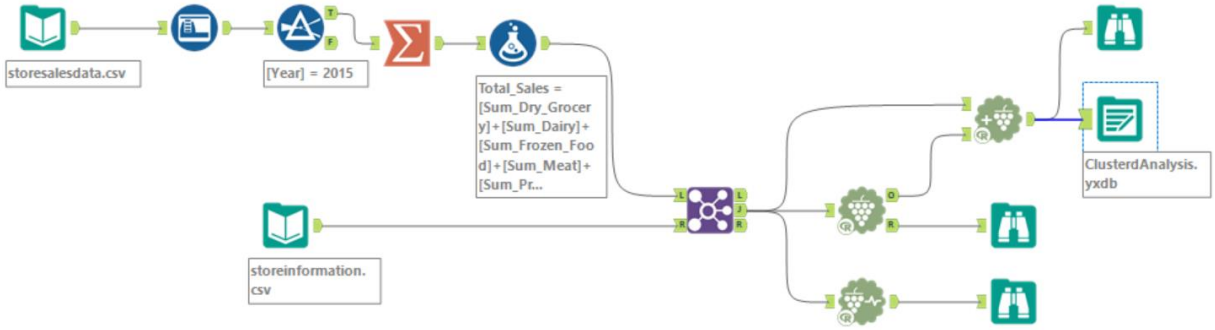
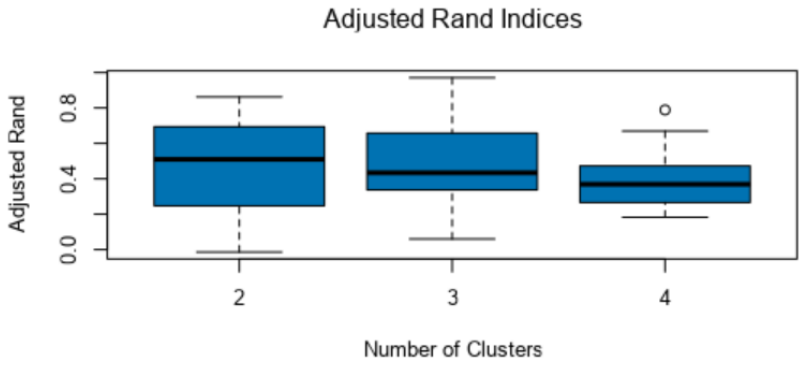
Project: Predictive Analytics Capstone

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 three. This is because it has high mean values within the Adjusted Rand.



2. **How many stores fall into each store format?**

   Cluster 1 has 23 store, Cluster 2 has 29 stores and Cluster 3 has 33 stores.

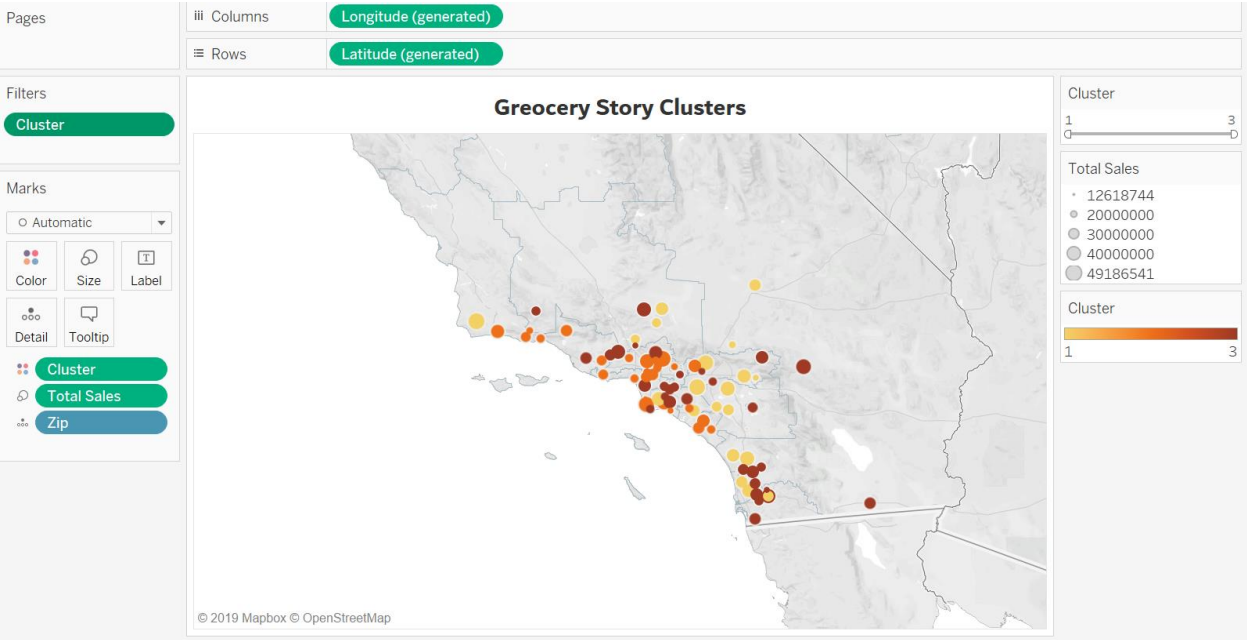
Cluster Information:

Cluster	Size	Ave Distance	Max Distance	Separation
1	23	2.320539	3.55145	1.874243
2	29	2.540086	4.475132	2.118708
3	33	2.115045	4.9262	1.702843

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

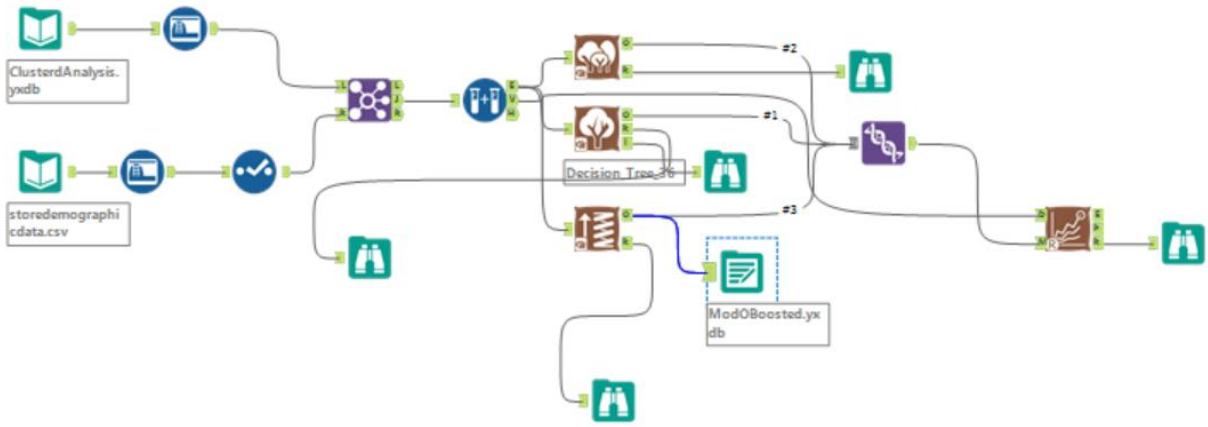
   Stores that fall under cluster 2 for example would want an increase in Meat inventory as compared to stores that fall under cluster 1 & 3.

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**

5. **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.)**



To figure out which model to use, we will union the 3 models and use a model comparison tool to generate a report for us.

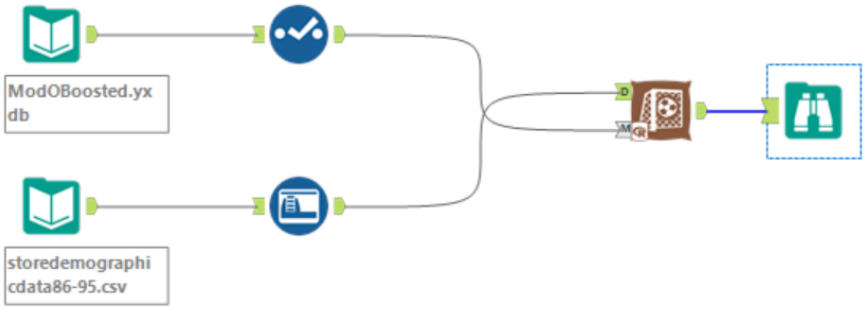
Model Comparison Report

Fit and error measures

Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Boosted_Model	0.8235	0.8543	0.8000	0.6667	1.0000
Decision_Tree	0.7059	0.7327	0.6000	0.6667	0.8333
Forest_Model	0.8235	0.8251	0.7500	0.8000	0.8750

The model comparison report shows comparison matrix of Decision Tree, Forest Model and Boosted Model.

Boosted Model is chosen despite having same accuracy as Forest Model due to higher F1 value.



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

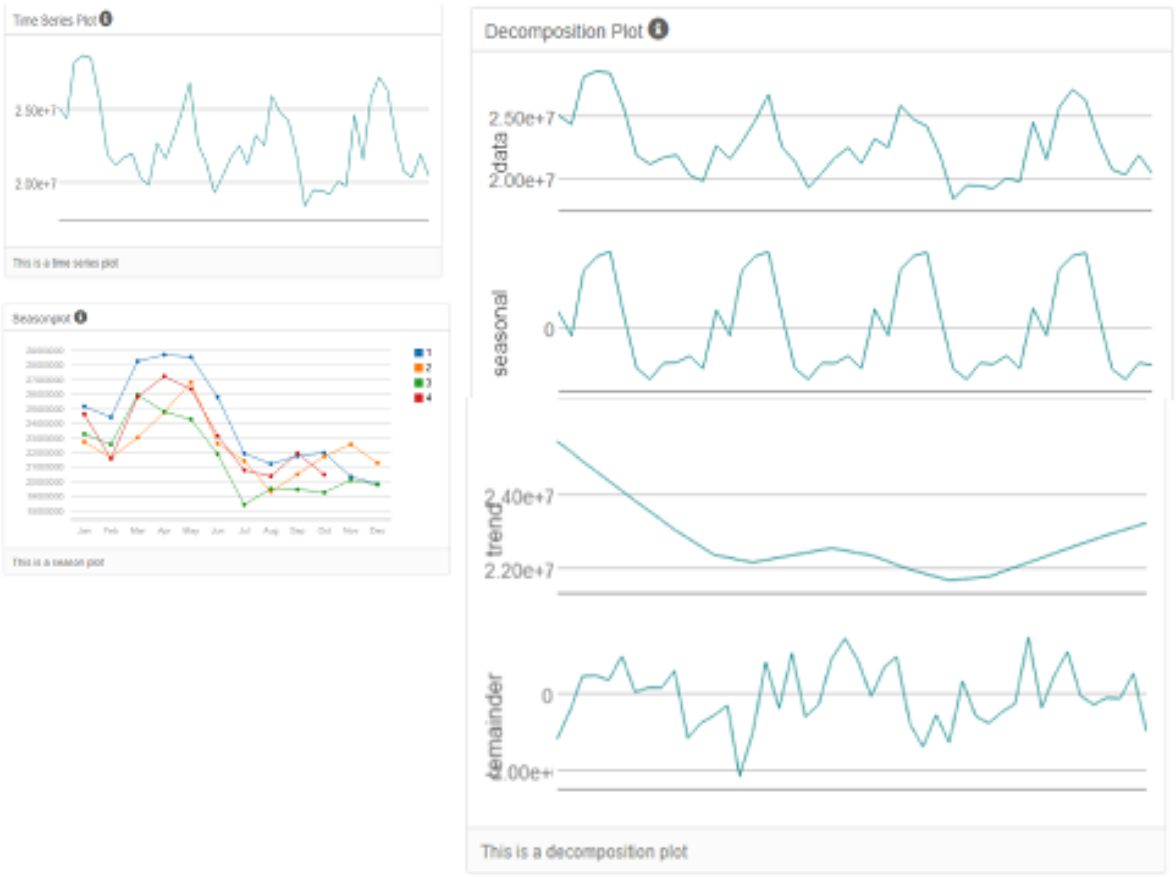
Store Number	Segment
S0086	3
S0087	2
S0088	1
S0089	2
S0090	2
S0091	1
S0092	2
S0093	1
S0094	2
S0095	2

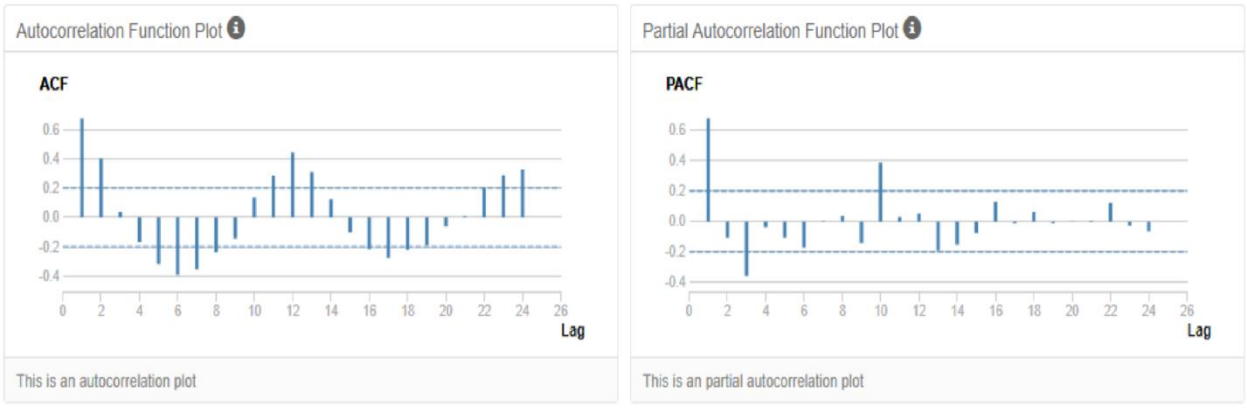
Task 3: **Predicting Produce Sales**

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?**

ETS(M,N,M) with no dampening is used for ETS model.

The seasonality shows increasing trend and should be applied multiplicatively. The trend is not clear and nothing should be applied. Its error is irregular and should be applied multiplicatively.





ETS model's accuracy is higher when compared to ARIMA model. A holdout sample of 6 months data is used. Its RMSE of **663,707**is lower than ARIMA's **1,042,209** while its MASE is **0.32** compared to ARIMA's **0.54**. ETS also has a higher AIC at **1,279** while ARIMA's AIC is **880**.

Method:
ETS(M,N,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3502.9443415	969051.6076376	787577.7006835	-0.1381187	3.4677635	0.4396486	0.0077488

Information criteria:

AIC	AICc	BIC
1279.4203	1299.4203	1304.7535

Accuracy Measures:

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

Information Criteria:

AIC	AICc	BIC
880.4445	881.4445	884.4411

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-102530.8325034	1042209.8528363	738087.5530941	-0.5465069	3.3006311	0.4120218	-0.1854462

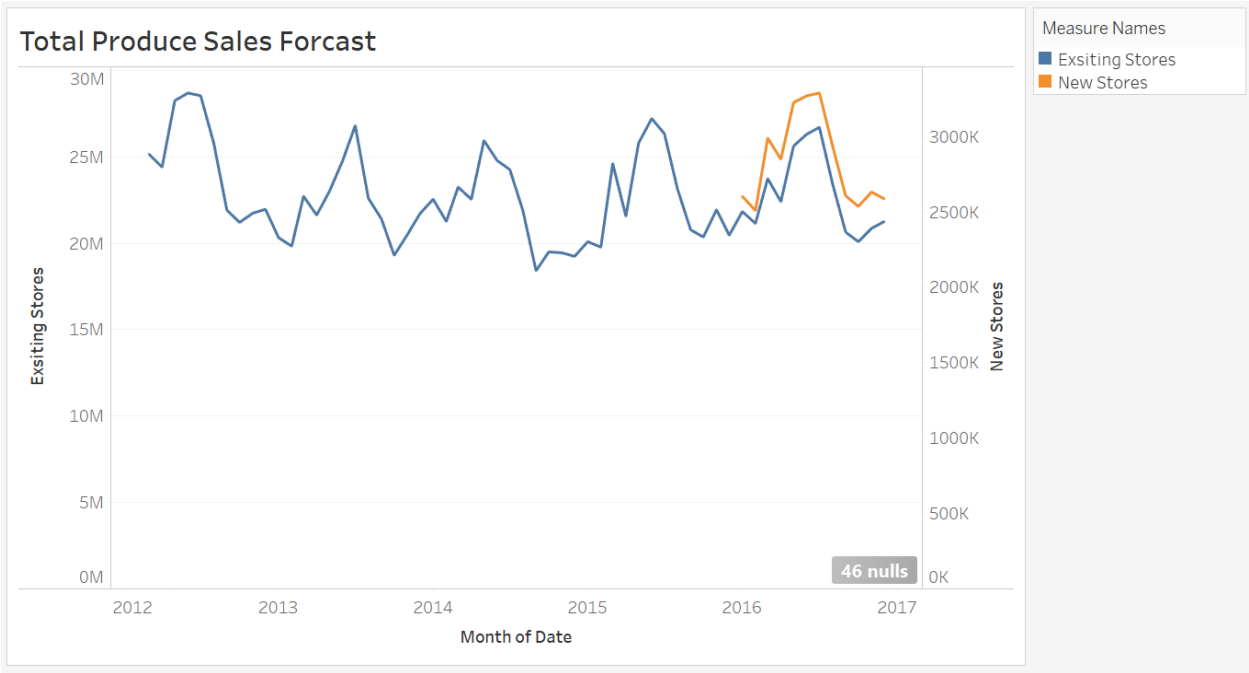
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-604232.3	1050239	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.**

Table below shows the forecast sales for existing stores and new stores.

Year	Month	New Stores	Existing Stores
2016	Jan	2603262	21829060
2016	Feb	2508878	21146330
2016	Mar	2989458	23735687
2016	Apr	2849287	22409515
2016	May	3224711	25621829
2016	Jun	3269623	26307858
2016	Jul	3288334	26705093
2016	Aug	2937302	23440761
2016	Sep	2606592	20640047
2016	Oct	2536270	20086270
2016	Nov	2631293	20858120
2016	Dec	2586562	21255190



The chart above shows the historical and forecast sales for existing stores and new stores over the period from Mar-12 to Dec-16.