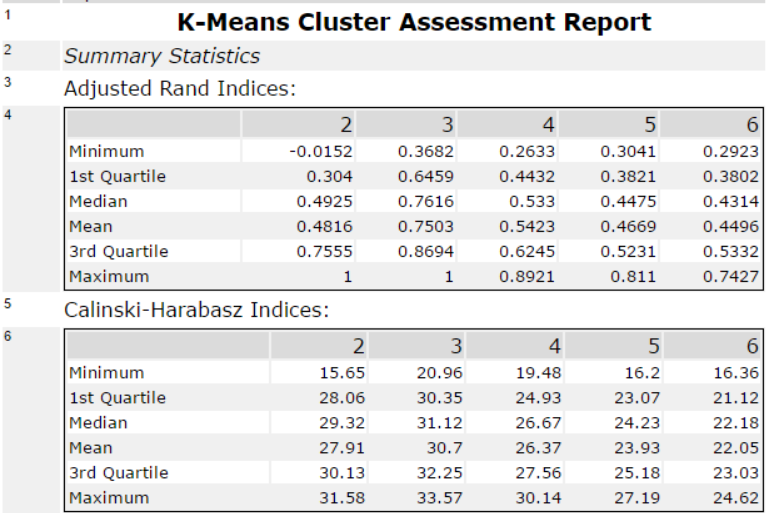
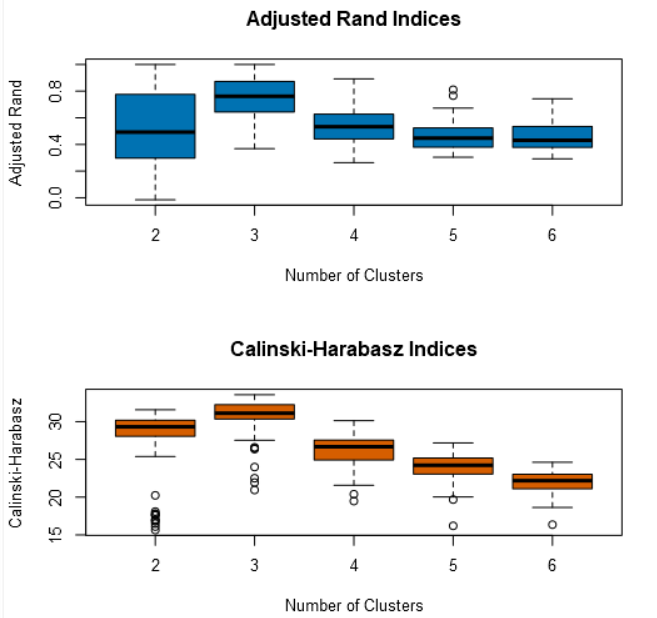
## Task 1: Determine Store Formats for Existing Stores

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

In order to determine the best number of store formats to use. A k-centroids analysis was done using k-means clustering method for k=2,3,4,5,6

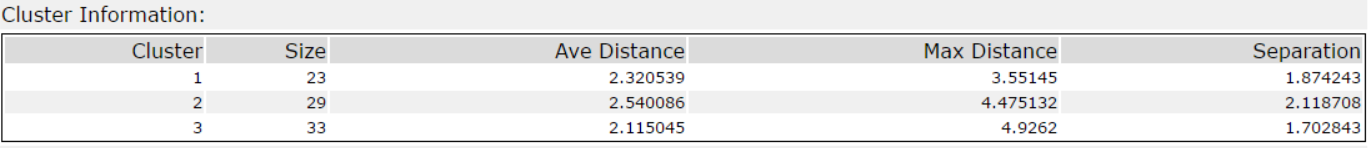




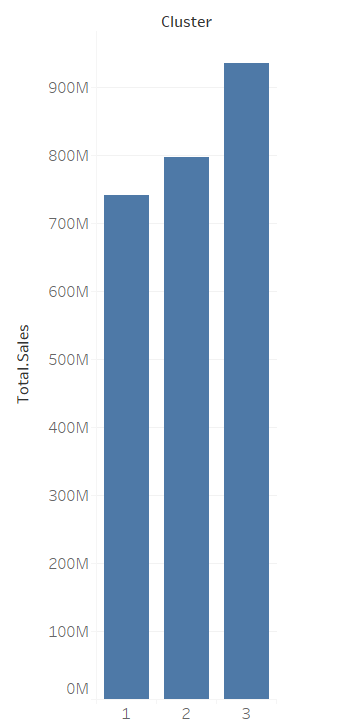
From the above diagnostics, cluster 3 seems to be the best cluster and therefore I will use cluster 3 as my base to compare the number of k terms. Comparing cluster 3, between all the differing k terms, it looks like using k=3 for my analysis offers the best results with cluster 3 having the tightest range and highest mean when k=3.

The optimal number of store formats in my opinion is 3.

1. **How many stores fall into each store format?**



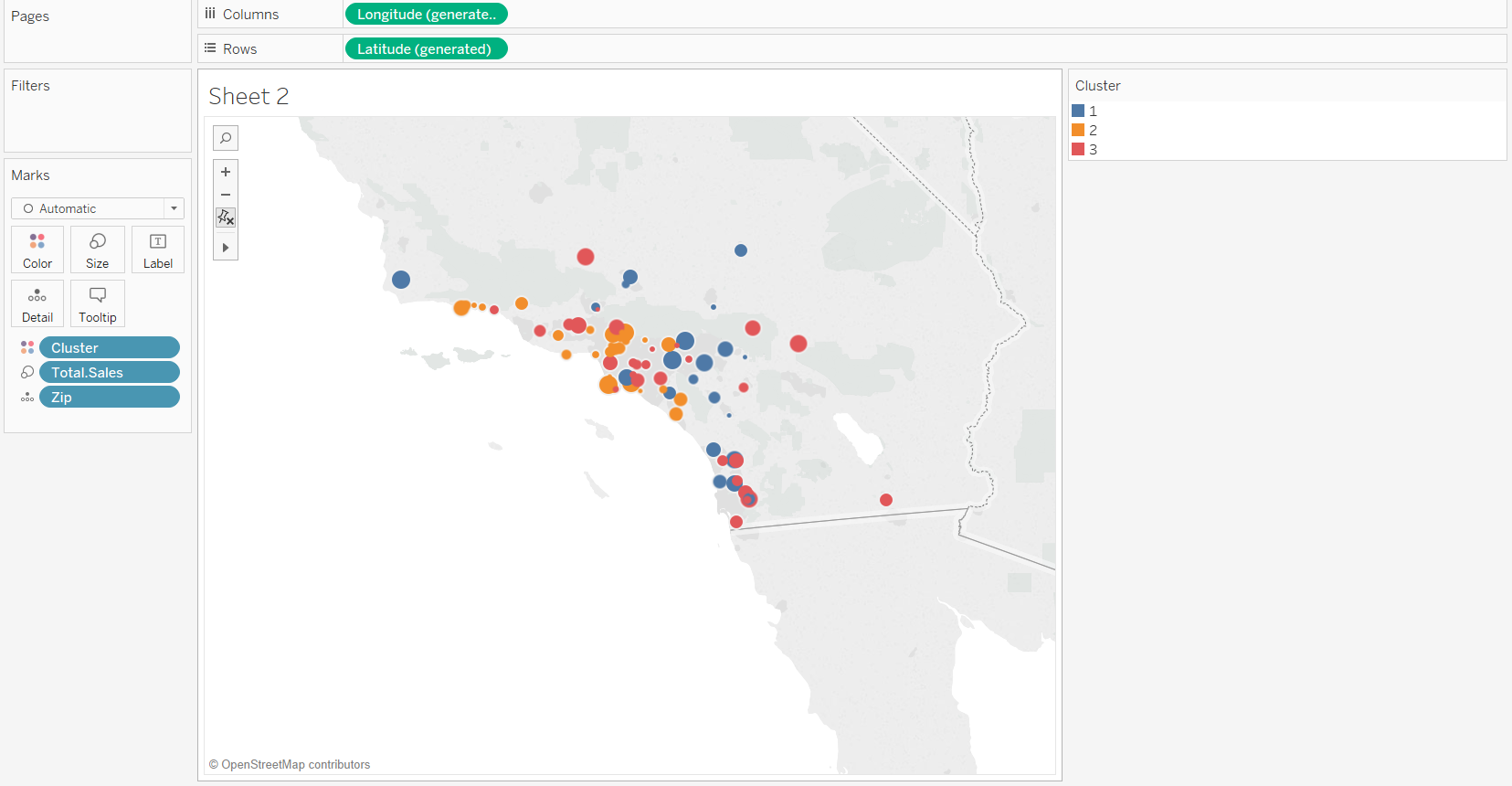
1. **Based on the results of the clustering model, what is one way that the clusters differ from one another?**



Overall, cluster 3 sold more in total in comparison to the other two clusters.

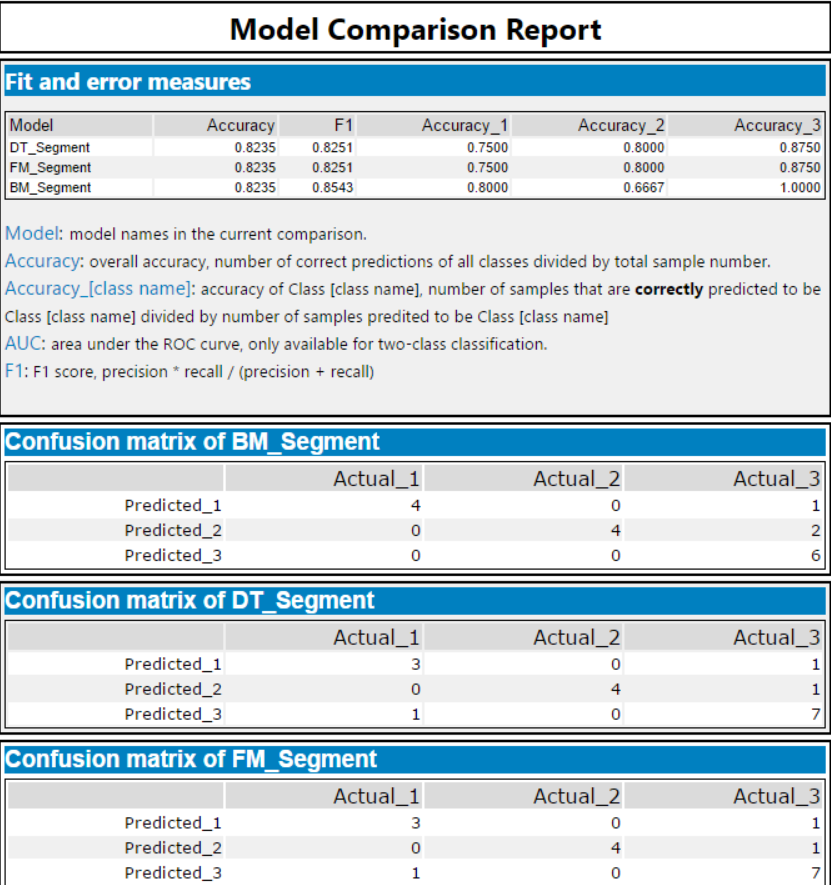
1. **Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.**

https://public.tableau.com/views/Book2\_20234/Sheet2?:embed=y&:display\_count=yes

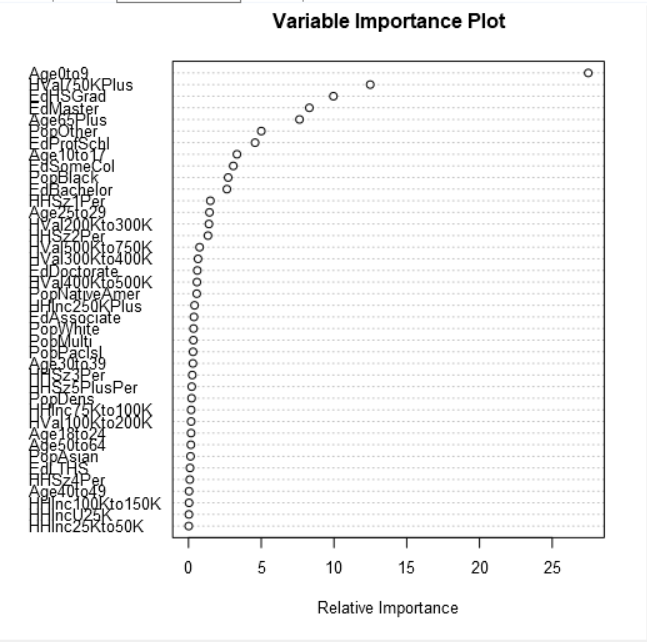
The size of the circle represents the amount of total sale for each store

## 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.)**



Based on the above results, I have decided to use the Boosted, even though the forest model and boosted model both have the same accuracy score, I have used the higher F1 score of the Boosted model as my deciding factor.



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

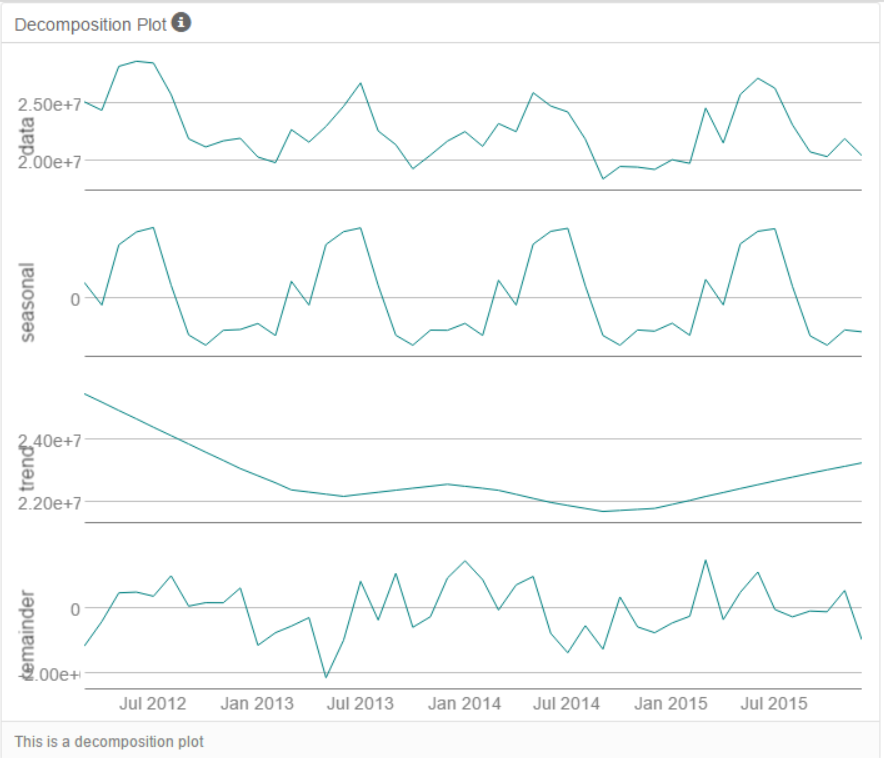
|  |  |
| --- | --- |
| Store Number | Segment |
| S0086 | 3 |
| S0087 | 2 |
| S0088 | 1 |
| S0089 | 2 |
| S0090 | 2 |
| S0091 | 1 |
| S0092 | 2 |
| S0093 | 1 |
| S0094 | 2 |
| S0095 | 2 |

## 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?**

Both ETS and ARIMA models were run for comparison. Analysis of the initial time series decomposition plots below allowed further analysis of model parameters to be established.

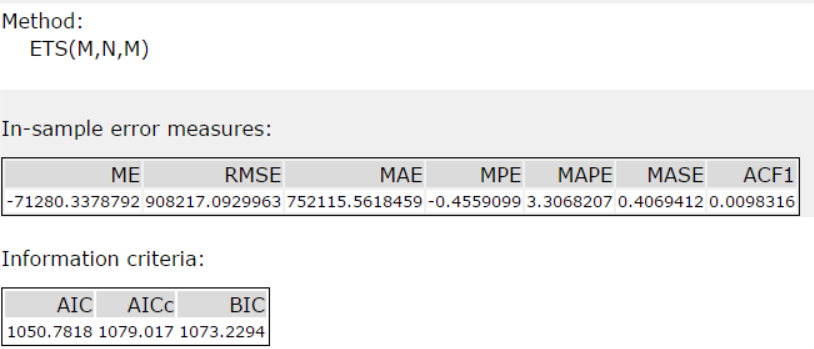
The data used here is sales for produce only per month for all stores aggregated.

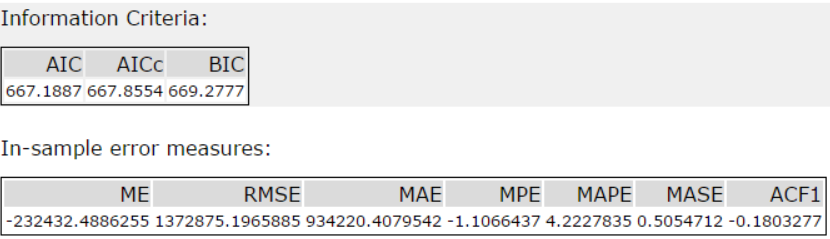


From the above decomposition plots, I can see that the Error element is increasing, Trend element is non-existent and the Seasonal element is also increasing, therefore an ETS(M,N,M) will be used. As for the ARIMA model, I have set the model to calculate the elements automatically.

For comparison, a holdout period of 12 periods was used to validate the ETS and ARIMA model.

Below is the ETS(M,N,M) in-sample summary.





The ETS(M,N,M) will be used for forecasting due to the model having lower error values compared to the ARIMA model.

The monthly total produce of all the new stores are calculated as follows. The average sale of each segment of each month is calculated by diving the forecasted sum produce sales of each segment by the number of existing stores in that segment. The monthly average sale of each segment is then multiplied by the number of new stores in that segment. The monthly total produce sale for all the new stores are then calculated by summing the total sales of all segments in each month.

1. **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.**

<https://public.tableau.com/views/Task3_180/Task3?:embed=y&:display_count=yes&publish=yes>

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Month | Existing Stores Sales | New Stores Sales |
| 2016 | 1 | 21539936.007499 | 2761958 |
| 2016 | 2 | 20413770.60136 | 2656665.25 |
| 2016 | 3 | 24325953.097628 | 3099057.75 |
| 2016 | 4 | 22993466.348585 | 2873607.25 |
| 2016 | 5 | 26691951.419156 | 3327835.25 |
| 2016 | 6 | 26989964.010552 | 3356062 |
| 2016 | 7 | 26948630.764764 | 3391942.75 |
| 2016 | 8 | 24091579.349106 | 2991382.5 |
| 2016 | 9 | 20523492.408643 | 2664295.25 |
| 2016 | 10 | 20011748.6686 | 2588209.75 |
| 2016 | 11 | 21177435.485838 | 2702838.25 |
| 2016 | 12 | 20855799.10961 | 2761943.25 |

