k-Means — Tue Mar 19 24, 16:28:00

Number of clusters: 8
Optimization: initialize with KMeans++, 10 re-runs limited to 300 steps

Data

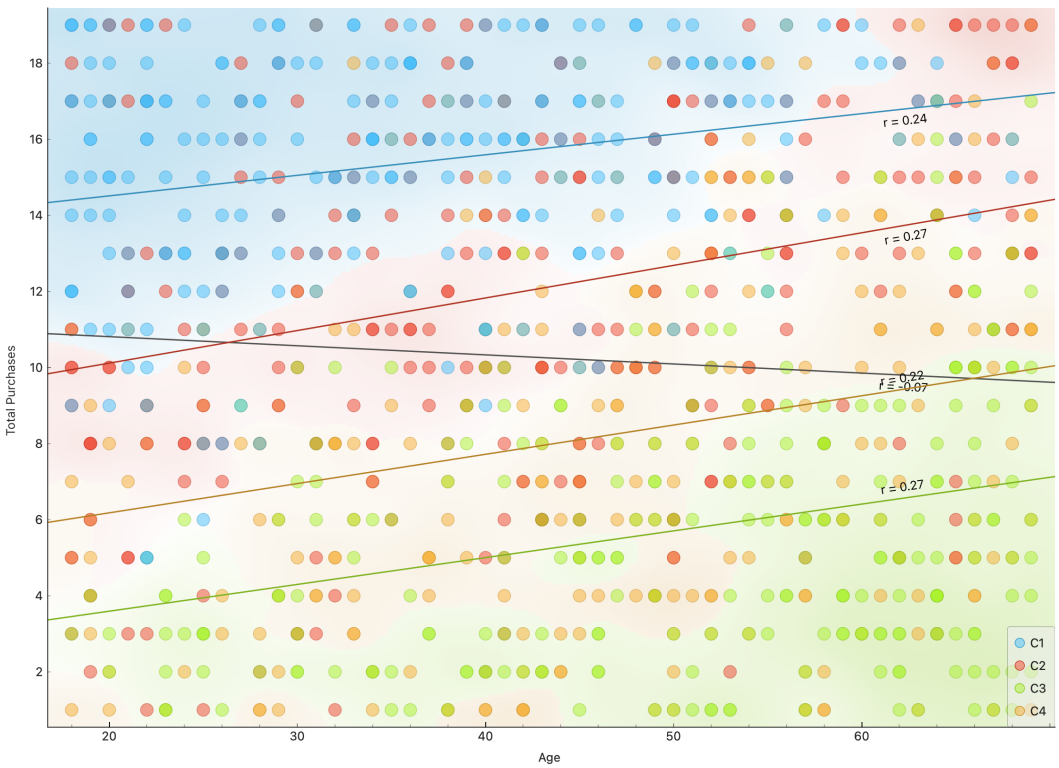
Data instances: 1000
Features: CustomerID, Age, Annual Income (k$), Spending Score (1-100), Gender, Total Purchases, Category Most Purchased, Purchased Recently

Silhouette scores for different numbers of clusters

2	0.111
3	0.104
4	0.112
5	0.109
6	0.111
7	0.109
8	0.113

When using K-Means it can be difficult to choose how many clusters, especially as our data does not show many patterns. This is further proved as the spread of Silhouette scores for recommended number of clusters is also spread very thin. With there being no preference to the number of clusters, I am going to choose 4 as the number of clusters in the next step.

Scatter Plot — Tue Mar 19 24, 16:36:15



Color: Cluster

This graph looks at age compared to total purchases with the colour showing the different clusters. Unfortunately there does not seem to be any real segmentation that is occuring which suggests that differences in age do not act as a good variable to determine total purchases.