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 the store is 3 clusters. Because 3 clusters have high median and smaller spread.

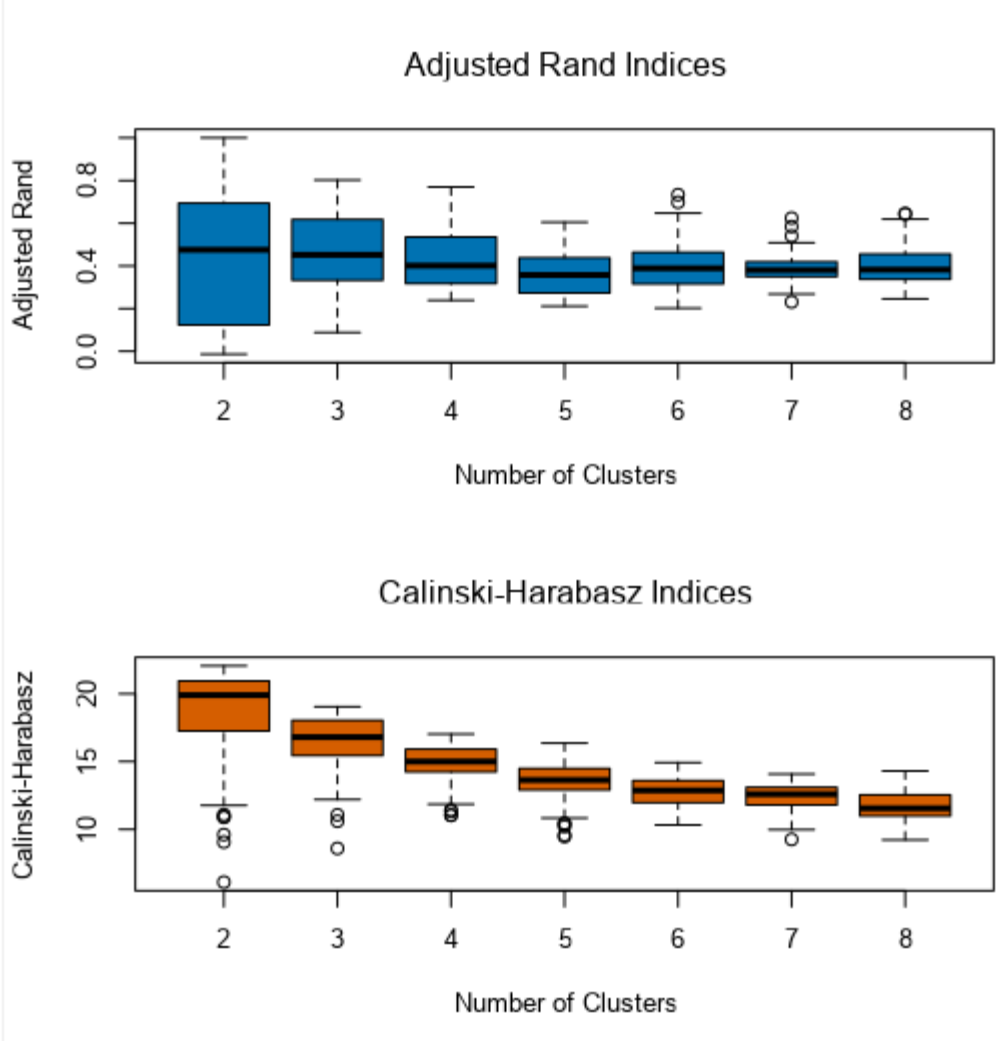

Figure 1: Adjusted Rand & Calinski_Harabasz Indices

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

   Cluster 1: 23 stores
   Cluster 2: 29 stores
   Cluster 3: 33 stores

Cluster Information:

Cluster	Size	Ave Distance	Max Distance	Separation
1	23	2.320539	3.551451	1.874244
2	29	2.540086	4.475132	2.118708
3	33	2.115045	4.926201	1.702844

Figure 2: Cluster Info.

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

   Cluster 3 sold more meat and dairy than cluster 1 and 2

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/profile/hashim8020#!/vizhome/GroceryStoreCluster_15817905570730/Sheet1

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.)**

   Boosted Model has been chosen, even though all models have same accuracy but Boosted Model has higher F1 rate than the others.

Model Comparison Report

Fit and error measures

Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
ForestModel	0.8235	0.8426	0.7500	1.0000	0.7778
TreeModel	0.8235	0.8426	0.7500	1.0000	0.7778
BoostedModel	0.8235	0.8889	1.0000	1.0000	0.6667

Figure 3: Comparison among the models

2. **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) have been chosen for forecasting. After comparing both ETS and ARIMA, ETS is more accurate than ARIMA.

Record Report

1

Comparison of Time Series Models

2

Actual and Forecast Values:

Actual	ETS1	ARIMA1
19444753.17	20954549.498	21031463.85798
21936906.81	20899853.78763	21165512.05495
21962976.75	23342005.09054	21286462.81556
21715706.67	20921264.24179	21395595.84997
19240384.75	20382324.73577	21494065.8318
20462899.3	22044587.46326	21582914.61506

3

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS1	-630159.6	1271062	1240658	-3.2204	6.0156	0.6604
ARIMA1	-532064.6	1291405	1121404	-2.8793	5.5696	0.5969

4

Figure 4: Comparison among ETS & ARIMA

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 Store
Jan-16	**2,414,195**	**18,718,638**
Feb-16	**2,467,286**	**19,576,817**
Mar-16	**2,721,922**	**21,277,761**
Apr-16	**2,360,616**	**18,988,119**
May-16	**2,442,805**	**19,495,186**
Jun-16	**2,473,840**	**20,160,316**
Jul-16	**2,765,064**	**21,916,423**
Aug-16	**2,471,756**	**20,266,649**
Sep-16	**2,391,350**	**19,368,335**
Oct-16	**2,579,295**	**20,793,990**
Nov-16	**2,800,597**	**21,277,864**
Dec-16	**2,493,159**	**19,072,851**

https://public.tableau.com/profile/hashim8020#!/vizhome/HistoricalandForcastedsalesforProduce/Sheet1

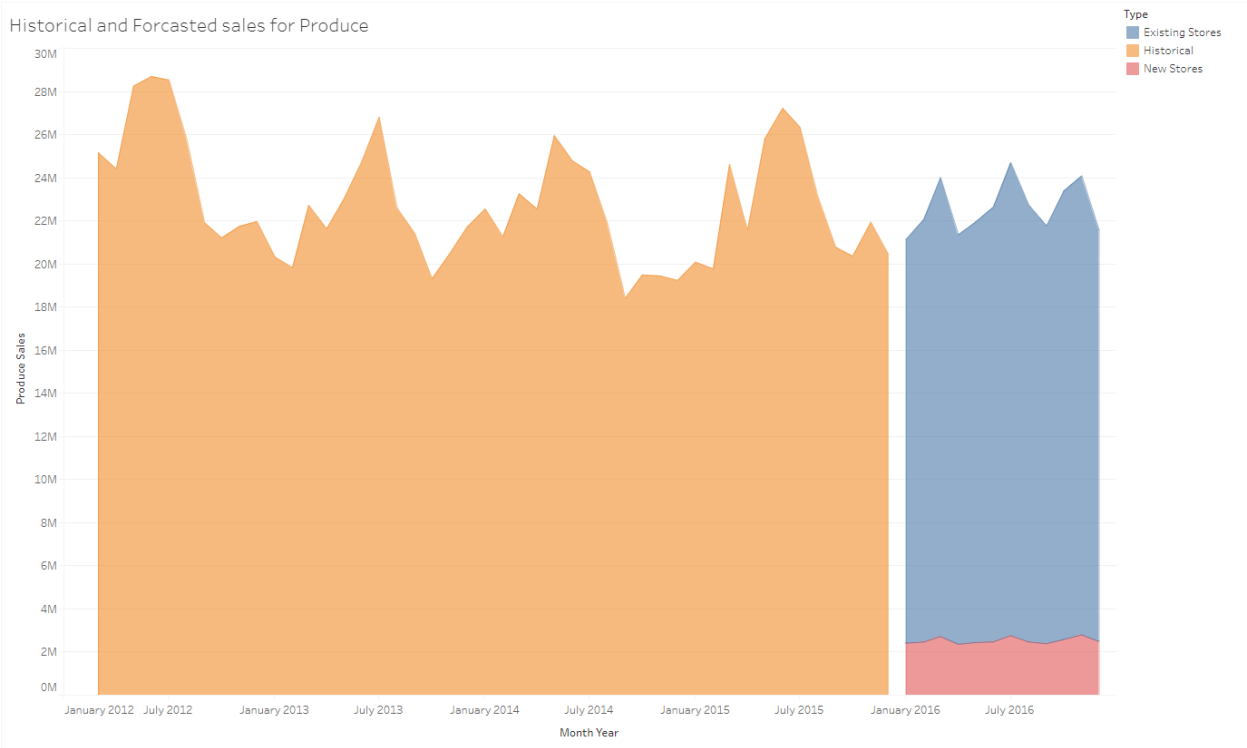

Figure 5: Visualization

Alteryx Workflows:-

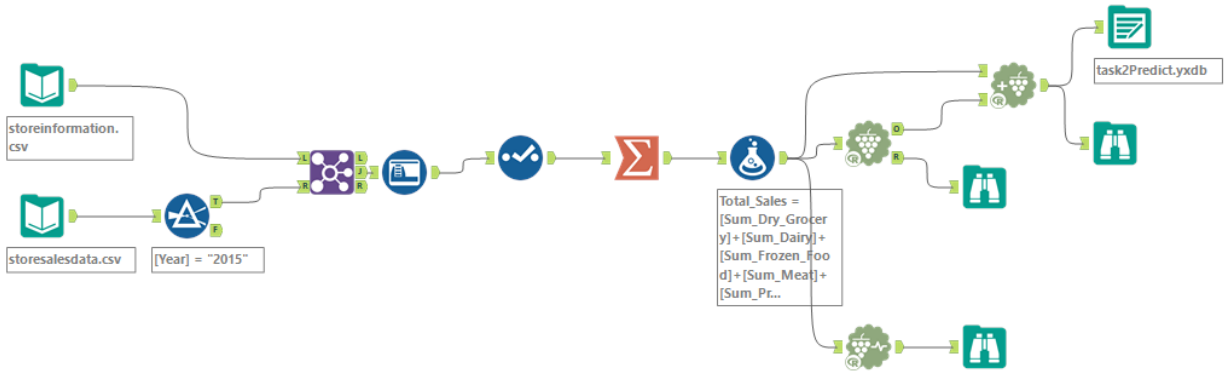

Figure 6: Task 1 Workflow

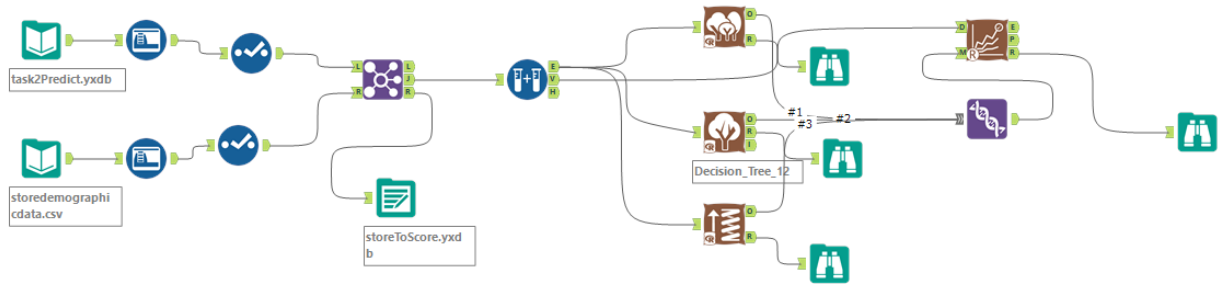

Figure 7: Task 2 Workflow (1/2)

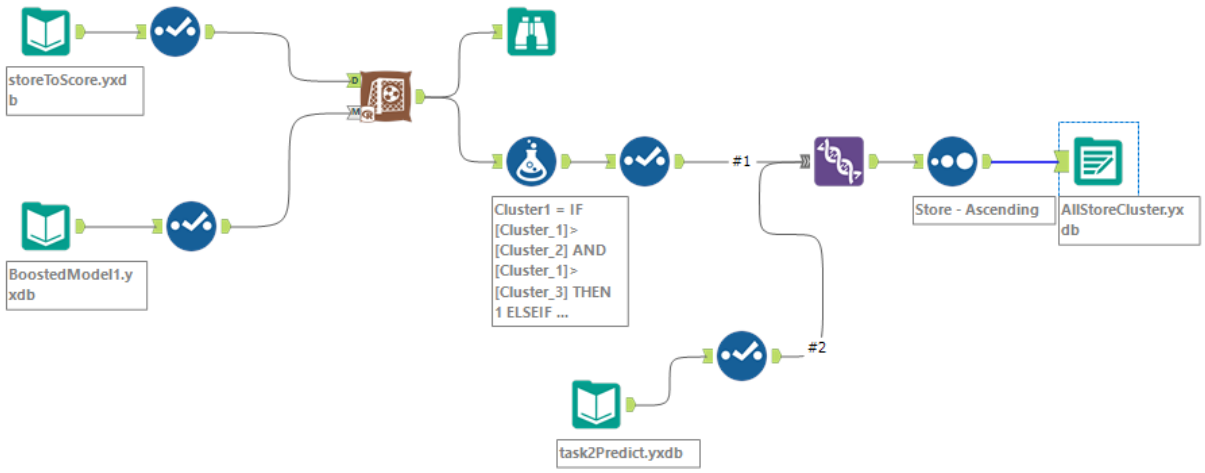

Figure 8: Task 2 Workflow (2/2)

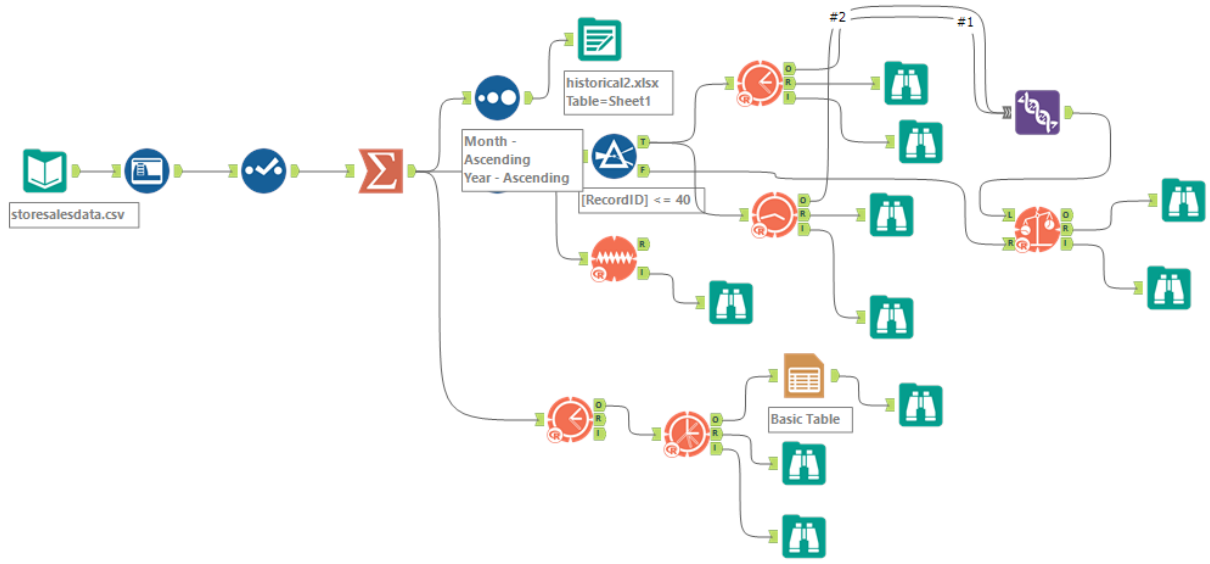

Figure 9: Task 3 Workflow (1/2)

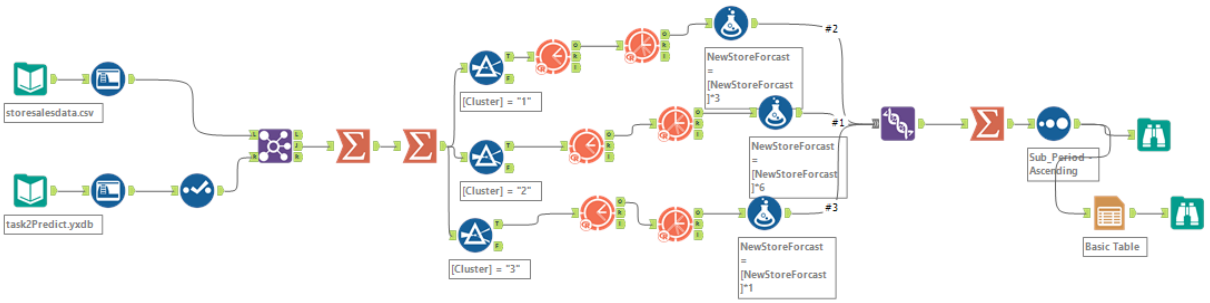

Figure 10: Task 3 Workflow (2/2)