

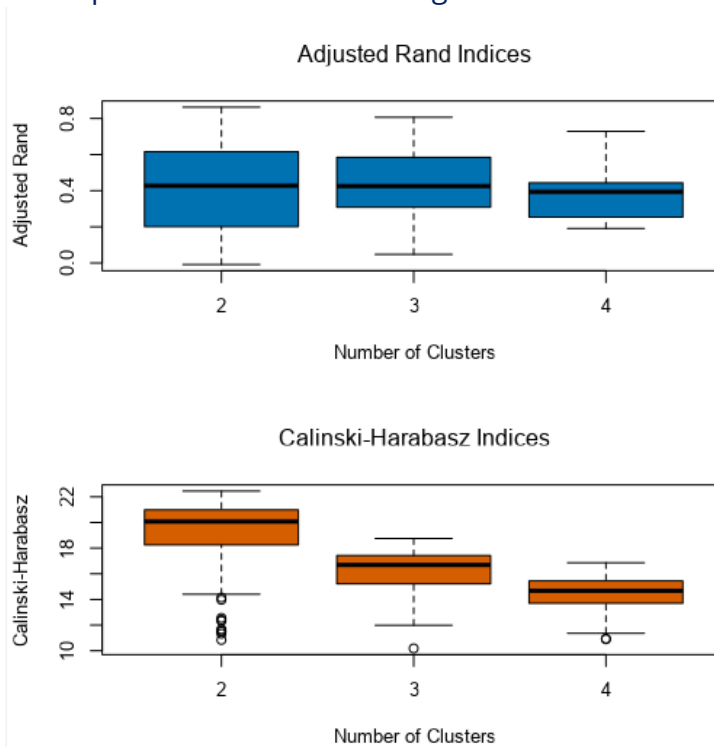
Project: Predictive Analytics Capstone

Task 1: Determine Store Formats for Existing Stores

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

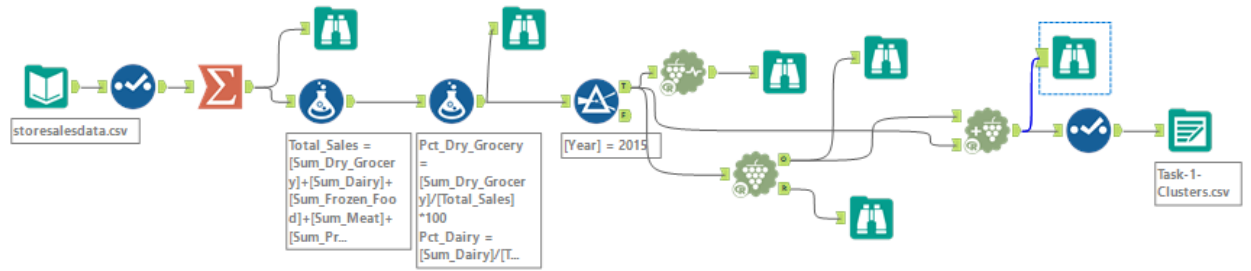
Ans: The optimal number of store formats is **3**. The way I reached to this conclusion is explained below:

- Used the **Store Sales Data** to find the total sales and percentage sales of each category.
- Filtered the sales for the year 2015 and used the K-Centroids Diagnostic tool to find the optimal number of formals.
- The report showed the following results:



As clearly shown in the AR Indices plot, 3 number of clusters has a high median value as well as it has a lesser spread, making it the best candidate for optimal number of store formats.

Workflow:



2. How many stores fall into each store format?

Ans: Segment 1: 25
 Segment 2: 35
 Segment 3: 25

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

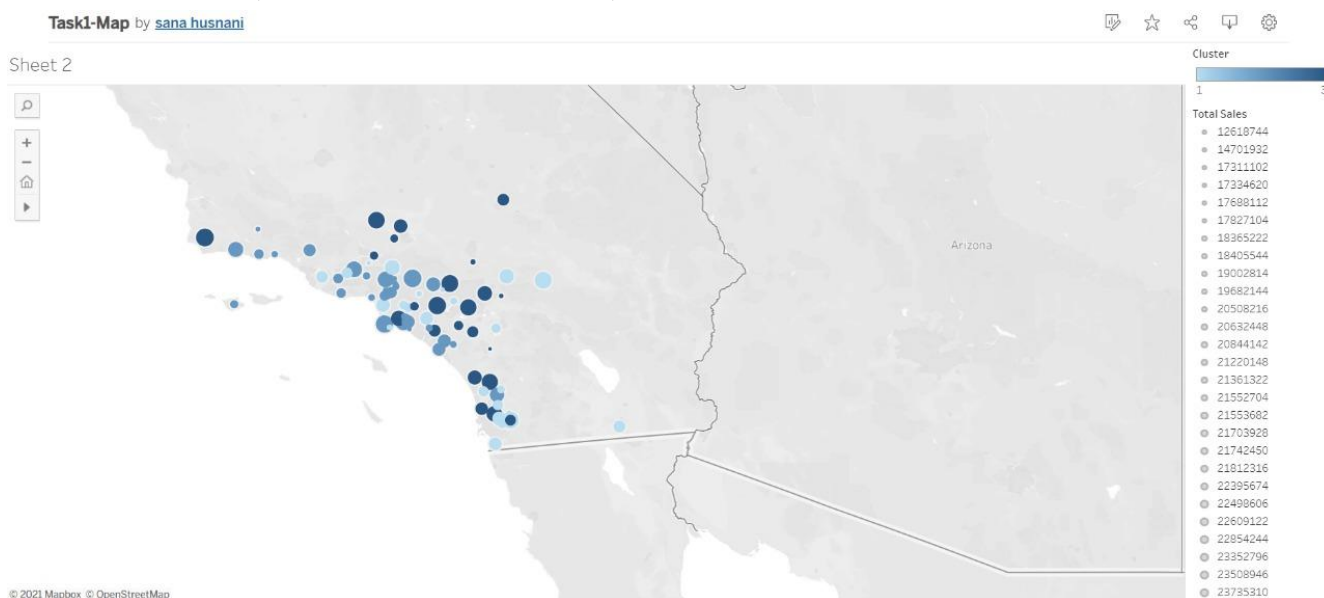
Ans:

	Pct_Dry_Grocery	Pct_Dairy	Pct_Frozen_Food	Pct_Meat	Pct_Produce	Pct_Floral	Pct_Deli
1	0.528249	-0.215879	-0.261597	0.614147	-0.655027	-0.663872	0.824834
2	-0.594802	0.655893	0.435129	-0.384631	0.812883	0.71741	-0.46168
3	0.304474	-0.702372	-0.347583	-0.075664	-0.483009	-0.340502	-0.178482
	Pct_Bakery	Pct_GM					
1	0.428226	-0.674769					
2	0.312878	-0.329045					
3	-0.866255	1.135432					

As shown in the table above top selling category in cluster 1 is Deli and Meat. For cluster 2, top two selling categories are Produce and Floral. And in cluster 3, GM and Grocery are the top two selling categories.

4. 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.

Ans: Task1-Map by [sana husnani](#)



Task 2: Formats for New Stores

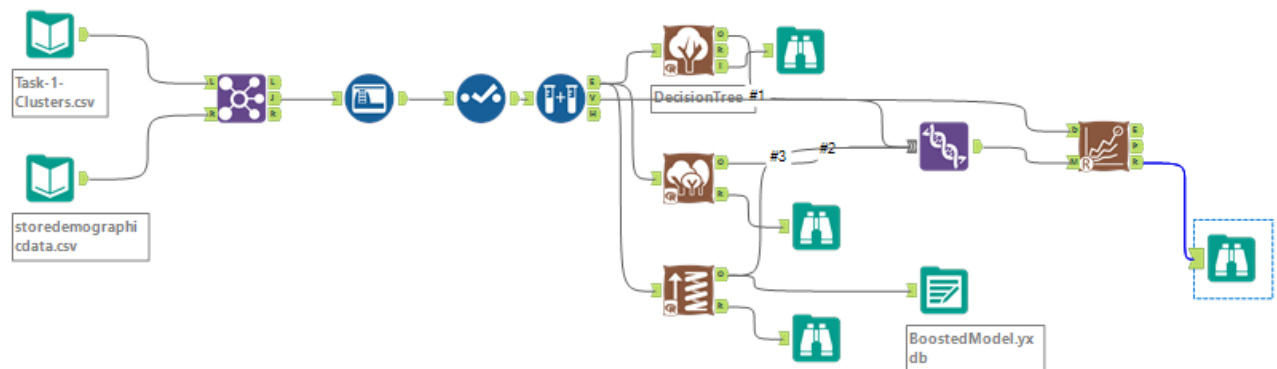
1. What methodology did you use to predict the best store format for the new stores? Why did you choose that methodology? (Remember to Use a 20% validation sample with Random Seed = 3 to test differences in models.)

Ans: The best classification model to predict the segments for the new stores is **Boosted Model**. I used Model Comparison tool to pick the best classification model out of Decision Tree, Random Forest and Boosted Model. The Model Comparison tool showed the results as shown below:

Model Comparison Report					
Fit and error measures					
Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
DecisionTree	0.6471	0.6667	0.5000	1.0000	0.5000
RandomForest	0.7059	0.7500	0.5000	1.0000	0.7500
BoostedModel	0.7647	0.8333	0.5000	1.0000	1.0000

As clearly seen in the screenshot above, boosted model has the highest accuracy and the highest F1 score and hence it is our best pick for classification of the new stores.

Workflow:

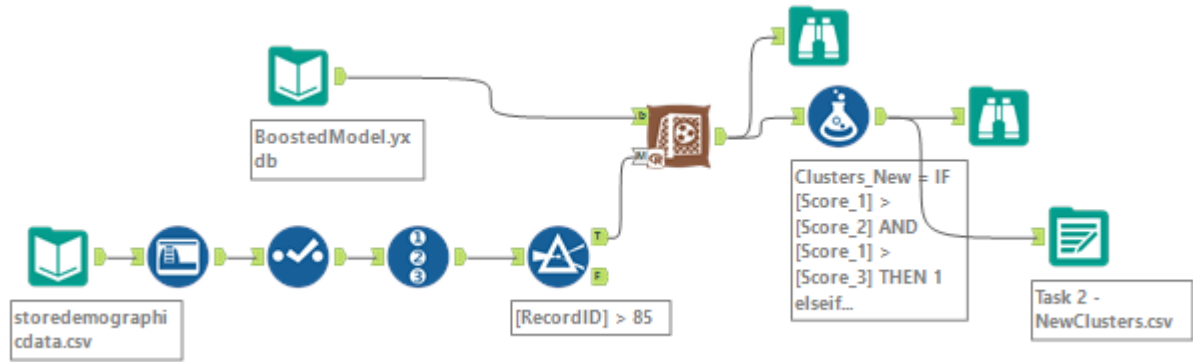


2. What format do each of the 10 new stores fall into? Please fill in the table below.

Store Number	Segment
S0086	1
S0087	2
S0088	3
S0089	2
S0090	2
S0091	3
S0092	2

S0093	3
S0094	2
S0095	2

Workflow.

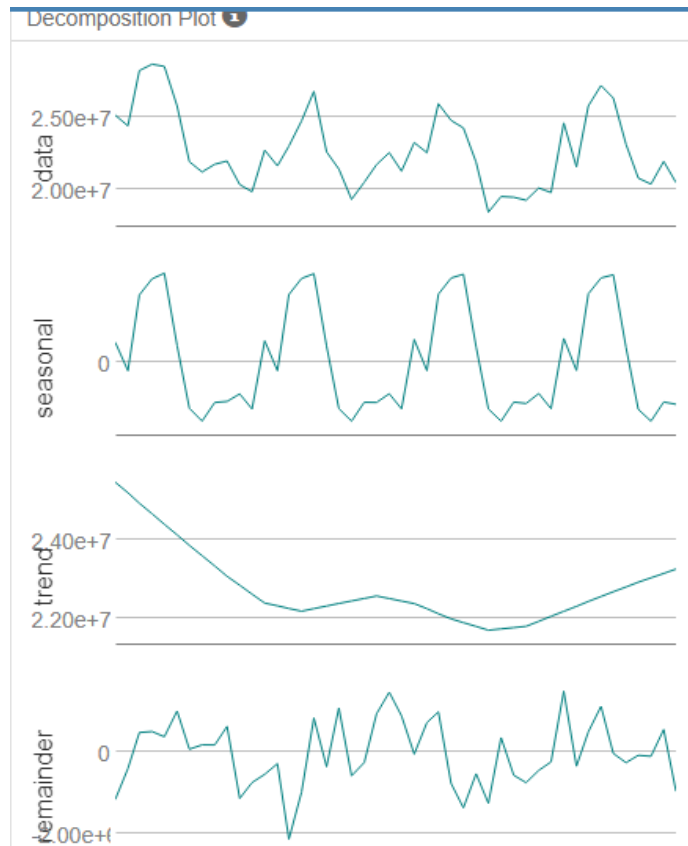


Task 3: Predicting Produce Sales

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?

Ans: I used **ETS(M,N,M)** model to make the forecast.

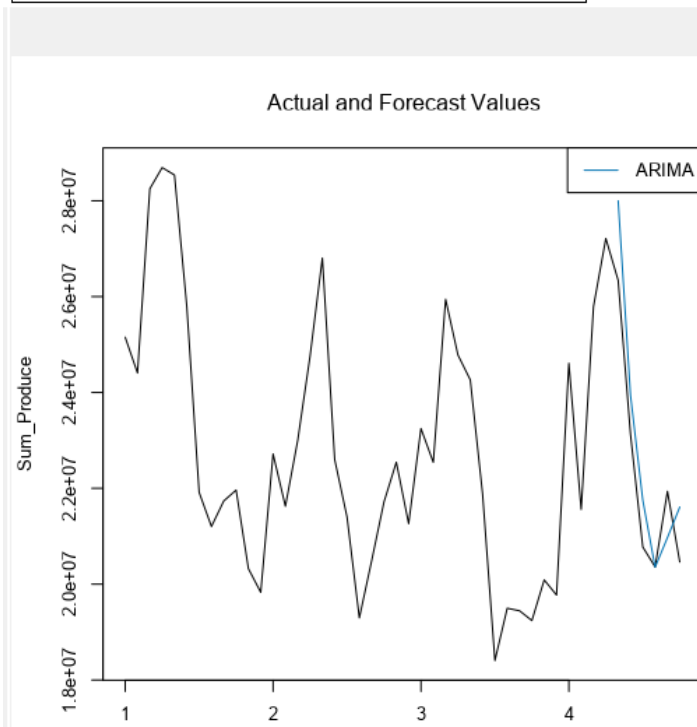
I first studied the trend, seasonality and the error of the data.



As seen in the decomposition plot, the seasonality is multiplicative, there is no trend and the error is multiplicative too. I used ETS (M,N,M) vs ARIMA (1,0,0)(1,1,0) and compared the performance of both using the TS Compare tool. The results of which are shown below:

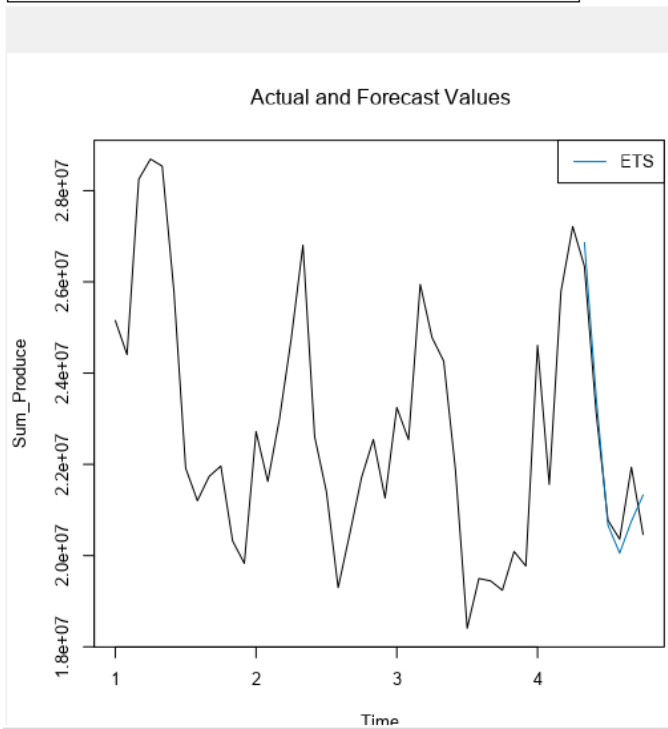
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-604232.3	1050239	928412	-2.6156	4.0942	0.5463



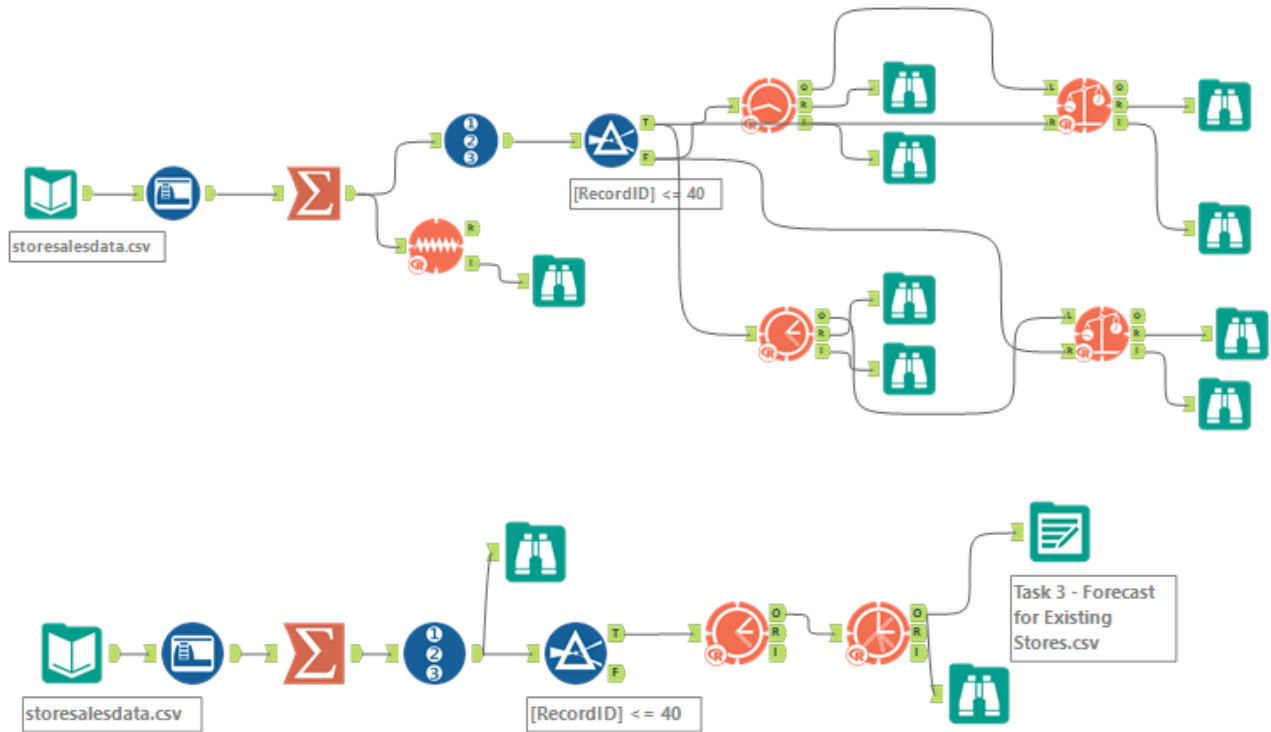
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-21581.13	663707.2	553511.5	-0.0437	2.5135	0.3257



The graphs below show that ETS performs better, has lower error ratio and better accuracy. Hence, I used ETS(M,N,M) to make the forecast.

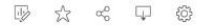
Workflow:



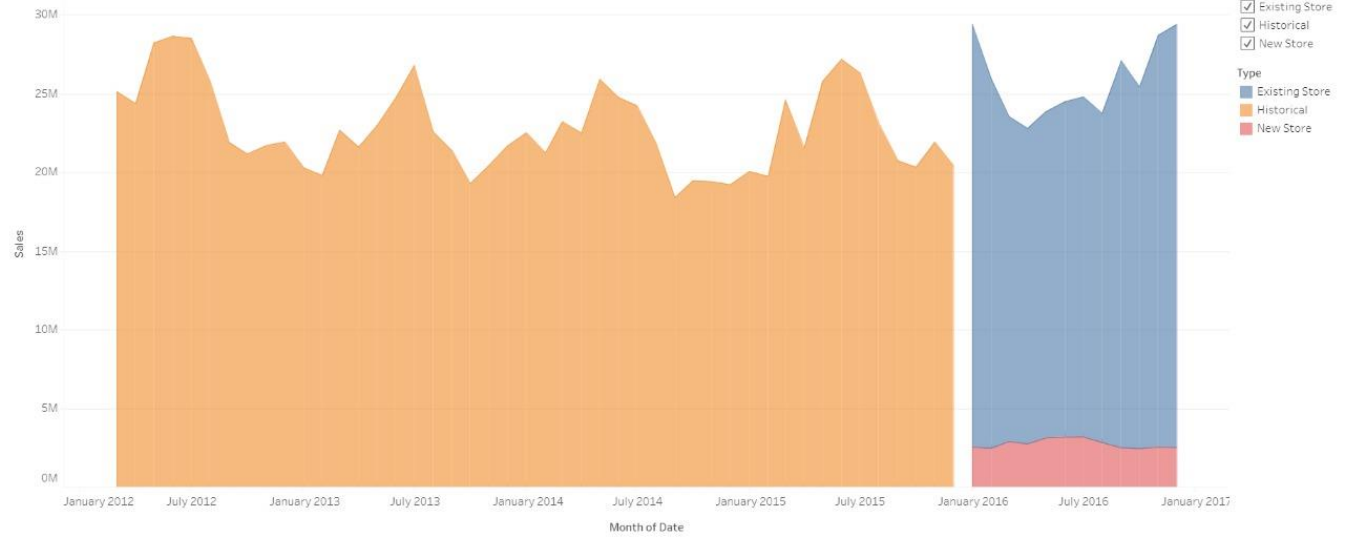
3. Please provide a table of your forecasts for existing and new stores. Also, provide visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.

Month	New Stores	Existing Stores
Jan – 16	2563357.91	26860639.57
Feb – 16	2483924.728	23468254.5
Mar – 16	2910944.146	20668464.64
Apr – 16	2764881.87	20054544.08
May – 16	3141305.867	20752503.52
Jun – 16	3195054.204	21328386.81
Jul – 16	3212390.954	21611877.98
Aug – 16	2852385.769	20931380.13
Sep – 16	2521697.187	24588621.43
Oct – 16	2466750.894	22974656.79
Nov – 16	2557744.588	26185910.65
Dec - 16	2530510.805	26879542.76

Task 3- Viz by sana husnani



Sheet 2



Workflow:

