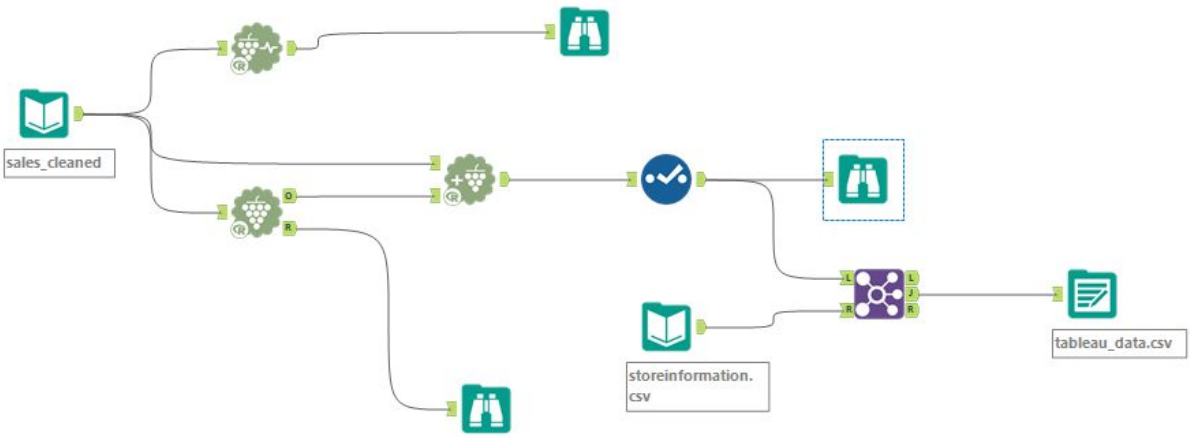
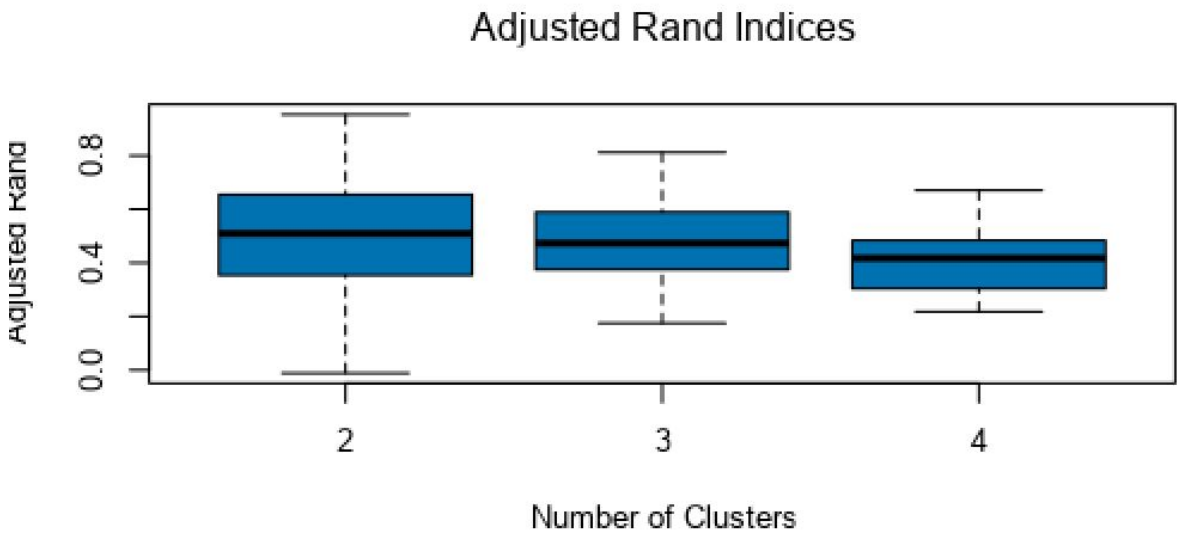
Project: Predictive Analytics Capstone

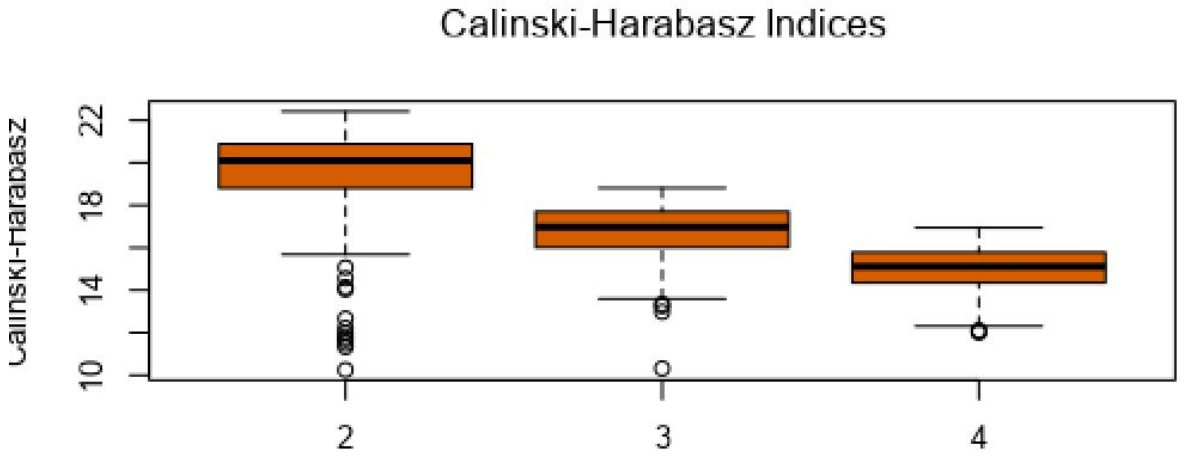
Complete each section. When you are ready, save your file as a PDF document and submit it here:
https://coco.udacity.com/nanodegrees/nd008/locale/en-us/versions/1.0.0/parts/7271/project

Task 1: Determine Store Formats for Existing Stores

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



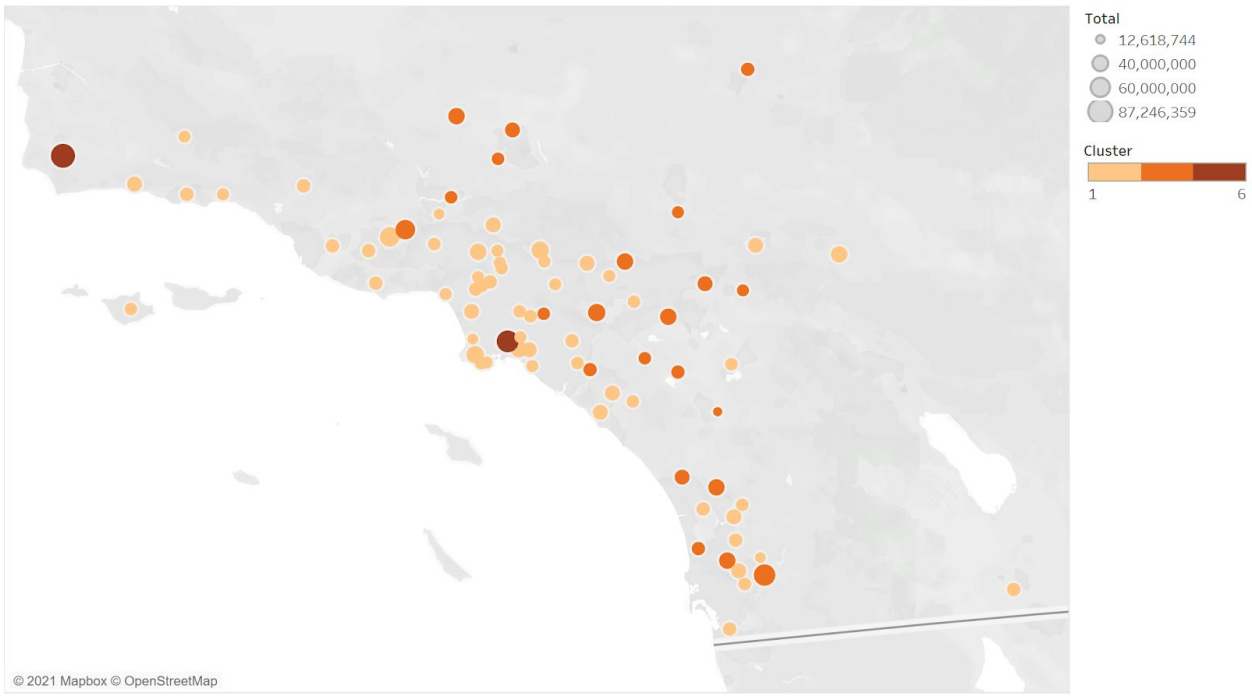




3 - as we want to make sure the median is as high as possible and spread as small as possible, 3 cluster looks like the optimal solution.

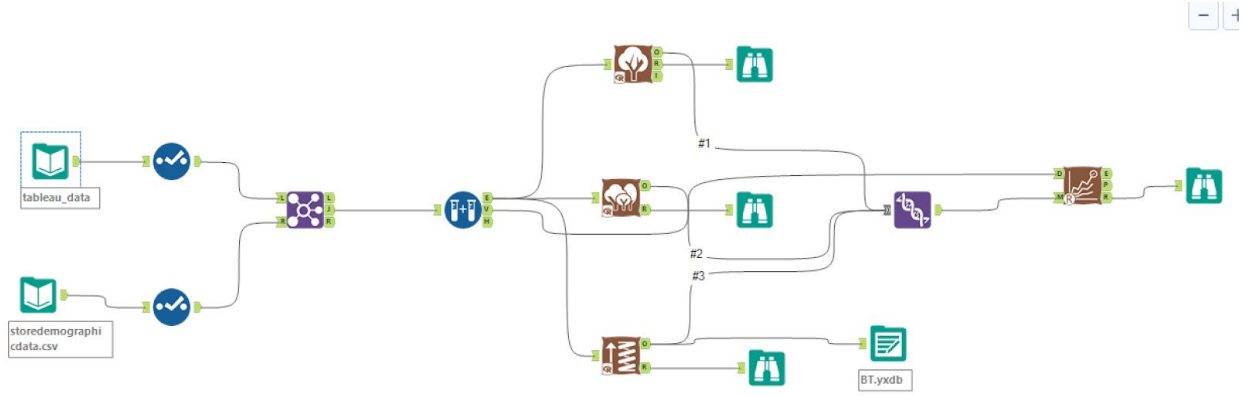
2. How many stores fall into each store format?
   25, 35, 25
3. Based on the results of the clustering model, what is one way that the clusters differ from one another?
   Each cluster sells more certain products - for example, cluster 1 sells more meat than the other two clusters.
4. Please provide 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.

Store Segment



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Cluster. Size shows sum of Total. Details are shown for Zip.

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.)

Fit and error measures

Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
forest	0.7059	0.7500	0.5000	1.0000	0.7500
boosted	0.7647	0.8333	0.5000	1.0000	1.0000
DT	0.6471	0.6667	0.5000	1.0000	0.5000

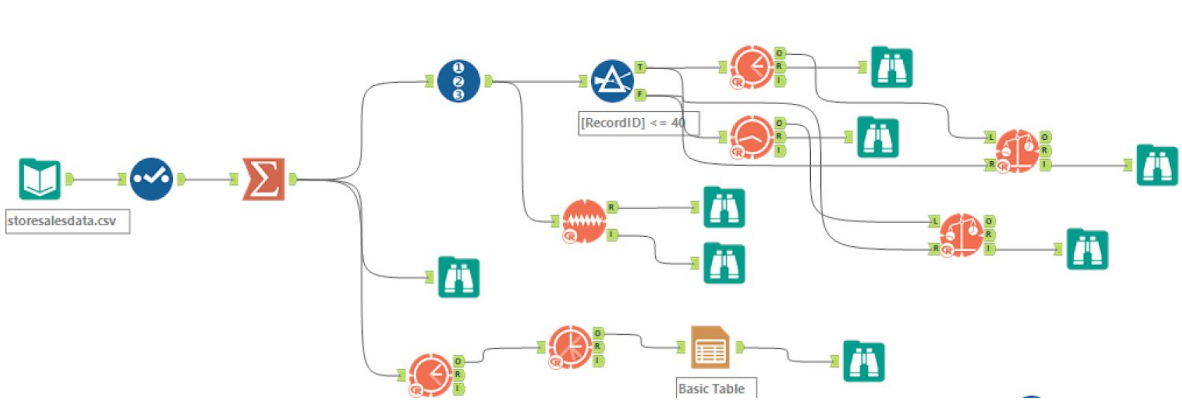
Boosted model - as it has the highest accuracy and F1 score.

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	3
S0089	2
S0090	2
S0091	3
S0092	2
S0093	3
S0094	2
S0095	2

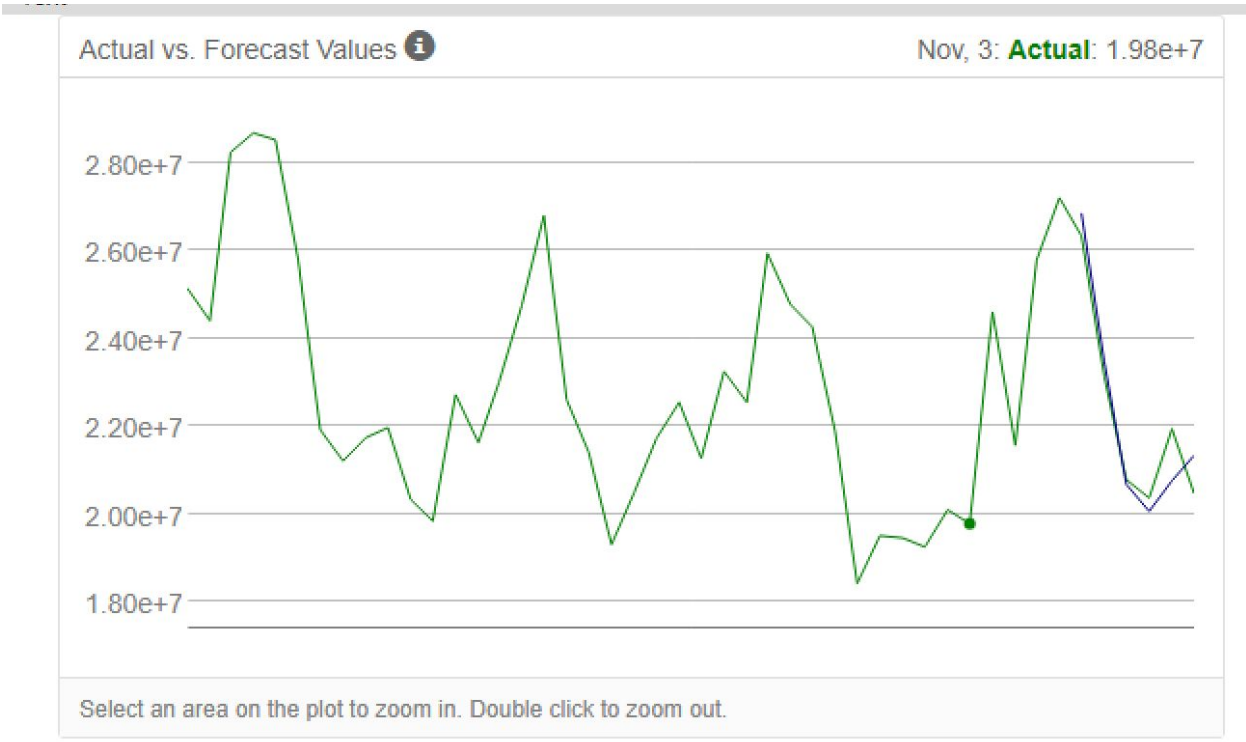
Task 3: Predicting Produce Sales

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?

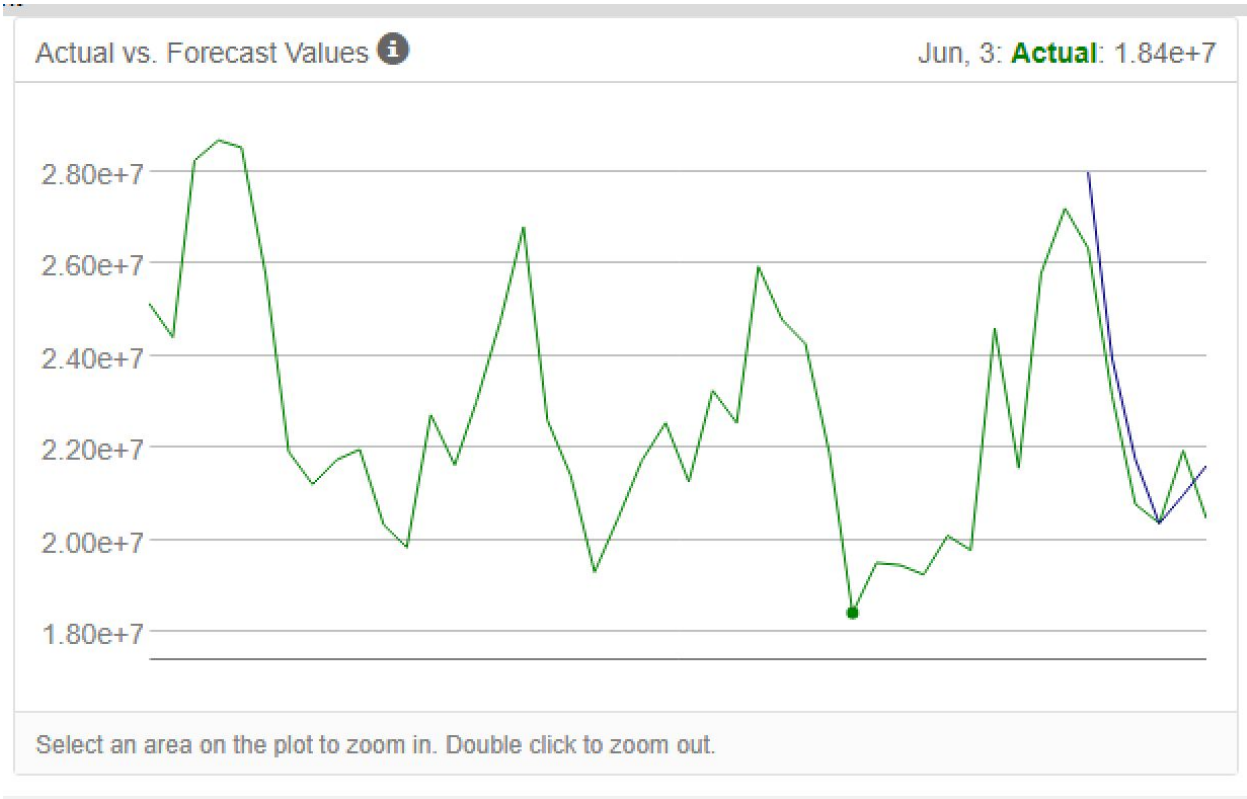


I chose ETS(M,N,M) for the following reasons:

- ETS model has higher accuracy visually
  ETS model:



ARIMA model:

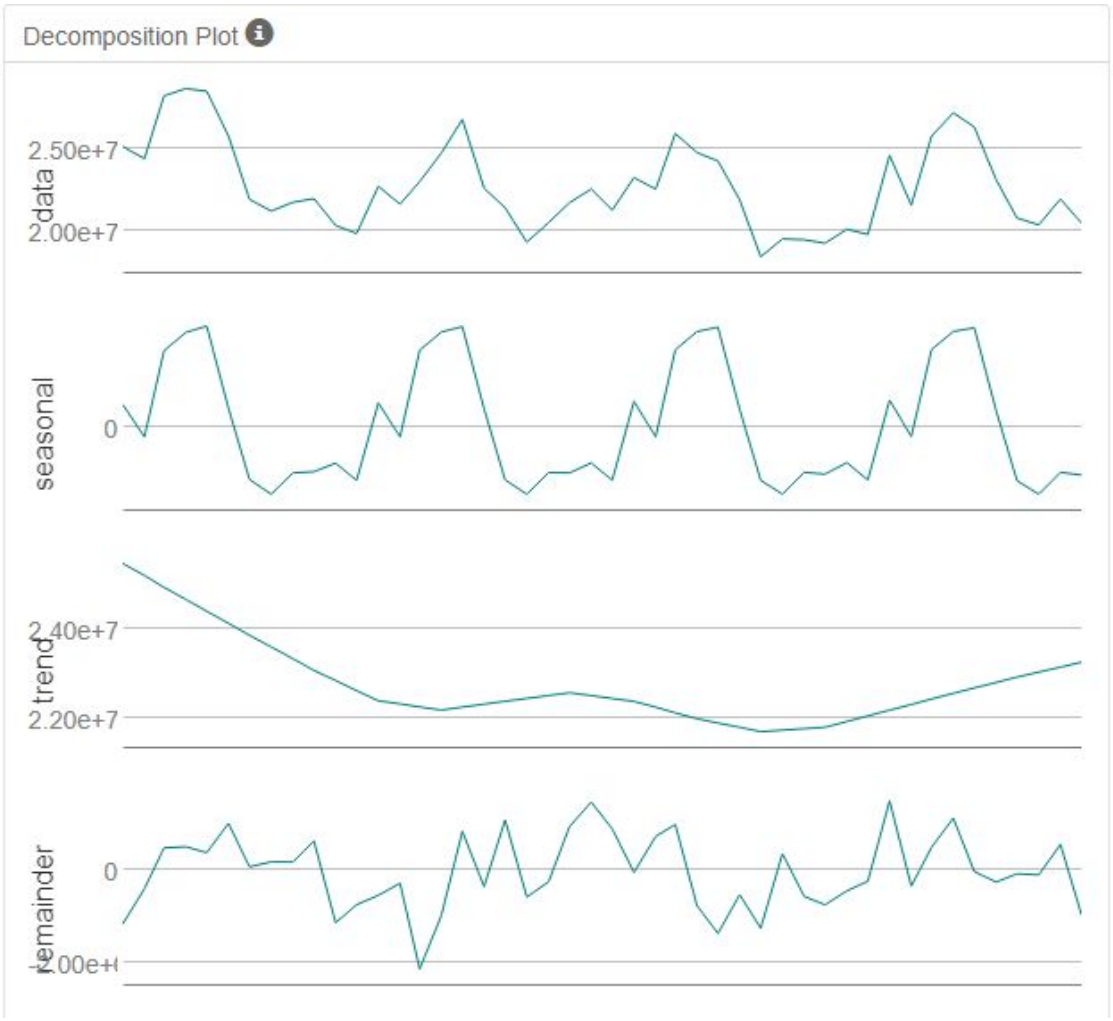


Lastly, we can compare the measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-21,581.1252	663,707.1529	553,511.4848	-0.0437	2.5135	0.3257
AR	-604,232.2943	1,050,239.1848	928,412.0244	-2.6156	4.0942	0.5463

We can see that ETS model has higher accuracy with the holdout sample as its error values are all smaller than AR's.

- Looking at the decomposition plots, we can see that there is no trend, seasonal and error are both multiplicative. Thus I chose ETS(M,N,M)



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.

Year	Month	Existing	New
2016	1	21136641.78	2563357.91
2016	2	20507039.12	2483924.728
2016	3	23506565.98	2910944.146

2016	4	22208405.76	2764881.87
2016	5	25380147.77	3141305.867
2016	6	25966799.47	3195054.204
2016	7	26113792.57	3212390.954
2016	8	22899285.77	2852385.769
2016	9	20499583.91	2521697.187
2016	10	19971242.82	2466750.894
2016	11	20602665.92	2557744.588
2016	12	21073222.08	2530510.805