Project: Predictive Analytics Capstone

## Task 1: Determine Store Formats for Existing Stores

1. What is the optimal number of store formats? How did you arrive at that number?

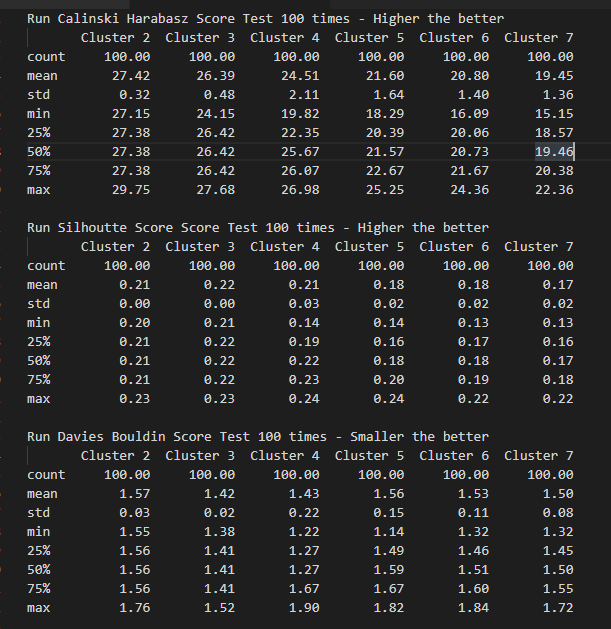
The optimal number of store formats is three. The the test used to select optimal number of store format.

Davies Bouldin Score Test show 3 is optimal store format number

Silhoutte Score Test show 3 is optimal store format number

Calinski Harabasz Score Test show 2 is optimal store format number.

Overall I chose 3 as the optimal number of store formats



1. How many stores fall into each store format?

Store Format 0 has 22 stores

Store Format 1 has 32 stores

Store Format 2 has 31 stores

1. Based on the results of the clustering model, what is one way that the clusters differ from one another?
2. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.

## Task 2: Formats for New Stores

1. What methodology did you use to predict the best store format for the new stores? Why did you choose that methodology? (Remember to Use a 20% validation sample with Random Seed = 3 to test differences in models.)
2. What format do each of the 10 new stores fall into? Please fill in the table below.

|  |  |
| --- | --- |
| Store Number | Segment |
| S0086 |  |
| S0087 |  |
| S0088 |  |
| S0089 |  |
| S0090 |  |
| S0091 |  |
| S0092 |  |
| S0093 |  |
| S0094 |  |
| S0095 |  |

## Task 3: Predicting Produce Sales

1. What type of ETS or ARIMA model did you use for each forecast? Use ETS(a,m,n) or ARIMA(ar, i, ma) notation. How did you come to that decision?

2. Please provide a table of your forecasts for existing and new stores. Also, provide visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.