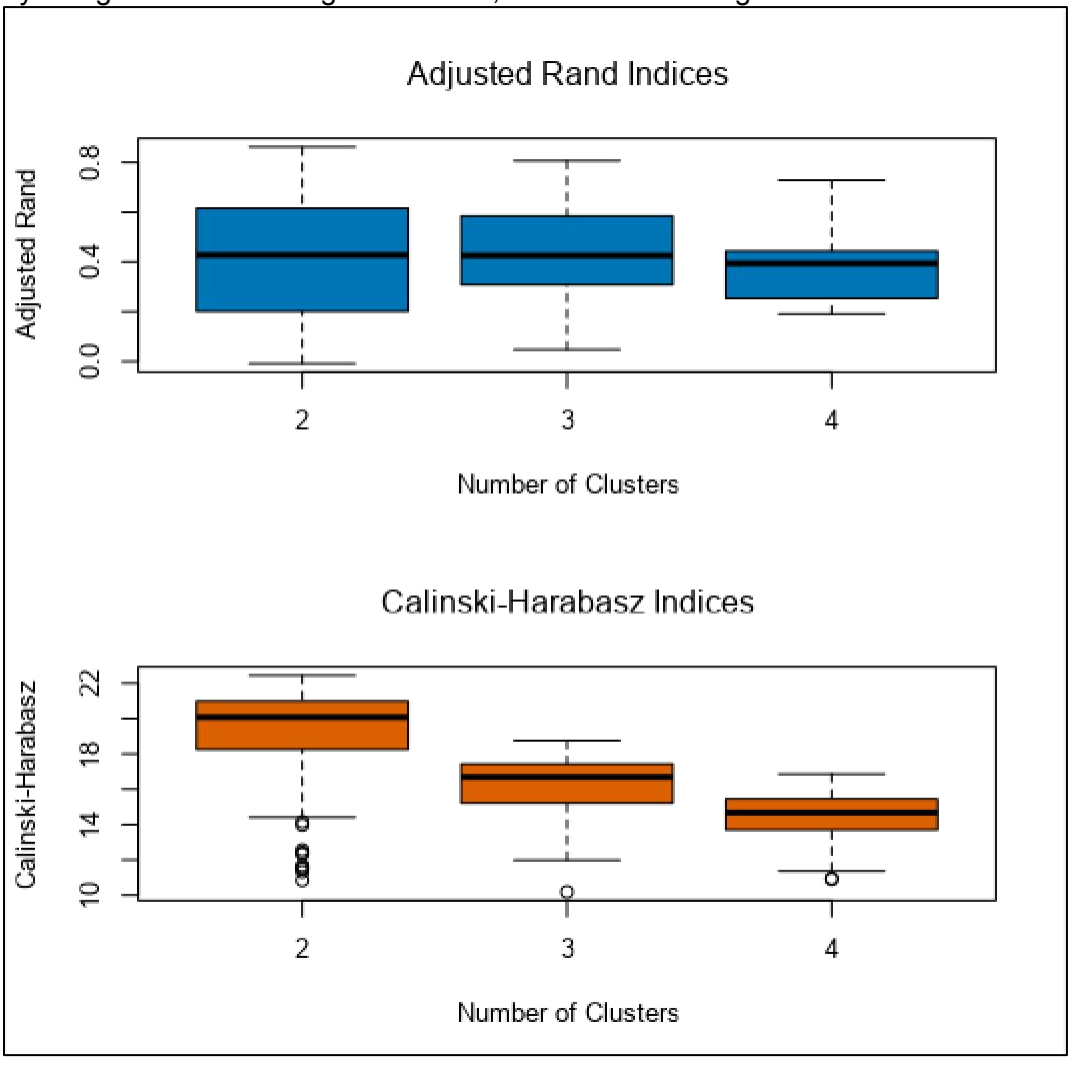
Project: Predictive Analytics Capstone

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

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

3 clusters
by using K-centroids Diagnostics tool ,K-means clustering method



AR and CH show that 3 cluster higher median and fairly compact spread

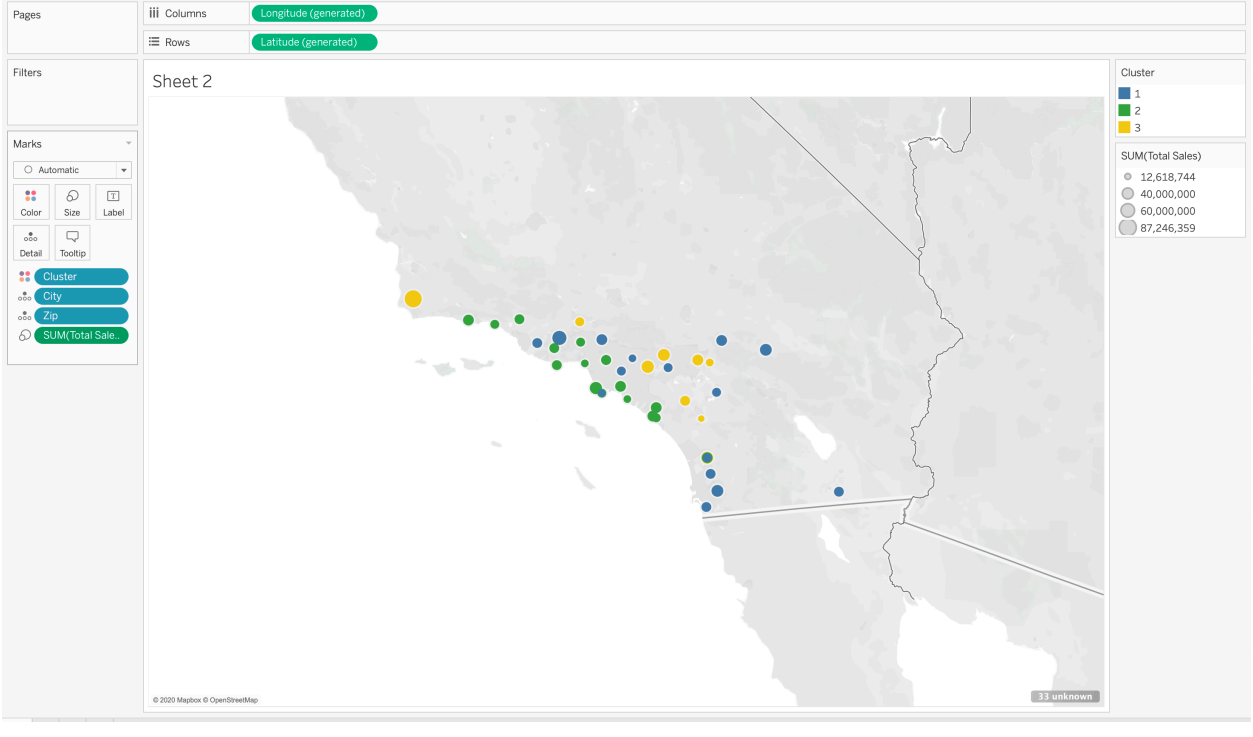
- How many stores fall into each store format?

Cluster	Size
1	25
2	35
3	25

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

stores in cluster1 have high Meat percentage of total sales,while
cluster2 and cluster3 have the lower percentage of total sales for Meat.
but for General_Merchandise cluster1 has lower percentage of total sales than cluster2 and cluster3.

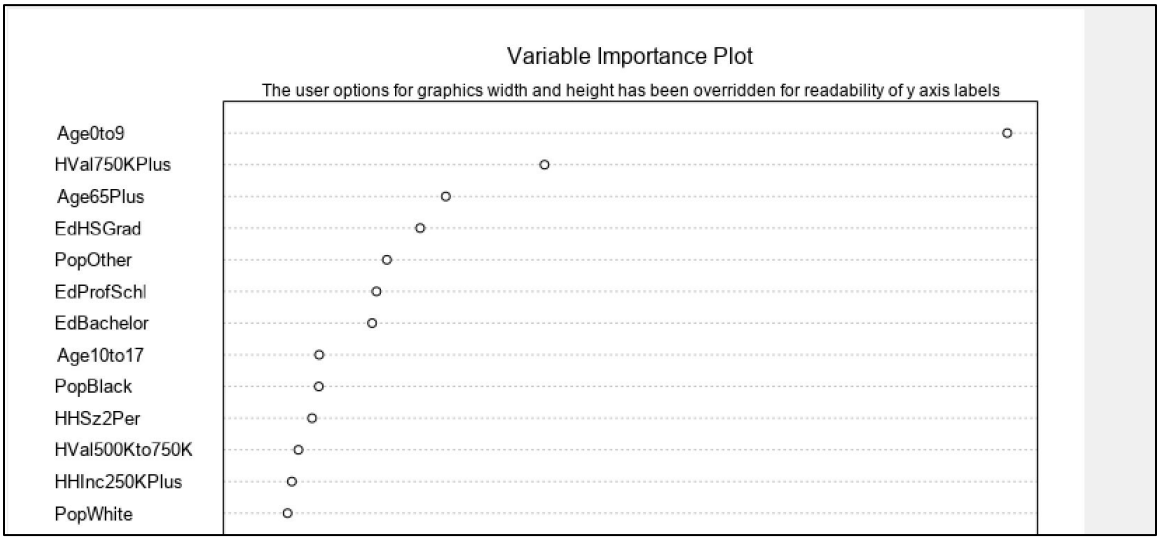
- Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.



https://public.tableau.com/views/clusters_16085737729580/Sheet2?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Task 2: Formats for New Stores

- 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.)
  forest model has the higher accuracy ,and the three most important variables are Age0to9, HVal750KPlus and Age65Plus.



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

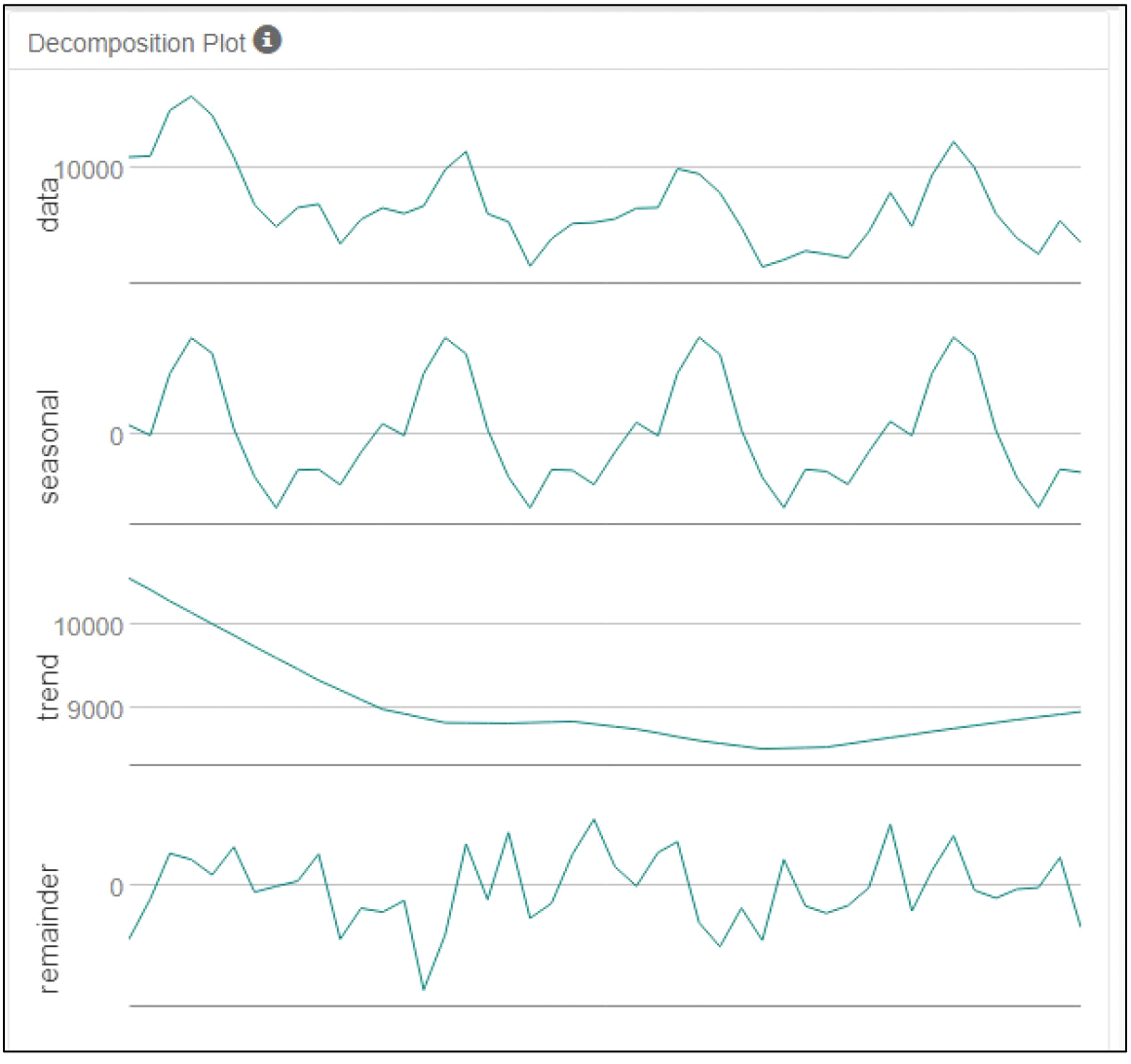
Store Number	Segment
S0086	1
S0087	2
S0088	3
S0089	2
S0090	2
S0091	3
S0092	2

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?

using TS Plot tool



- The Decomposition plot showsseasonality is growing slightly over time then apply it multiplicatively. The ETS tool's auto selection also indicates multiplicative seasonality.
- For the remainder(error) The ETS tool's auto selection Multiplicative.
- No Trend observed,The ETS tool's auto selection None

forecast values using ARIMA and ETS and compare these forecasted results in the TS Compare tool.

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-604232.29	1050239.2	928412	-2.6156	4.0942	0.5463
ETS	-21581.13	663707.2	553511.5	-0.0437	2.5135	0.3257

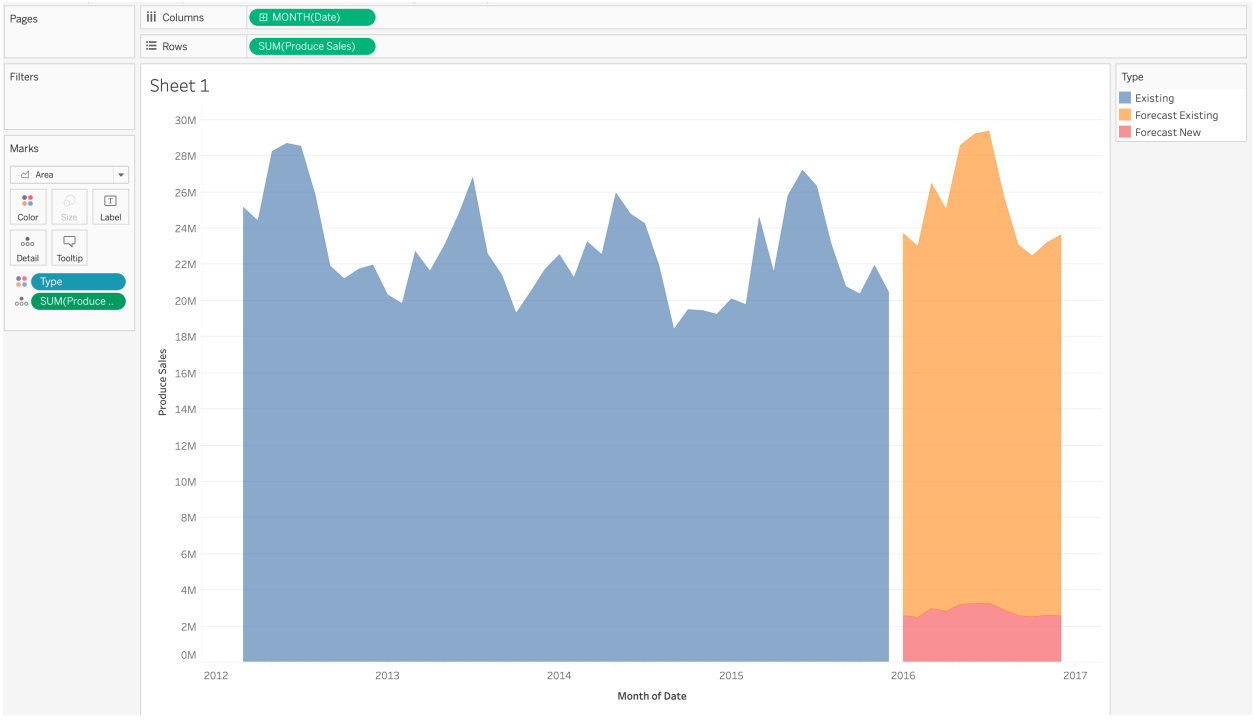
The average percent error of theETS model is smaller than that of the ARIMA model .
The MASE in ETS model is smaller than ARIMA model ,ETS model should be used to forecast future values.

ETS(M,N,M) will be used

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

Year	Month	Existing_Store	New_Stores_Sales
2016	1	21136641.781775	2574250.039917
2016	2	20507039.12384	2483373.311944
2016	3	23506565.982355	2972112.509008
2016	4	22208405.755153	2823370.749941
2016	5	25380147.771963	3197153.435276
2016	6	25966799.465113	3249051.683797
2016	7	26113792.565116	3259755.568084
2016	8	22899285.769116	2904985.757996
2016	9	20499583.908226	2574827.142126
2016	10	19971242.820704	2506604.028212
2016	11	20602665.916965	2601483.152137
2016	12	21073222.081854	2559789.053104

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



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.