Creating Customer Segments
Project Rubric

Overview:

This rubric is here to help you understand the specifications for how your project will be evaluated. It is the same rubric you should share with others who give you feedback. You should look at the rubric before you begin working on this project and before you submit it.

Before you begin:

1. Read the project instructions carefully.

Before you submit:

1. Read the rubric below in detail and do your best to evaluate where your project stands.
2. If you think your project does not meet specifications for any criterion, make necessary changes so that it "meets specifications".
3. When you are confident that your project meets or exceeds specifications in each criterion, share it with others for feedback

The Rubric:

Criteria	Meets Specifications
Functionality	
TRANSITION: Implementation	Code in required implementation sections produce the correct results. Results do not conflict with the answers provided.
Data Exploration	
OPTIONAL: Selecting Samples	Three separate samples of the data are chosen and their establishment representations are proposed and justified.
OPTIONAL: Feature Relevance	A prediction score for the removed feature is accurately reported. Justification is made for whether the removed feature is relevant.
OPTIONAL: 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	
OPTIONAL: Outlier Detection	Student identifies extreme outliers and discusses whether the outliers should be removed. Justification is made for any data points removed.
Feature Transformation	
PCA	The total variance explained for two and four dimensions of the data from PCA is accurately reported, or a number of dimensions is chosen based on the total variance explained by PCA. The first four dimensions are correctly interpreted as a representation of customer spending.
Clustering	

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.
Creating Clusters	The cluster visualization provided produces the optimal number of clusters based on the clustering algorithm chosen. Optionally, several silhouette scores are accurately reported, and the optimal number of clusters is chosen based on the best reported score.
Data Recovery	The establishments represented by each customer segment are discussed with justification.
OPTIONAL: Sample Predictions	Sample points are correctly identified by customer segment, and the predicted cluster for each sample point is discussed.

Conclusion

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.
Predicting Additional Features	Student discusses with justification how the clustering data can be used in a supervised learner for predicting new features.
OPTIONAL: 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.