Project Overview
Cluster Analysis – Online retailers

- Summary

This exploratory data analysis report provides insights into the customers of an online retailer, which offers a range of products . The objective of the analysis is to investigate whether there are distinct segments in the market, based on customer behavior and characteristics. Through descriptive and cluster analysis, we aim to identify patterns in the data that support the existence of three customer segments. The findings of this report can inform the online retailer's marketing strategies and improve their understanding of their customer base

- Data Set

The datasets in this project contain 1 Excel file. (Link)

- Tools

R Programming Language

- Analytical Skills

EDA by R
Cluster Analysis (K-Means) by R
Data Visualization by R
R Codes : (Link)

Prepared by
Hamed Movahedizadeh

Exploratory Data Analysis

First, let´s have a look at a few records of the dataset

Heads

First few records

Customer	Income (in $10,000s)	Average Purchase (in $100s)	Last year purchases	Number of Dependents
1	7.8	1.59	5	0
2	1.1	0.63	1	0
3	7.2	1.03	6	2
4	8	1.43	5	1
5	1	0.35	2	0
6	1.5	0.66	2	0
7	9.1	1.96	5	2
8	0.6	0.41	1	0
9	0.8	0.6	2	0
10	8	1.54	9	2

Exploratory Data Analysis

Tails

Last few records

Customer	Income (in $10,000s)	Average Purchase (in $100s)	Last year purchases	Number of Dependents
293	6.6	0.83	6	3
294	9.4	1.46	9	2
295	1	0.38	2	0
296	1.4	0.74	2	0
297	3	0.53	5	3
298	9.6	1	9	1
299	0.8	0.38	1	0
300	1.3	0.31	2	0
Average	4.43933333333334	1.00356666666667	4.43666666666667	0.99
Std Dev	2.8466937148047	0.487748508537775	2.41497326592643	1.03608796000608

Exploratory Data Analysis

Overview

Data Structures

Data Structures

Divisions	Metrics	Values
size	observations	302
size	variables	5
size	values	1,510
size	memory size (KB)	0
duplicated	duplicate observation	0
missing	complete observation	302
missing	missing observation	0
missing	missing variables	0
missing	missing values	0

Data Types

Divisions	Metrics	Values
data type	numerics	4
data type	integers	0
data type	factors/ordered	0
data type	characters	1
data type	Dates	0
data type	POSIXcts	0
data type	others	0

Job Informations

Divisions	Metrics	Values
dataset	dataset	Segmentation
dataset	dataset type	tbl_df
job	samples	302 / 302 (100%)
job	created	2023-04-22 11:40:19
job	created by	dlookr

Exploratory Data Analysis

Summary

It seems that we do not need to clean the data and dataset is ready for further investigation.

- For “Number of Dependents, it is normal that some customers have no dependents (Zero)

Variables

Diagnostic overview of individual variables

	Variables	Types	Missing	Cardinality	Zero	Negative	Outlier
▶	Customer	character	✓	🔑 identifier	-	-	-
▶	Income (in $10,000s)	numeric	✓	✓	✓	✓	✓
▶	Average Purchase (in $100s)	numeric	✓	✓	✓	✓	✓
▶	Last year purchases	numeric	✓	✓	✓	✓	✓
▶	Number of Dependents	numeric	✓	✓	⚠	✓	✓

Cluster Analysis

- We assume that there are 3 clusters
- Therefore => K= 3
- Also, we should normalize our numeric values. Normalizing the data ensures that each variable contributes equally to the distance calculation, by putting them on the same scale.

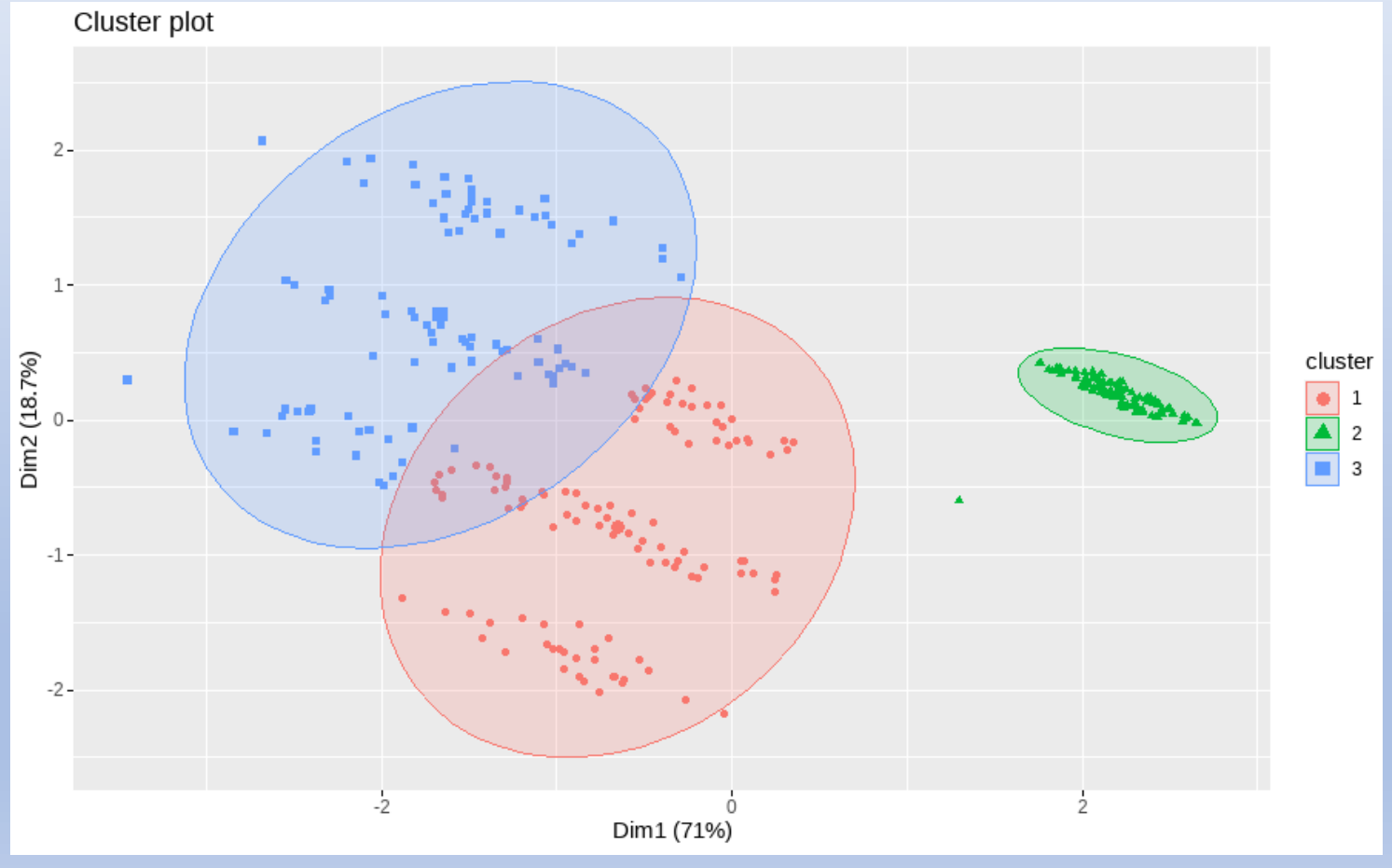
- First, we visualize the outcome of the cluster analysis

Cluster Analysis

Visualization

The x and y axes representing two of the principal components of the data. The principal components are calculated based on the variance of the original variables, and the first two principal components are chosen as the x and y axes for the plot.

In this case, the x-axis represents the first principal component, which accounts for 71% of the variance in the data, while the y-axis represents the second principal component, which accounts for 18.7% of the variance. This means that the two axes together capture almost 90% of the total variance in the data. Therefore, the outcome is significantly meaningful.



Cluster Analysis

Outcome

Based on the Cluster analysis outcomes and historical assumption, we can conclude that probably there are 3 segments:

- Cluster 1 : Family

With average income and higher number of dependents

- Cluster 2: Students

With lower income, less purchases and lower number of dependents

- Cluster 3: Professionals

With higher income and more frequent purchases

Cluster	Income (in $10,000s)	Average Purchase (in $100s)	Last year purchases	Number of Dependents
1	0.2651851	0.02973073	0.2718305	0.9603889
2	-1.2074740	-1.03679348	-1.2039072	-0.9489183
3	1.0279925	1.12684478	1.0157937	-0.1195824

Recommendation

Based on the cluster analysis outcomes and historical assumptions, we have identified three distinct customer segments: Family, Students, and Professionals. To effectively target each segment, we recommend the following marketing strategies:

- Develop detailed customer personas: The marketing team should create detailed customer personas that capture their demographics, psychographics, behaviors, and preferences. This will enable the team to develop more targeted marketing strategies and personalize messaging.
- Leverage data analytics tools: The marketing team should leverage data analytics tools such as Google Analytics, Tableau, or Power BI to monitor and analyze the performance of each segment. These tools can help the team track customer behavior, identify trends, and make data-driven decisions.
- Conduct market research: To gain deeper insights into each segment, the marketing team should conduct market research such as surveys, focus groups, or interviews. This will enable the team to better understand their motivations, pain points, and preferences, and tailor marketing strategies accordingly.

- Optimize website and online presence: The marketing team should optimize the company's website and online presence to improve the online experience for each segment. For example, they can create landing pages that are tailored to each segment or use social media platforms that are popular among each segment.

Recommendation

- Provide personalized recommendations: The marketing team should provide personalized recommendations to each segment based on their past behavior and preferences. For example, they can send personalized product recommendations via email or mobile app notifications.
- Test and refine marketing strategies: The marketing team should conduct A/B tests or multivariate tests to identify the most effective approach. This will enable the team to refine marketing strategies and improve ROI.
- Continuously measure and improve customer satisfaction: The marketing team should continuously measure and monitor customer satisfaction metrics such as Net Promoter Score (NPS) or Customer Satisfaction Score (CSAT). This will enable the team to identify areas for improvement and make data-driven decisions to enhance the customer experience.

By incorporating these recommendations into their marketing strategy, the marketing team can showcase their ability to tailor marketing efforts to the unique needs and preferences of each customer segment and improve overall customer satisfaction and business performance.