TARGET CONSUMER PROFILING PLAN: CUSTOMER SEGMENTATION

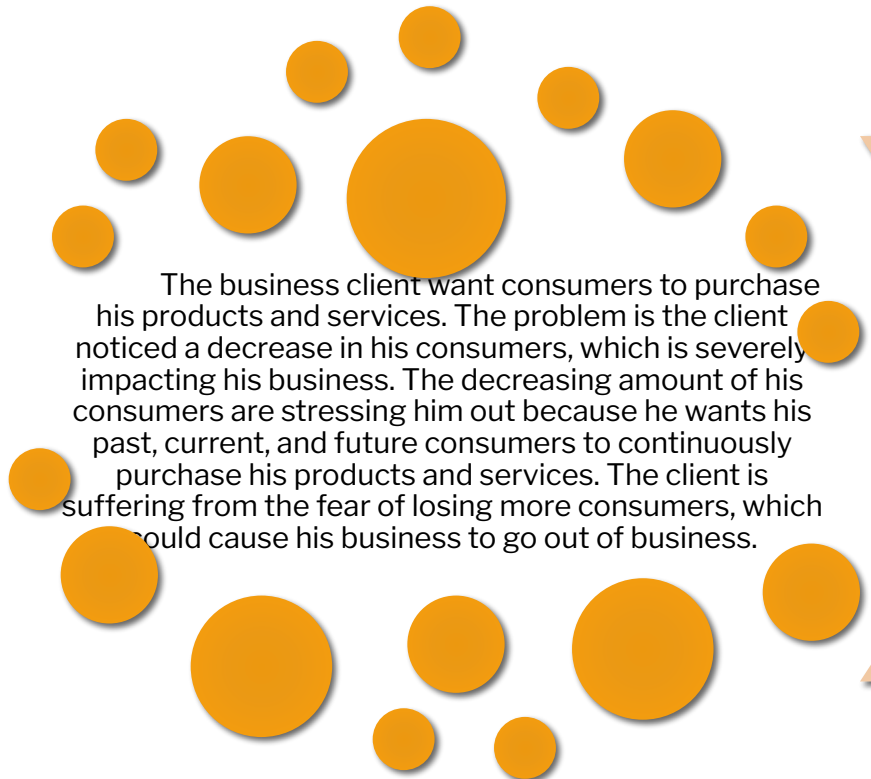LaChandra Ash

November 7, 2021

Abstract

The constant emergence of multiple business opponents has provoked a large sum of unadorned oppositions among the businesses, while obtaining new consumers and retaining current consumers. Due to the large amounts of competition; products, services, and consumer reports must be chosen wisely to maintain the revolving presence of the consumers. Every business must possess the ability to understand their consumers' wants, needs, and concerns if they wish to gain superior leverage over their competitors. Understanding the consumers will allow the businesses to develop and utilize the best programs for their marketing strategies to attract more consumers and retain current consumers.
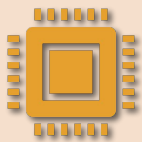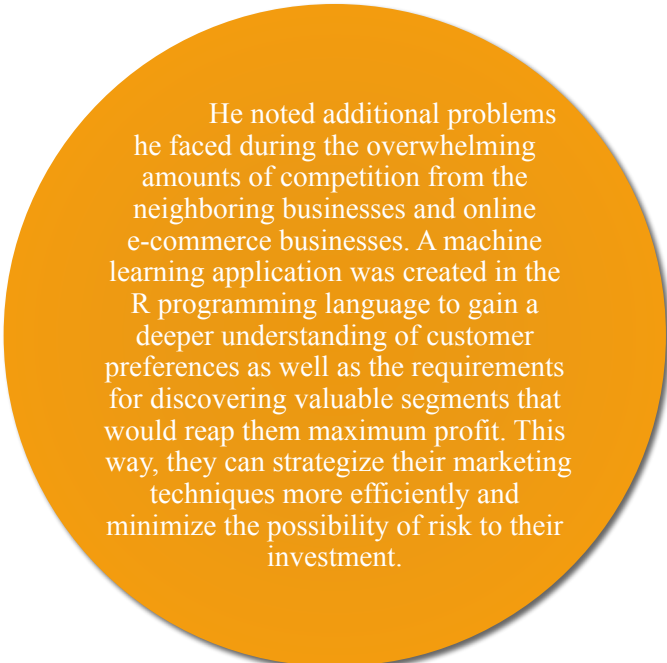
The use of systematic consumer segmentation can provide each business with an improved understanding about their consumers. Every segment will encompass consumers that possess mutual similarities and characteristics. The developments of machine learning and big data birthed and fueled an automated process that involves consumer segmentation. The k-means clustering type of algorithm is implemented during the machine learning processing of the consumers' data. The k-means algorithm program that was developed by MATLAB, was trained through the usage of the z-score dataset. The training data was acquired from specific retail businesses located within a mall.
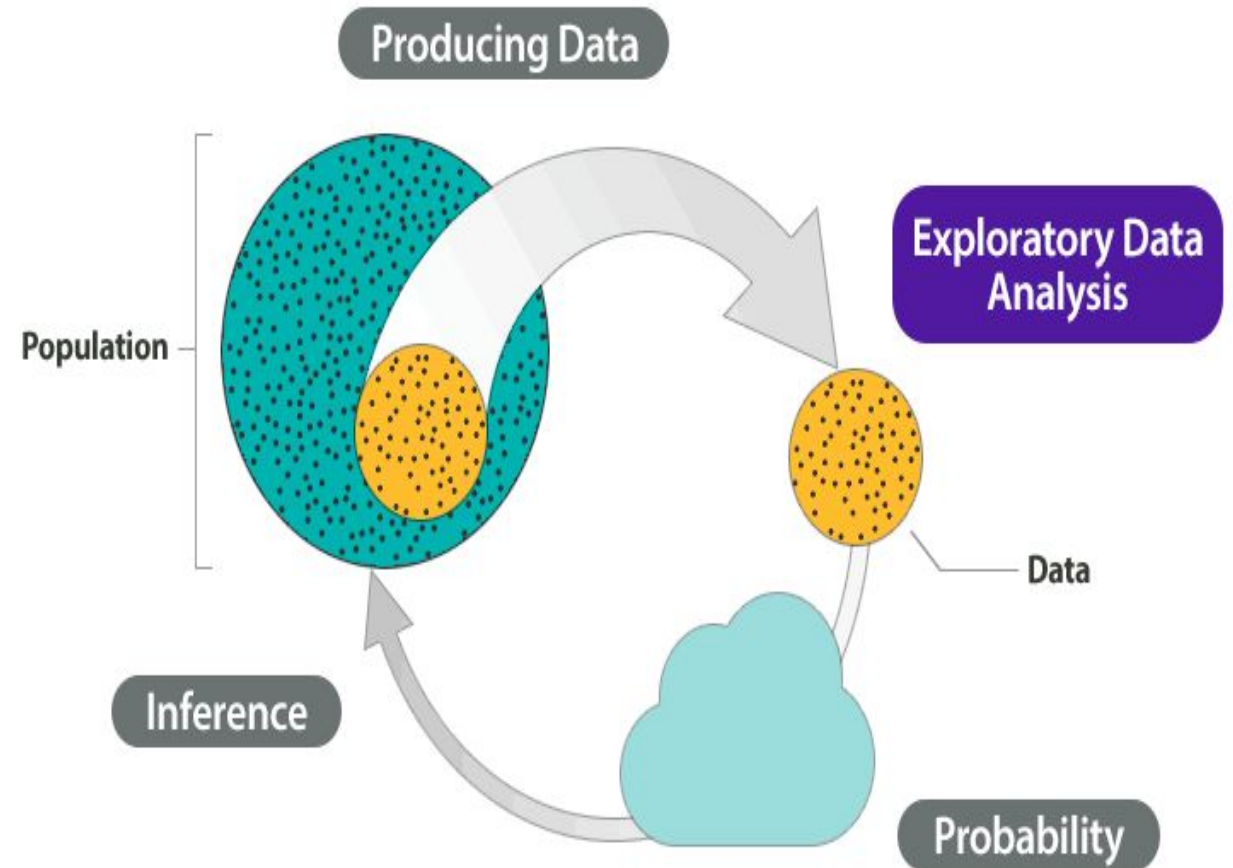
Introduction

The business client want consumers to purchase his products and services. The problem is the client noticed a decrease in his consumers, which is severely impacting his business. The decreasing amount of his consumers are stressing him out because he wants his past, current, and future consumers to continuously purchase his products and services. The client is suffering from the fear of losing more consumers, which could cause his business to go out of business.
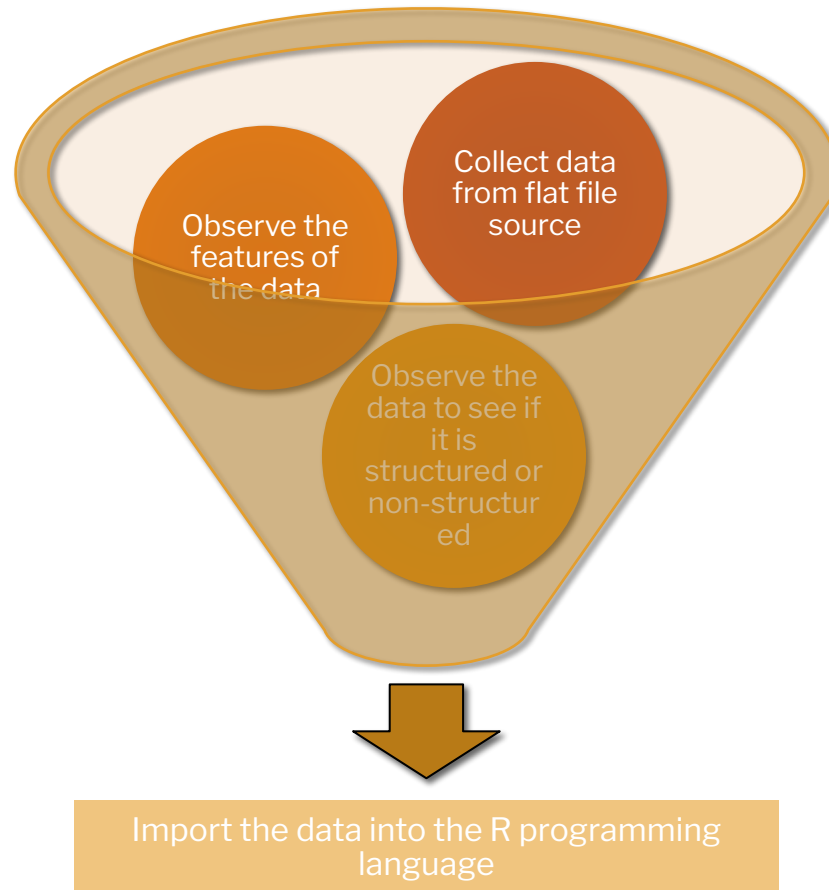
He noted additional problems he faced during the overwhelming amounts of competition from the neighboring businesses and online e-commerce businesses. A machine learning application was created in the R programming language to gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit. This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

Methods

- Data Exploratory
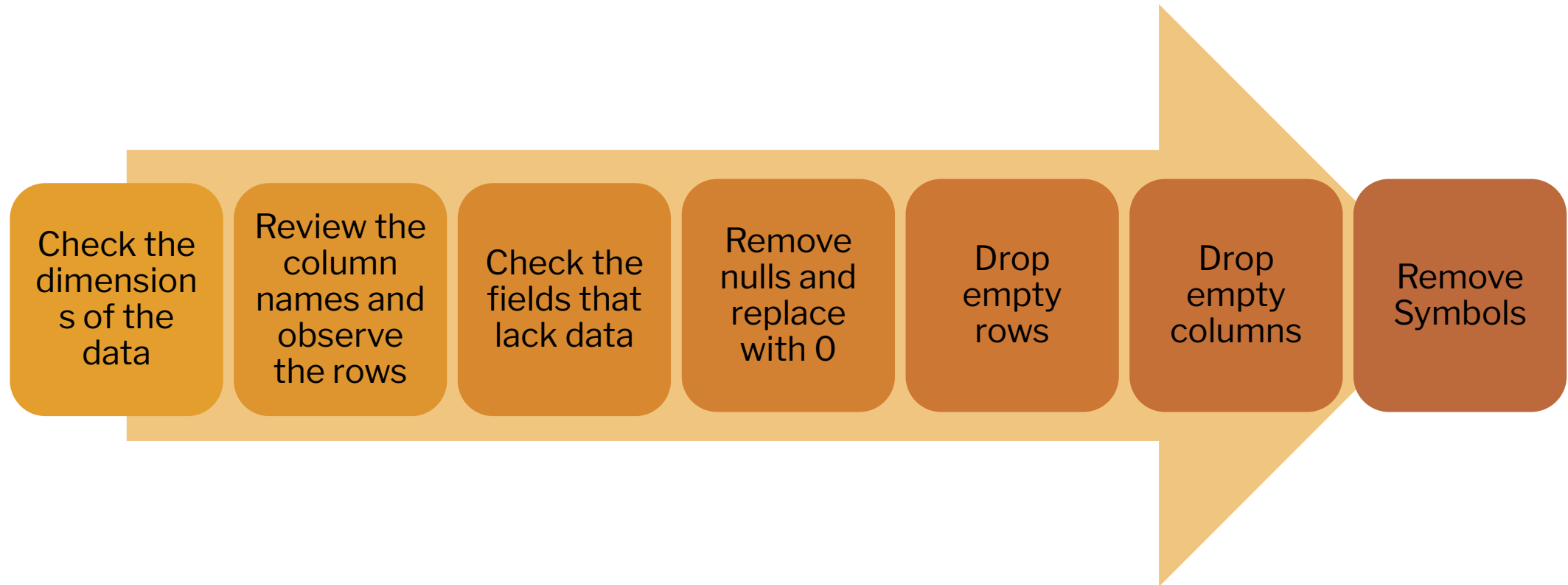- Visualization
- K-Means Algorithm

Data Exploratory

Observe the features of the data

Collect data from flat file source

Observe the data to see if it is structured or non-structured

Import the data into the R programming language

Data Exploratory (cont)

Check the dimensions of the data → Review the column names and observe the rows → Check the fields that lack data → Remove nulls and replace with 0 → Drop empty rows → Drop empty columns → Remove Symbols

Visualization

Use bar plots to display the comparison between genders

Use pie chart to observe female and male ratio distribution

Use histogram to observe customer ages

Use the boxplot to observe the descriptive analysis of the consumers' ages

Use histogram to display the annual income of the consumers

Use density plot to review and observe the annual income of the consumers

Analyze the consumers' spending scores

Use boxplot to observe the descriptive analysis of the spending score

Conclude the results from the visualizations

K-Means

Determine optimal clusters

Algorithm randomly detects K

Annotate clusters that are used

Reassign objects

Assign the clusters

Randomly select K objects

Observe number of clusters
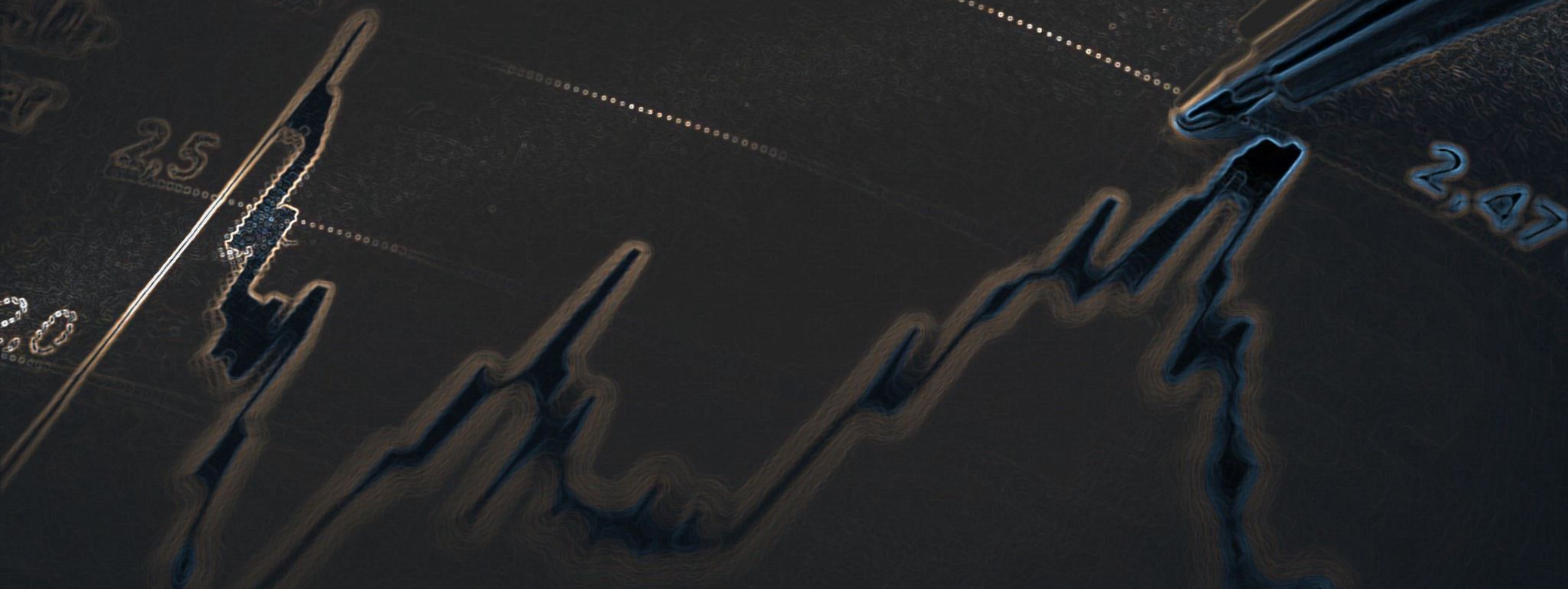
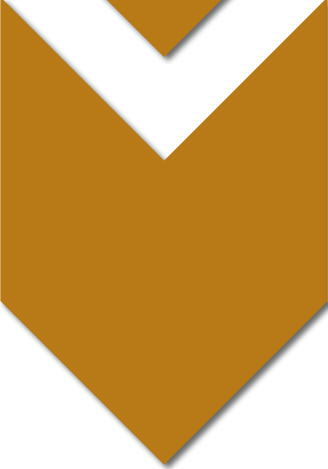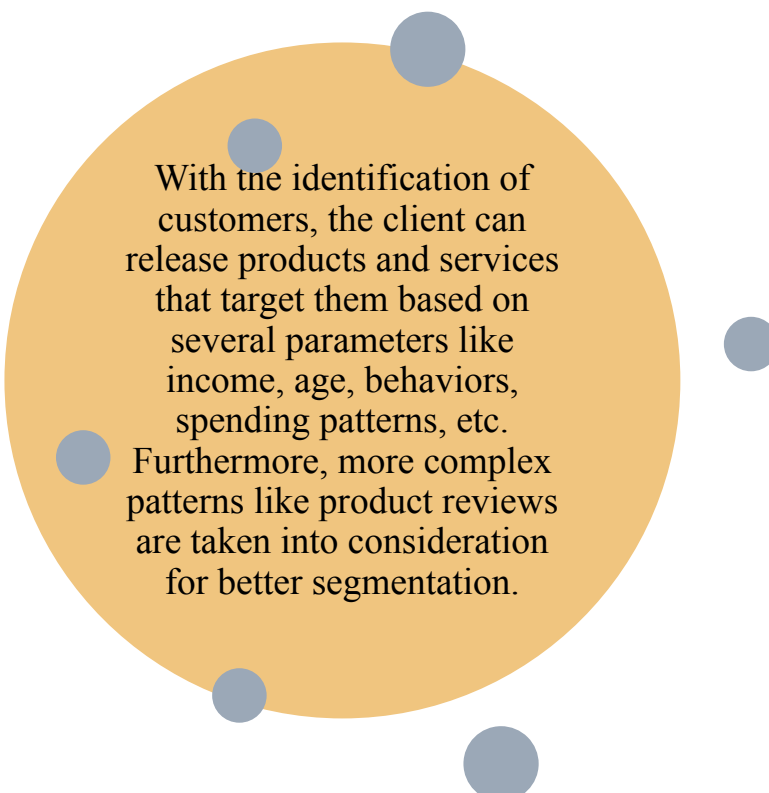# K-Means (cont)

- Use the gap statistic, elbow method, or silhouette method to determine optimal clusters
- Visualize the results from clustering

Results

- The number of females were higher than the males.
- I conclude that the percentage of females is **56%**, whereas the percentage of male in the customer dataset is **44%**.
- From the descriptive analysis, I concluded the min annual income of the customers is 15 and the max is 137.
- People earning average income of $70,000, had highest frequency count in distribution.

- The average salary of all the customers is 60.56.
- The min spending score is 1, max is 99 and the avg is 50.20.
- The descriptive analysis of spending score is min at 1, max and avg. is 50.20.
- I conclude that customers between class 40 and 50 have the highest spending score among all the classes.

Conclusion

With the identification of customers, the client can release products and services that target them based on several parameters like income, age, behaviors, spending patterns, etc. Furthermore, more complex patterns like product reviews are taken into consideration for better segmentation.