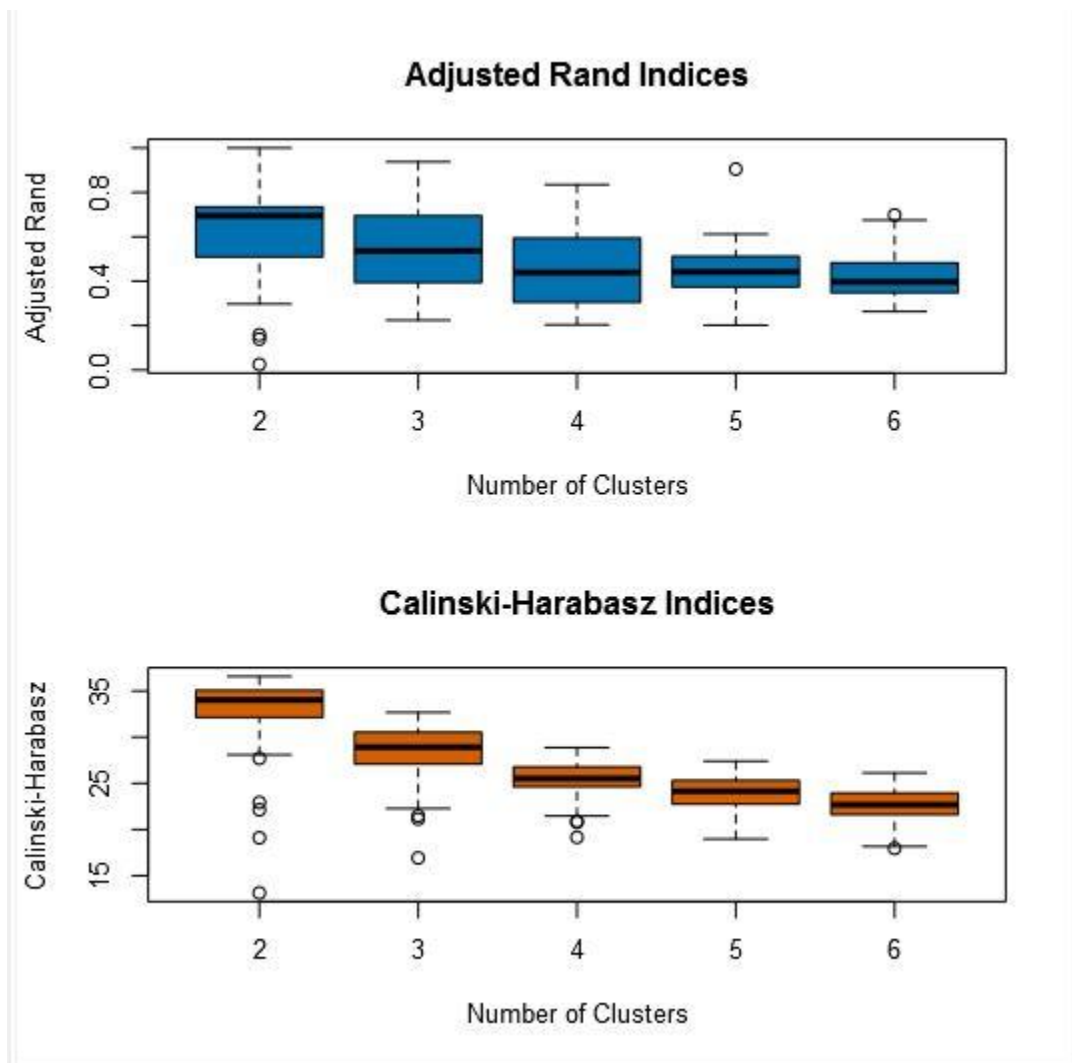


# Project: Predictive Analytics Capstone

## 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 3. The number of formats determined from analyzing the AR and CH indexes for K-means clustering algorithm. From the graph below, we can see that for cluster 3, the medians are high and the plots are compact and symmetrical, which is why cluster=3 is the correct number of formats required



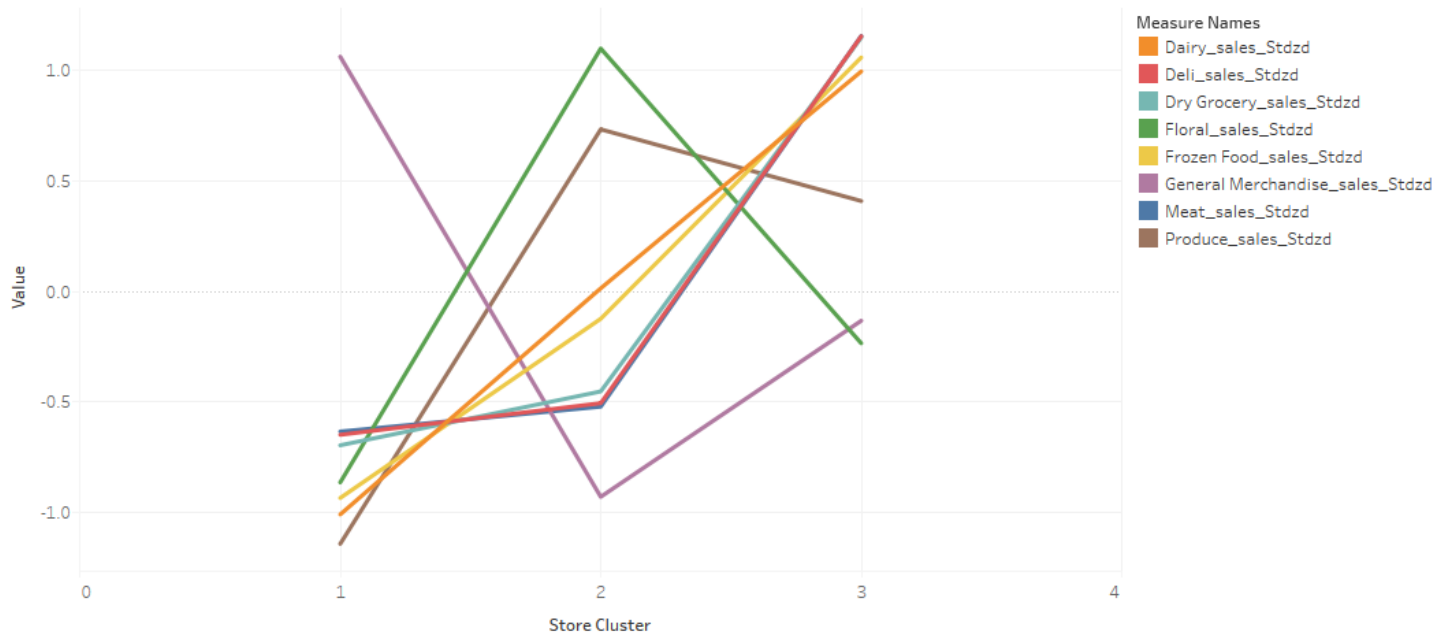
2. How many stores fall into each store format?

Count	Store_Cluster
23	1
29	2
33	3

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

One way the Clusters are different from one another is shown when they are standardized with z-along different clusters for each of the different product types.

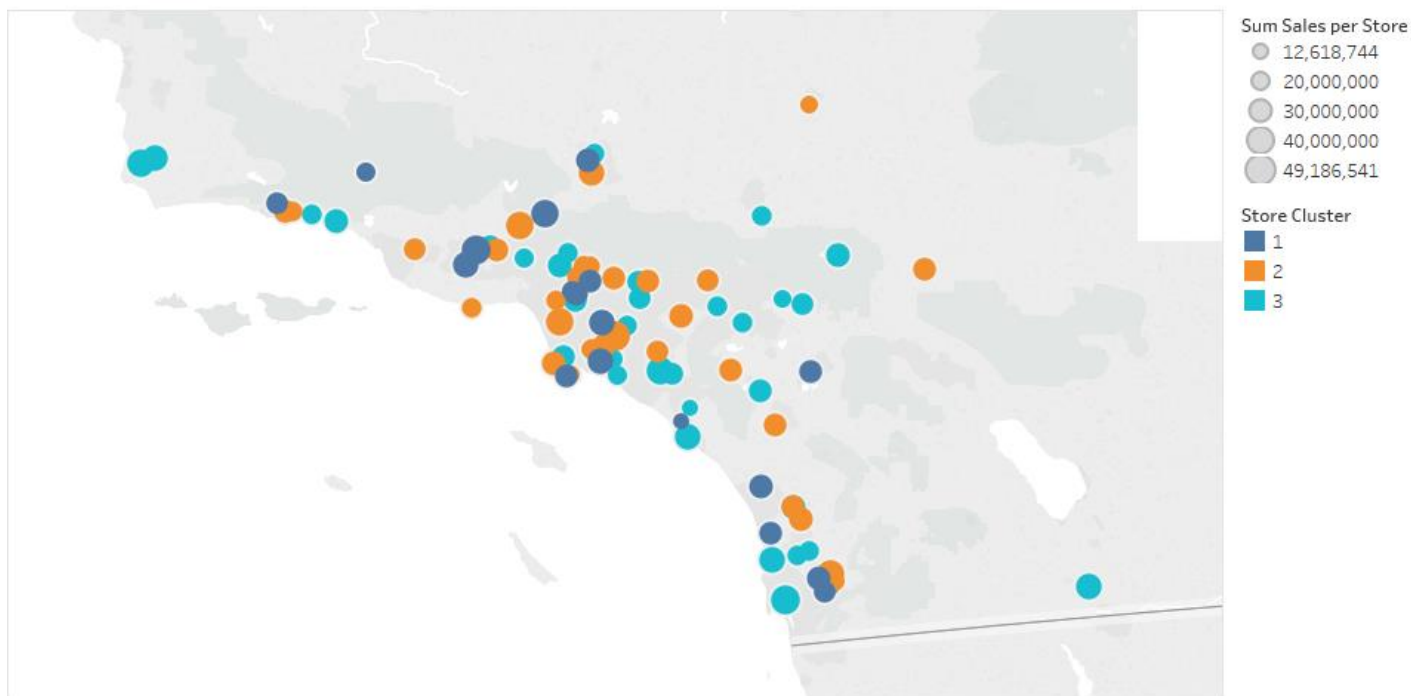
Store Cluster standardized



The trends of Dairy\_sales\_Stdzd, Deli\_sales\_Stdzd, Dry Grocery\_sales\_Stdzd, Floral\_sales\_Stdzd, Frozen Food\_sales\_Stdzd, General Merchandise\_sales\_Stdzd, Meat\_sales\_Stdzd and Produce\_sales\_Stdzd for Store Cluster. Color shows details about Dairy\_sales\_Stdzd, Deli\_sales\_Stdzd, Dry Grocery\_sales\_Stdzd, Floral\_sales\_Stdzd, Frozen Food\_sales\_Stdzd, General Merchandise\_sales\_Stdzd, Meat\_sales\_Stdzd and Produce\_sales\_Stdzd.

- 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.

Store clusters



Map based on average of LONGITUDE and average of LATITUDE. Color shows details about Store Cluster. Size shows sum of Sum Sales per Store. Details are shown for various dimensions.

[https://public.tableau.com/profile/homagni.bhattacharjee#!/vizhome/P\\_8\\_new/Sheet1](https://public.tableau.com/profile/homagni.bhattacharjee#!/vizhome/P_8_new/Sheet1)

## Task 2: Formats for New Stores

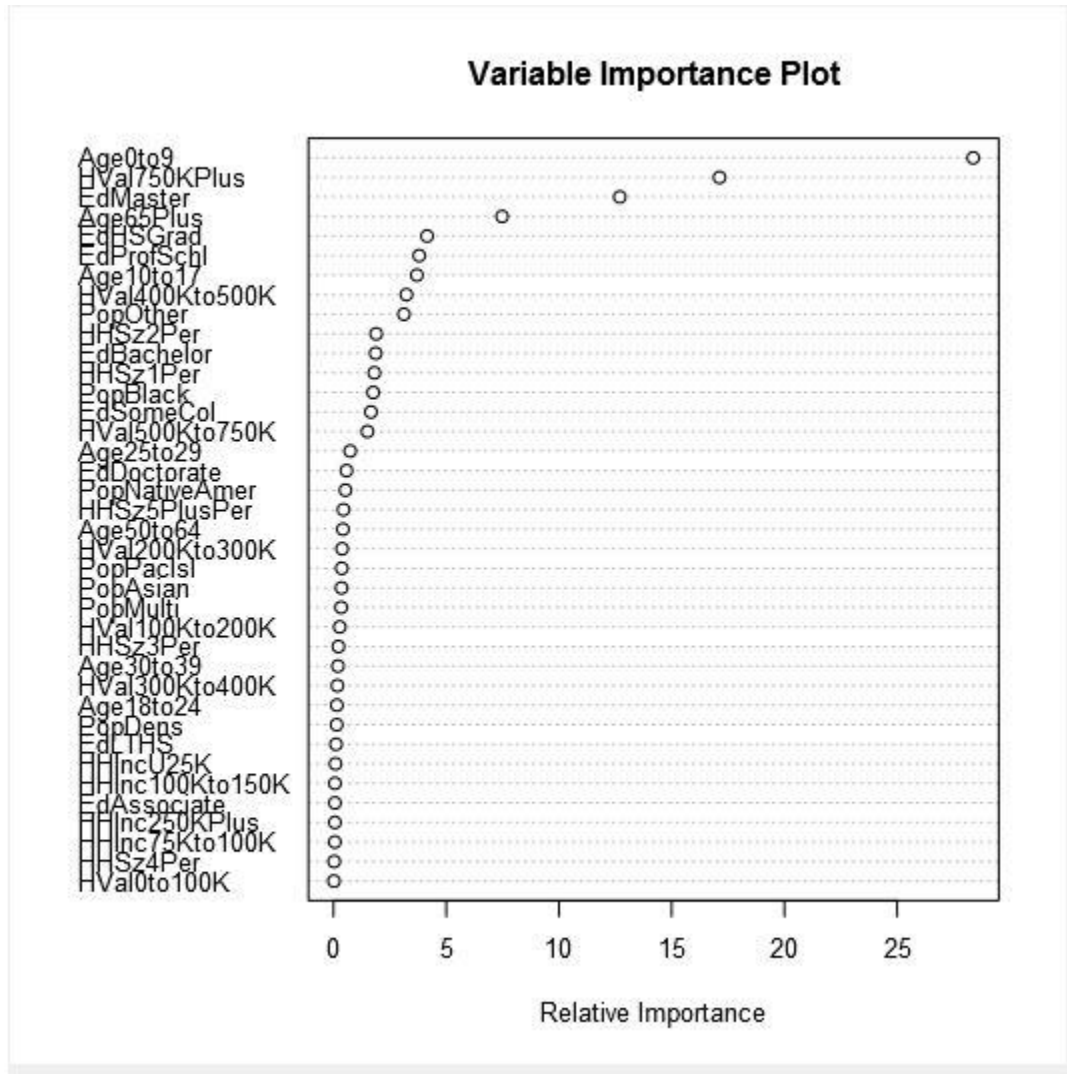
- What methodology did you use to predict the best store format for the new stores?

The boosted model was used to predict the best store format for the new stores because of the following reasons:

- It had the high accuracy.
- Highest F1 value of 0.8543.

Fit and error measures		
Model	Accuracy	F1
Task2_Boosted	0.8235	0.8543
Task2_Decision_Tree	0.8235	0.8251
Task2_Forest	0.8235	0.8251

2. What are the three most important variables that help explain the relationship between demographics indicators and store formats?



The 3 most important variables are Age Grp 0 to 9, Ed Master and High Value 750K plus.

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

The new stores were scored using the demographic predictor variables (all) of the 85 existing stores. The score results were in the form of 3 scores for each store, together adding up to 1. The highest value among these three score values corresponds to the corresponding store cluster as given below.

Store	Score_1	Score_2	Score_3	Store_Cluster
S0086	0.494734	0.012388	0.492878	1
S0087	0.063566	0.824404	0.11203	2
S0088	0.457695	0.060476	0.481829	3
S0089	0.02465	0.926534	0.048815	2
S0090	0.023626	0.918913	0.057461	2
S0091	0.927411	0.003404	0.069185	1
S0092	0.039425	0.886528	0.074047	2
S0093	0.890308	0.005433	0.104259	1
S0094	0.006694	0.965396	0.02791	2
S0095	0.109779	0.508989	0.381232	2

### Task 3: Predicting Produce Sales

- 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?

To forecast sales for existing stores sales across all stores by month was aggregated per month and separated based on clusters determined previously.

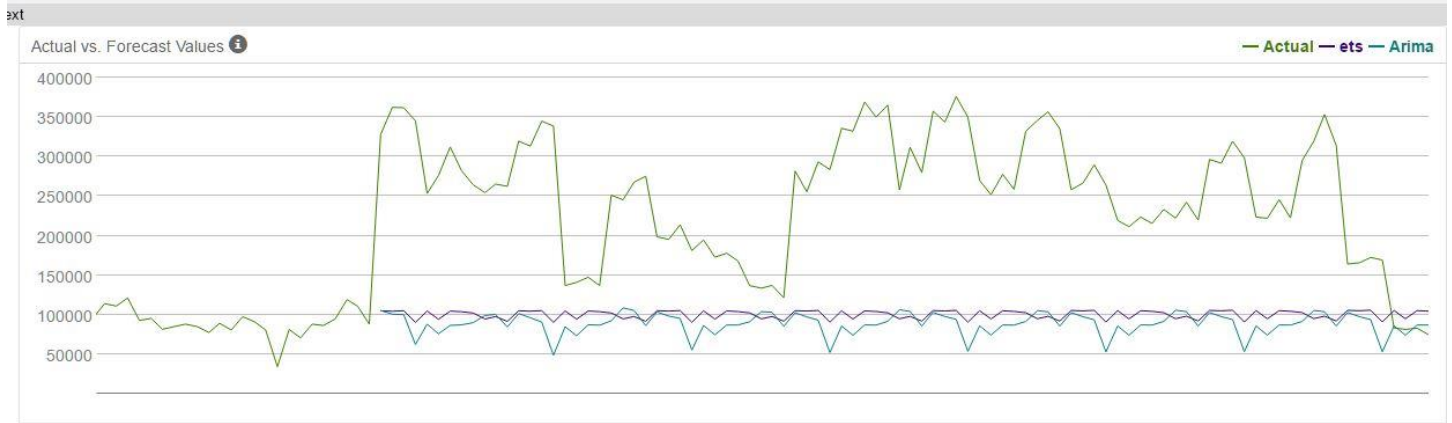
The two predictive models ETS and ARIMA were evaluated to find the best forecasting solution with the following configurations respectively for all clusters

ETS- MNM

ARIMA (1,0,0) (1,1,0)(12)

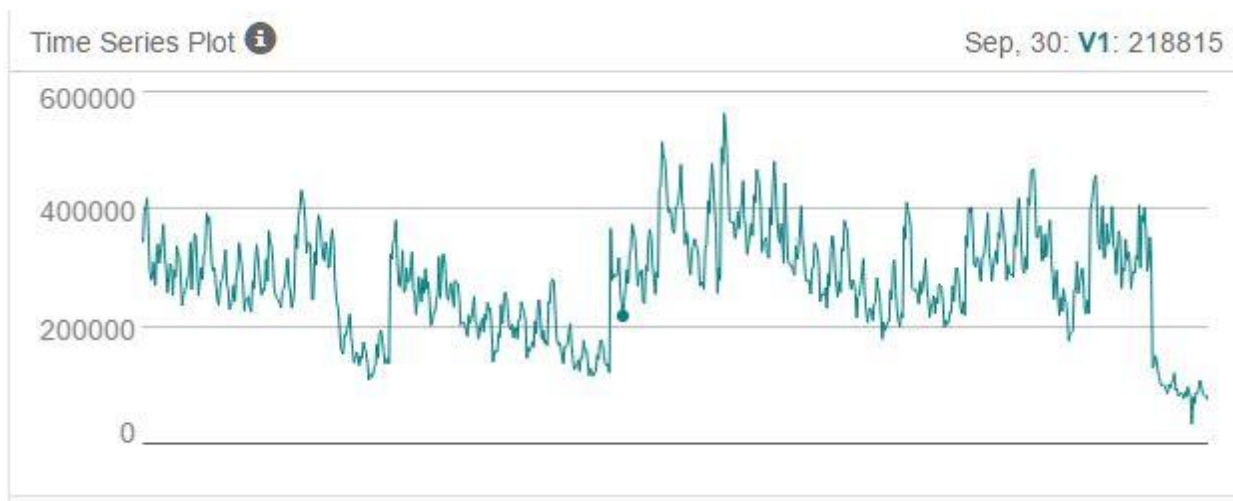
The two models are plotted against a validation sample of the last 6 periods in the dataset.

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ets	152992.3763	171010.3506	154890.2709	54.5142	56.8793	5.5657
Arima	164483.6649	181195.2387	164893.746	60.0445	60.5708	5.9252

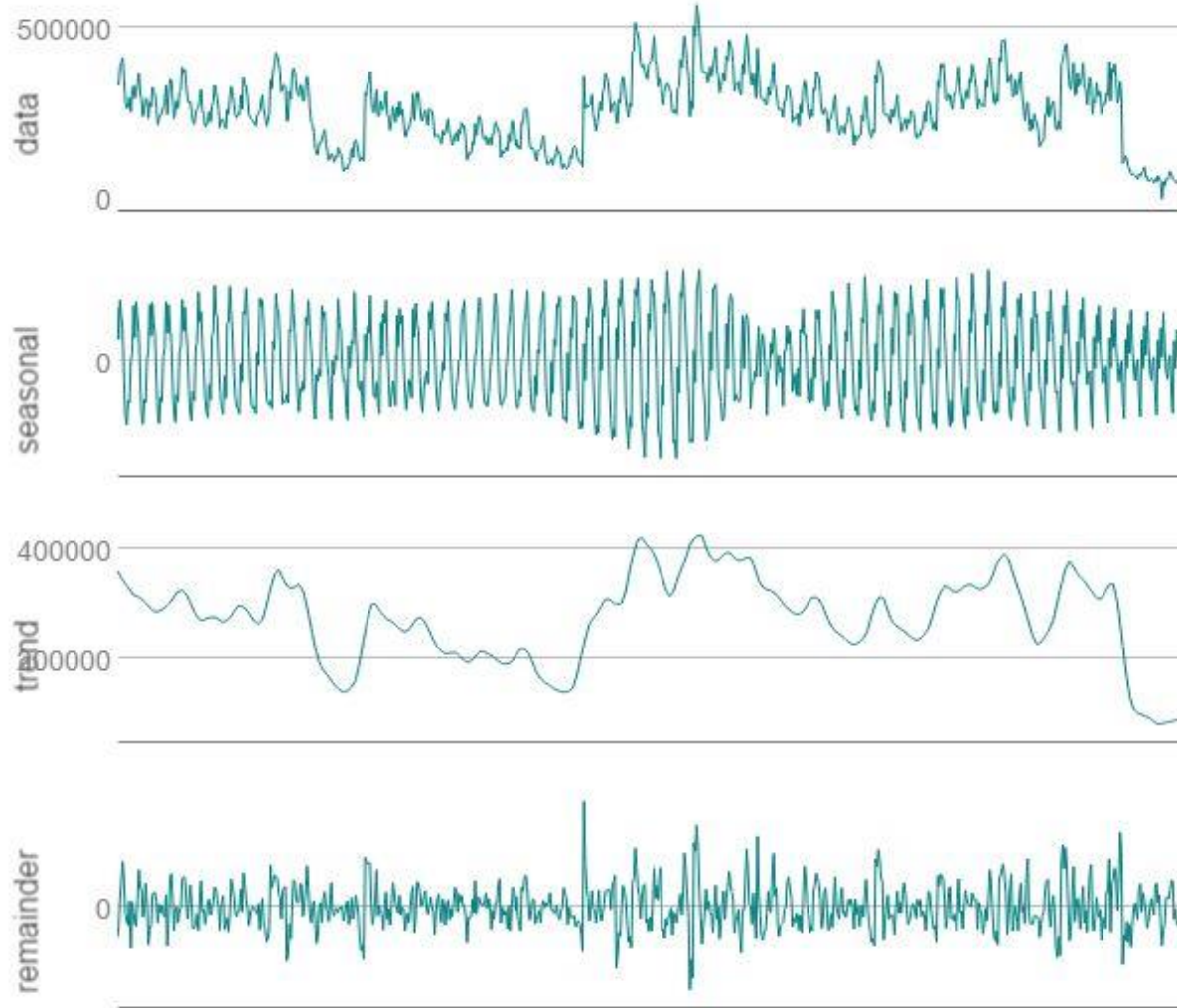


From the error terms, we can see that the ETS has lower error and hence the model ETS (M, N, M) is used to predict sales for stores of all clusters.

The terms M,N,M are chosen for the ETS due to the following reasons



### Decomposition Plot



Error: The error term changes varies over time and hence the error terms is multiplicative (M).

Trend: No discernable trend could be identified. Hence, the trend term chosen is None (N)

Seasonality: The seasonality term changes varies over time and hence the error terms is multiplicative (M). Hence, ETS configuration is MNM.

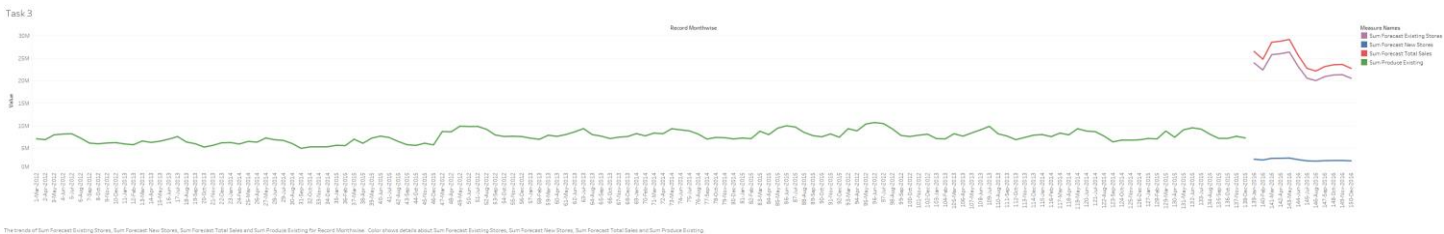
The model is then run by aggregating the daily sales data according to month and year for the stores for each cluster to generate the forecasts for the existing stores.

Now to determine the sales forecast for the new stores, the average store sales forecast per month per cluster is determined. It is then multiplied to the number of stores for each cluster in the new stores, thus giving us the total sale forecast for the new stores per cluster per month. These data are then aggregated to give the total sales per month for 2016.



RecordID	Month	Year	Sum_Forecast_Existing_Stores	Sum_Forecast_New_Stores	Sum_Forecast_Total_Sales
1	Jan	2016	24126450	2565916.5	26692366
2	Feb	2016	22622306	2410825.8125	25033132
3	Mar	2016	26003291	2768536.3125	28771827
4	Apr	2016	26402325.5	2808475.9375	29210802
5	May	2016	26657449.5	2839170.3125	29496620
6	Jun	2016	23440501.5	2495422.6875	25935924.5
7	Jul	2016	20673795.5	2201385.1875	22875181
8	Aug	2016	20141288.5	2142556.9375	22283845
9	Sep	2016	20869869	2218811.375	23088680
10	Oct	2016	21162678	2249924.25	23412602.5
11	Nov	2016	21489544.5	2285888.5625	23775433
12	Dec	2016	20745220	2208711.875	22953931.5

2. Please provide a Tableau Dashboard (saved as a Tableau Public file) that includes a table and a plot of the three monthly forecasts; one for existing, one for new, and one for all stores. Please name the tab in the Tableau file "Task 3".



[https://public.tableau.com/profile/homagni.bhattacharjee#!/vizhome/P\\_8\\_TASK\\_3/Sheet1](https://public.tableau.com/profile/homagni.bhattacharjee#!/vizhome/P_8_TASK_3/Sheet1)