

Artwork/Project Title

Market Clustering

Year Accomplished

2023

Role/Position

Data Scientist

Publication Link

<https://colab.research.google.com/drive/1h>

Artwork/Project Description

During my holiday, my friend and I made a machine learning project about clustering. We want to learn and practice unsupervised learning using grocery firm dataset. By using the Python language, we found that this data could be divided into 4 or 5 clusters, but in our experiments, we obtained better results in cluster 5. A lot of data preparation was carried out on this dataset, for handled missing values, renamed column, then we also did feature engineering on several attributes. After preparing the data, we did PCA, this is because the number of dimensions/attributes of this dataset is very large.

	count	mean	std	min	25%	50%	75%	max
ID	2240.0	5592.159621	3246.662198	0.0	2626.25	5458.5	8427.75	11191.0
Year_Birth	2240.0	1966.800804	11.984069	1893.0	1959.00	1970.0	1977.00	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	35303.00	51381.5	68522.00	666666.0
Kidhome	2240.0	0.444196	0.536398	0.0	0.00	0.0	1.00	2.0
Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
Recency	2240.0	49.109375	28.962453	0.0	24.00	49.0	74.00	99.0
MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0

Figure 1 Data Description

Age

```
[ ] #Feature Engineering
#Age of customer today
df_clean["Age"] = 2023 - df_clean["Year_Birth"]

<ipython-input-22-56c2f5270f72>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

Figure 2 Feature Engineering

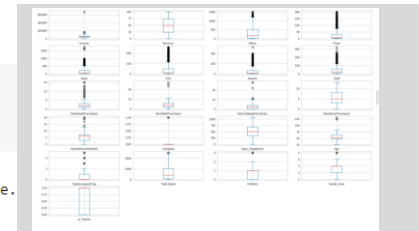


Figure 3 Outlier

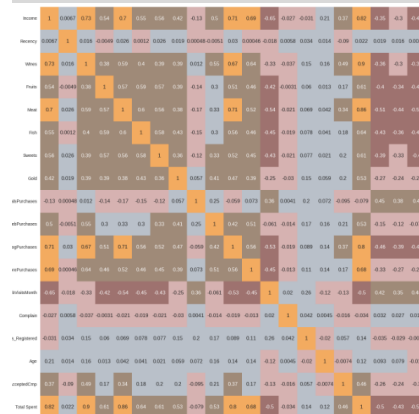


Figure 4 Correlation Matrix

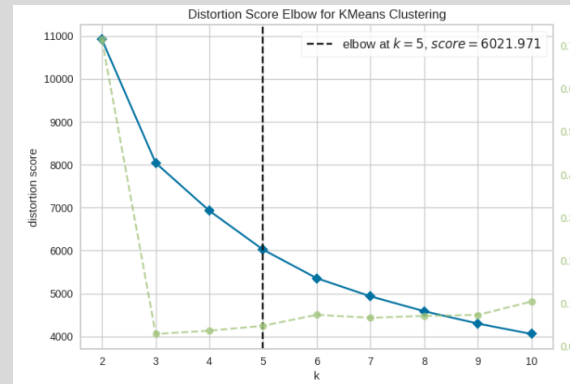


Figure 5 Elbow Score

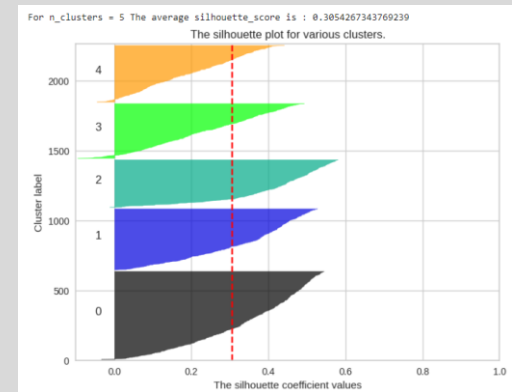


Figure 6 Silhouette Score

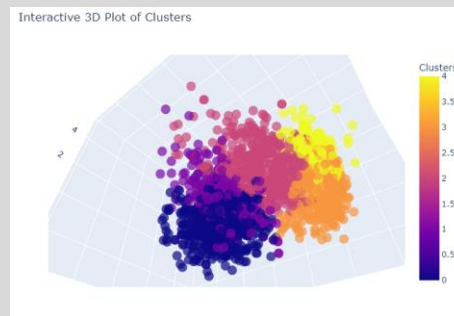


Figure 7 Clustering K-Means 5

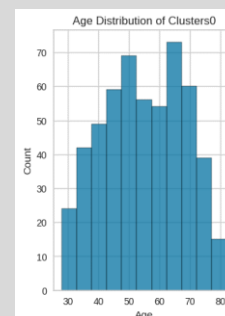


Figure 8 Profiling Age

```
CLUSTER 0
1 252
2 166
0 39
Name: Education, dtype: int64
0 246
1 211
Name: Is_Alone, dtype: int64
2 247
1 178
3 32
Name: Family_Size, dtype: int64
```

Figure 9 Profiling Education, Whether alone or not, and Family Size

Belinda Mutiara

BINUS Students 5th Semester

belinda.mutiara@binus.ac.id

phone: 0812 8308 7870

line: belinda.mutiara

linkedin: www.linkedin.com/in/belinda-mutiara

Portfolio Submission for

BINUS Internship Track 2024