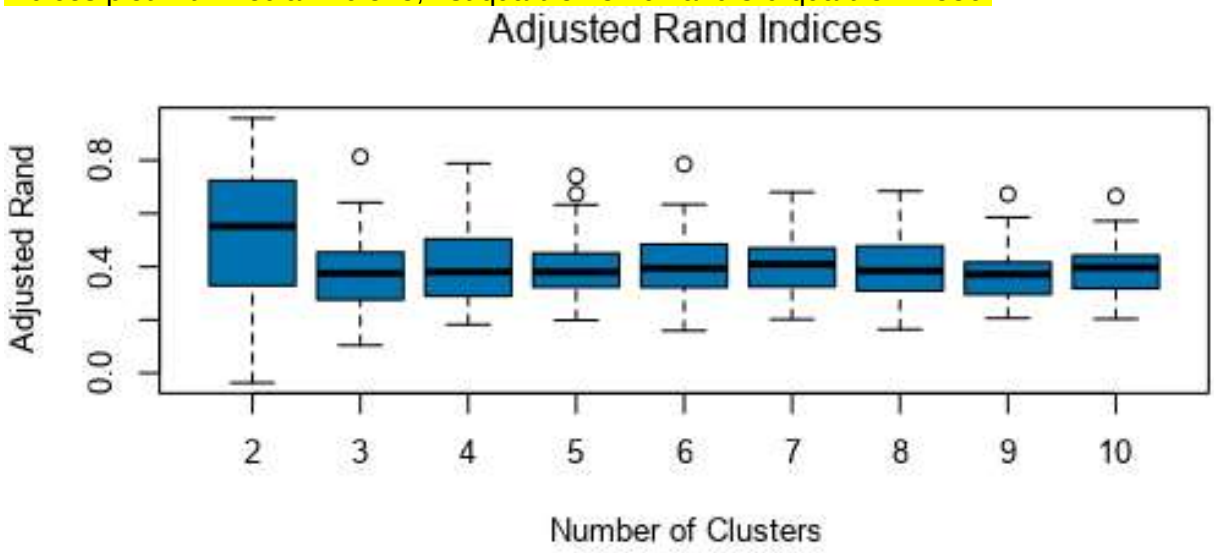
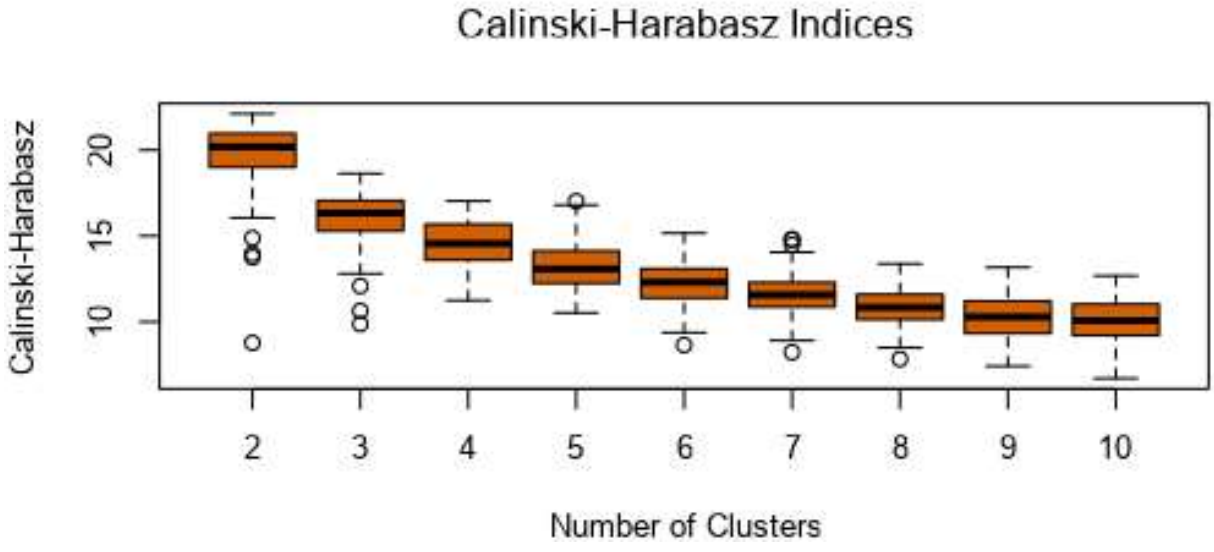
Project: Predictive Analytics Capstone

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://coco.udacity.com/nanodegrees/nd008/locale/en-us/versions/1.0.0/parts/7271/project

Task 1: Determine Store Formats for Existing Stores

1. What is the optimal number of store formats? How did you arrive at that number?
   3 store formats. K-Centroids Diagnostic tool running results Adjusted Rand Indices plot with median 0.355, 1st quartile 0.296 and 3rd quartile 0.495. It also results Calinski-Harabasz Indices plot with median 16.515, 1st quartile 15.202 and 3rd quartile 17.556.



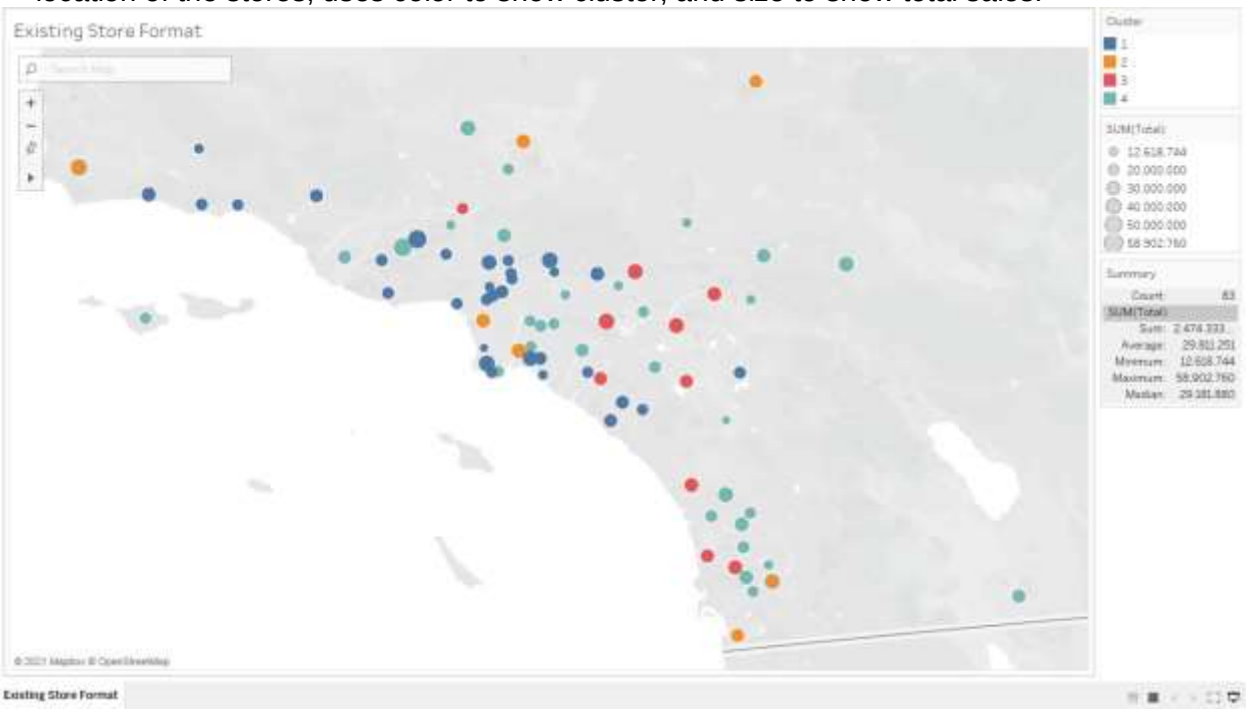


2. How many stores fall into each store format? 25 stores for cluster 1, 35 stores for cluster 2, and 25 stores for cluster 3.

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

Cluster	Store Format
1	Dry Grocery, Meat, Deli, Bakery
2	Dairy, Frozen Food, Produce, Floral
3	Dry Grocery, General Merchandise

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.)

Boosted Model. Because it provides best accuracy (0.7059) and F1 (0.75) score.

Model Comparison Report

Fit and error measures

Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Decision_Tree_Store_Format	0.6471	0.6667	0.5000	1.0000	0.5000
Forest_Store_Format	0.7059	0.7500	0.5000	1.0000	0.7500
Boosted_Store_Format	0.7059	0.7500	0.5000	1.0000	0.7500

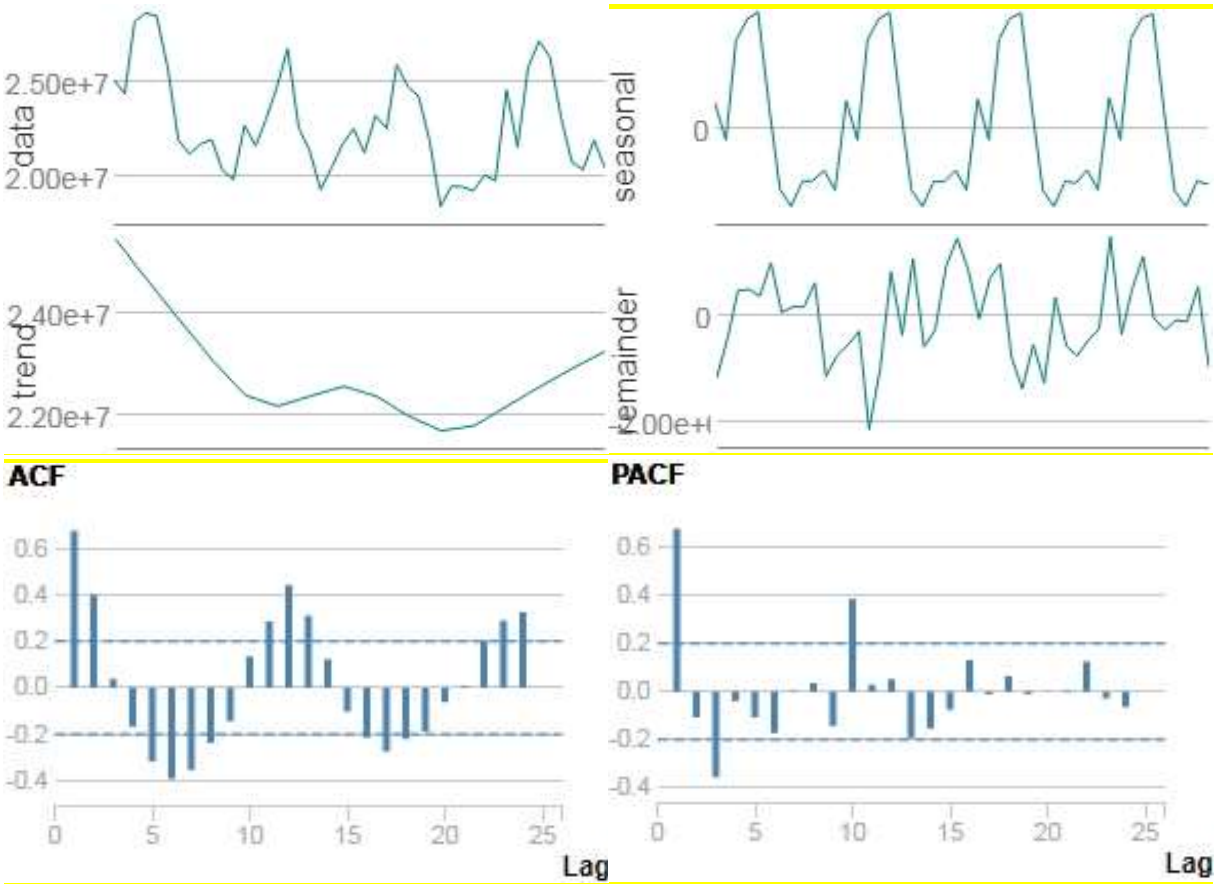
2. What format do each of the 10 new stores fall into? Please fill in the table below.

Store Number	Segment
S0086	1
S0087	2
S0088	1
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?

Here is TS Plot from existing store:



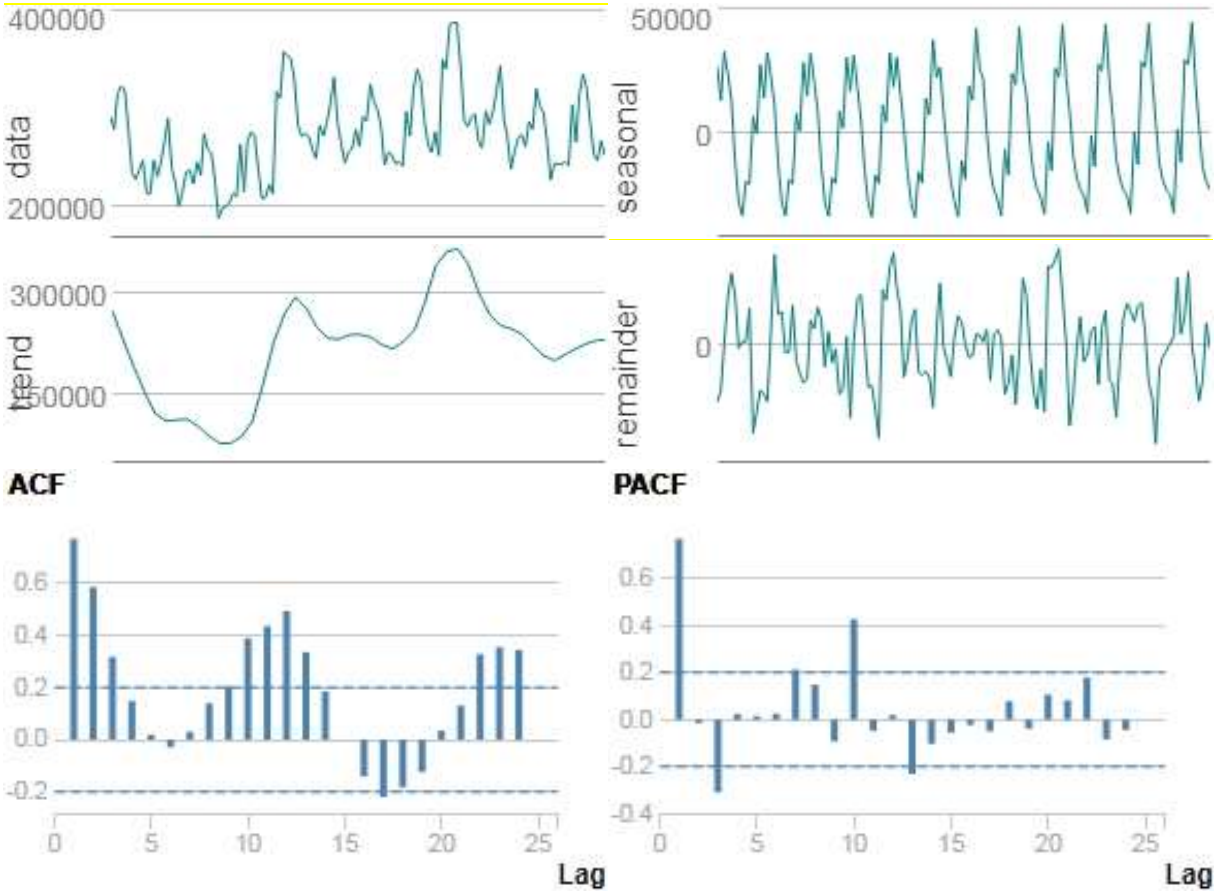
Based on the decomposition plot, there are fluctuative variance in error; increasing and decreasing trend line; and increasing sales from season to season; so I propose ETS (M, N, M). Also there is seasonal pattern with lag 2 in ACF and lag 1 in PACF that suggest the

proposal of ARIMA(1,0,0)(1,1,0)[12]. The AIC result of ETS model is 1279.42 and 880.45 for ARIMA model, hence the forecast will be done with ARIMA model.
The result of TC Compare Tool is shown here:

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_Current_Store	-141894.6	4374270	3675934	-2.5042	15.9008	2.1629
ARIMA_Current_Store	112812.6	4215678	3509555	-1.267	15.0978	2.065

For new store, the TS Plot are:



Based on the decomposition plot, there are fluctuative variance in error; increasing trend line; and increasing sales from season to season; so I propose ETS (M, N, M). Also there is seasonal pattern with lag 4 in ACF and lag 1 in PACF that suggest the proposal of ARIMA(0,1,1)(1,1,0)[12]. Since new store sales forecast are represented by store formats or segements or clusters, the results of TS Compare Tool per clusters are shown below:

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_New_Store_Cluster_1	8239.627	49628.14	40489.31	1.1717	16.0718	1.9443
ARIMA_New_Store_Cluster_1	-2671.713	50789.88	42594.78	-3.5224	17.4489	2.0454

Accuracy Measures:

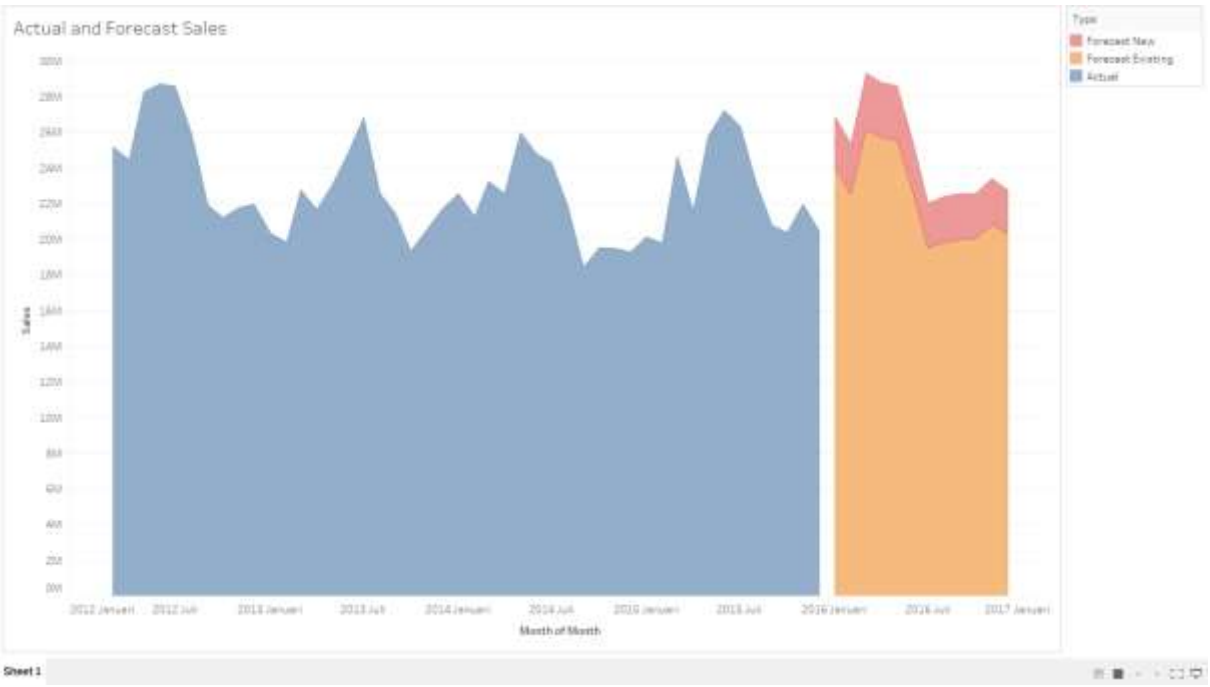
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_New_Store_Cluster_2	-7398.24	51162.69	44160.14	-4.2728	15.4501	2.2918
ARIMA_New_Store_Cluster_2	-41597.97	69111.46	57001.09	-16.4972	20.9629	2.9582

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_New_Store_Cluster_3	-9161.638	11401.14	9618.574	-3.3608	3.5317	0.4488
ARIMA_New_Store_Cluster_3	-21710.399	23645.81	21710.399	-8.0077	8.0077	1.013

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	3107530.966	23978213.22
Feb-16	2965687.344	22505883.58
Mar-16	3387224.112	26103951.89
Apr-16	3297107.188	25674792.11
May-16	3285083.857	25513322.61
Jun-16	3013062.907	22603853.1
Jul-16	2660880.352	19492822.89
Aug-16	2723862.81	19820599.89
Sep-16	2737859.791	19955643.57
Oct-16	2717612.537	19992630.42
Nov-16	2766827.309	20789788.35
Dec-16	2705755.668	20233469.91



Before you submit

Please check your answers against the requirements of the project dictated by the rubric. Reviewers will use this rubric to grade your project.