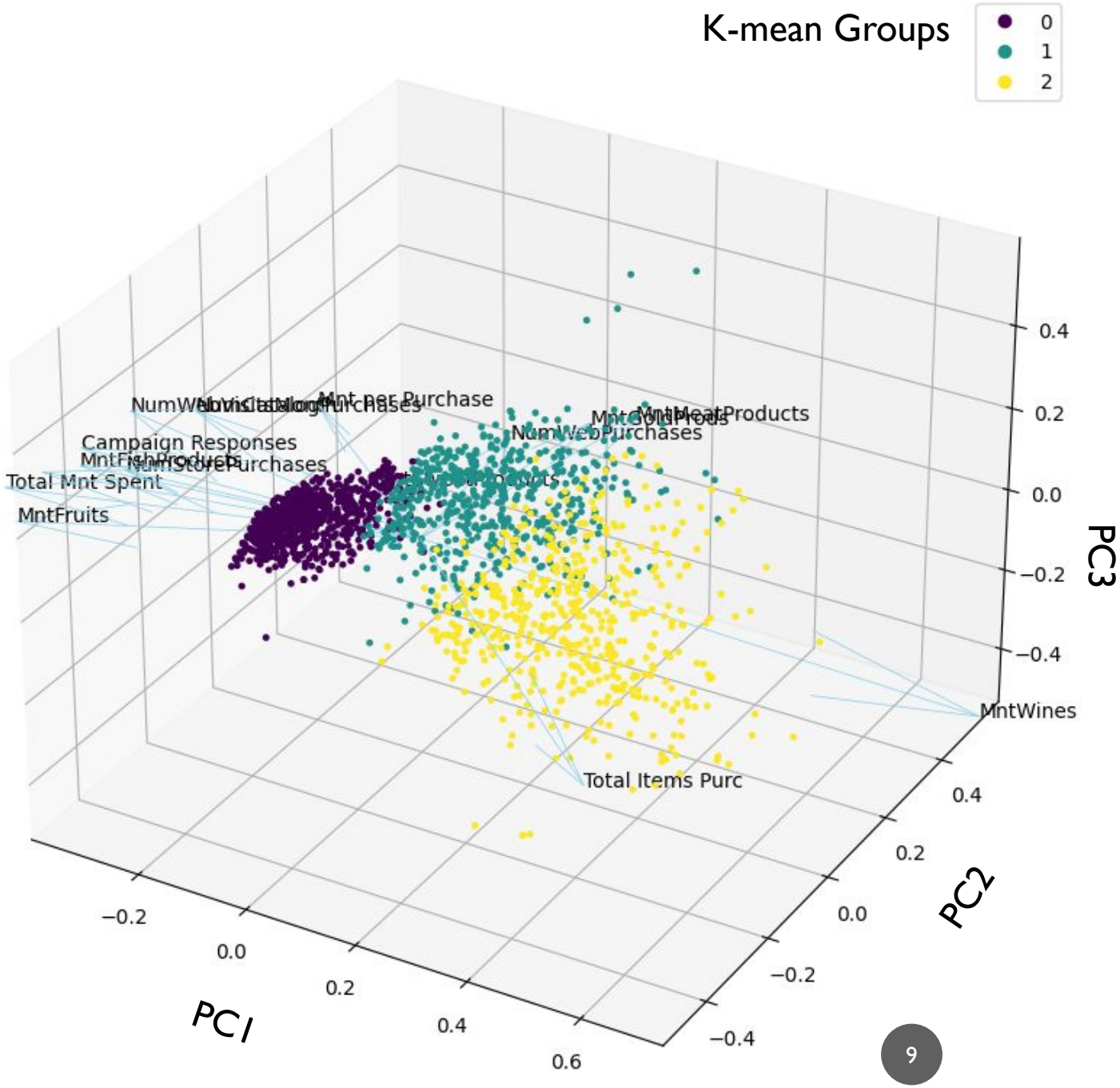
Highest Values

BIPLOT (PCA + FEATURE LOADINGS)

Restricted Customers
- Closer to the origin of the reduced dimensional space
- Dense number of features contributing to clustering

Loyal Customers
- Largest separation from customer base due to total items purchased and amount spent on wine

Untapped Opportunity
- Influenced by amount spent on meat product and number of web purchases made, amount per purchase, amount spent on gold



BUYING BEHAVIOUR & RECOMMENDATIONS

Restricted Customer (Group 0)

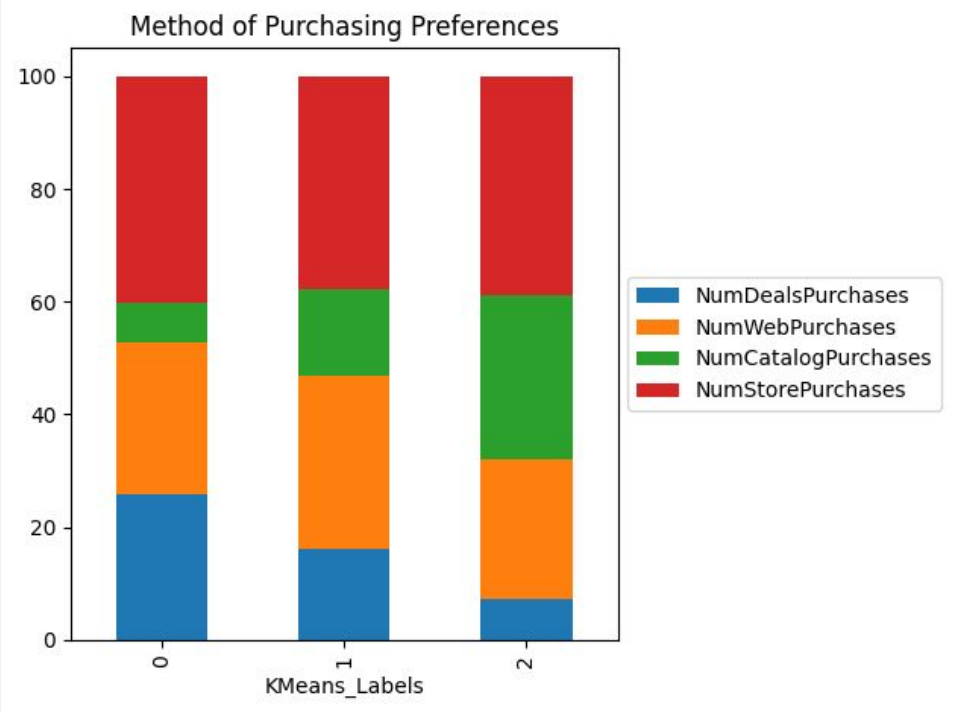
- Higher proportion of deals accepted
- Poor existing revenue and low market potential; low-cost marketing strategy, applied to largest group

Untapped opportunity (Group 1)

- High web visits and web purchases
- Medium existing revenue and good market opportunity; promote high cost per item categories; high-cost marketing for bigger returns

Loyal Customers (Group 2)

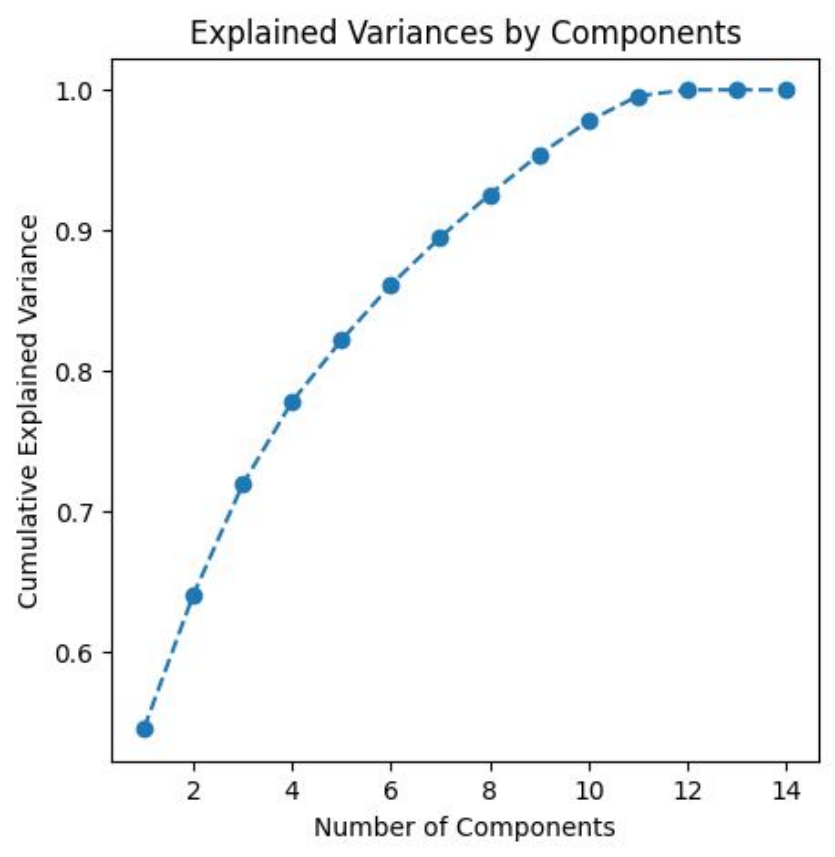
- Less captivated by deals, more use of the catalog
- High existing revenue and maintain interest; high purchases across categories, especially wine; more so access and brand awareness; influencer establishment

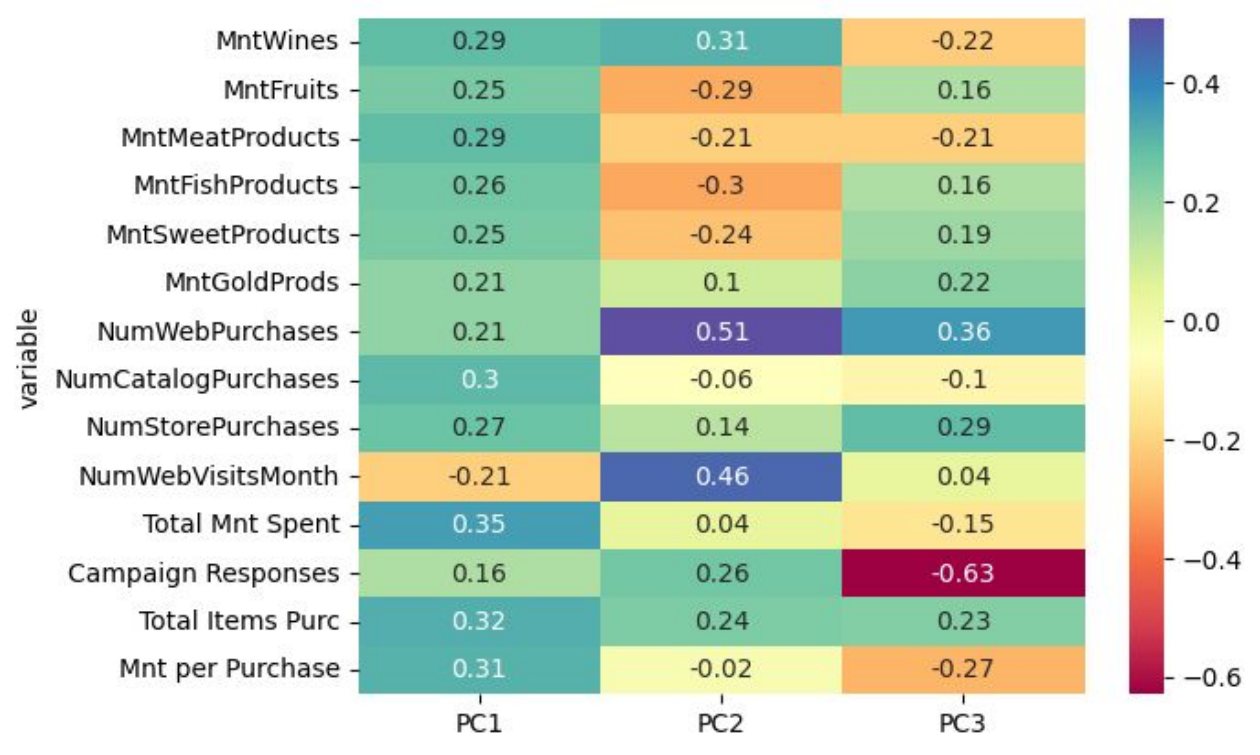


Collaborate with Marketing Team to find create solutions, based on established group parameters, for highest return of investment and brand promotion.

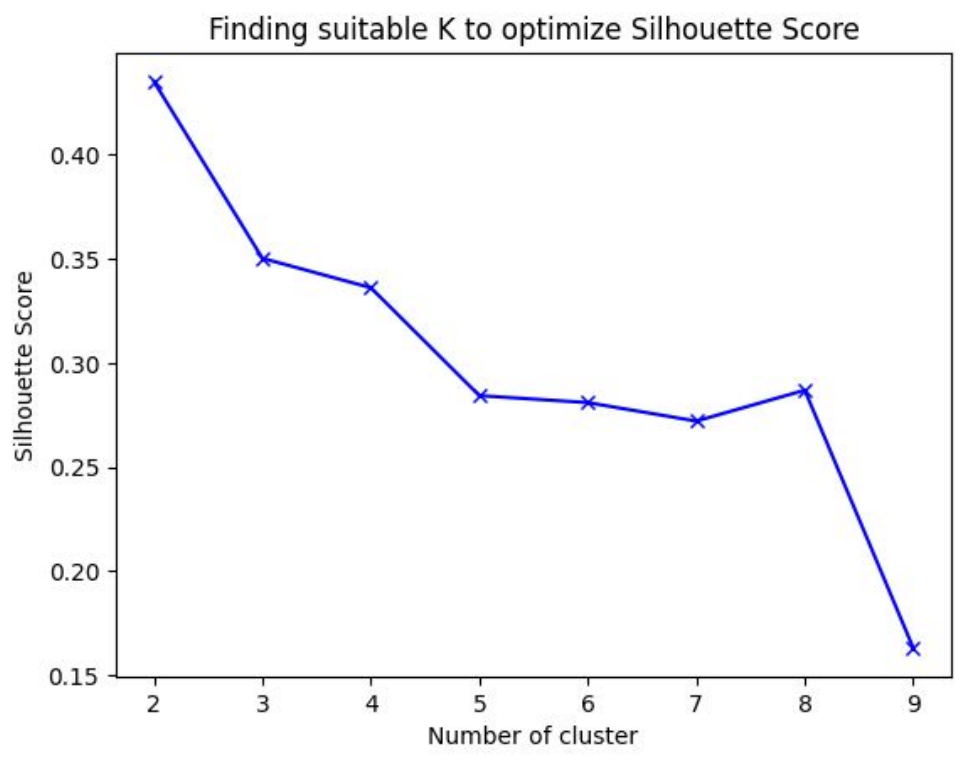
QUESTIONS & ADDITIONAL SLIDES

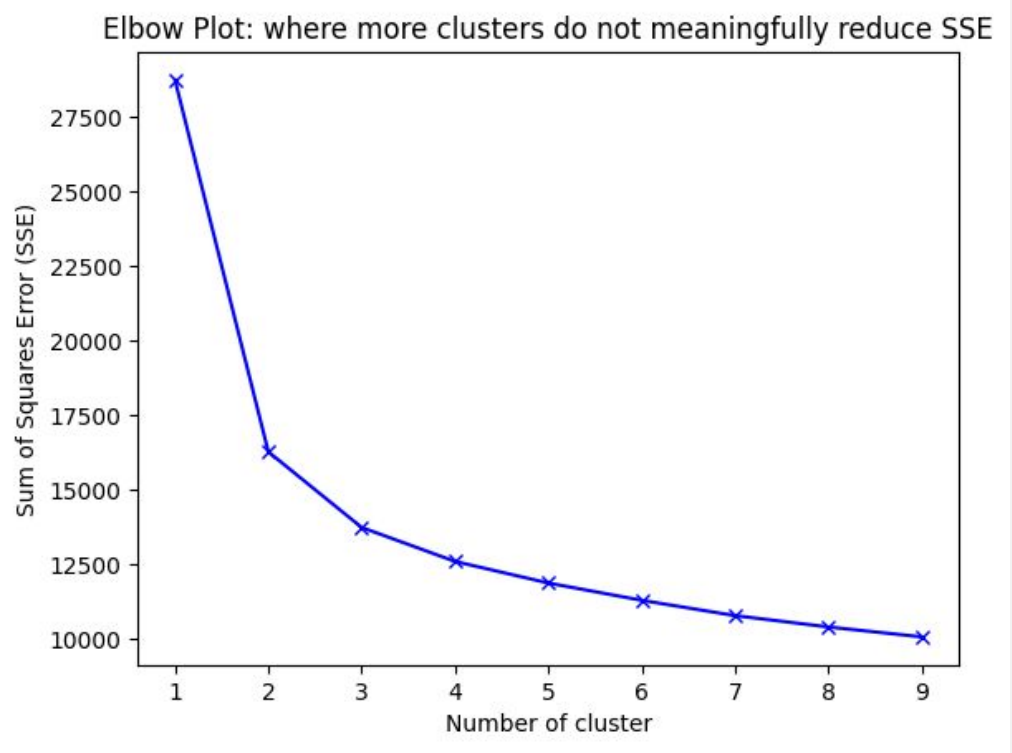
CHOOSING PRINCIPAL COMPONENTS





CHOOSING K CLUSTERS





LIMITATIONS & SECONDARY OPTION

- Assumptions for K-Means
  - Linear cluster boundaries
  - Clusters same size, similar number of points
  - Susceptible to outliers
- Gaussian Mixture Model (GMM)
  - Offered more similar sized clusters
  - Though lower silhouette score
  - Secondary choice
- Silhouette score misrepresentation
  - Non-spherical clusters
  - Outliers
  - Varying densities