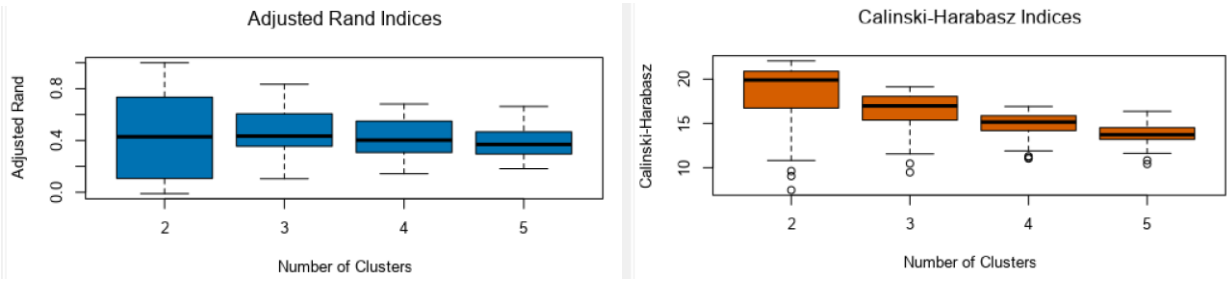
Project: Predictive Analytics Capstone

Task 1: Determine Store Formats for Existing Stores

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

   By performing cluster analysis methods using k-centroids diagnostic tool with number of clusters ranging from 2-5. We were able to see the below reports.

   K-Means Cluster Assessment Report



Based on the K-means report's Adjusted Rand and Calinski-Harabasz indices, number of clusters 3 has the highest median and considerably less variation. Hence the optimal number of store segments is 3.

2. How many stores fall into each store format?

Cluster	Store_Count
1	23
2	29
3	33

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

   Based on the clusters report

	X.Dry_Grocery	X.Dairy	X.Frozen_Food	X.Meat	X.Produce	X.Floral	X.Deli
1	0.327833	-0.761016	-0.389209	-0.086176	-0.509185	-0.301524	-0.23259
2	-0.730732	0.702609	0.345898	-0.485804	1.014507	0.851718	-0.554641
3	0.413669	-0.087039	-0.032704	0.48698	-0.53665	-0.538327	0.64952
	X.Bakery	X.General_Merchandise					
1	-0.894261	1.208516					
2	0.396923	-0.304862					
3	0.274462	-0.574389					

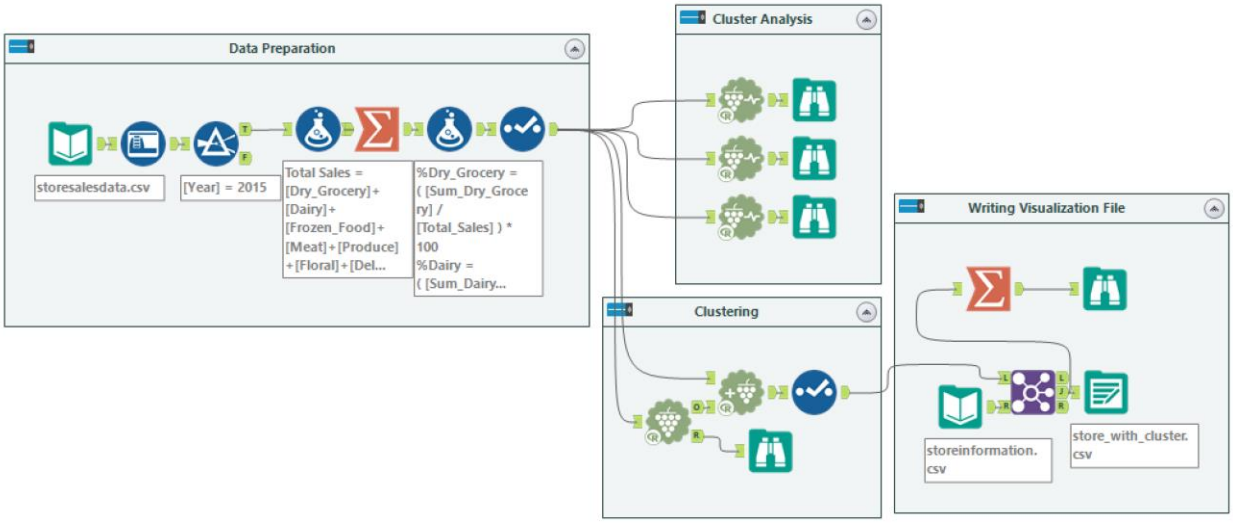
It is clear that,

- **Segment-1** mostly sold more dry groceries and general merchandise which has longer or no expiry period.

- **Segment-2** mostly sold more Dairy, Frozen food, Produce(farm-produced crops) and Floral which are FMCG(Fast Moving Consumable Food) products with a shorter expiry period compared to segment-1 products.
- **Segment-2** mostly sold more meat, deli and dry groceries which are meat and processed food.

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.

https://public.tableau.com/views/Store_Clustered/Sheet1?:display_count=y&publish=yes&:origin=viz_share_link



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?

   Based on the model performance report

Fit and error measures

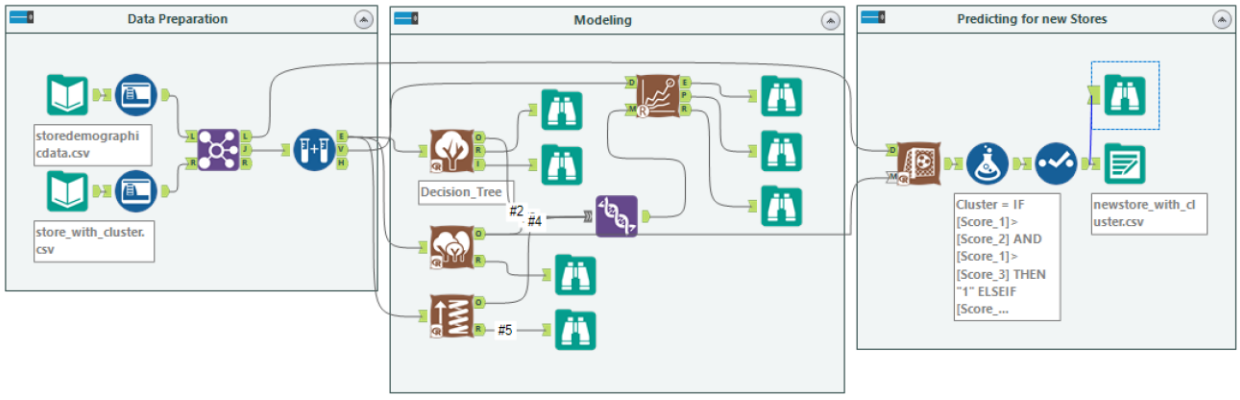
Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Forest_Model	0.8235	0.8251	0.7500	0.8000	0.8750
Boosted_Model	0.8235	0.8543	0.8000	0.6667	1.0000
Decision_Tree	0.7059	0.7327	0.6000	0.6667	0.8333

   Forest model shows the highest accuracy and very less biasness between categories. Hence selecting Forest Model for predicting segment for new stores.

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

   Segment interchangeably named as cluster.

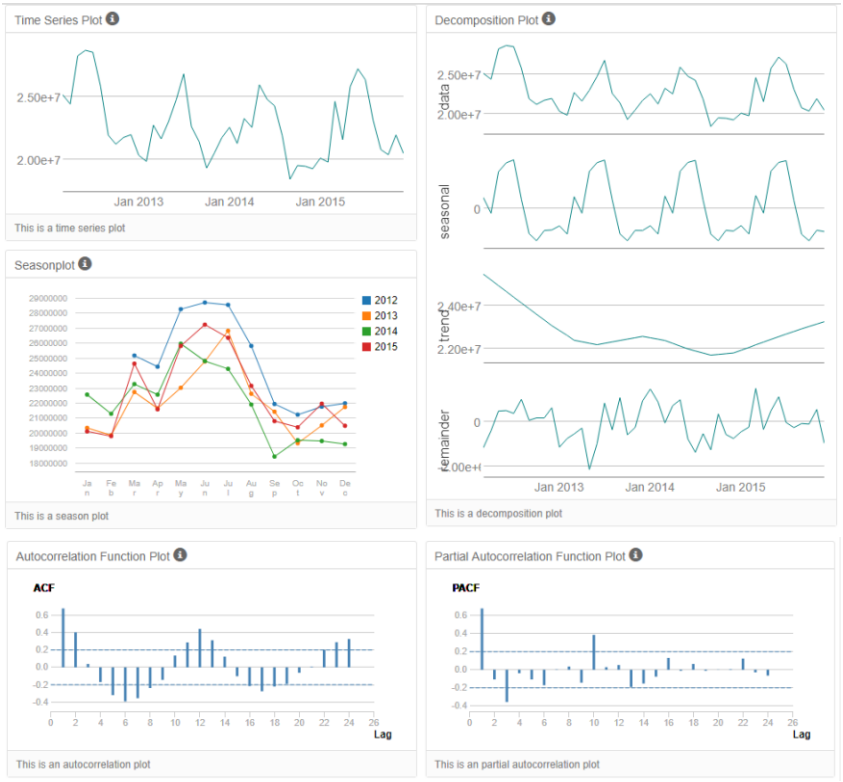
Store	Cluster
S0086	3
S0087	2
S0088	3
S0089	2
S0090	2
S0091	1
S0092	2
S0093	1
S0094	2
S0095	2



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?

By referring to the TS Plot report



There is an existent seasonality in multiplicative behavior, no prominent behavior trend and existent an error element in multiplicative behavior. Hence ETS model will be (M,N,M). ARIMA is auto configured.

Holding 6 months of data for validation as suggested.

Below are the reports after validating ETS and ARIMA models on the hold out data.

- **ETS**

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

- **ARIMA**

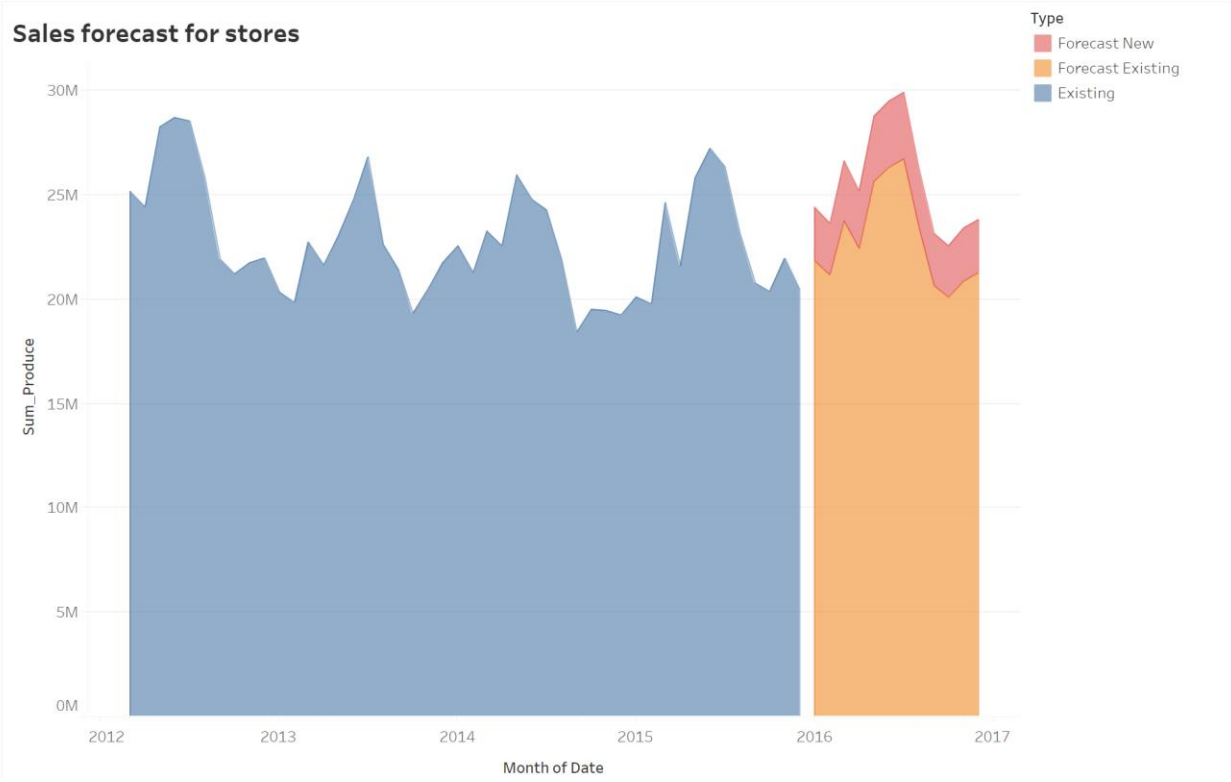
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-604232.3	1050239	928412	-2.6156	4.0942	0.5463

ETS model has considerably low RMSE and other error ratings hence choosing ETS model for forecasting.

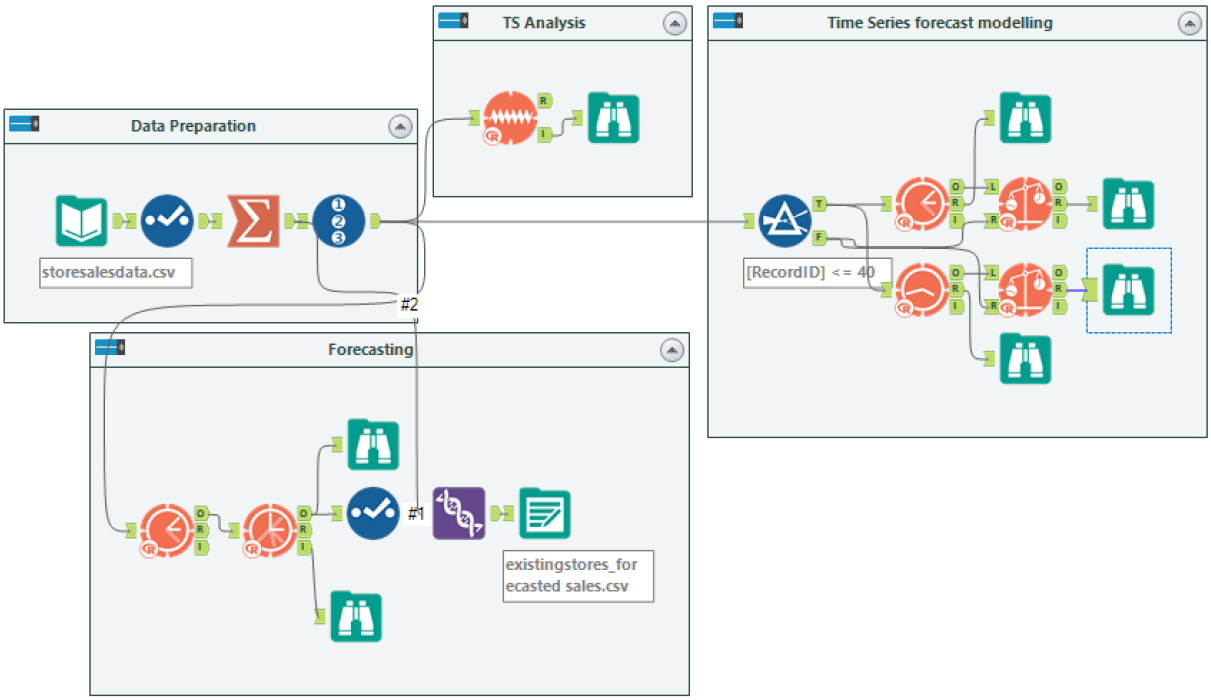
2. Please provide a table of your forecasts for existing and new stores. Also, provide

visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.

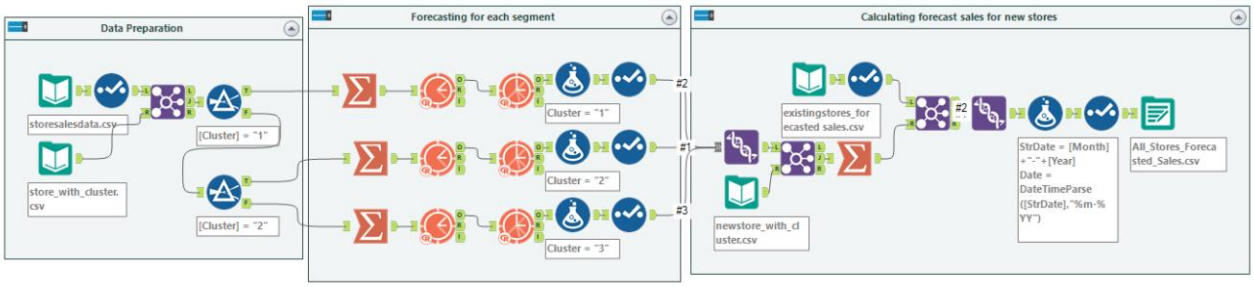
Month of Date	Forecast Existing	Forecast New
January 2016	21,829,060	2,558,029
February 2016	21,146,330	2,469,380
March 2016	23,735,687	2,879,303
April 2016	22,409,515	2,751,396
May 2016	25,621,829	3,115,792
June 2016	26,307,858	3,166,788
July 2016	26,705,093	3,192,489
August 2016	23,440,761	2,828,498
September 2016	20,640,047	2,504,011
October 2016	20,086,270	2,451,215
November 2016	20,858,120	2,545,169
December 2016	21,255,190	2,533,848



https://public.tableau.com/views/ForecastedSales_15823049577120/Sheet1?:display_count=y&publish=yes&:origin=viz_share_link



Total produce sales forecast for exiting stores



Produce sales forecast for new stores