Project: Predictive Analytics Capstone

Ans Jayan

[ansjayan@msn.com](mailto:ansjayan@msn.com)

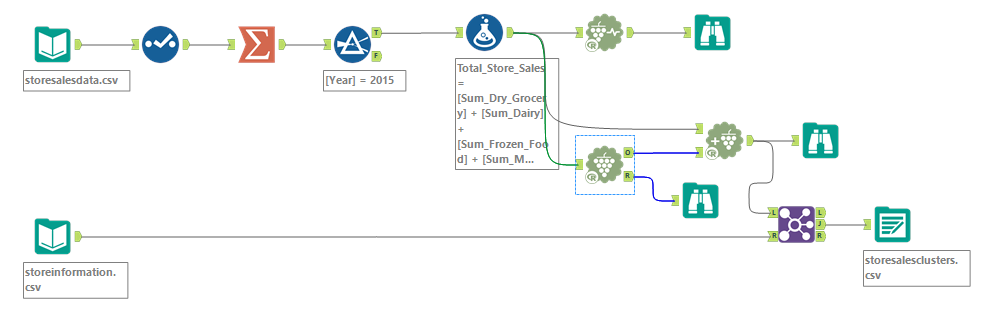
The company runs 85 grocery stores currently. All stores use the same format for selling products and company ships the same amount of product to all stores. Some stores have product surplus and some has shortage. Company is planning to open 10 more stores. So, the company want to make decisions about store formats and planning inventory.

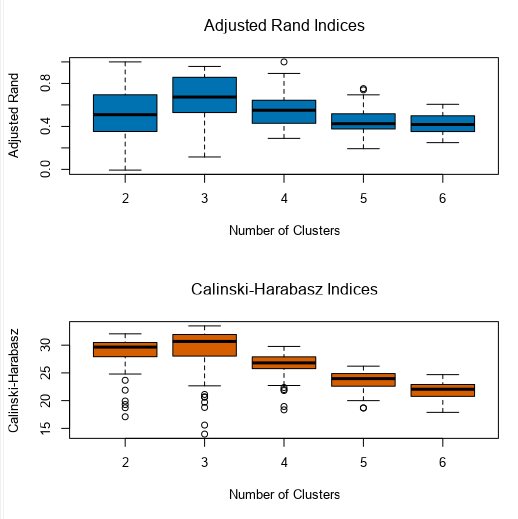
## Task 1: Determine Store Formats for Existing Stores

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

The optimal number of store format is 3. Using given data, a k-means clustering

model is created.





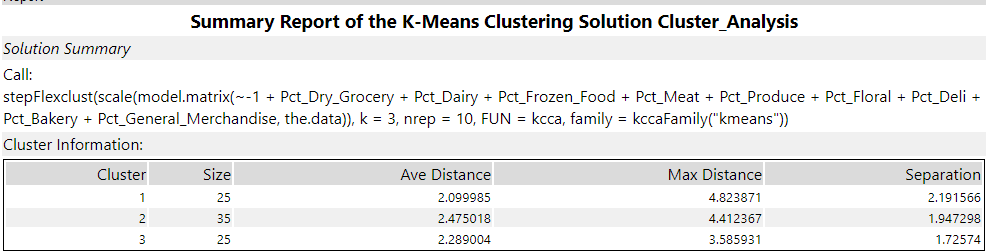
From the adjusted rand indices, higher the index better the stability. For calinski-harabasz indices, higher the index better the distinctness and compactness of clusters. Required is Median high and spread minimized. Cluster 3 has high median and fairly ok spread.

1. How many stores fall into each store format?

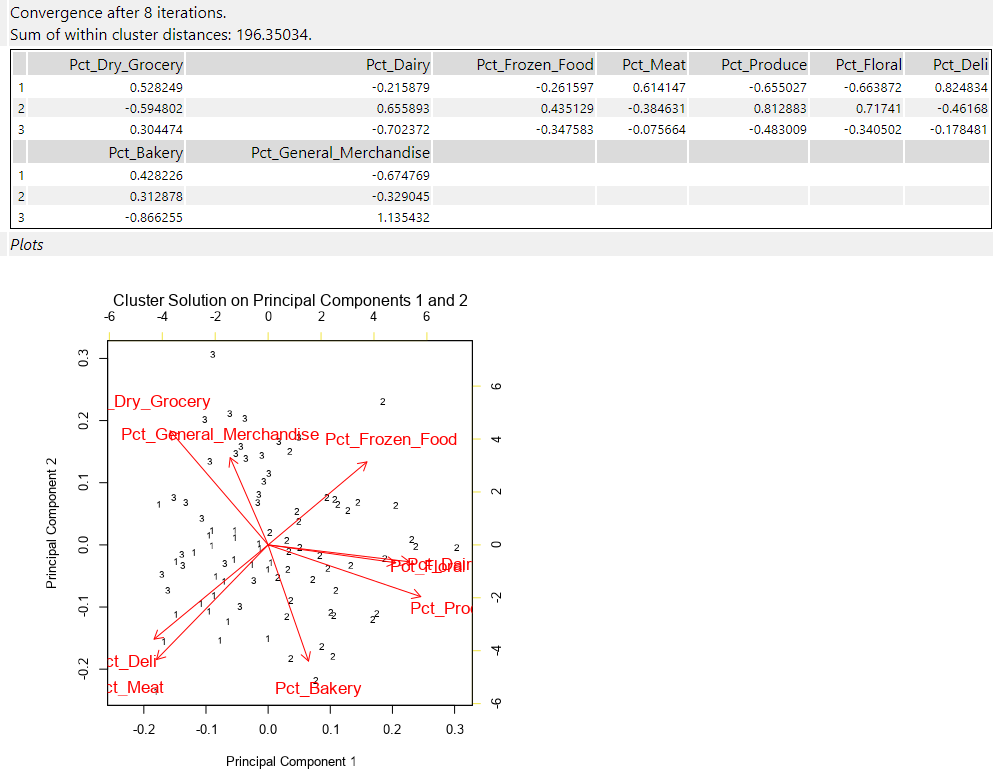
Cluster 1: - 25

Cluster 2: - 35

Cluster 3: - 25

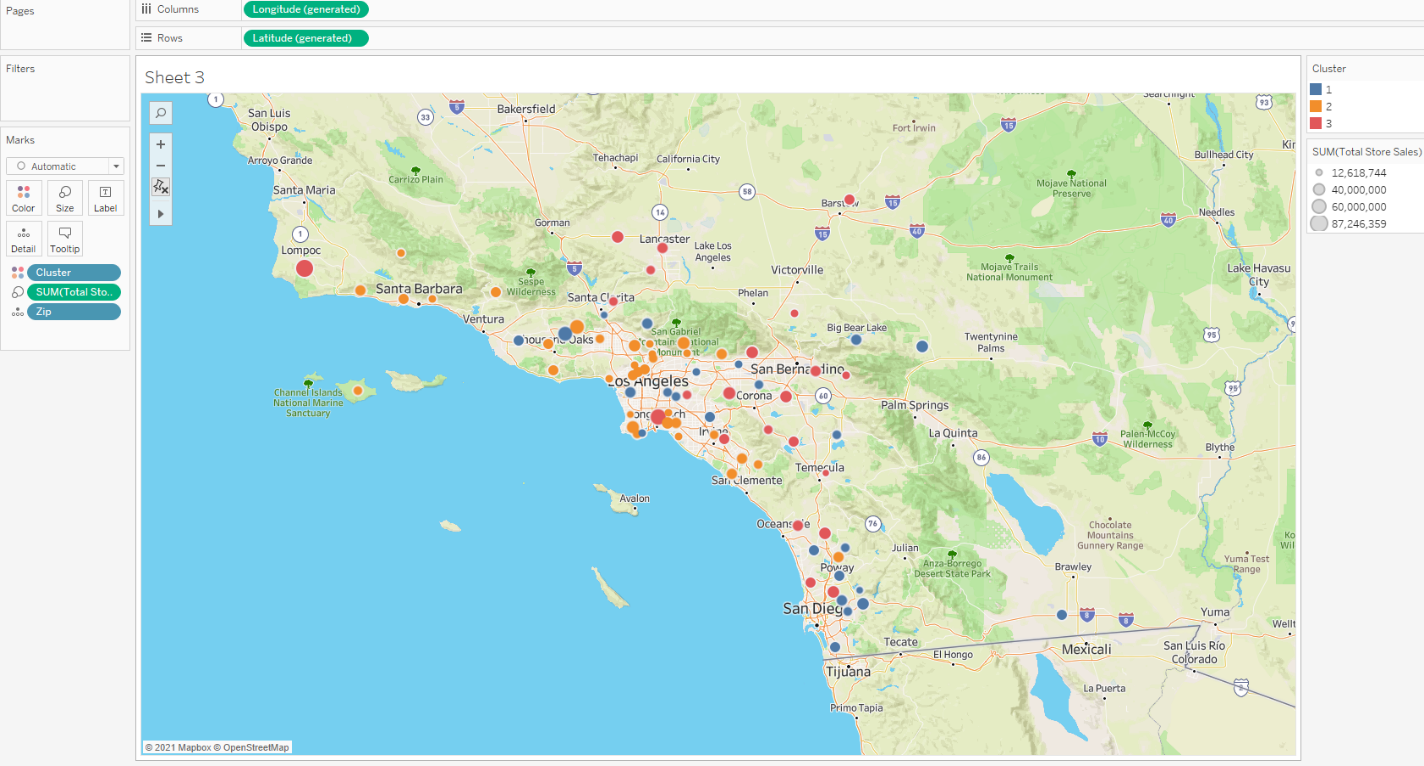


1. Based on the results of the clustering model, what is one way that the clusters differ from one another?



The clusters differ from one another based on the percentage of sales in different categories of each store. The cluster 1 sells Meat and Deli more, cluster 2 sells Produce and Floral more and cluster 3 sells General Merchandise more.

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.

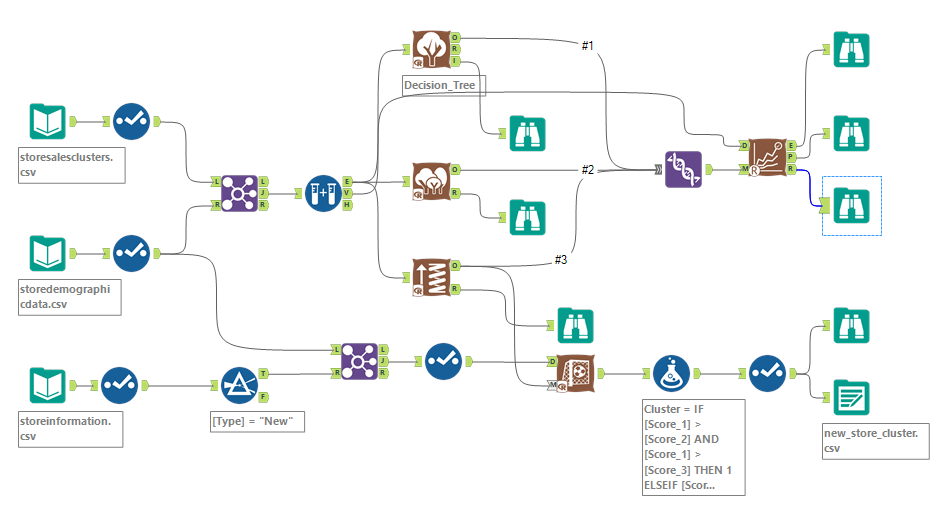


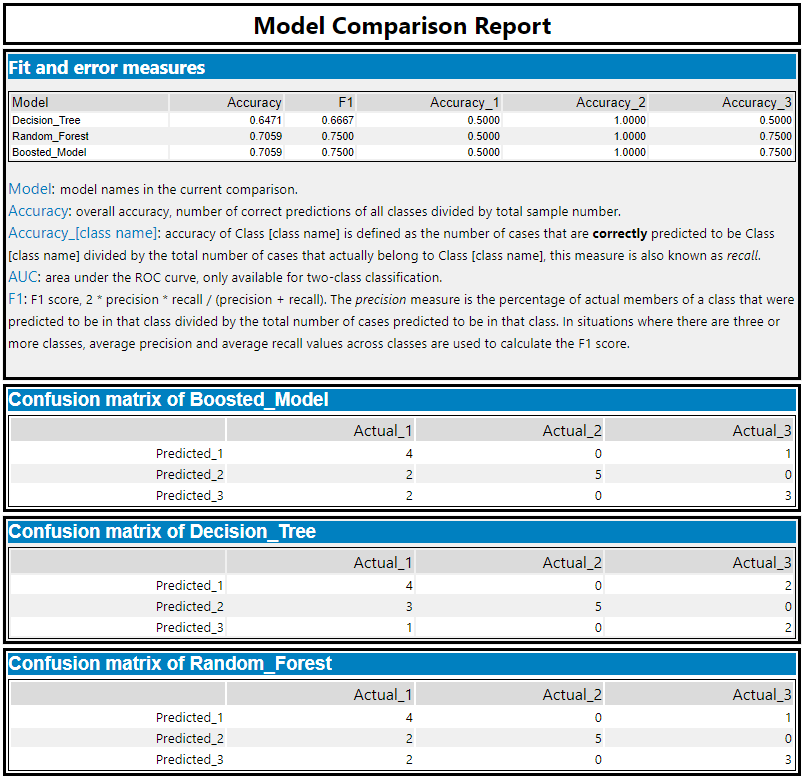
## Task 2: Formats for New Stores

The company is opening 10 new stores. Using the given demographic data, we can find the store format by modeling the classification methods like random forest, boosted models or decision tree.

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.)

I used Boosted model to predict the best store format for the new stores.





From the model comparison report we can see that both boosted model and random forest model has similar performances and better than decision tree.

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

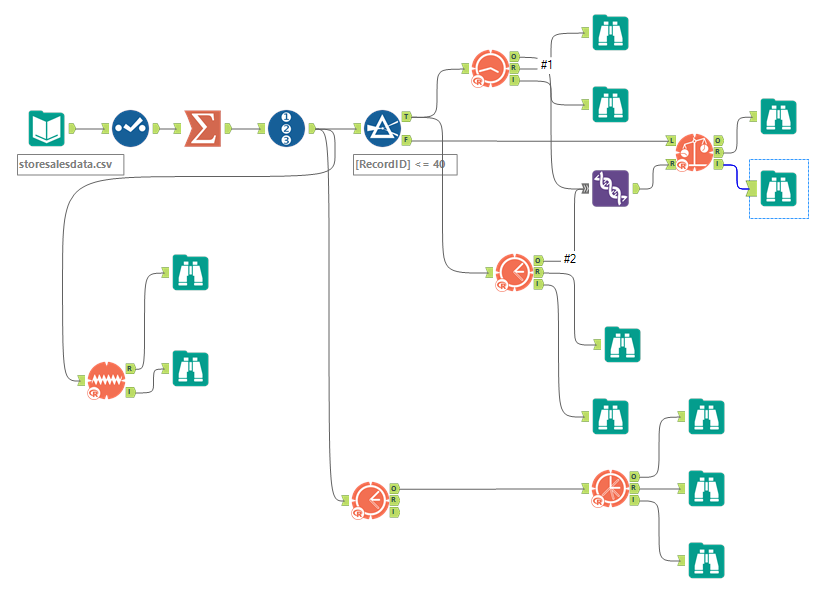
|  |  |
| --- | --- |
| Store Number | Segment |
| S0086 | 1 |
| S0087 | 2 |
| S0088 | 3 |
| S0089 | 2 |
| S0090 | 2 |
| S0091 | 3 |
| S0092 | 2 |
| S0093 | 3 |
| S0094 | 2 |
| S0095 | 2 |

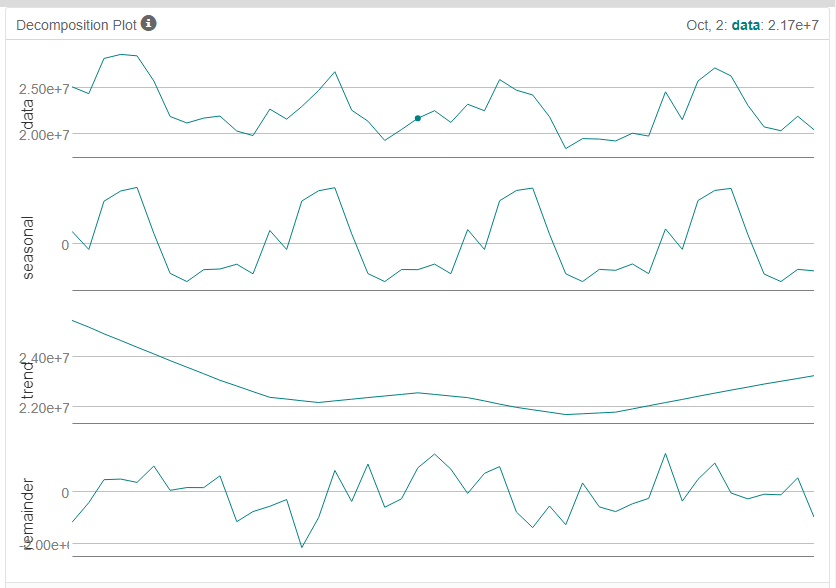
## Task 3: Predicting Produce Sales

A monthly forecast of Produce sales for one year of 2016 for both existing and new stores have to be prepared.

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?

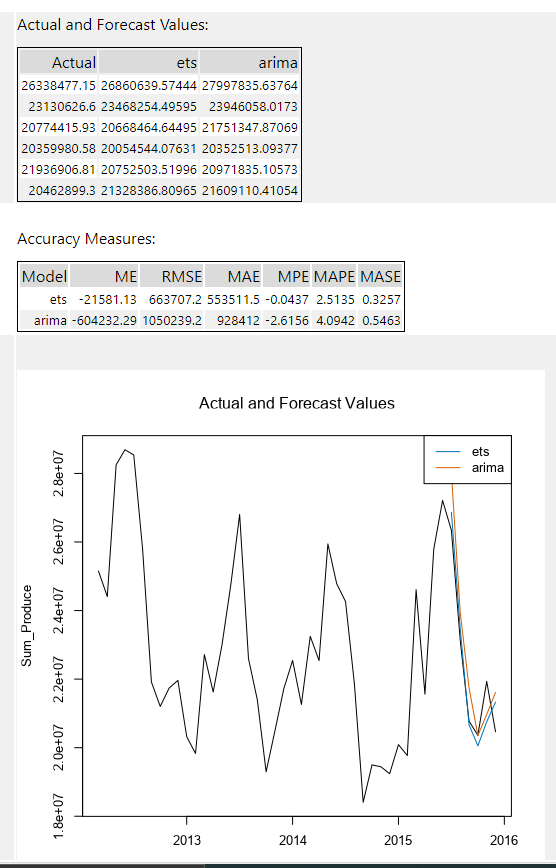
ETS (M, N, M) model is used for forecast.





From the above decomposition plot, Error component is multiplicative, Trend is none, and

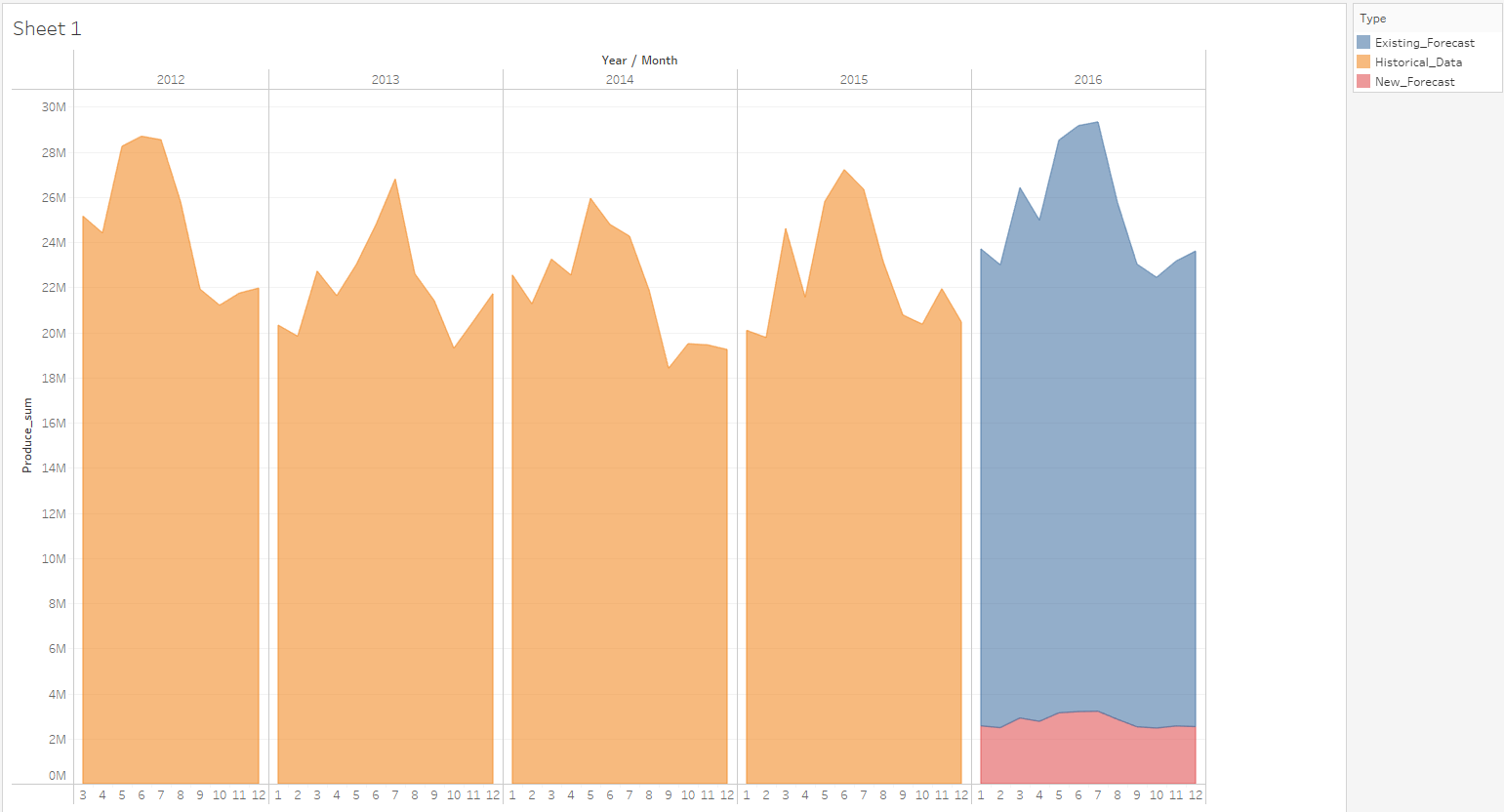
Seasonal component is multiplicative.

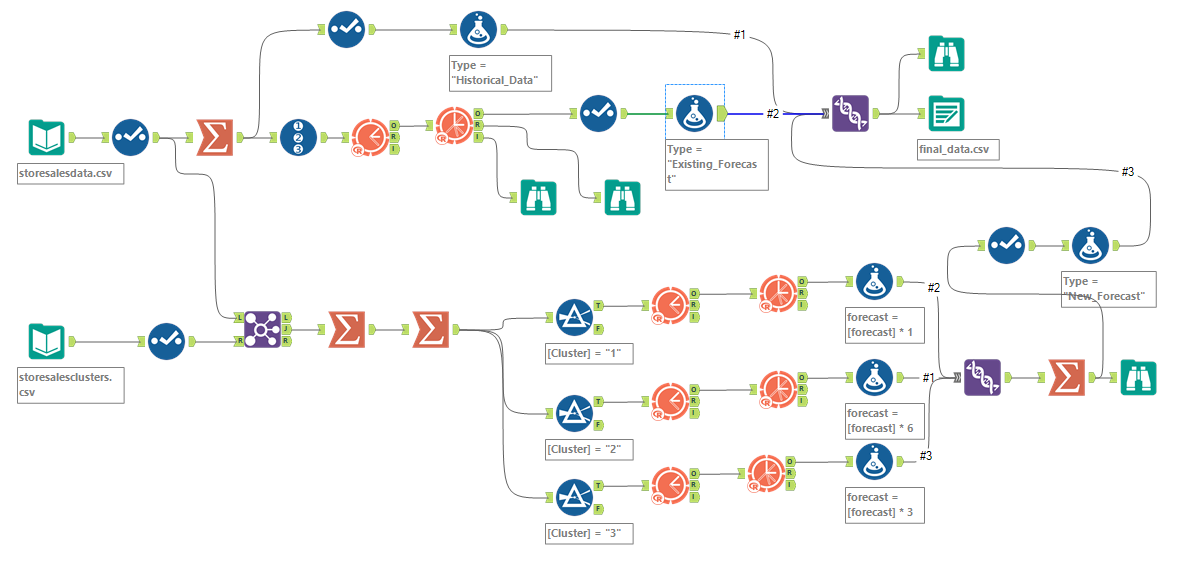


From above summary of ETS and ARIMA, the error measures are low in ETS than in ARIMA, so ETS method is used for forecasting.

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 | Existing | New |
| 2016 | 01 | 21136641.78 | 2563357.91 |
| 2016 | 02 | 20507039.12 | 2483924.72 |
| 2016 | 03 | 23506565.98 | 2910944.14 |
| 2016 | 04 | 22208405.75 | 2764881.86 |
| 2016 | 05 | 25380147.77 | 3141305.86 |
| 2016 | 06 | 25966799.46 | 3195054.20 |
| 2016 | 07 | 26113792.56 | 3212390.95 |
| 2016 | 08 | 22899285.76 | 2852385.76 |
| 2016 | 09 | 20499583.90 | 2521697.18 |
| 2016 | 10 | 19971242.82 | 2466750.89 |
| 2016 | 11 | 20602665.91 | 2557744.58 |
| 2016 | 12 | 21073222.08 | 2530510.80 |





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.