UDACITY

PROJECT SPECIFICATION

Creating Customer Segments

Data Exploration

CRITERIA	MEETS SPECIFICATIONS
Question 1 Selecting Samples	Three separate samples of the data are chosen and their establishment representations are proposed based on the statistical description of the dataset.
Question 2 Feature Relevance	A prediction score for the removed feature is accurately reported. Justification is made for whether the removed feature is relevant.
Question 3 Feature Distributions	Student identifies features that are correlated and compares these features to the predicted feature. Student further discusses the data distribution for those features.

Data Preprocessing

CRITERIA	MEETS SPECIFICATIONS
Feature Scaling	Feature scaling for both the data and the sample data has been properly implemented in code.

CRITERIA	MEETS SPECIFICATIONS
Question 4 Outlier Detection	Student identifies extreme outliers and discusses whether the outliers should be removed. Justification is made for any data points removed.

Feature Transformation

CRITERIA	MEETS SPECIFICATIONS
Question 5 Principal Component Analysis	The total variance explained for two and four dimensions of the data from PCA is accurately reported. The first four dimensions are interpreted as a representation of customer spending with justification.
Dimensionality Reduction	PCA has been properly implemented and applied to both the scaled data and scaled sample data for the two-dimensional case in code.

Clustering

CRITERIA	MEETS SPECIFICATIONS
Question 6 Clustering Algorithm	The Gaussian Mixture Model and K-Means algorithms have been compared in detail. Student's choice of algorithm is justified based on the characteristics of the algorithm and data.

CRITERIA	MEETS SPECIFICATIONS
Question 7 Creating Clusters	Several silhouette scores are accurately reported, and the optimal number of clusters is chosen based on the best reported score. The cluster visualization provided produces the optimal number of clusters based on the clustering algorithm chosen.
Question 8 Data Recovery	The establishments represented by each customer segment are proposed based on the statistical description of the dataset. The inverse transformation and inverse scaling has been properly implemented and applied to the cluster centers in code.
Question 9 Sample Predictions	Sample points are correctly identified by customer segment, and the predicted cluster for each sample point is discussed.

Conclusion

CRITERIA	MEETS SPECIFICATIONS
Question 10 A/B Test	Student correctly identifies how an A/B test can be performed on customers after a change in the wholesale distributor's service.
Question 11 Predicting Additional Data	Student discusses with justification how the clustering data can be used in a supervised learner for new predictions.

CRITERIA	MEETS SPECIFICATIONS
Question 12 Comparing Customer Data	Comparison is made between customer segments and customer 'Channel' data. Discussion of customer segments being identified by 'Channel' data is provided, including whether this representation is consistent with previous results.

Student FAQ