

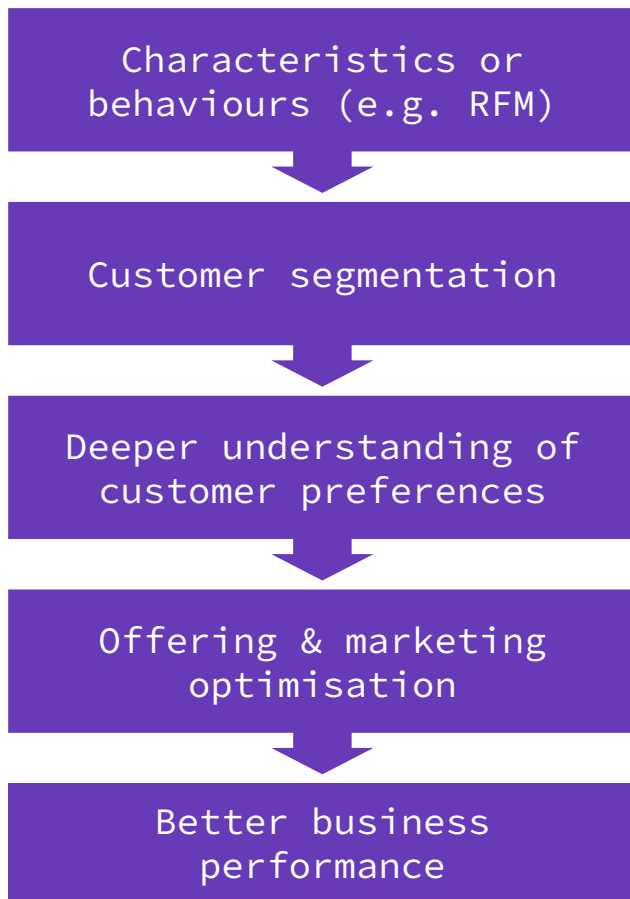
Customer Segmentation in E-commerce Using Clustering Techniques

Darya Solomenko - 16th of August, 2024

Customer Segmentation

— — —

- The global e-commerce market is expected to total **\$6.3 trillion in 2024**.
- 20.1% of retail purchases are expected to take place online.



Data set

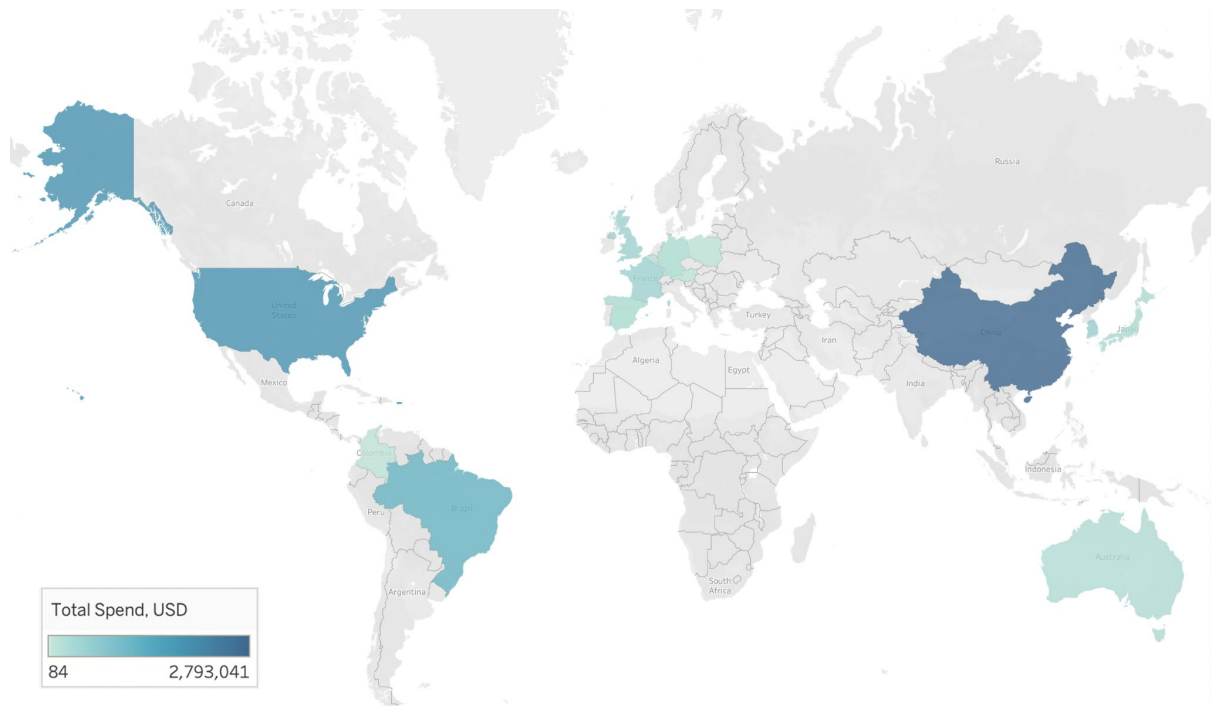
— — —

TheLook a fictitious e-commerce clothing store, operating on a global scale.

37,060 rows, 17 features.

2022 - 2023 customer and sales data for.

+ Complimentary demographic dataset.



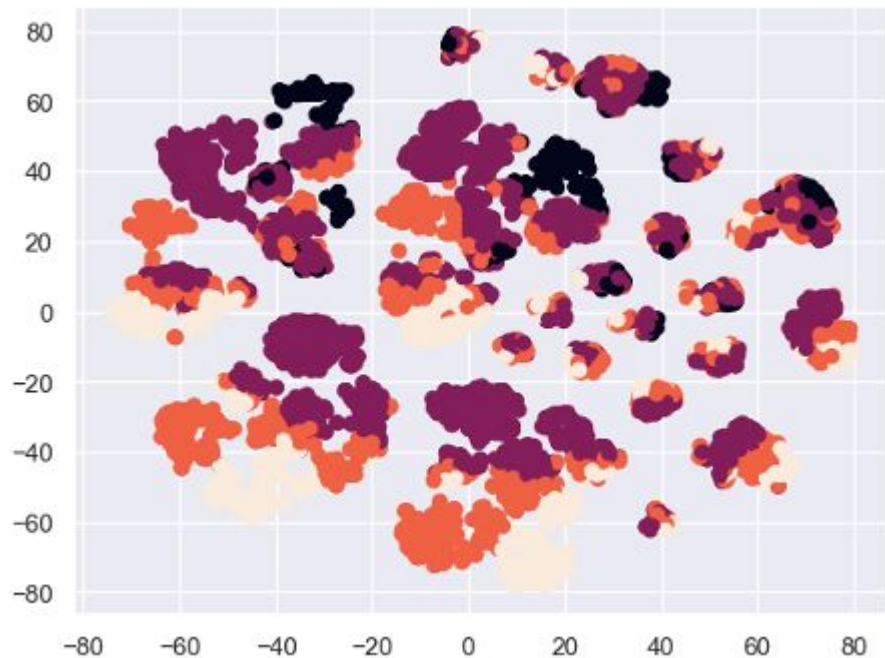
t-SNE

— — —

Dimensionality reduction technique applied for an easier visualisation of potential clusters in the 2D space.

Indication that optimal number of clusters for all features will be **greater than ten**.

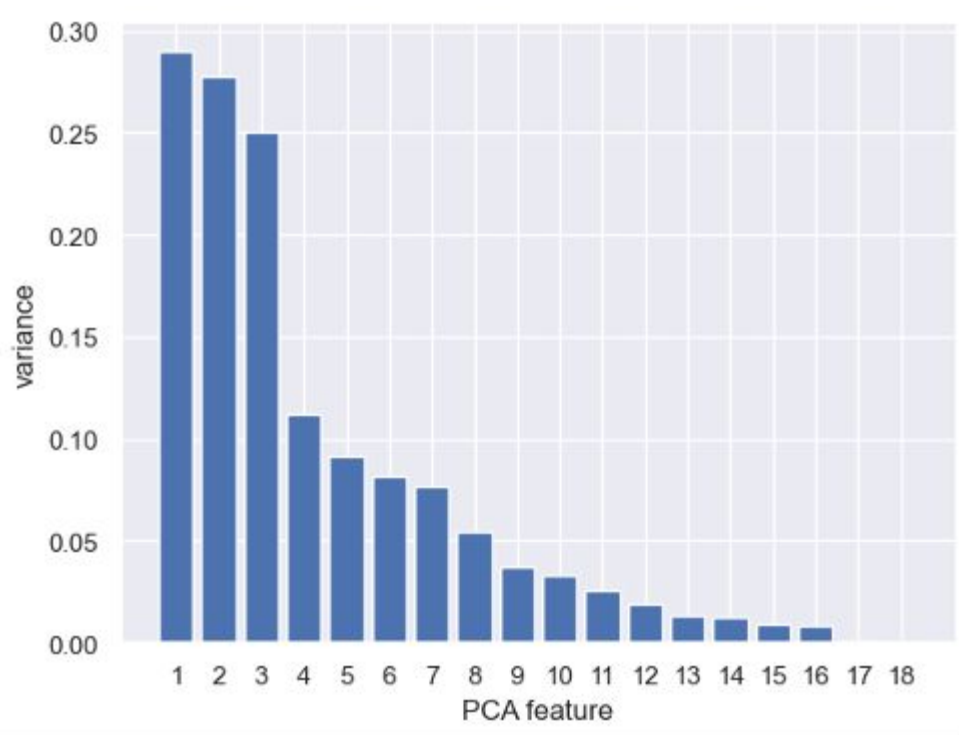
T-SNE scatter plot for all features
coloured by income category



PCA

— — —
First 3 three principal components captured 82% of the variance in the complete dataset.

Explained variance only 25% when considering numeric features alone.



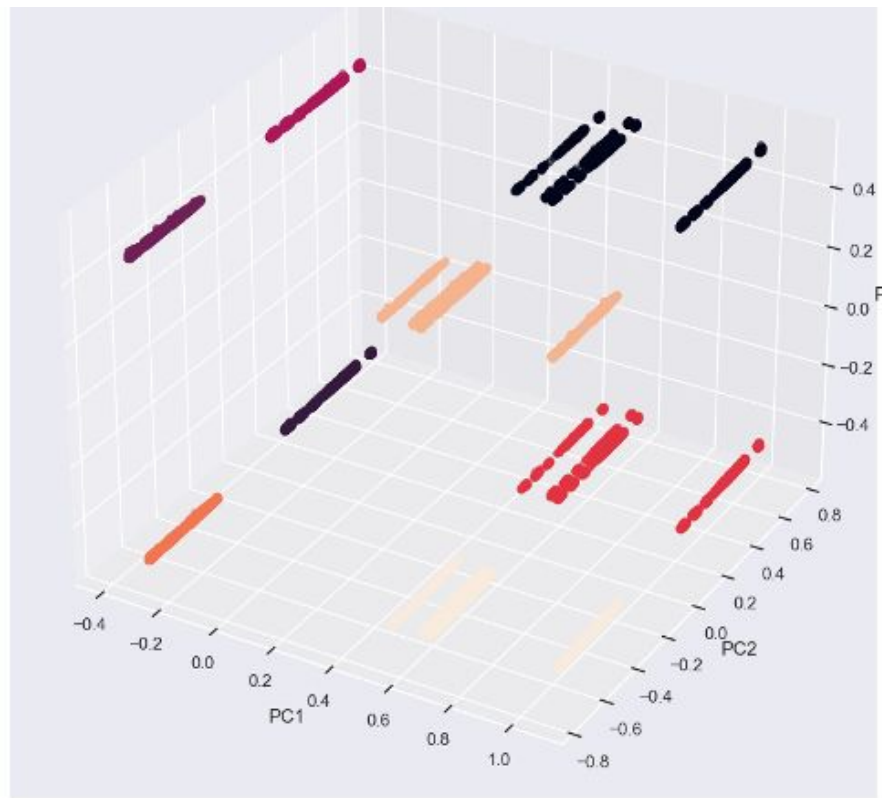
Modelling - PC features

4 models:

- K-means (8 clusters)
- Agglomerative ('average' linkage)
- MeanShift (bandwidth 0.4)
- DBSCAN (epsilon 0.3)

Models	Clusters	Silhouette score	DBI	Execution time, s
K-Means	8	0.77	0.44	0.05
Agglomerative	2	0.39	1.29	32.09
MeanShift	12	0.77	0.46	18.56
DBSCAN	12	0.77	0.46	1.00

3D plot of principal components
coloured by cluster label (K-means)



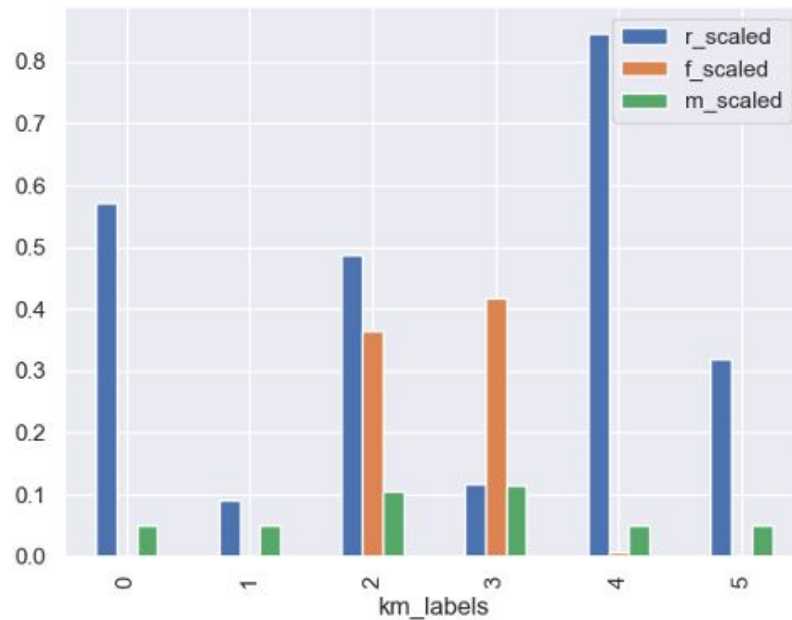
Modelling - RFM features

2 models:

- K-means (6 clusters)
- DBSCAN (epsilon 0.3)

Models	Clusters	Silhouette score	DBI	Execution time, s
K-Means	6	0.48	113.34	0.06
DBSCAN	4	0.33	1.08	4.29

Scaled RFM features by label (K-means)



Conclusion

— — —

- **Best performance:** The K-means algorithm with principal component features. **But:** segments lacked clear business relevance and reflected categorical data patterns.
- **Best business value:** RFM-based clustering -> two high-value customer segments comprising 19% of the customer base. **But:** produced lower quality clusters.

Future research:

- Validate findings using real-world data.
- Expand the RFM model to incorporate customer lifetime (LRFM) and granular product category information.

Thank you!