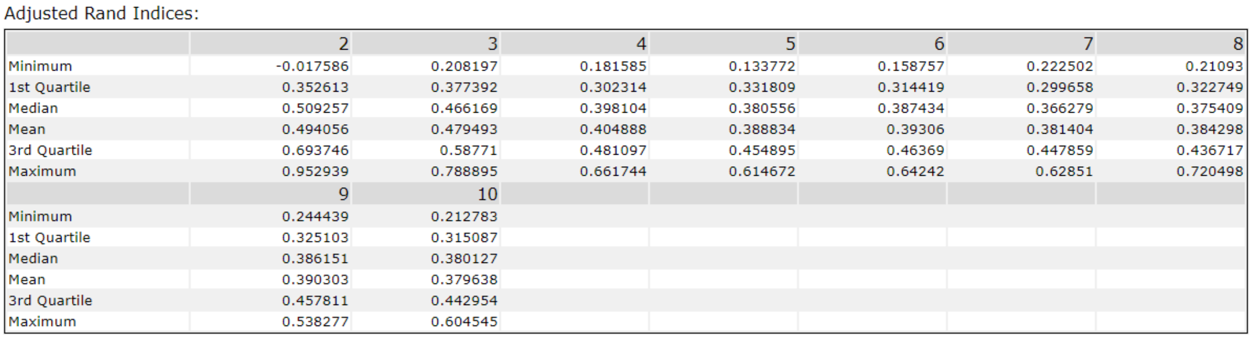
Project: Predictive Analytics Capstone

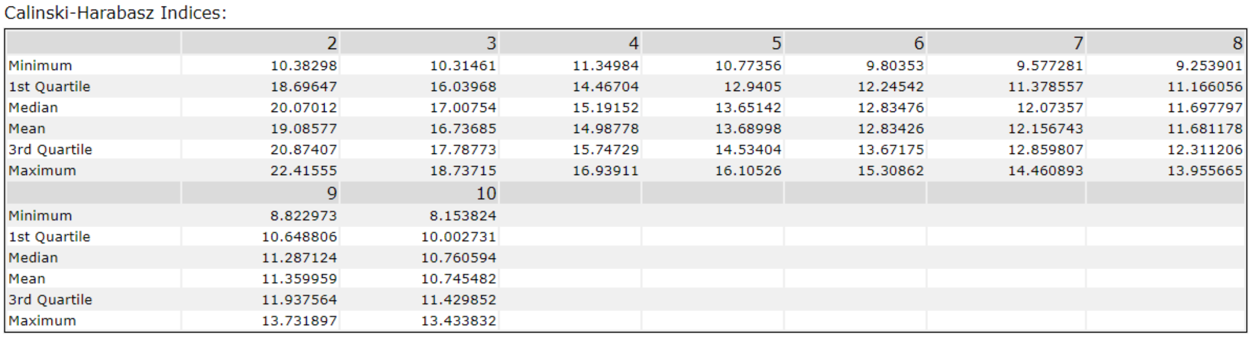
Task 1: Determine Store Formats for Existing Stores

1. What is the optimal number of store formats? How did you arrive at that number?
   Optimal number of store formats is 3. As Cluster 3 has relatively high median and compact spread.



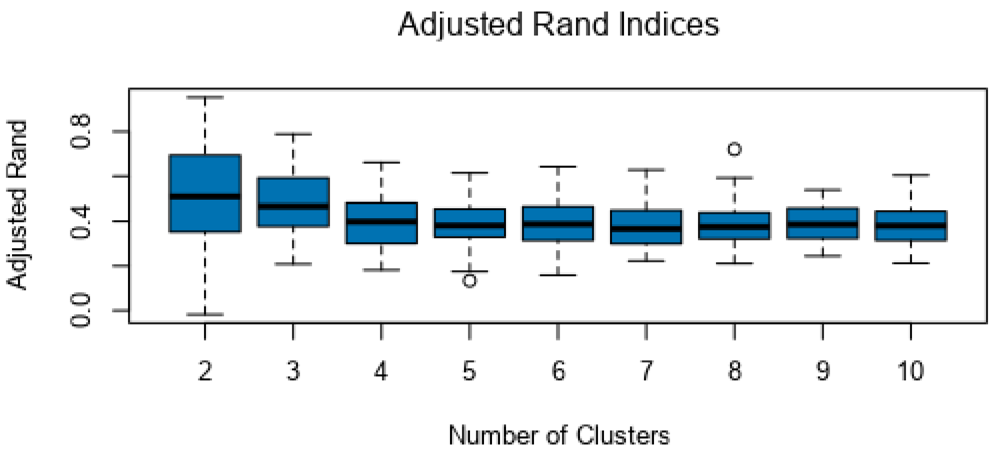
Adjusted Rand Indices:

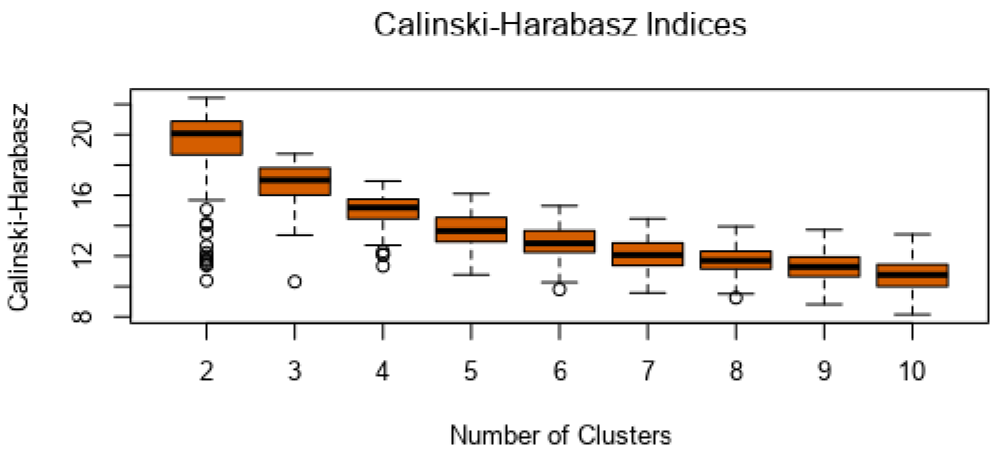
	2	3	4	5	6	7	8	9	10
Minimum	-0.017586	0.208197	0.181585	0.133772	0.158757	0.222502	0.21093	0.244439	0.212783
1st Quartile	0.352613	0.377392	0.302314	0.331809	0.314419	0.299658	0.322749	0.325103	0.315087
Median	0.509257	0.466169	0.398104	0.380556	0.387434	0.366279	0.375409	0.386151	0.380127
Mean	0.494056	0.479493	0.404888	0.388834	0.39306	0.381404	0.384298	0.390303	0.379638
3rd Quartile	0.693746	0.58771	0.481097	0.454895	0.46369	0.447859	0.436717	0.457811	0.442954
Maximum	0.952939	0.788895	0.661744	0.614672	0.64242	0.62851	0.720498	0.538277	0.604545



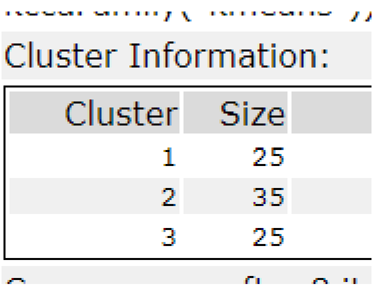
Calinski-Harabasz Indices:

	2	3	4	5	6	7	8	9	10
Minimum	10.38298	10.31461	11.34984	10.77356	9.80353	9.577281	9.253901	8.822973	8.153824
1st Quartile	18.69647	16.03968	14.46704	12.9405	12.24542	11.378557	11.166056	10.648806	10.002731
Median	20.07012	17.00754	15.19152	13.65142	12.83476	12.07357	11.697797	11.287124	10.760594
Mean	19.08577	16.73685	14.98778	13.68998	12.83426	12.156743	11.681178	11.359959	10.745482
3rd Quartile	20.87407	17.78773	15.74729	14.53404	13.67175	12.859807	12.311206	11.937564	11.429852
Maximum	22.41555	18.73715	16.93911	16.10526	15.30862	14.460893	13.955665	13.731897	13.433832





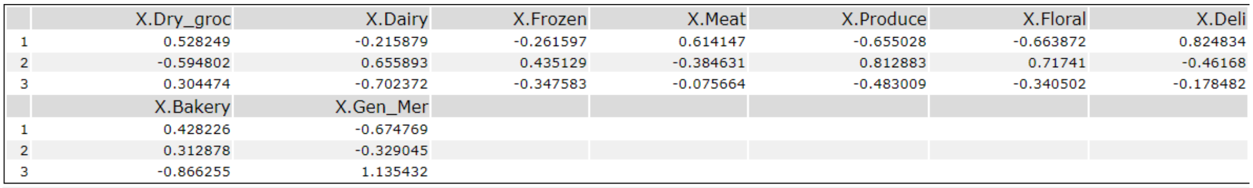
2. How many stores fall into each store format?
   Cluster 1: 25
   Cluster 2: 35
   Cluster 3: 25



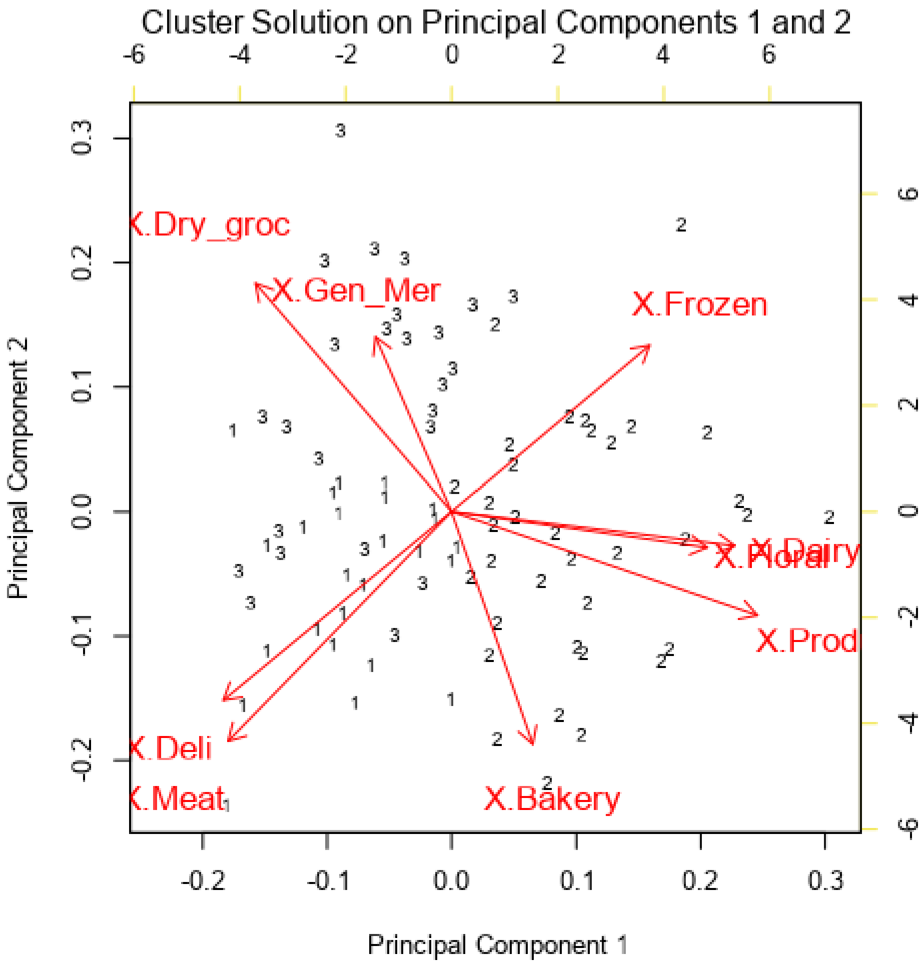
Cluster Information:

Cluster	Size	
1	25	
2	35	
3	25	

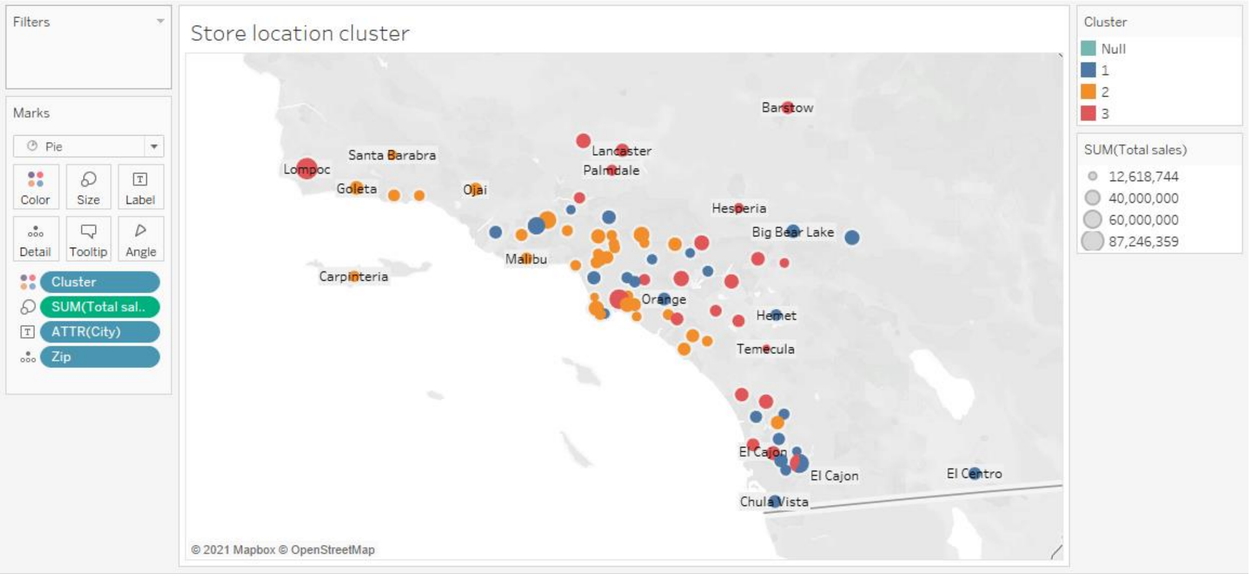
3. Based on the results of the clustering model, what is one way that the clusters differ from one another?
   One way to differentiate clusters is that store 1 sold more Deli, store 2 sold more floral and produce categories, while store 3 sold more General merchandize



	X.Dry_groc	X.Dairy	X.Frozen	X.Meat	X.Produce	X.Floral	X.Deli	X.Bakery	X.Gen_Mer
1	0.528249	-0.215879	-0.261597	0.614147	-0.655028	-0.663872	0.824834	0.428226	-0.674769
2	-0.594802	0.655893	0.435129	-0.384631	0.812883	0.71741	-0.46168	0.312878	-0.329045
3	0.304474	-0.702372	-0.347583	-0.075664	-0.483009	-0.340502	-0.178482	-0.866255	1.135432

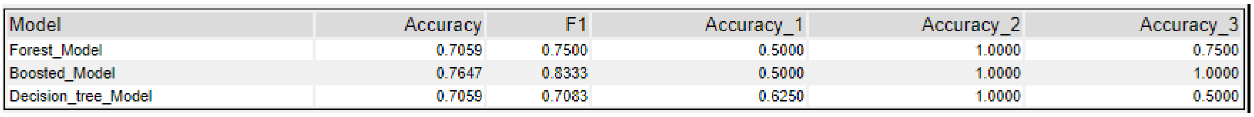


4. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.



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.)
   I will go with Boosted model, because of highest Accuracy and F1 score.



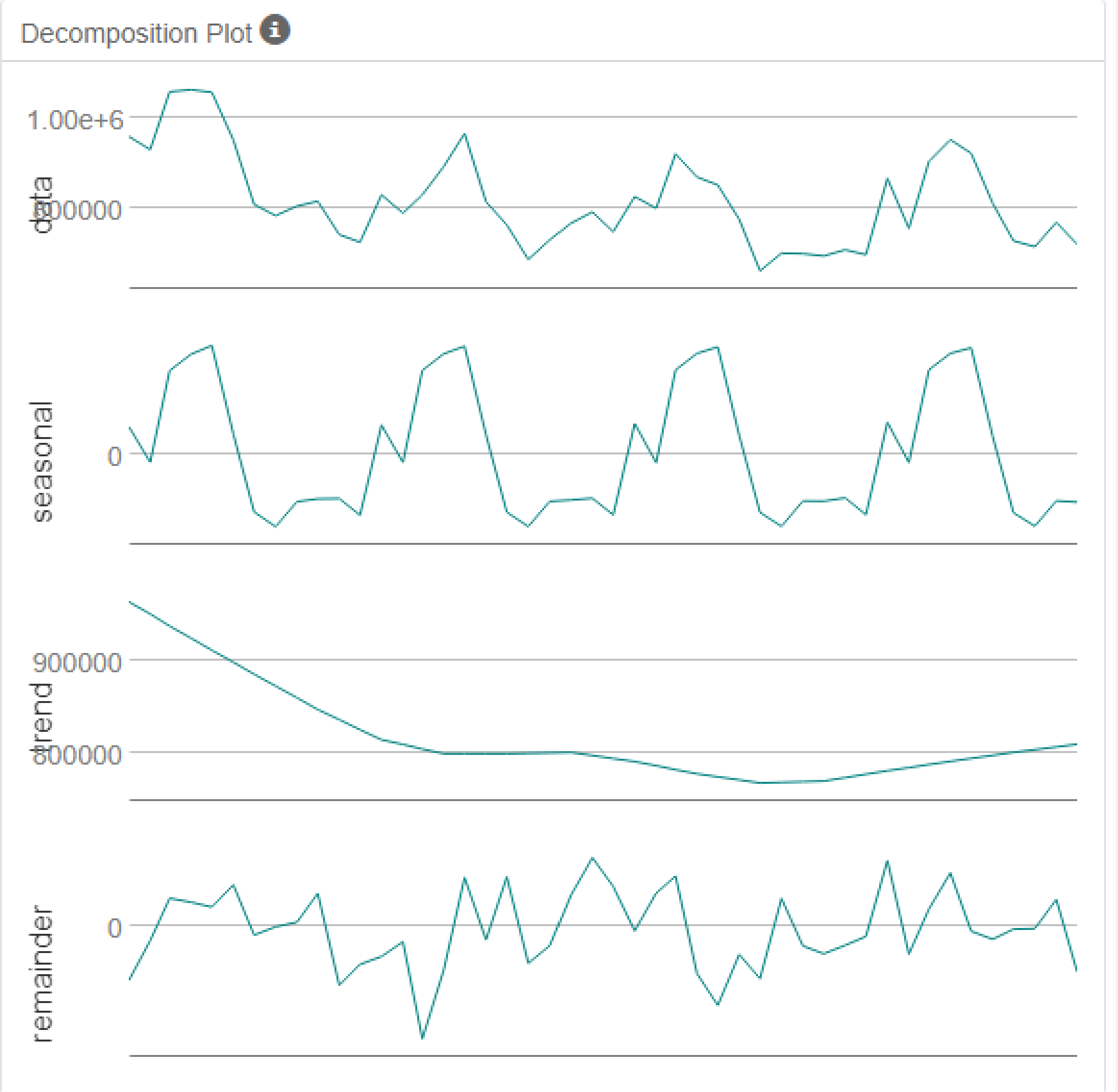
Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Forest_Model	0.7059	0.7500	0.5000	1.0000	0.7500
Boosted_Model	0.7647	0.8333	0.5000	1.0000	1.0000
Decision_tree_Model	0.7059	0.7083	0.6250	1.0000	0.5000

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

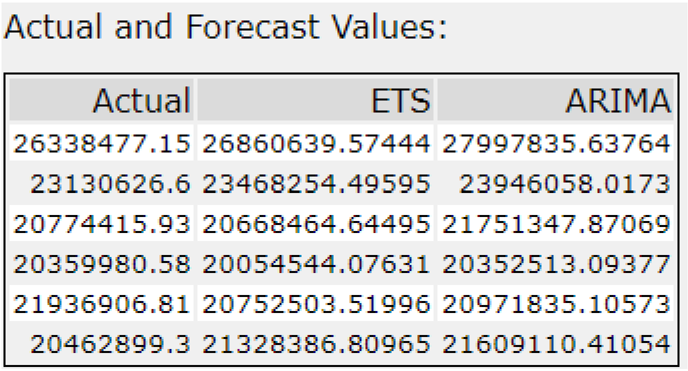
Store Number	Segment
S0086	1
S0087	2
S0088	3
S0089	2
S0090	2
S0091	3
S0092	2
S0093	3
S0094	2
S0095	2

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?
I used ETS with (M,N,M) configuration: Multiplicative error, No trend and Multiplicative season which can be observed in the figure below.
Based on greater accuracy, RMSE, and MASE values ETS model performed better than Arima.

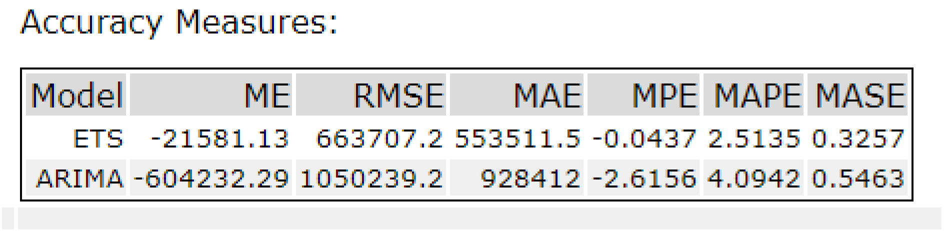


Decomposition Plot



Actual and Forecast Values:

Actual	ETS	ARIMA
26338477.15	26860639.57444	27997835.63764
23130626.6	23468254.49595	23946058.0173
20774415.93	20668464.64495	21751347.87069
20359980.58	20054544.07631	20352513.09377
21936906.81	20752503.51996	20971835.10573
20462899.3	21328386.80965	21609110.41054

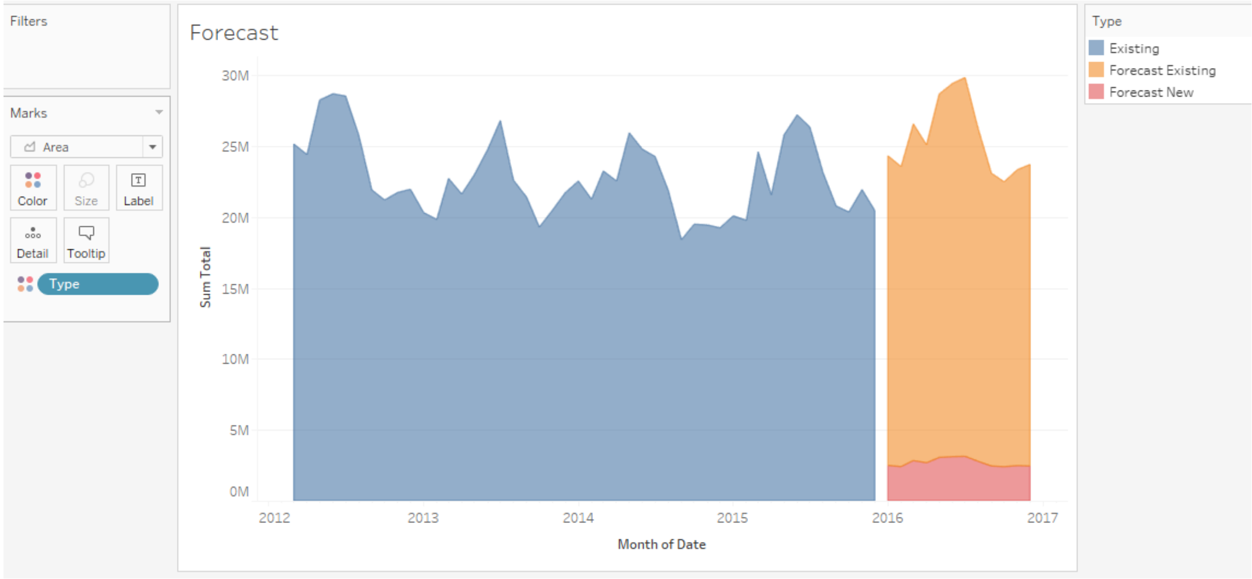


Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-21581.13	663707.2	553511.5	-0.0437	2.5135	0.3257
ARIMA	-604232.29	1050239.2	928412	-2.6156	4.0942	0.5463

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.

Month	**New Stores**	**Existing Stores**
Jan-16	2,491,319	21,829,060
Feb-16	2,408,385	21,146,330
Mar-16	2,833,157	23,735,687
Apr-16	2,679,433	22,409,515
May-16	3,054,886	25,621,829
Jun-16	3,106,152	26,307,858
July-16	3,132,699	26,705,093
Aug-16	2,776,154	23,440,761
Sep-16	2,451,566	20,640,047
Oct-16	2,401,772	20,086,270
Nov-16	2,477,302	20,858,120
Dec-16	2,452,170	21,255,190

Filters
Marks
Area
Color
Size
Label
Detail
Tooltip
Type
Forecast
30M
25M
20M
15M
10M
5M
0M
Sum Total
2012
2013
2014
2015
2016
2017
Month of Date
Type
Existing
Forecast Existing
Forecast New