K-MEANS Model

Method

Using PC analysis for dimension reduction. I will be able to select the select the variables that will enable the best exploratory analysis and the ability to predict credit card fraudulent activities with other relational features.

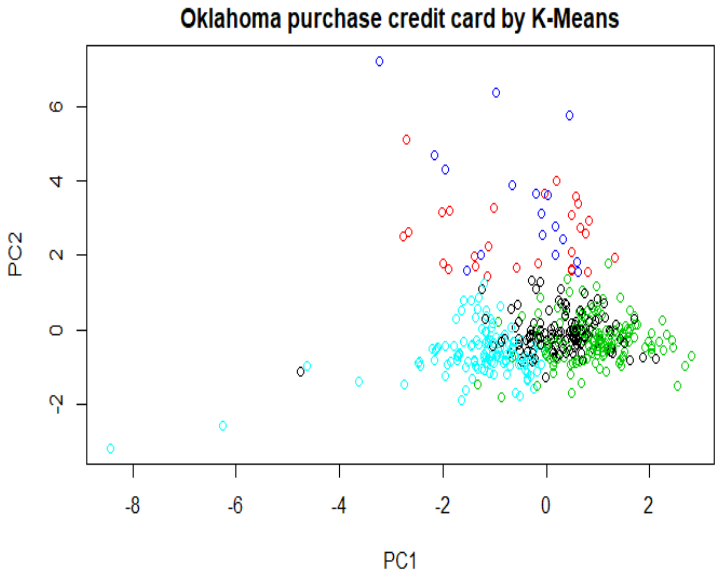
Model

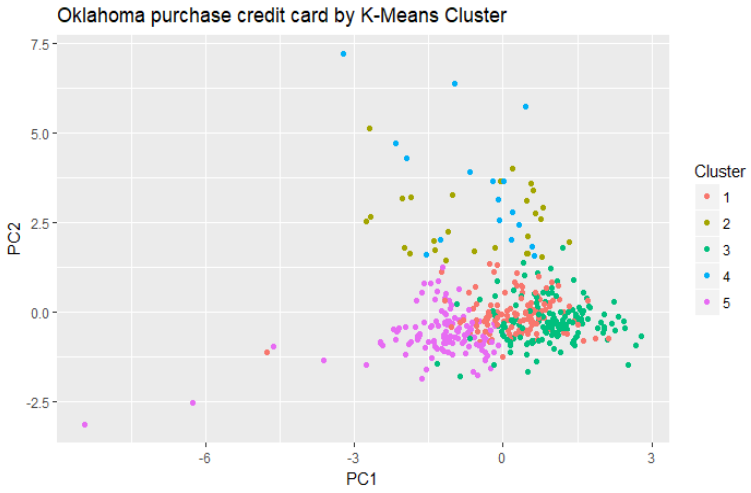
K-Means clustering known as the nearest neighbor approach due to all points connecting to each other based on the centroid location and grouping together based on recalculating the mean for that point, which becomes the new centroid.

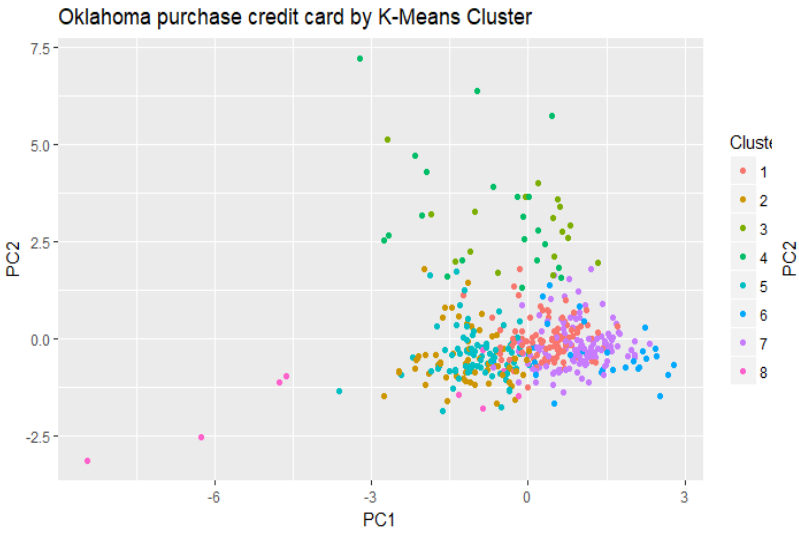
The PCA method required hyper-parameters where 1 percent was used as a test due to the large amount of rows or when transposed the columns increased dramatically. Also, I increased the clusters (5, 8, 12) for each model to observe more groups outliers and relationships concerning the transaction distribution or spread.

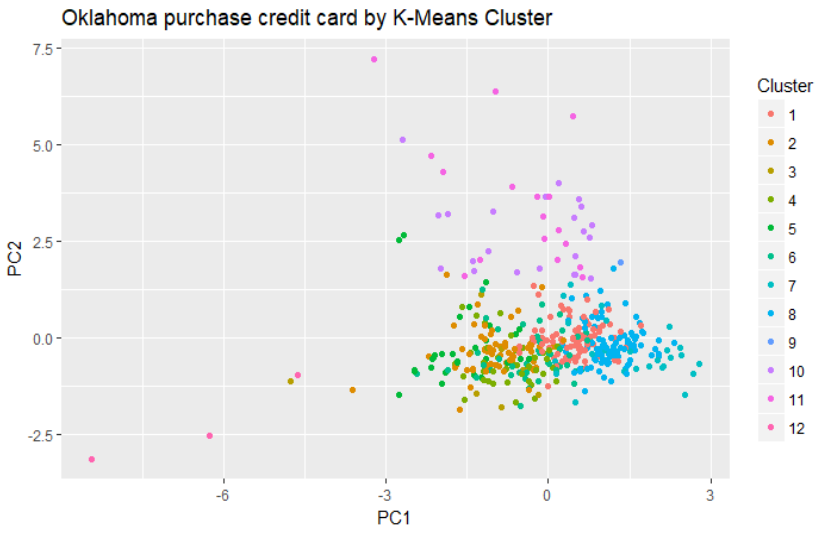
Insights

The outliers are located at very low average and extremely high purchases, can indicate rare type of purchases from certain agencies. There are more clustered outliers available or visible to observe transaction irregularities.









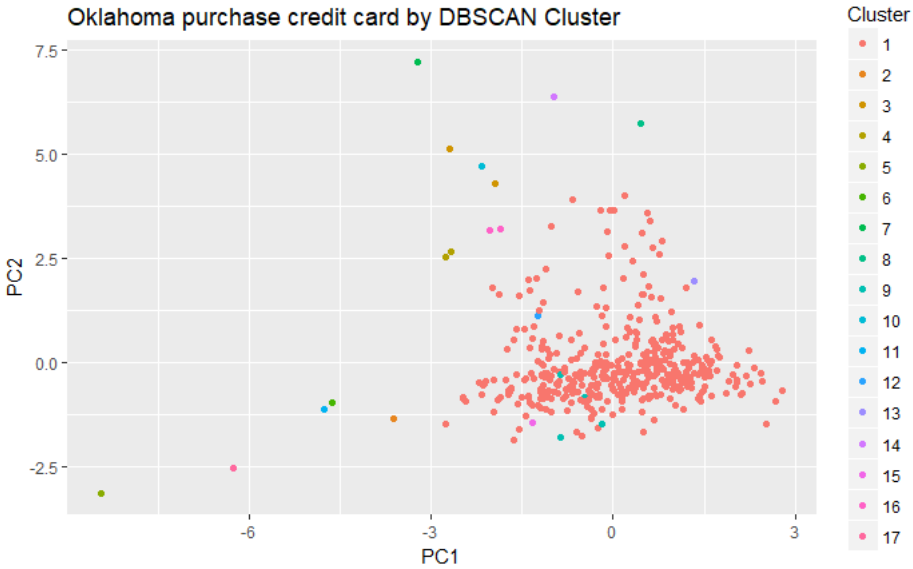
DBSCAN Model

Method

Using PC analysis for dimension reduction. I will be able to select the select the variables that will enable the best exploratory analysis and the ability to predict credit card fraudulent activities with other relational features.

Model

This DBSCAN model groups the variances that are closely related to each other in terms of distance and those points that is too far away are considered outliers. The noise follows familiar activities and ignore the irregular ones (outliers). The PCA method required hyper-parameters where 1 percent was used as a test due to the large amount of rows or when transposed the columns increased dramatically.



Insights

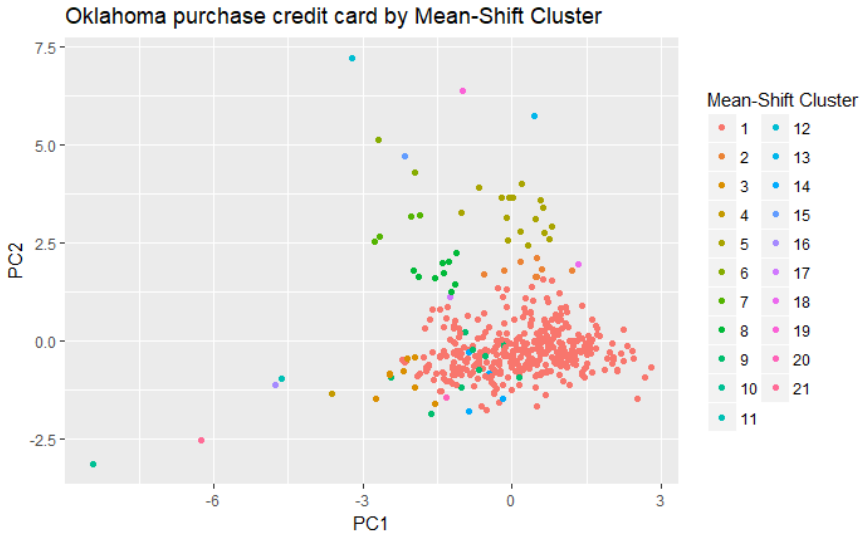
This model shows a higher correlations of relational clusters. There are fewer clustered outliers available or visible to observe transaction irregularities.

Mean-Shift Model

Method
Using PC analysis for dimension reduction. I will be able to select the select the variables that will enable the best exploratory analysis and the ability to predict credit card fraudulent activities with other relational features.

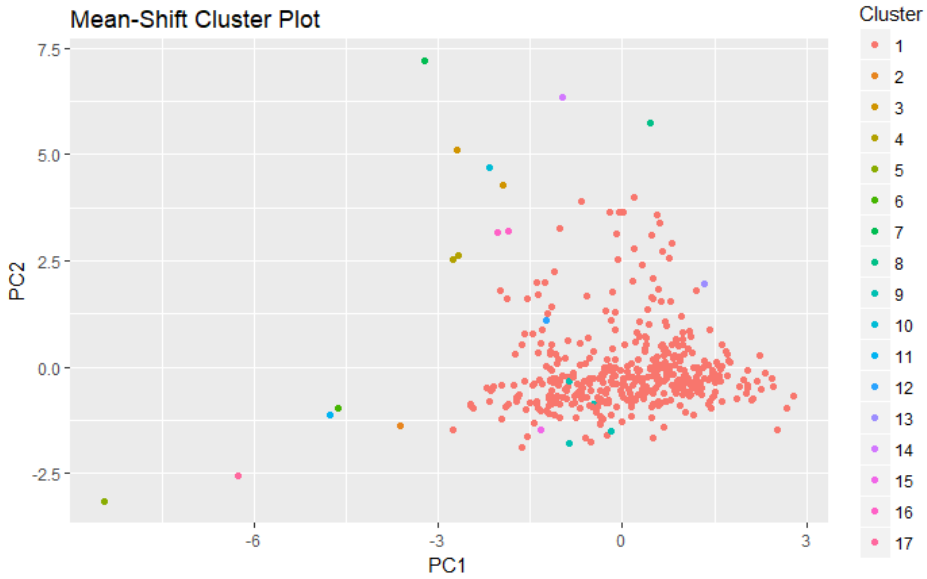
Model
This Mean-Shift model groups the variances on top of each other until it hits a peak in terms of the highest point of variances. The PCA method required hyper-parameters where 1 percent was used as a test due to the large amount of rows or when transposed the columns increased dramatically.

A. Blur Version Mean-Shift Model
Provides more cluster and a much better analysis concerning granularity of cluster Relationships



B. Mean-Shift Cluster Plot

Very similar to the DBSCAN model visual that has two extremes concerning outliers or irregularities of clusters that points out possible rare activities that may be worth investigating



Insights

This model shows a higher correlations of relational clusters. There are many clustered spatial outliers available or visible to observe transaction irregularities.