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?

Based on the index graphs from the k-means analysis tool, I decided to divide the stores in three segments. I discarded two categories due to the Rand index, and four categories based on Calinski-Harabasz index.

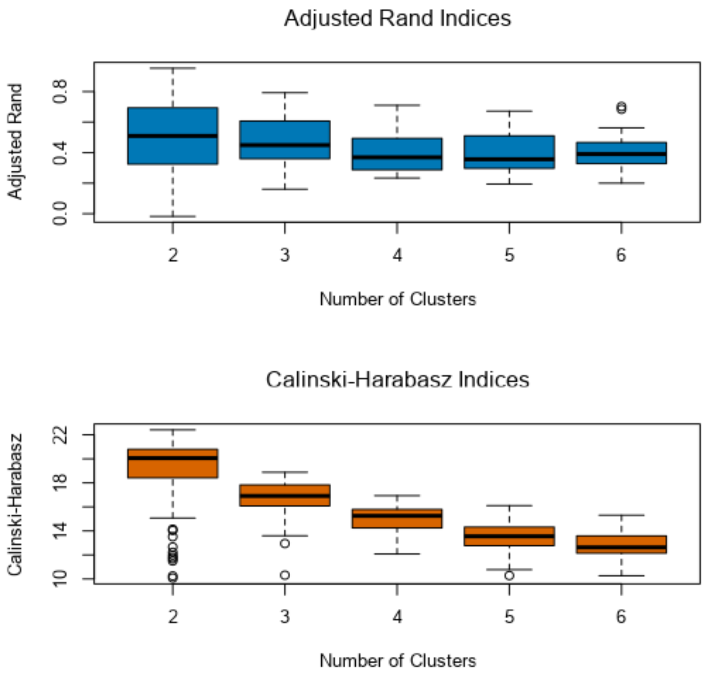
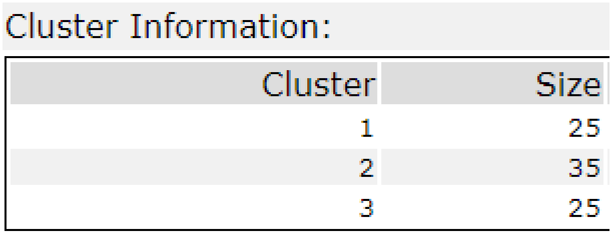


Illustration 1: K-means indexes

2. How many stores fall into each store format?

Based on the segmentation results page, there are 25 stores in segment 1, 35 in segment 2 and 25 in segment 3.



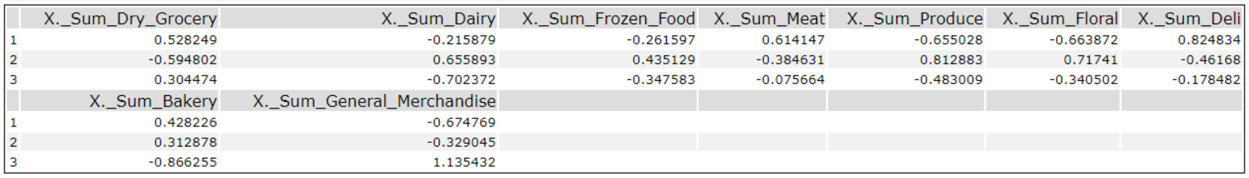
Cluster Information:

Cluster	Size
1	25
2	35
3	25

Illustration 2: Cluster sizes

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

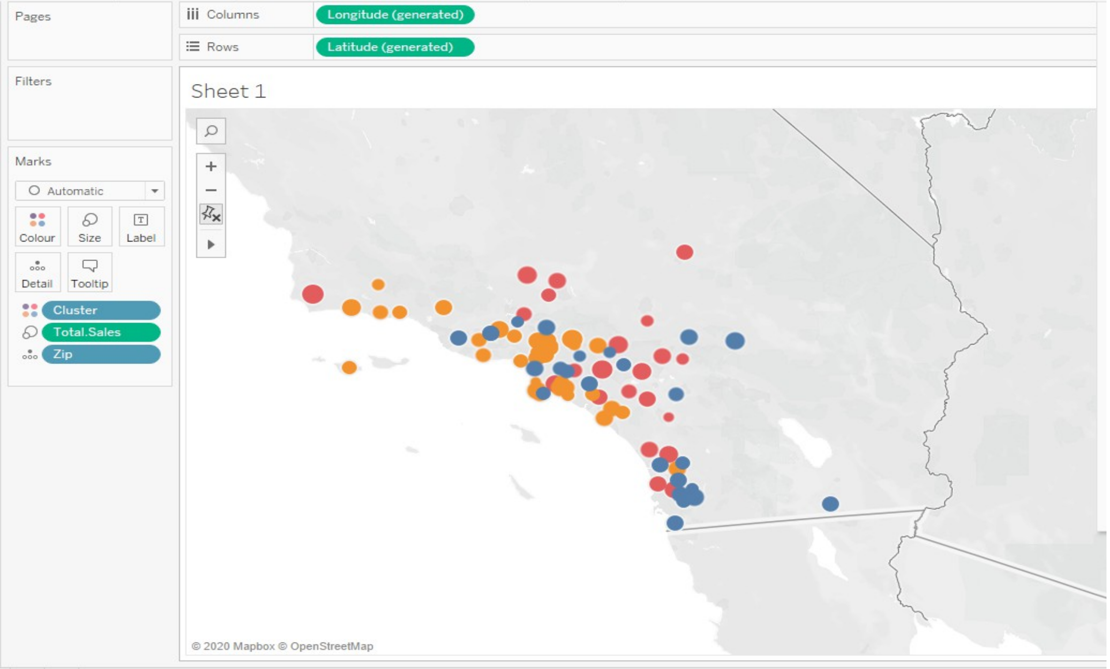
In all the product areas, there is a clear discrepancy between the different groups, and in each area one groups is clearly distinct, as demonstrated in the screen capture. For instance, segment 1 has a relatively high demand for dry groceries and would benefit from larger inventories. At the same time, segment 2 has relatively low demand for dry groceries, and would likely have product over unless changes are implemented.



	X._Sum_Dry_Grocery	X._Sum_Dairy	X._Sum_Frozen_Food	X._Sum_Meat	X._Sum_Produce	X._Sum_Floral	X._Sum_Deli	X._Sum_Bakery	X._Sum_General_Merchandise
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

Illustration 3: Segmentation results

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.**



Columns: Longitude (generated)

Rows: Latitude (generated)

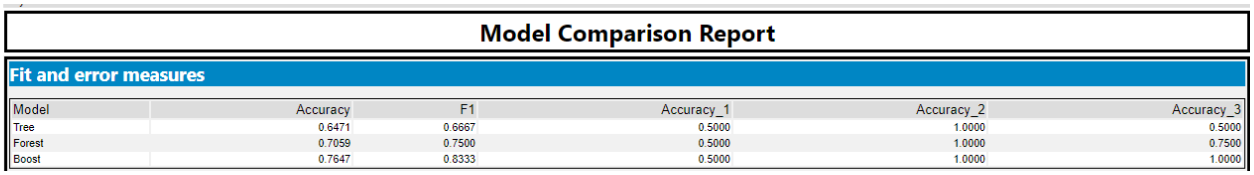
Illustration 4: Tableau visualization

https://public.tableau.com/profile/santeri1008#!/vizhome/udacycapstone1/Sheet1?publish=yes

Task 2: Formats for New Stores

1. **What methodology did you use to predict the best store format for the new stores?**

Based on the model, I chose to use the boosted model. It ranked as the most accurate model based on overall accuracy, F1 and ability to correctly predict segment 2 and 3. It tied in predicting the segment 1.



Model Comparison Report

Fit and error measures

Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Tree	0.6471	0.6667	0.5000	1.0000	0.5000
Forest	0.7059	0.7500	0.5000	1.0000	0.7500
Boost	0.7647	0.8333	0.5000	1.0000	1.0000

Illustration 5: Model comparison

2. **What format do each of the 10 new stores fall into?**

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?

Based on the TS Plot tool, we can see that there is no distinguishable trend, there is a seasonal pattern but it is constant, and the error seems more or less constant. However, the error is a little difficult to estimate just by looking at the graph, so I ran both ETS and ARIMA on auto to see how Alteryx would assign the models.

I ended up with the following parameters: ETS(M,N,M) and ARIMA(1,0,0)(1,1,0)[12].

Based on accuracy measures in graph 7, I prefer to work with ETS model. ETS has lower value in nearly all the fields, indicating a higher overall predicting power.

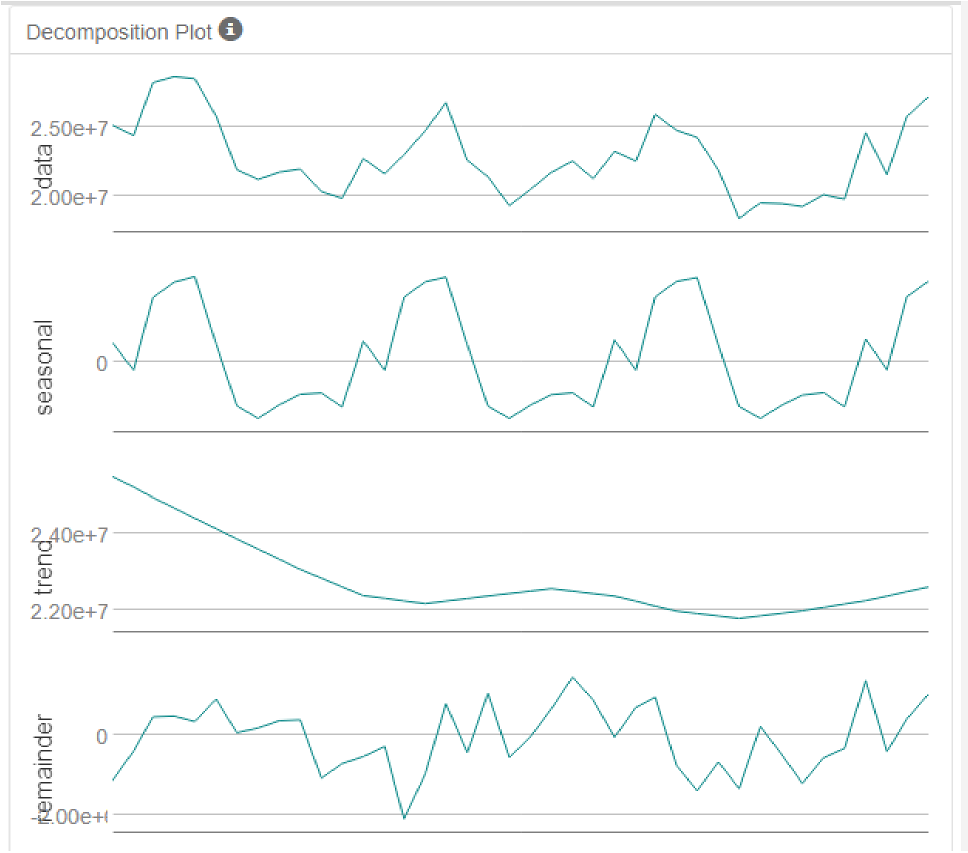
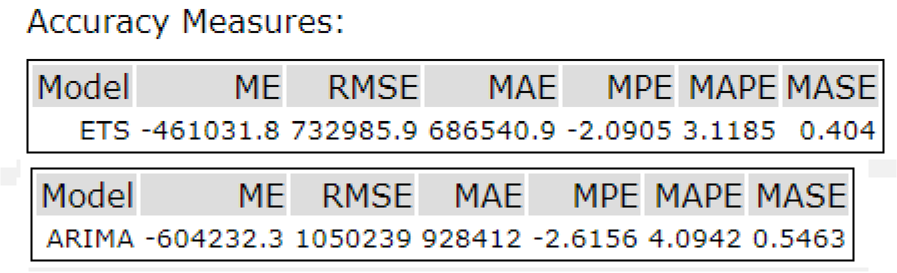


Illustration 6: TS Plot



Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-461031.8	732985.9	686540.9	-2.0905	3.1185	0.404

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

Illustration 7: Accuracy measures

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	1,819,799	21,108,810
Feb-16	1,706,554	20,479,209
Mar-16	1,772,585	23,478,735
Apr-16	1,643,907	22,180,570
May-16	1,760,897	25,300,265
Jun-16	1,681,693	25,918,633
Jul-16	1,846,298	26,085,949
Aug-16	1,751,055	22,943,818
Sep-16	1,846,292	20,471,768
Oct-16	1,784,366	19,943,420
Nov-16	1,849,332	20,574,839
Dec-16	1,765,635	21,045,390

https://public.tableau.com/profile/santeri1008#!/vizhome/udacitycapstone2/Sheet1?publish=yes

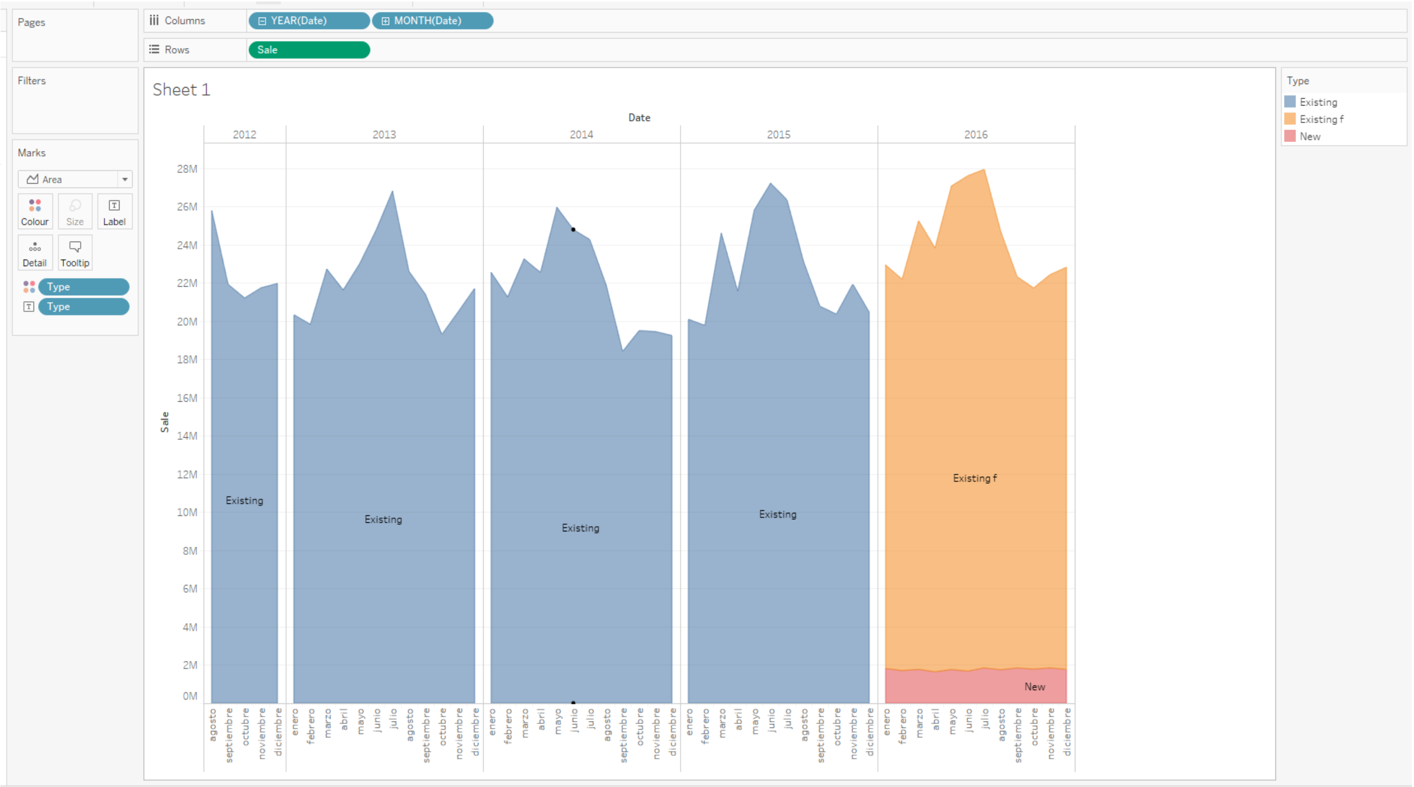


Illustration 8: Visualization