

Main objective:

This analysis aims to provide customer segmentation to understand the main Target Customers so that the results can be given to the marketing team to plan the marketing strategy accordingly.

Data description :

The dataset represents some basic data about Mall customers like Customer ID, age, gender, annual income and spending score.

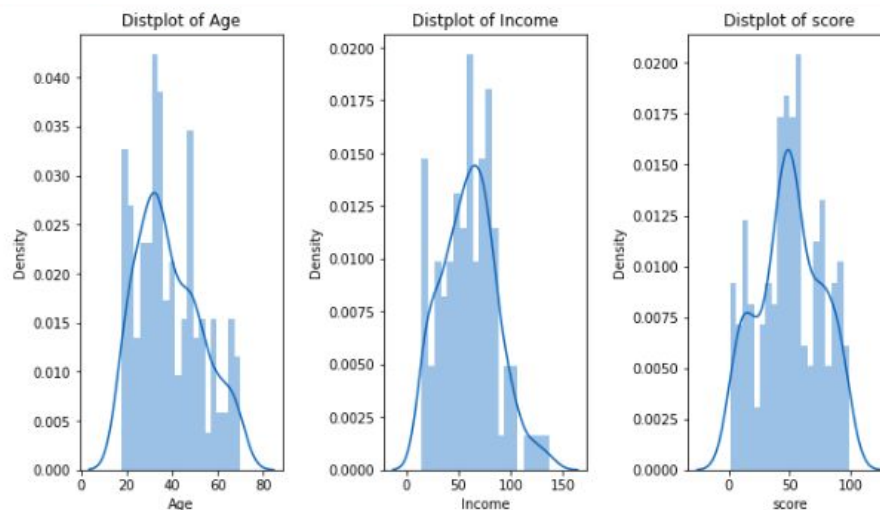
Spending Score is something you assign to the customer based on your defined parameters like customer behavior and purchasing data.

This dataset contains 200 rows and 5 columns

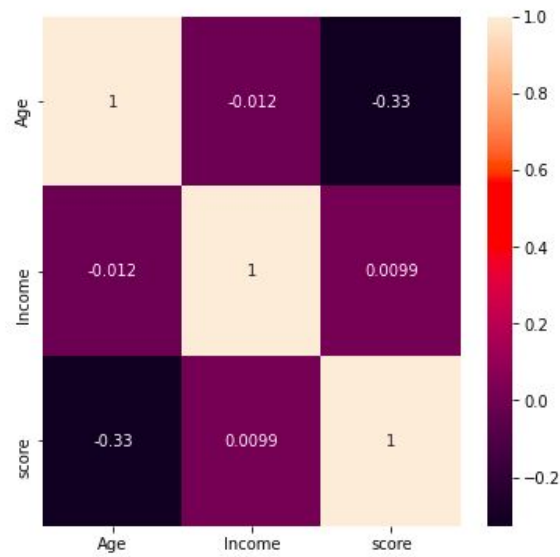
CustomerID int64
Gender object
Age int64
Annual Income (k\$) int64
Spending Score (1-100) int64

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

- Distribution of numeric columns:



- Correlation between data:



Data exploration and Data cleaning:

- We can drop the customer Id column as it's not effective for customer clustering.
- Transforming Gender column to binary values as it contains only 2 classes (Male ,Female)
- Renaming columns :
 'Annual Income (k\$)' : 'Income',
 'Spending Score (1-100)' : 'score'
- Dropping the Gender feature to build strategy on Age, Income,and Spending score.
- Scaling the data using MinMaxScaler
- Now data looks like this :

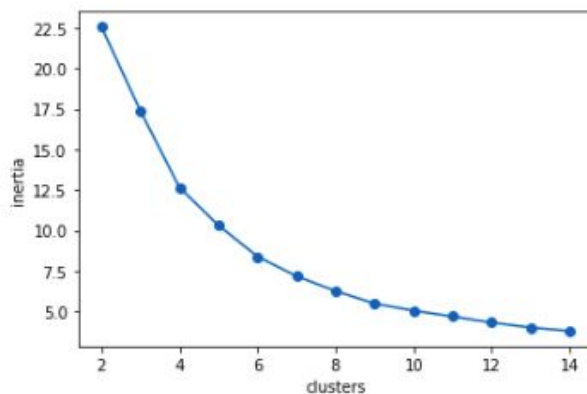
	Income	Age	score
0	0.000000	0.019231	0.387755
1	0.000000	0.057692	0.816327
2	0.008197	0.038462	0.051020
3	0.008197	0.096154	0.775510
4	0.016393	0.250000	0.397959
...
195	0.860656	0.326923	0.795918
196	0.909836	0.519231	0.275510
197	0.909836	0.269231	0.744898
198	1.000000	0.269231	0.173469
199	1.000000	0.230769	0.836735

200 rows × 3 columns

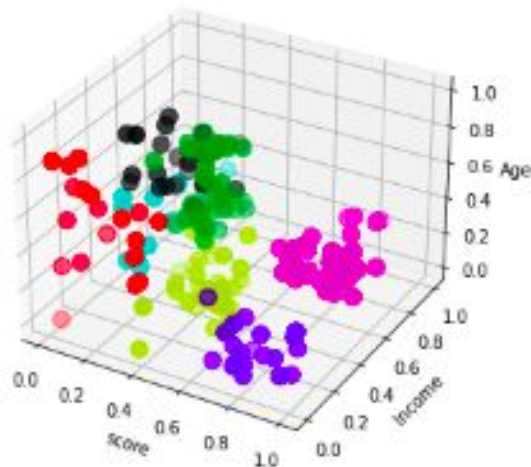
Unsupervised clustering Models:

1-Kmeans Model:

- To determine the number of clusters we try n_clusters in range (2-10)
- From the figure below 7 looks a good number for clustering



- Figure below shows the 7 clusters based on Age,Income,and score



- Statistics for the 7 clusters:

	Age	Income	score
count	20.000000	20.000000	20.000000
mean	0.543269	0.096311	0.177041
std	0.222689	0.059932	0.121788
min	0.038462	0.008197	0.020408
25%	0.360577	0.040984	0.051020
50%	0.557692	0.094262	0.137755
75%	0.677885	0.147541	0.290816
max	0.942308	0.196721	0.357143

	Age	Income	score
count	44.000000	44.000000	44.000000
mean	0.737325	0.317250	0.493738
std	0.164436	0.067536	0.061154
min	0.480769	0.188525	0.346939
25%	0.596154	0.262295	0.456633
50%	0.692308	0.319672	0.489796
75%	0.908654	0.385246	0.551020
max	1.000000	0.426230	0.602041

	Age	Income	score
count	22.000000	22.000000	22.000000
mean	0.139860	0.087928	0.799629
std	0.101097	0.062022	0.107185
min	0.000000	0.000000	0.612245
25%	0.062500	0.034836	0.734694
50%	0.105769	0.077869	0.775510
75%	0.225962	0.141393	0.864796
max	0.326923	0.196721	1.000000

	Age	Income	score
count	39.000000	39.000000	39.000000
mean	0.282544	0.586381	0.827839
std	0.071705	0.133709	0.095556
min	0.173077	0.442623	0.632653
25%	0.230769	0.495902	0.750000
50%	0.269231	0.524590	0.836735
75%	0.336538	0.655738	0.908163
max	0.423077	1.000000	0.979592

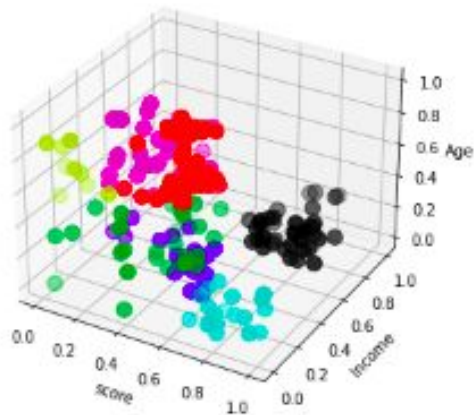
	Age	Income	score
count	14.000000	14.000000	14.000000
mean	0.252747	0.608899	0.126093
std	0.139056	0.147381	0.098900
min	0.019231	0.475410	0.000000
25%	0.168269	0.516393	0.048469
50%	0.307692	0.565574	0.102041
75%	0.346154	0.651639	0.204082
max	0.423077	1.000000	0.316327

	Age	Income	score
count	40.000000	40.000000	40.000000
mean	0.171154	0.324795	0.486224
std	0.134566	0.108022	0.081027
min	0.000000	0.000000	0.285714
25%	0.052885	0.266393	0.418367
50%	0.163462	0.356557	0.494898
75%	0.254808	0.395492	0.551020
max	0.423077	0.500000	0.612245

	Age	Income	score
count	21.000000	21.000000	21.000000
mean	0.590659	0.594457	0.192906
std	0.118880	0.128027	0.101113
min	0.442308	0.459016	0.040816
25%	0.500000	0.508197	0.132653
50%	0.557692	0.573770	0.163265
75%	0.692308	0.680328	0.255102
max	0.788462	0.909836	0.387755

2- AgglomerativeClustering:

- Using 7 clusters and ward linkage:



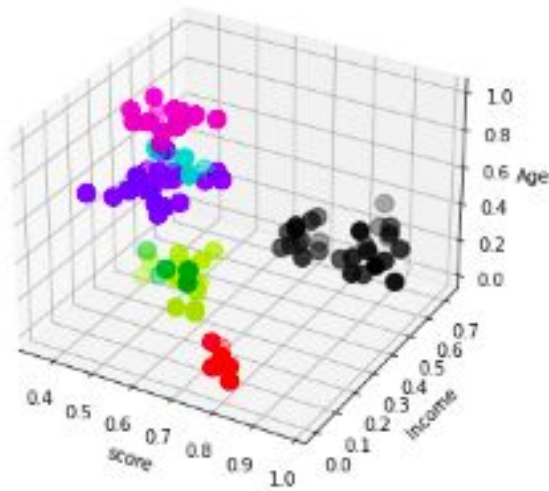
- Statistics for the 7 clusters:

	Age	Income	score
count	47.000000	47.000000	47.000000
mean	0.731997	0.324555	0.476118
std	0.160516	0.083410	0.077018
min	0.480769	0.188525	0.275510
25%	0.596154	0.258197	0.428571
50%	0.692308	0.319672	0.479592
75%	0.903846	0.389344	0.535714
max	1.000000	0.524590	0.602041
	Age	Income	score
count	30.000000	30.000000	30.000000
mean	0.319872	0.208470	0.409184
std	0.122304	0.133696	0.157634
min	0.019231	0.000000	0.051020
25%	0.250000	0.106557	0.346939
50%	0.326923	0.204918	0.418367
75%	0.413462	0.319672	0.525510
max	0.519231	0.442623	0.612245
	Age	Income	score
count	26.000000	26.000000	26.000000
mean	0.073964	0.392182	0.425432
std	0.062420	0.075541	0.172102
min	0.000000	0.254098	0.040816
25%	0.019231	0.348361	0.400510
50%	0.057692	0.385246	0.494898
75%	0.129808	0.444672	0.548469
max	0.192308	0.540984	0.591837
	Age	Income	score
count	28.000000	28.000000	28.000000
mean	0.497940	0.625293	0.159985
std	0.160328	0.138652	0.091566
min	0.269231	0.459016	0.000000
25%	0.360577	0.516393	0.099490
50%	0.490385	0.590164	0.153061
75%	0.572115	0.692623	0.216837
max	0.788462	1.000000	0.387755

	Age	Income	score
count	21.000000	21.000000	21.000000
mean	0.130952	0.087041	0.808552
std	0.094333	0.063411	0.101113
min	0.000000	0.000000	0.653061
25%	0.057692	0.032787	0.734694
50%	0.096154	0.073770	0.775510
75%	0.211538	0.147541	0.877551
max	0.326923	0.196721	1.000000
	Age	Income	score
count	9.000000	9.000000	9.000000
mean	0.728632	0.086521	0.105442
std	0.133042	0.046748	0.085678
min	0.538462	0.032787	0.020408
25%	0.653846	0.040984	0.030612
50%	0.692308	0.081967	0.132653
75%	0.807692	0.122951	0.132653
max	0.942308	0.147541	0.285714
	Age	Income	score
count	39.000000	39.000000	39.000000
mean	0.282544	0.586381	0.827839
std	0.071705	0.133709	0.095556
min	0.173077	0.442623	0.632653
25%	0.230769	0.495902	0.750000
50%	0.269231	0.524590	0.836735
75%	0.336538	0.655738	0.908163
max	0.423077	1.000000	0.979592

3-DBSCAN Model:

- Using $\text{eps} = 0.1$ and $\text{min_samples} = 5$ the model find 7 clusters :



- Statistics of the clusters:

	Age	Income	score
count	6.000000	6.000000	6.000000
mean	0.775641	0.297814	0.535714
std	0.037815	0.040044	0.053511
min	0.711538	0.229508	0.469388
25%	0.759615	0.286885	0.494898
50%	0.788462	0.303279	0.535714
75%	0.802885	0.319672	0.576531
max	0.807692	0.344262	0.602041
	Age	Income	score
count	19.000000	19.000000	19.000000
mean	0.067814	0.356342	0.513426
std	0.061274	0.050923	0.057728
min	0.000000	0.254098	0.408163
25%	0.019231	0.319672	0.484694
50%	0.057692	0.368852	0.520408
75%	0.105769	0.389344	0.551020
max	0.173077	0.426230	0.591837
	Age	Income	score
count	28.000000	28.000000	28.000000
mean	0.280220	0.538934	0.836735

	Age	Income	score
count	7.000000	7.000000	7.000000
mean	0.087912	0.043326	0.771137
std	0.029074	0.043040	0.029957
min	0.057692	0.000000	0.734694
25%	0.067308	0.012295	0.750000
50%	0.076923	0.040984	0.775510
75%	0.105769	0.057377	0.785714
max	0.134615	0.122951	0.816327

	Age	Income	score
count	5.000000	5.000000	5.000000
mean	0.246154	0.234426	0.546939
std	0.043853	0.028747	0.060626
min	0.173077	0.196721	0.469388
25%	0.250000	0.221311	0.510204
50%	0.250000	0.229508	0.540816
75%	0.269231	0.254098	0.602041
max	0.288462	0.270492	0.612245

	Age	Income	score
count	23.000000	23.000000	23.000000
mean	0.591137	0.317177	0.475599
std	0.046465	0.074857	0.058758
min	0.480769	0.196721	0.357143
25%	0.567308	0.245902	0.433673
50%	0.596154	0.319672	0.459184
75%	0.615385	0.381148	0.510204
max	0.692308	0.426230	0.591837

	Age	Income	score
count	14.000000	14.000000	14.000000
mean	0.940934	0.330211	0.499271
std	0.042311	0.065133	0.055091
min	0.865385	0.237705	0.408163
25%	0.923077	0.270492	0.464286
50%	0.942308	0.340164	0.505102
75%	0.961538	0.393443	0.543367
max	1.000000	0.409836	0.591837

Recommendation:

Using the DBSCAN model eliminates most of the noise in clusters so it's recommended to use that model.

key findings :

From studying the above models, as shown in the numbers the clusters with the highest mean score are :

- Young Age with low income
- Medium Age with medium to high income

So Marketing can focus more on these two groups.

for Kmeans Model:

	Age	Income	score
count	22.000000	22.000000	22.000000
mean	0.13986	0.087928	0.799629

	Age	Income	score
count	39.000000	39.000000	39.000000
mean	0.282544	0.586381	0.827839

Agglomerative Clustering Model

	Age	Income	score
count	21.000000	21.000000	21.000000
mean	0.130952	0.087041	0.808552

	Age	Income	score
count	39.000000	39.000000	39.000000
mean	0.282544	0.586381	0.827839

for DBSCAN Model:

	Age	Income	score
count	7.000000	7.000000	7.000000
mean	0.087912	0.043326	0.771137

	Age	Income	score
count	28.000000	28.000000	28.000000
mean	0.28022	0.538934	0.836735

next steps:

This Dataset contains data about 200 customers only,so getting more data will help providing a better insight on customers categories and best ways to approach them.