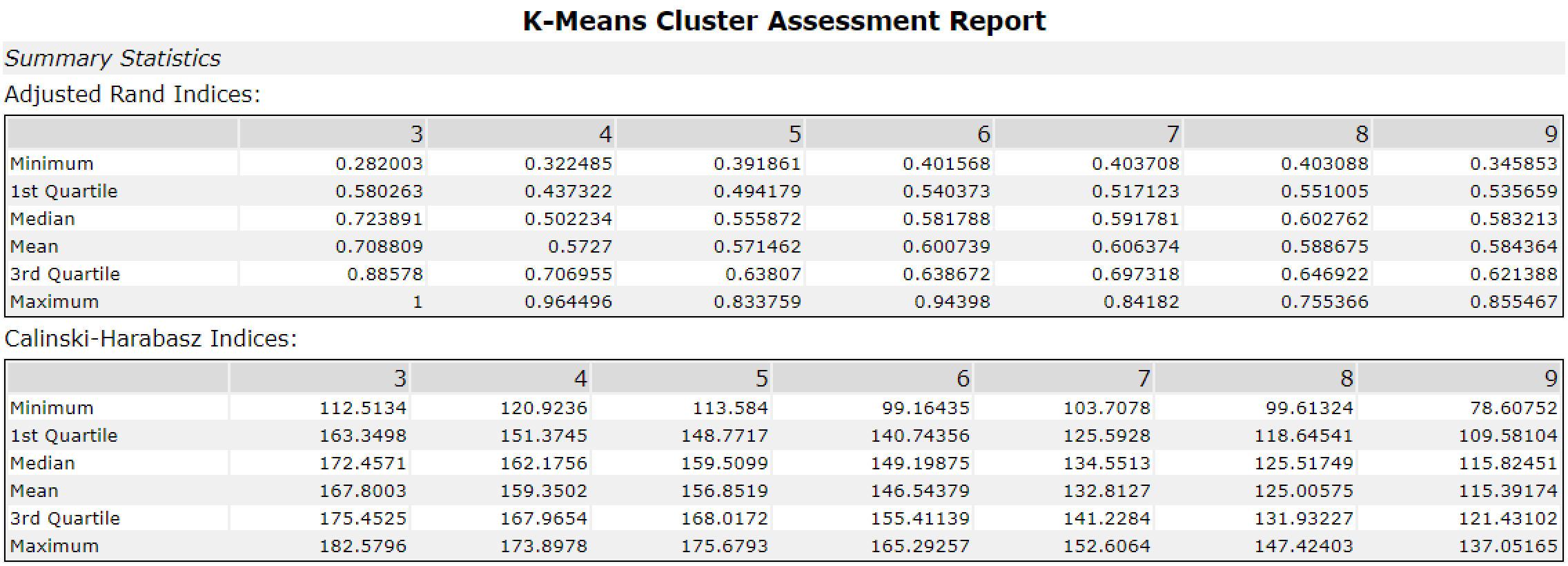
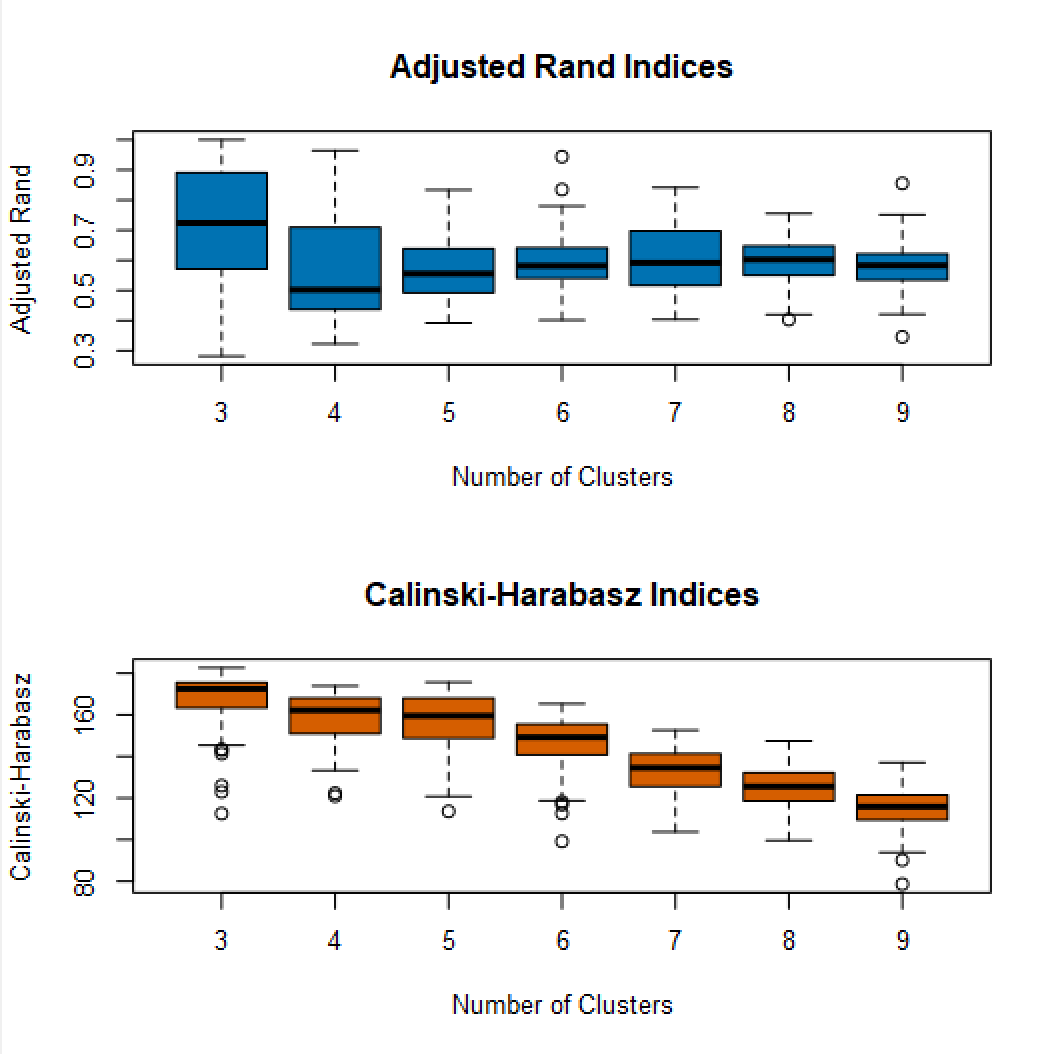
Project: Predictive Analytics Capstone

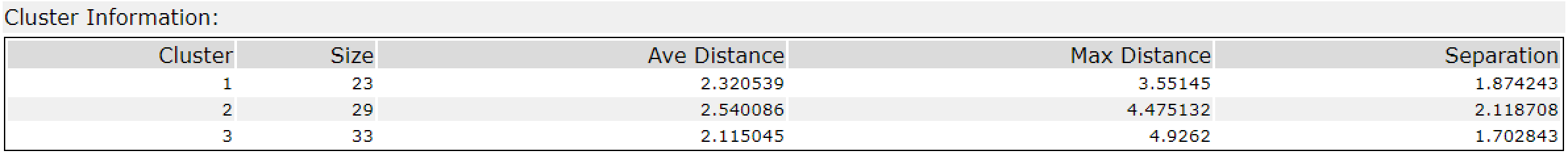
## Task 1: Determine Store Formats for Existing Stores

1. What is the optimal number of store formats? How did you arrive at that number?
   1. The optimal number of store formats is based on the K-Means report as well as AR Indices and CH Indices is 3.

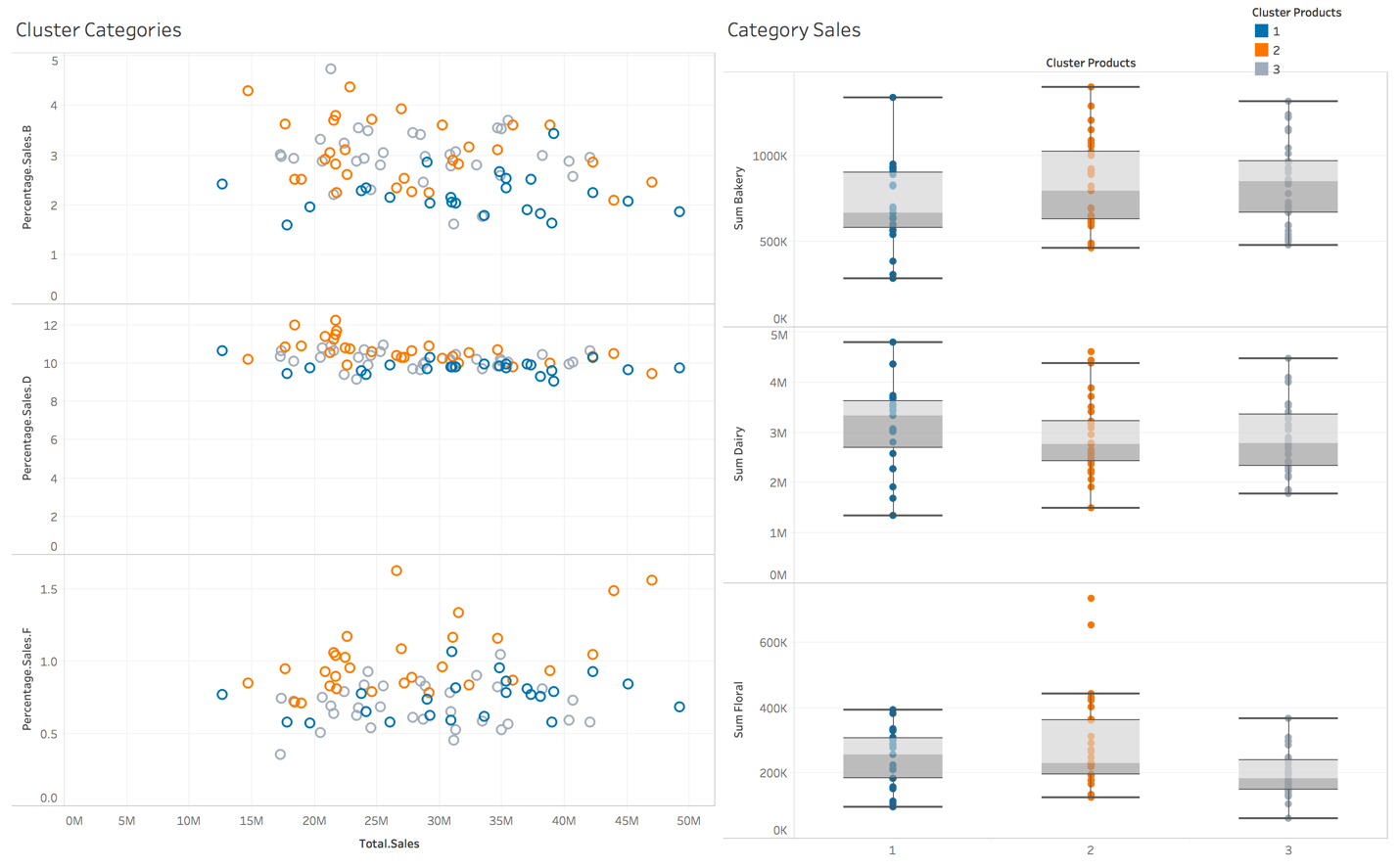




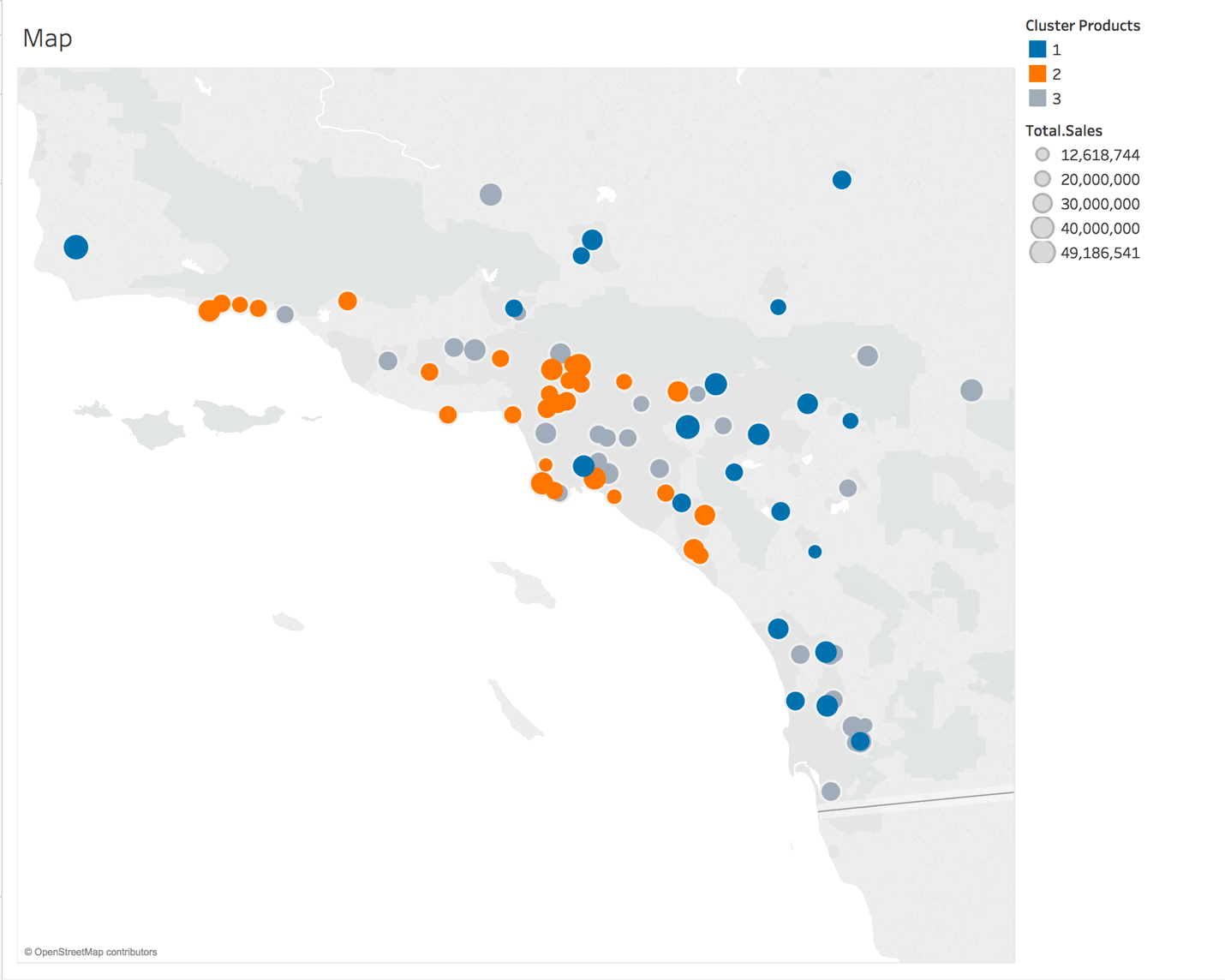
1. How many stores fall into each store format?
   1. Based on K-Centroids Cluster Analysis, cluster 1 has 23 stores, cluster 2 has 29 stores, and cluster 3 has 33 stores.



1. Based on the results of the clustering model, what is one way that the clusters differ from one another?
   1. Cluster 2 had slightly more percentage sales in bakery compared to cluster 1 and 3.
   2. In terms of sales, cluster 2 has large sales in Bakery and Floral compared to cluster 1 and 3.



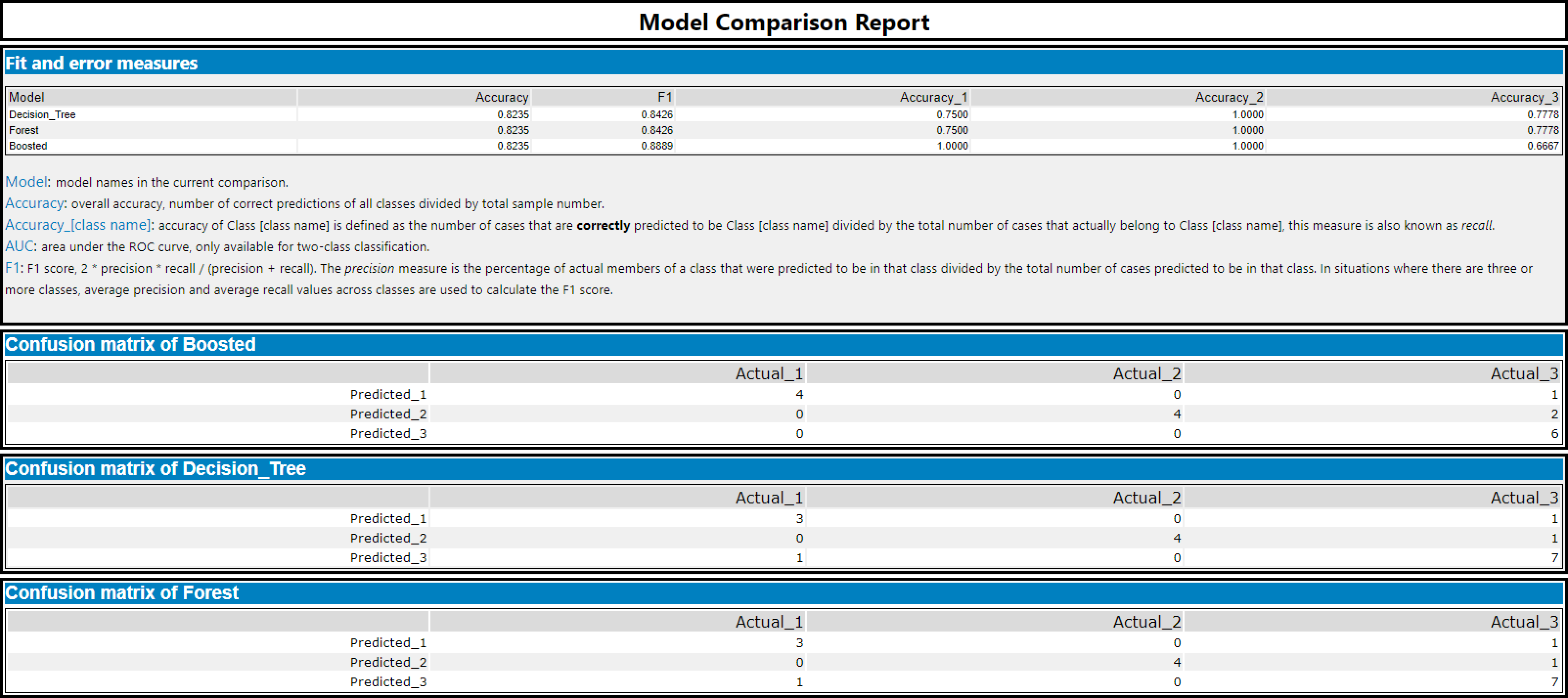
1. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.



* 1. [Link to Tableau public](https://public.tableau.com/profile/abhinav.konagala#!/vizhome/Task1_199/Dashboard2?publish=yes)

## 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.)
   1. Out of the three models (Decision Tree, Forest Model, and Boosted Model), I chose Boosted model as it had the higher F1 percentage compared to Decision Tree and Forest Model.



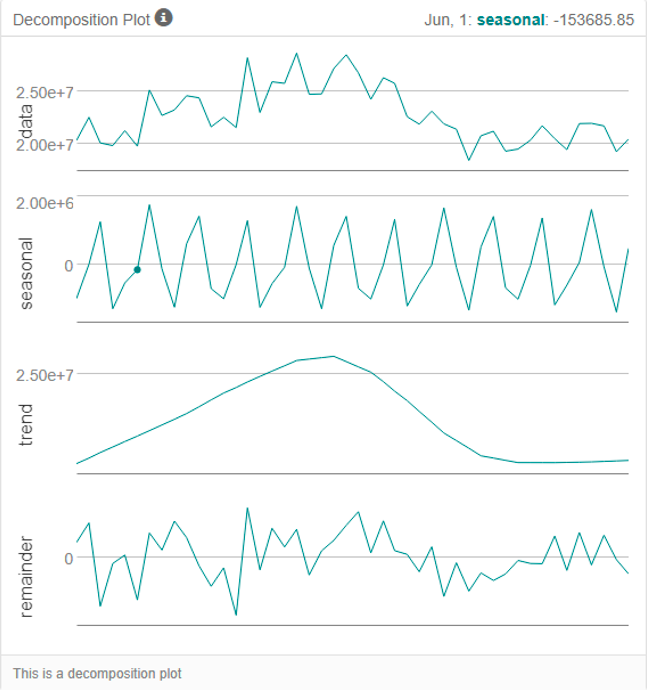
1. What format do each of the 10 new stores fall into? Please fill in the table below.

|  |  |
| --- | --- |
| 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?

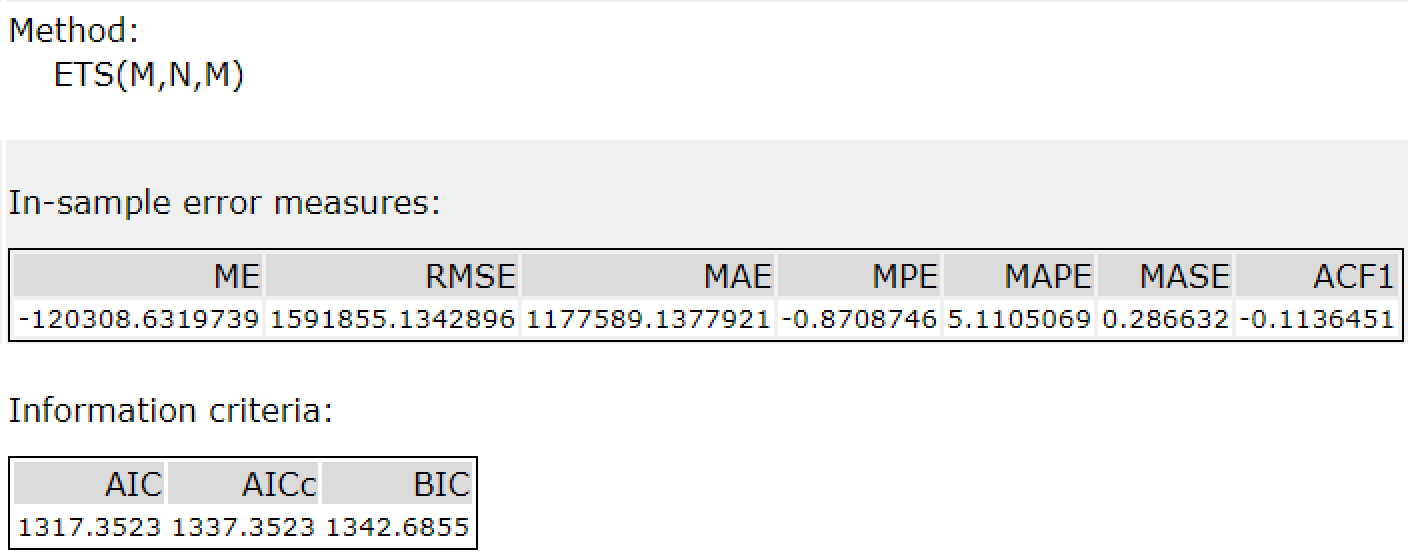
1. For the ETS, the error is not consistent so it should be applied multiplicatively, the trend isn’t linear so nothing should be applied, the seasonality is not consistent so it should be applied multiplicatively.

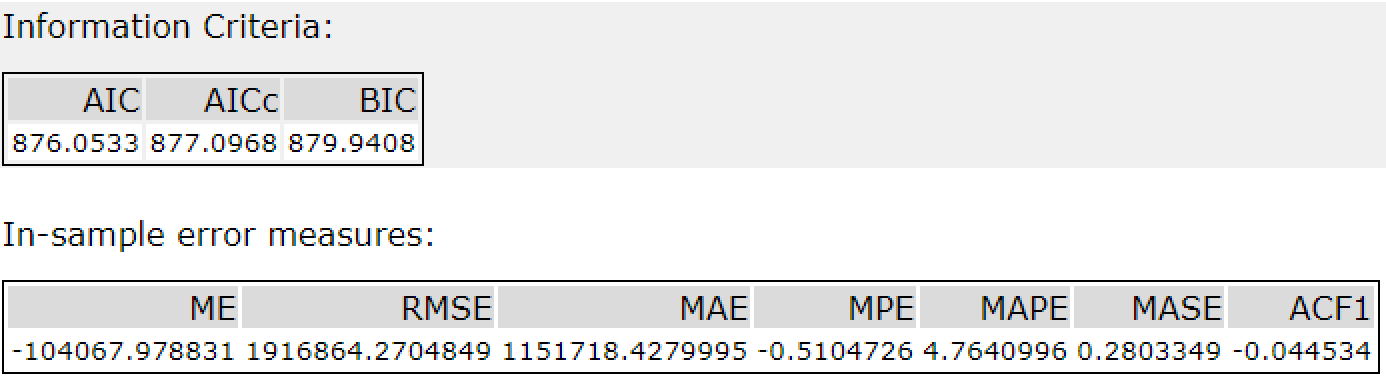


1. For the ARIMA (0,1,2) (0,1,0) is used as seasonal difference and first difference is performed.

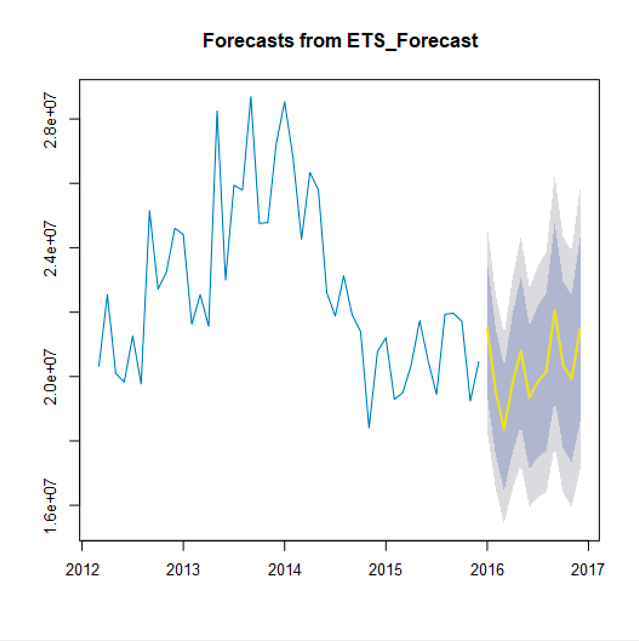


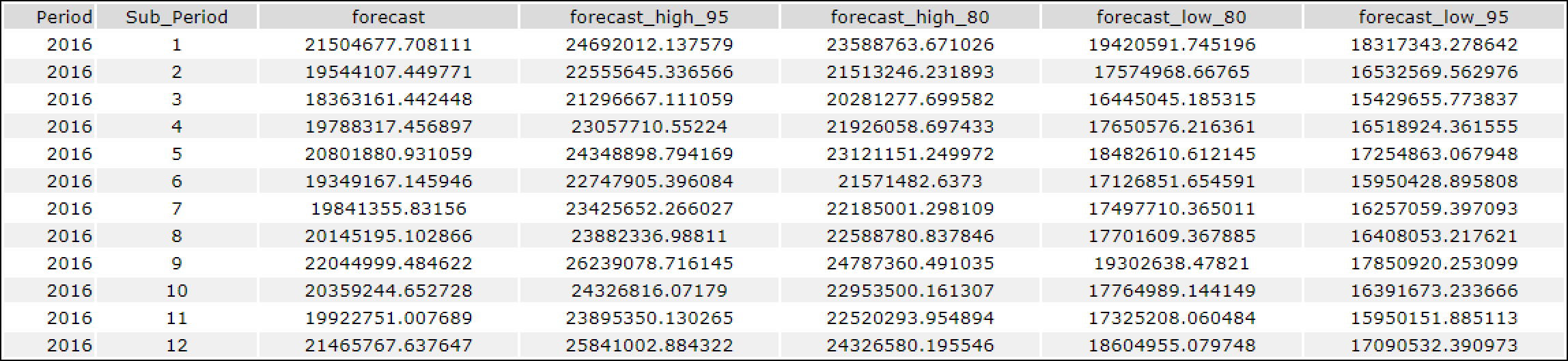
1. ETS model is more accurate compared to the ARIMA model. The RMSE value of ETS is much lower compared to the RMSE value of ARIMA. The MASE value of ETS is also low compared to the MASE value of ARIMA.
2. Picked the ETS model because of high AIC value compared to the ARIMA model.





1. For the TS Forecast, the percentage value of the larger confidence interval is 95%, the percentage value of the smaller confidence interval is 80%, and the number of periods is 12.





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.

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.