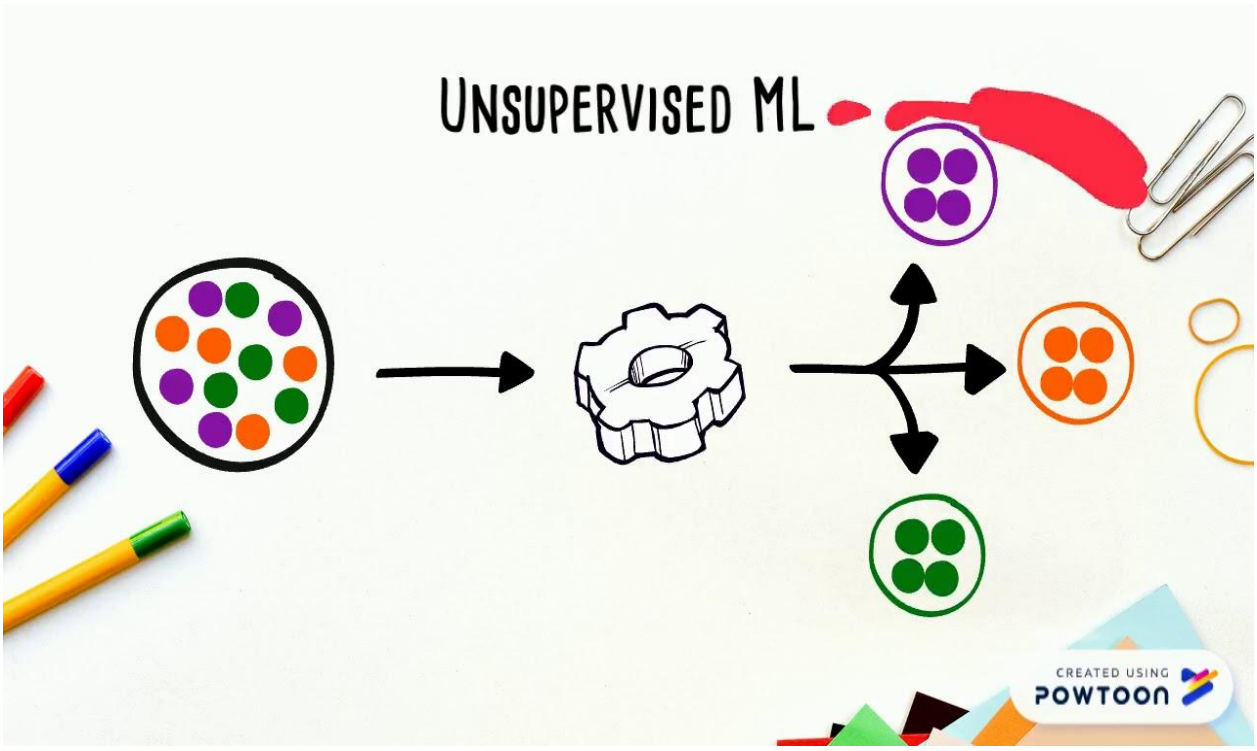


Unsupervised Learning Project

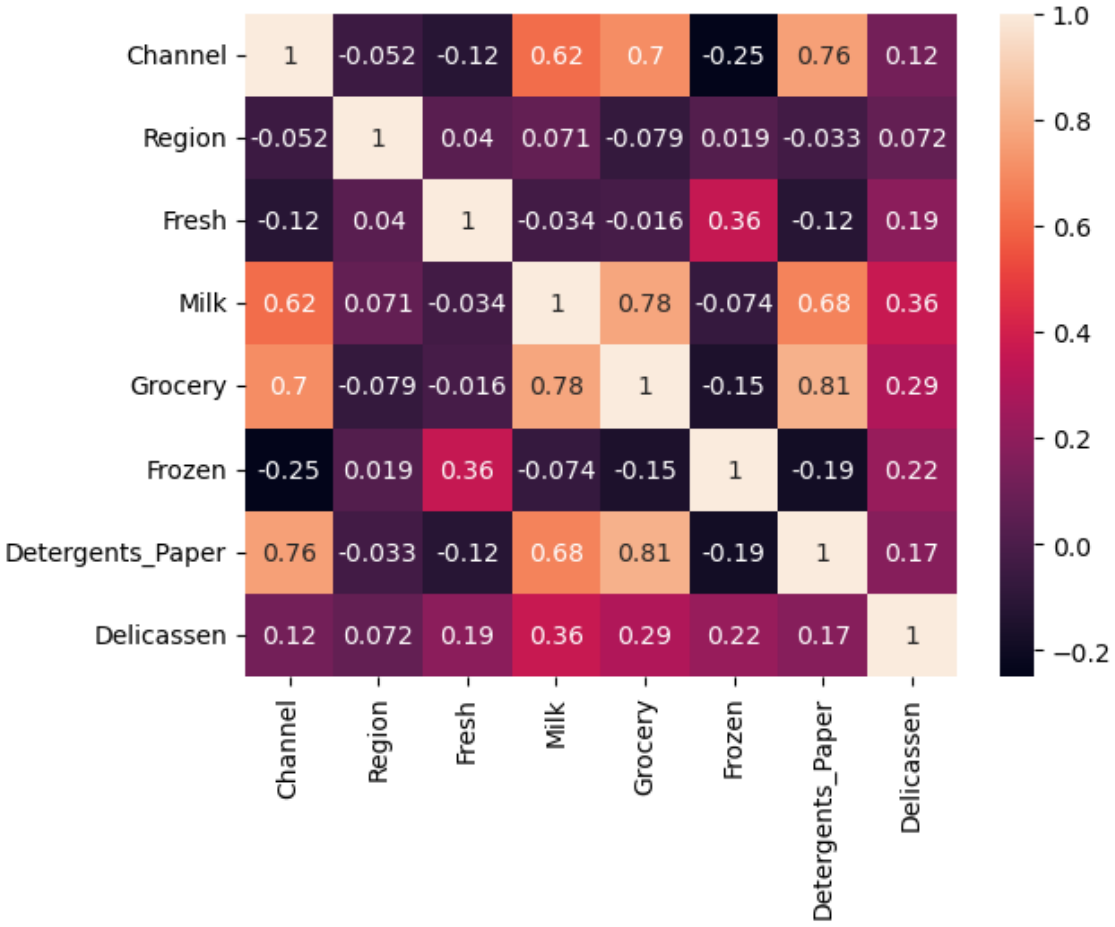
Zeynab Akolade

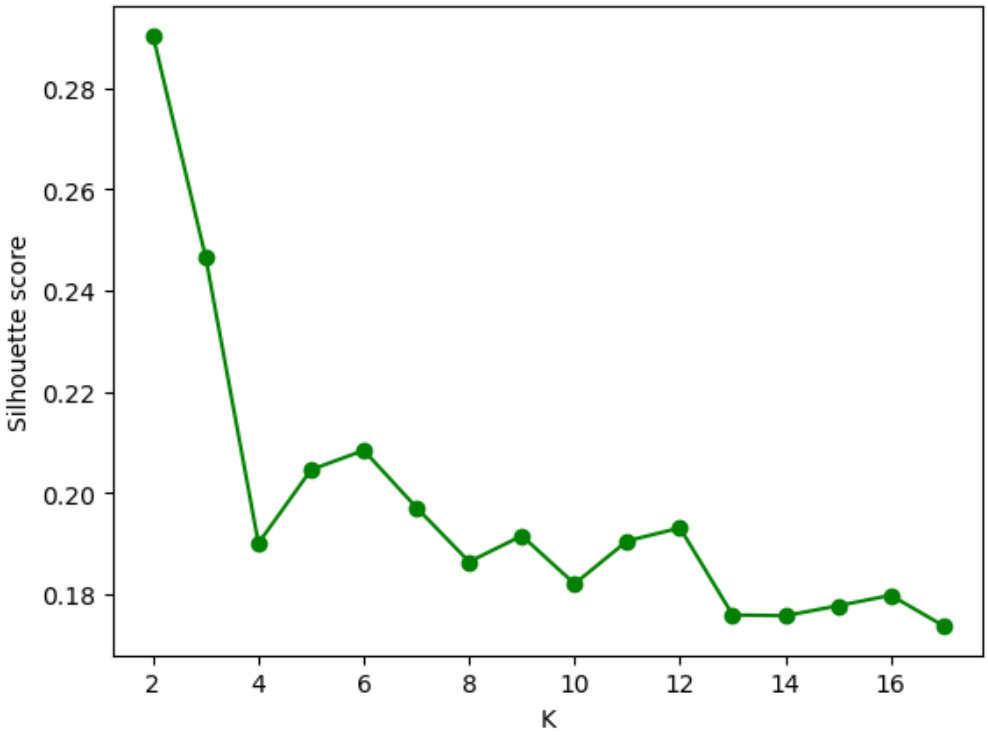
The dataset pertains to clients of a wholesale distributor and contains information about their annual expenditures in monetary units across various product categories.

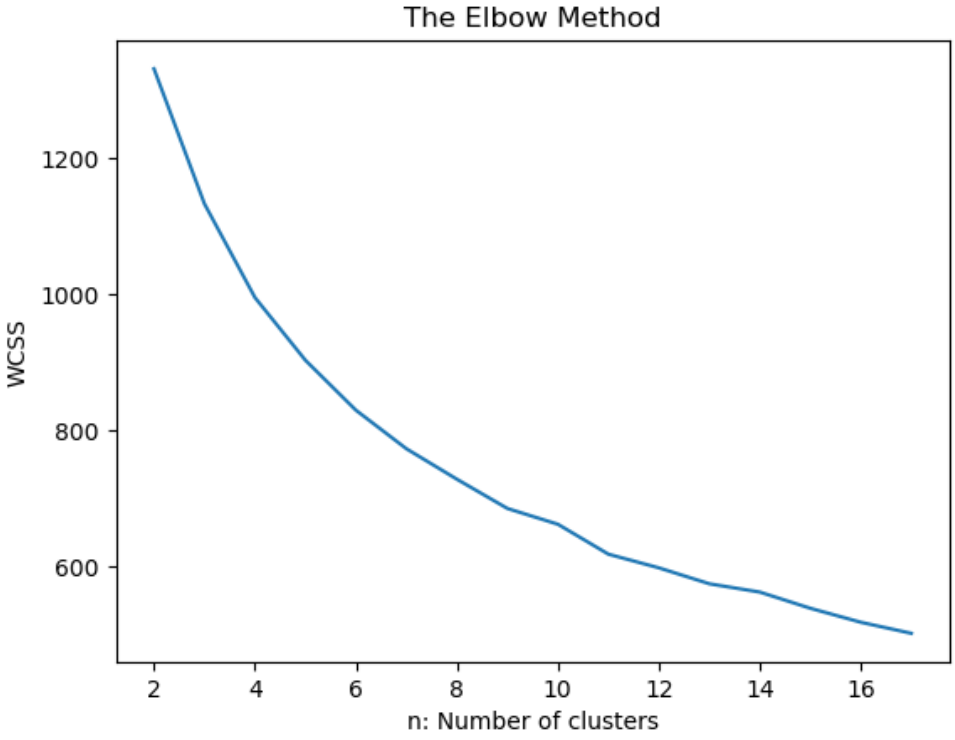
Conclusions:

1. Hierarchical clustering demonstrated superior performance in grouping customers.
2. Hierarchical clustering provided a more streamlined outcome, yielding two clusters as opposed to eleven in the K-means clustering model.
3. Visualization of K-means clustering was not very informative due to the two-dimensional nature of the data.
4. Determining the optimal number of clusters in the K-means clustering model was challenging, as the elbow method did not exhibit a clear elbow point.
5. PCA analysis revealed that only three components were necessary to effectively describe customers, as these components maximized the variance.
6. For future analyses, one option is to explore alternative methods for handling outliers, as the current IQR-based approach resulted in the removal of a significant number of rows. Subsequently, the models can be reevaluated with this revised dataset.
7. Another potential avenue for future analysis is to experiment with different ways of segregating the data, such as further column removal or the addition of dummy columns, to potentially yield more insights.

Some graphical representations created:




Silhouette score
0.28
0.26
0.24
0.22
0.20
0.18
2
4
6
8
10
12
14
16
K


The Elbow Method
WCSS
1200
1000
800
600
2
4
6
8
10
12
14
16
n: Number of clusters

```
<AxesSubplot:ylabel='Density'>
```