Project: Predictive Analytics Capstone

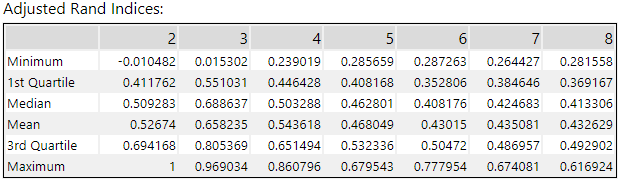
Complete each section. When you are ready, save your file as a PDF document and submit it here: <https://coco.udacity.com/nanodegrees/nd008/locale/en-us/versions/1.0.0/parts/7271/project>

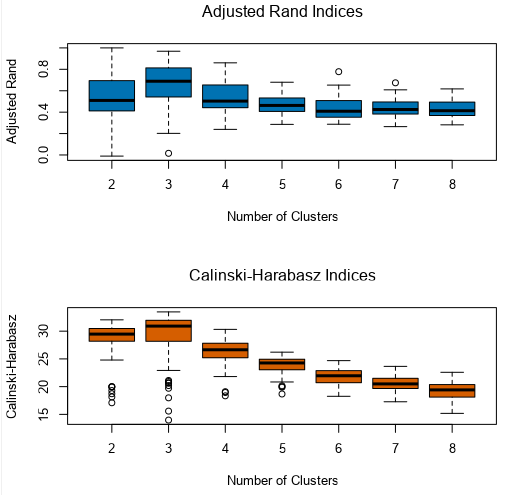
## Task 1: Determine Store Formats for Existing Stores

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

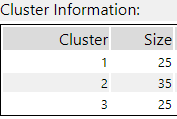
The optimal number of store formats is 3. I arrived this number by using K-Centroids Cluster Analysis and K-Centroids Diagnostics Tools with K-Means Clustering Method.

As cluster 3 median in ARI is the largest





1. How many stores fall into each store format?

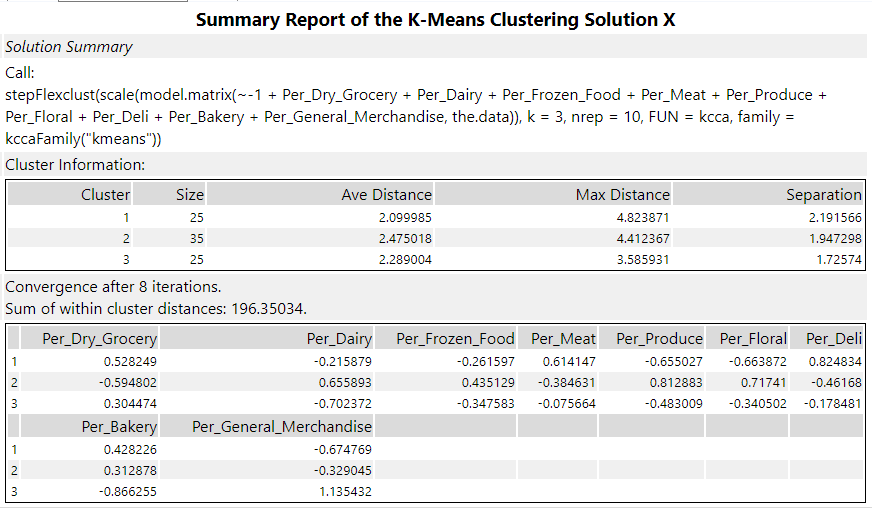


1. Based on the results of the clustering model, what is one way that the clusters differ from one another?

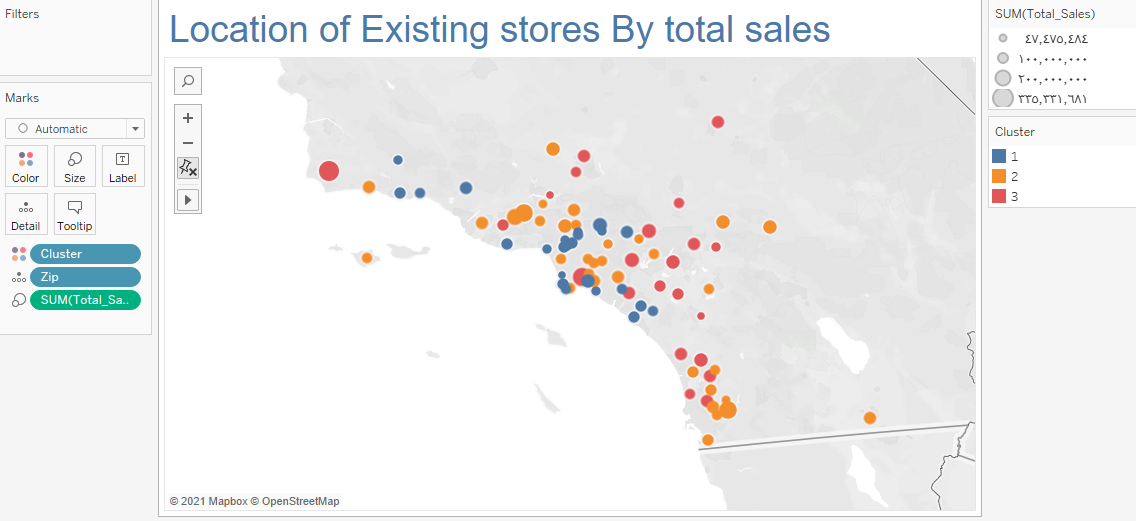
Cluster 1: lowest in average total sale and largest in (Dry\_grocery, Meat,deli,Bakery)

Cluster 2: largest in (Dairy-frozen\_food-produce-floral-)

Cluster 3: Largest in average total sale, largest in (General Merchandise)

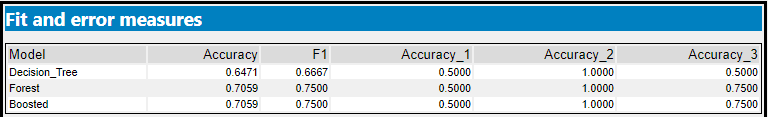


1. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.



## 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.)

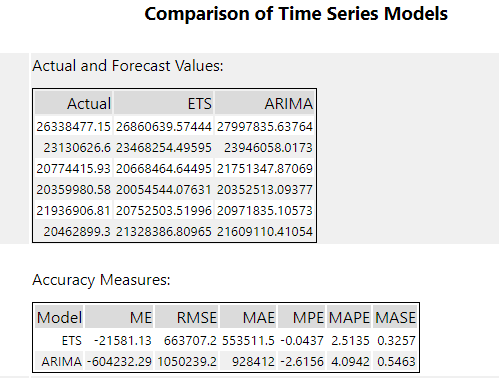
In the following table, any of the three methodologies can be used, However, I will eliminate the decision tree as it is the worst accurate.

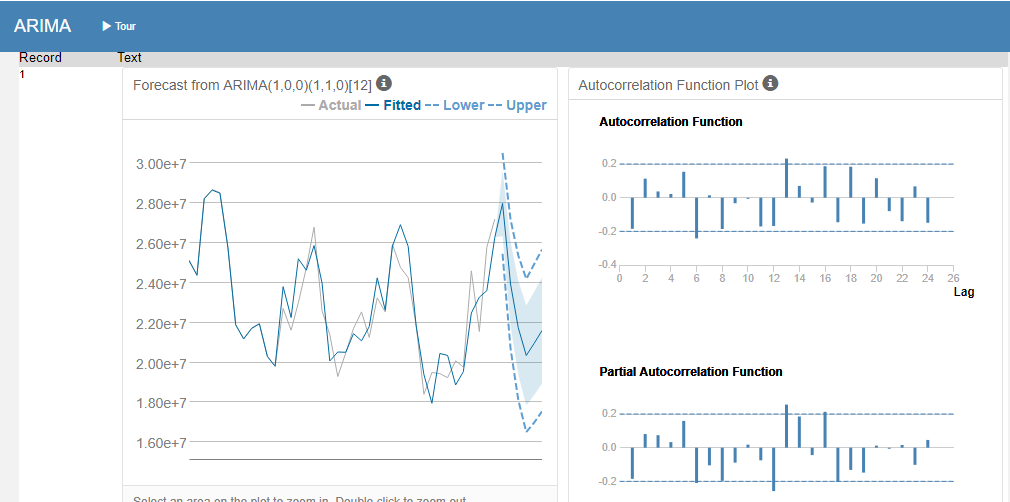
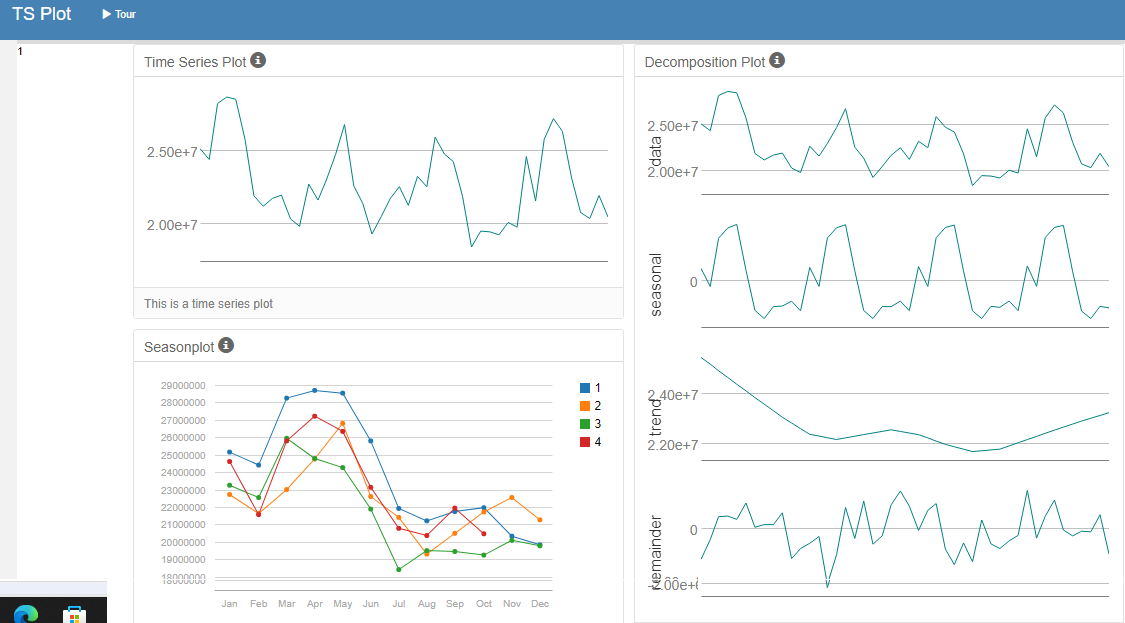
1. What format do each of the 10 new stores fall into? Please fill in the table below.

|  |  |
| --- | --- |
| Store Number | Segment |
| S0086 | 1 |
| S0087 | 2 |
| S0088 | 3 |
| S0089 | 2 |
| S0090 | 2 |
| S0091 | 3 |
| S0092 | 2 |
| S0093 | 3 |
| S0094 | 2 |
| S0095 | 2 |

## 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?

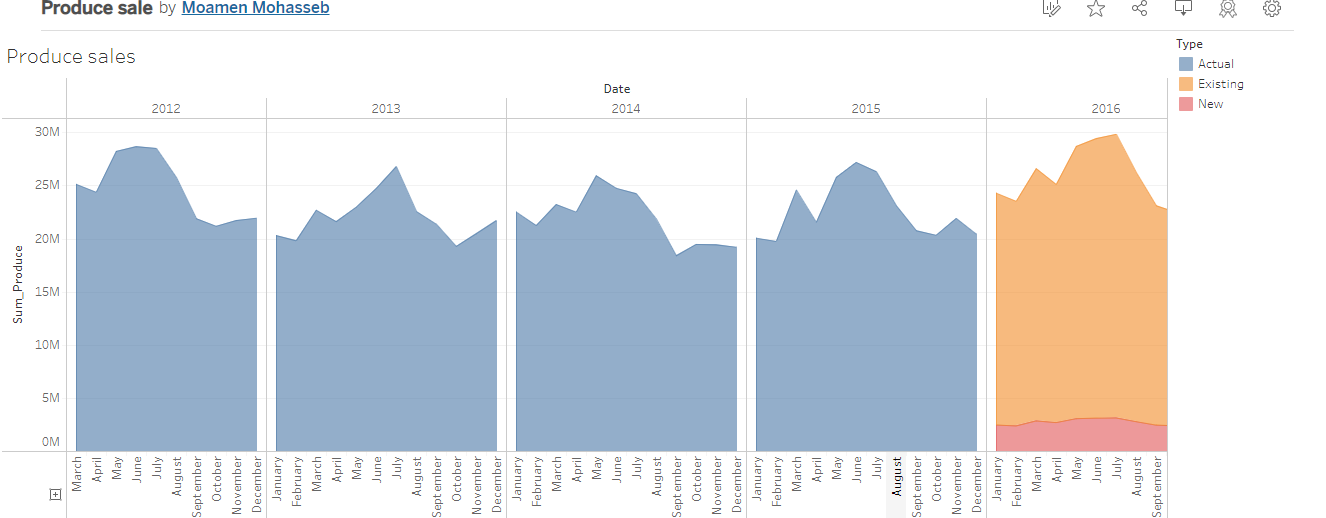
I used ETS model for forecast. I came to this decision after comparing between ETS and ARIMA and using TS Plot tool. Therefore, the forecast should be ETS(M, N, M)



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.

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Month | Forecast\_Integer | New\_Stores\_Sales |
| 2016 | 1 | 21829060 | 2493697 |
| 2016 | 2 | 21146330 | 2405584 |
| 2016 | 3 | 23735687 | 2879417 |
| 2016 | 4 | 22409515 | 2720393 |
| 2016 | 5 | 25621829 | 3089903 |
| 2016 | 6 | 26307858 | 3139497 |
| 2016 | 7 | 26705093 | 3155160 |
| 2016 | 8 | 23440761 | 2807733 |
| 2016 | 9 | 20640047 | 2482456 |
| 2016 | 10 | 20086270 | 2420097 |
| 2016 | 11 | 20858120 | 2510816 |
| 2016 | 12 | 21255190 | 2480120 |

<https://public.tableau.com/app/profile/moamen.mohasseb/viz/Producesale/Sheet1?publish=yes>



Before you submit

Please check your answers against the requirements of the project dictated by the rubric. Reviewers will use this rubric to grade your project.