

HA-5 Report

Task-1 (Clustering)

a) Screenshot of product features

Data columns (total 20 columns) :

#	Column	Non-Null Count	Dtype
0	avg_price	620 non-null	float64
1	bought_by	620 non-null	int64
2	sale_by	620 non-null	int64
3	Lunch	620 non-null	float64
4	Morning	620 non-null	float64
5	Afternoon	620 non-null	float64
6	Night	620 non-null	float64
7	wd_2	620 non-null	float64
8	wd_4	620 non-null	float64
9	wd_6	620 non-null	float64
10	wd_0	620 non-null	float64
11	wd_5	620 non-null	float64
12	wd_1	620 non-null	float64
13	wd_3	620 non-null	float64
14	season	620 non-null	int64
15	family_name	620 non-null	int32
16	spring_sale	620 non-null	float64
17	summer_sale	620 non-null	float64
18	autumn_sale	620 non-null	float64
19	winter_sale	620 non-null	float64

Features from 0 to 13 are as per Seminar plan and 14 to 19 newly generated features.

b) Explain (in a similar way I explain them in the plan) every single feature (you may skip features from the seminar plan) you use.

Explanation about newly generated features apart from the seminar plan.

season: In which season product had maximum sale.

Note: 0 indicates spring, 1 indicates summer,

2 indicates autumn and 3 indicates winter

family_name: family name of the product

spring_sale: total sale of the product in the spring season.

summer_sale: total sale of the product in the summer season.

autumn_sale: total sale of the product in the autumn season.

winter_sale: total sale of the product in the winter season.

c) Cluster's information: how many clusters do you have, how many objects are in these clusters.

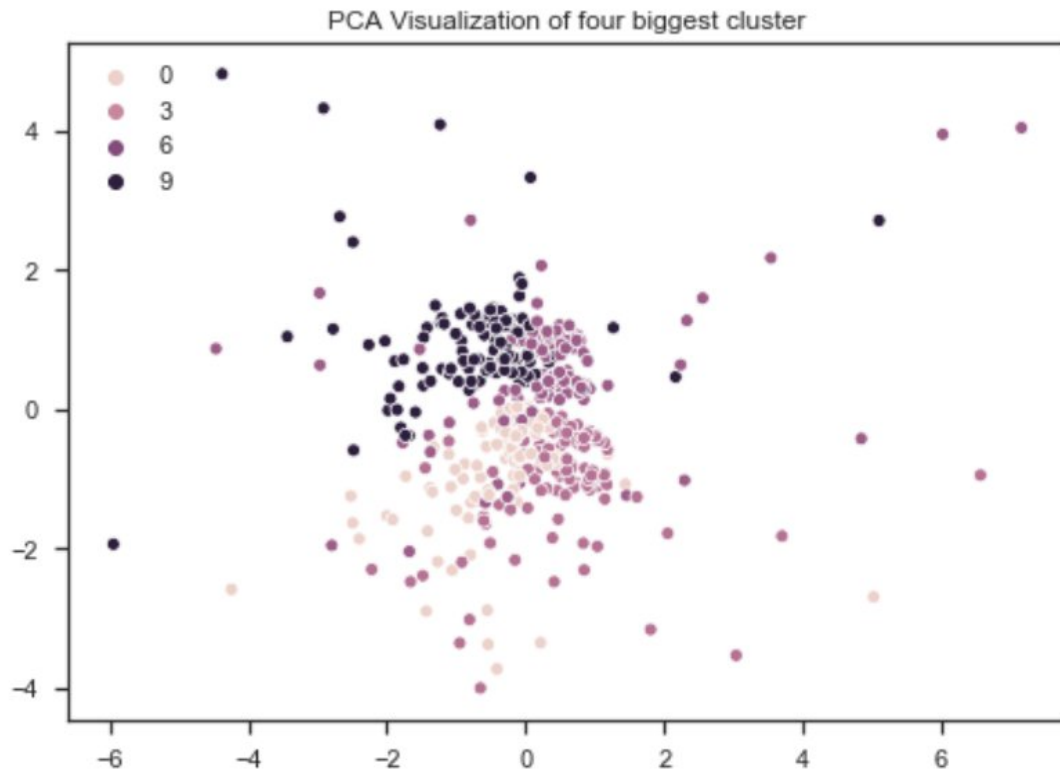
After applying elbow method, we got 11 clusters as an optimal value for our product data .

Below table depict the size of all the eleven clusters.

Cluster_Size	
Cluster_Number	
0	115
1	4
2	123
3	5
4	141
5	22
6	1
7	141
8	49
9	1
10	18

d) Cluster's interpretation. Try to provide an interpretation of every single cluster (or groups of clusters) you end up.

We have total eleven clusters. Let's interpret four biggest clusters:



Cluster_0 : It has high sale during the spring and bought in the Afternoon and Lunch.

Cluster_3: It has high sale during the spring & also summer and bought in the Afternoon and Lunch.

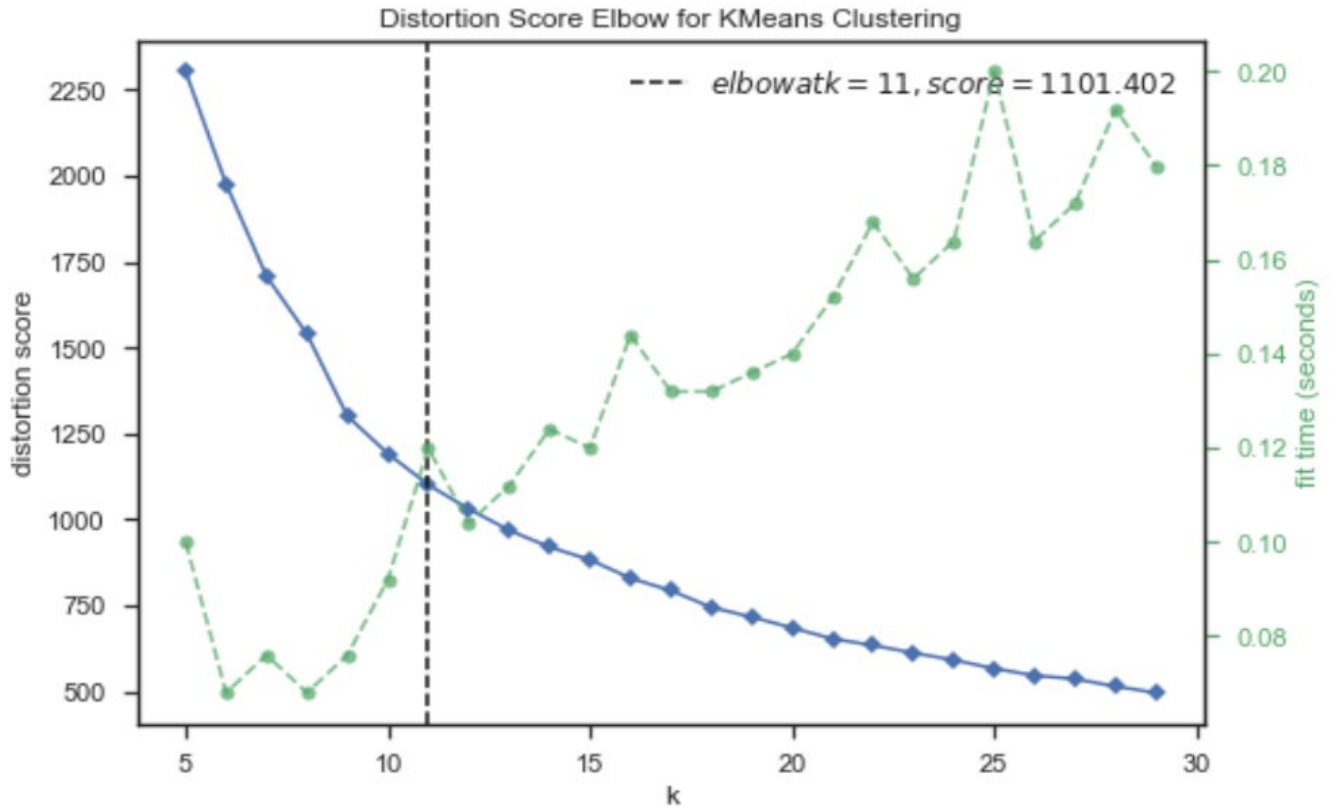
Cluster_6 : It has high sale during the winter and bought in the Night and also Afternoon, Lunch.

Cluster_9 : It has high sale during the autumn and bought in the Morning, Afternoon and also Lunch.

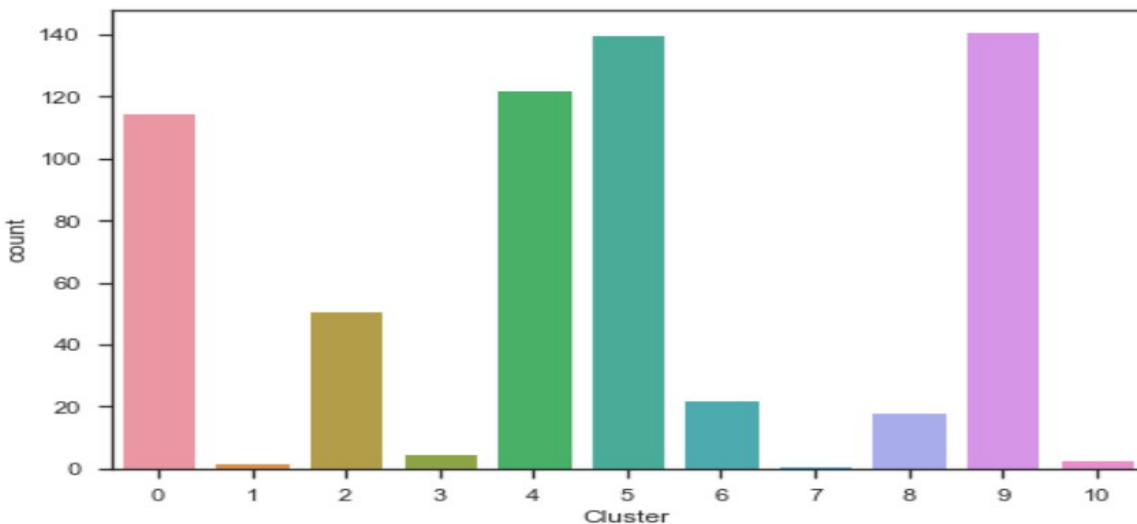
Our clusterization, gave the importance to season and hour features in the data.

e) You may include any visualization you find necessary, e.g.: colored PCA components, histogram or pie charts of cluster's sizes, “elbows” used for selection number of clusters (if you have used it).

Elbow method:



Count plot for Cluster Size:



Task-2 Recommender System

a. Report computed performance scores.

We considered the recommendation system problem as regression, below are the different regression performance scores of the model.

RMSE	6.830152
MSE	46.650975
MAE	2.658632
R^2	0.183562

Table: Performance Metrics

b. Elaborate on the quality of your recommendations.

asd fas

c. Provide 3-5 examples of `good` recommendations suggested by your recommender system.

	uid	iid	rui	est	details	err	lu	Ui
407	57	373	0.3	0.297916	{'actual_k': 6, 'was_impossible': False}	0.002084	22	6
365	60	617	0.3	0.294722	{'actual_k': 6, 'was_impossible': False}	0.005278	53	6
658	72	192	1.5	1.494464	{'actual_k': 30, 'was_impossible': False}	0.005536	61	38

Table-1: Three Best Recommendations

d. Provide 3-5 examples of 'bad' recommendations suggested by your recommender system.

	uid	iid	rui	est	details	err	Iu	Ui
840	19	74	72.4	13.658394	{'actual_k': 30, 'was_impossible': False}	58.741606	18	60
205	2	88	66.8	0.717059	{'actual_k': 10, 'was_impossible': False}	66.082941	13	10
140	96	72	83.3	2.608585	{'actual_k': 30, 'was_impossible': False}	80.691415	21	48

Table-2: Three Worst Recommendations

Iu: number of items bought by given user

Ui: number of users that have bought given item

e. Observations

Best Recommendations: In above table-1, Ui value is very small means that very few users have bought the target item which implies that set of neighbors are very few.

Worst Recommendations: In the above table-2, our model estimated the score very low but the user, item set had the high score. So, the model performed worst.

f. Number of Users bought the item-10

