

## 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>

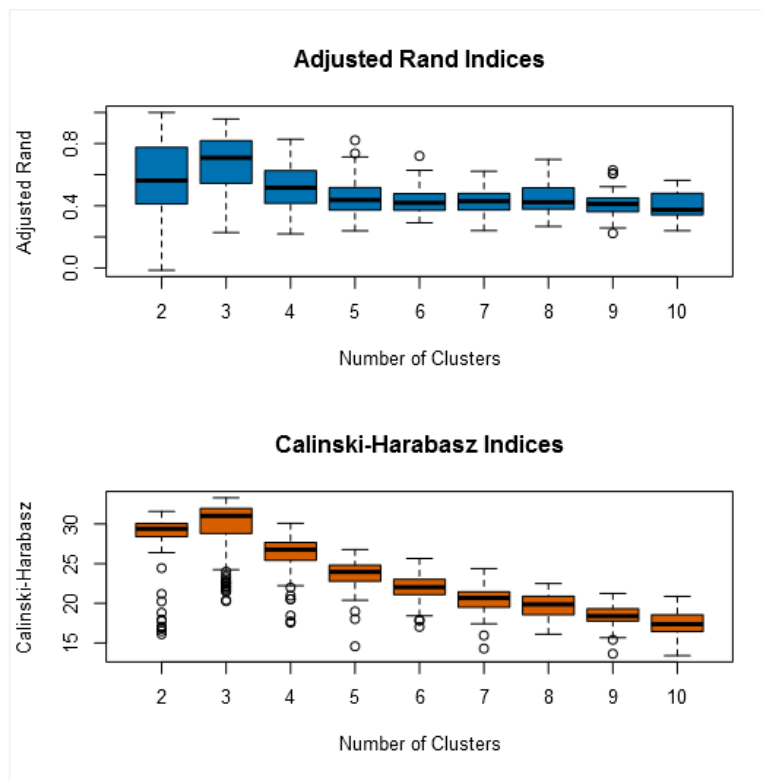
### Task 1: Determine Store Formats for Existing Stores

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

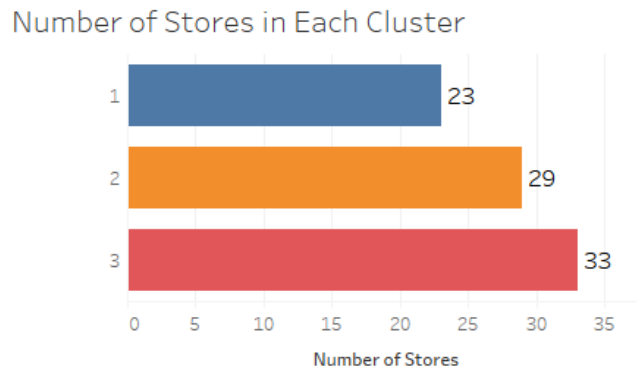
*The optimal number of store formats is three.*

*I arrived at this data with the following steps:*

- *I removed all data not relating to the year 2015 and I used only existing stores.*
- *I summarized each sales category by store.*
- *I found the percentage of each specific category to total sales.*
- *I used those percentage categories in the K-Centroids Diagnostics tool. In that tool, I selected each of the category percentage fields, standardized them, and applied each of the three clustering methods to them (K-means, K-Medians, Neural Gas) for 2 to 10 clusters.*
- *I found that each of the clustering methods were similar in their results and went with the K-means approach.*
- *I assessed the Adjusted Rand Indices and the Calinski-Harabasz Indices and found that three clusters were the optimal clusters because it was the most stable and more tightly packed than the other clusters. See Image below.*

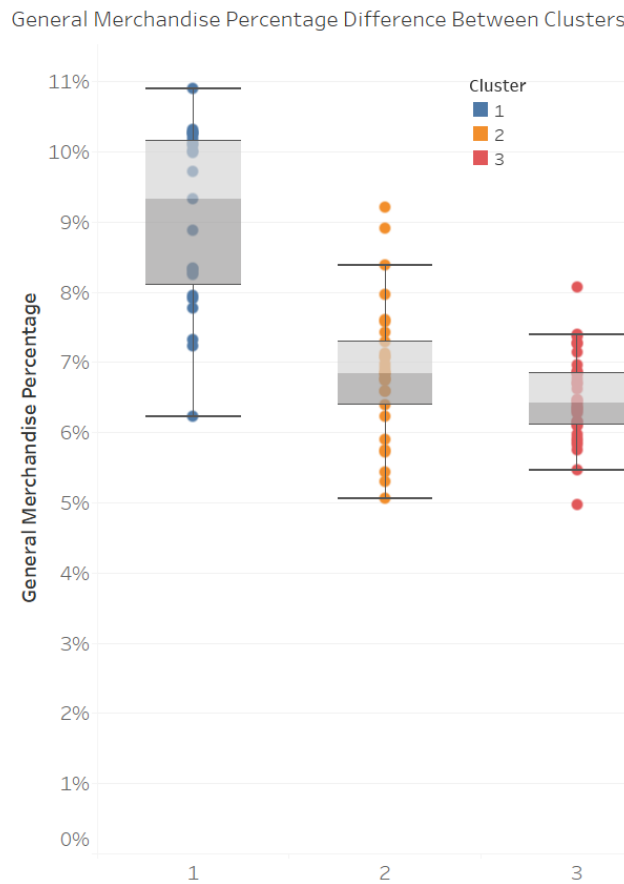


2. How many stores fall into each store format?



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

*One difference between the clusters is the percentage of general merchandise they sell. Cluster one has a median of 9.3% while cluster two has median of 6.8% and cluster three has a 6.4%. This means general merchandise is a greater percentage of sales for cluster one than cluster two and three. See chart below.*

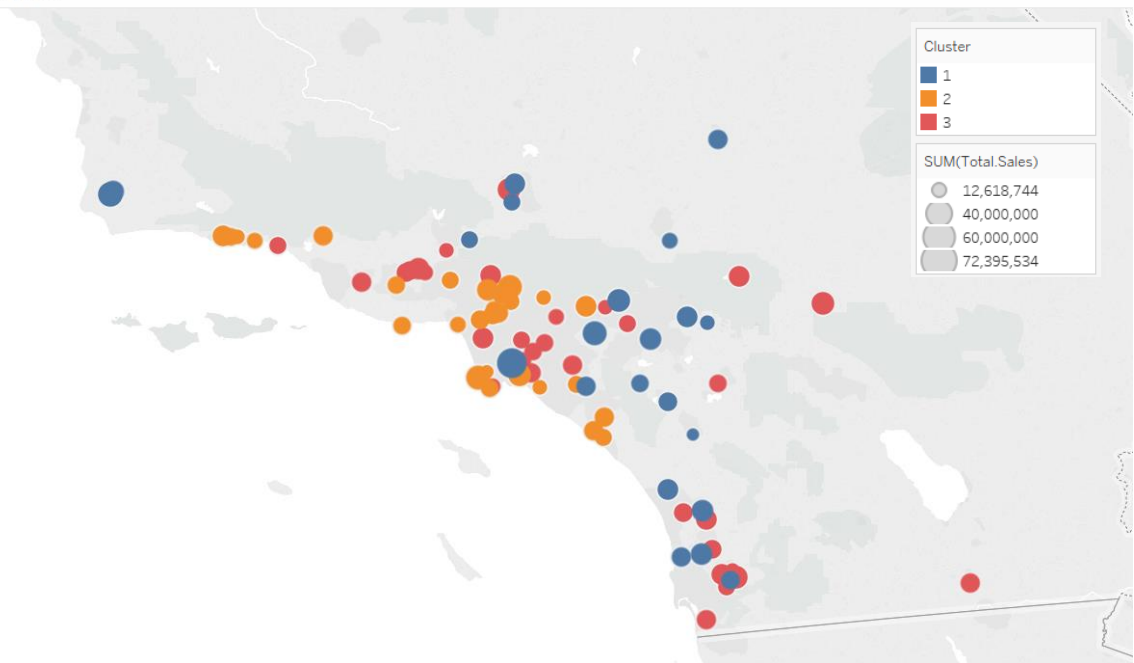


4. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.

*Note: I geocoded the addresses.*

[https://public.tableau.com/views/FinalProjectUdacityJamesSykesMay72017/MapofStores?:e\\_mbed=y&:display\\_count=yes](https://public.tableau.com/views/FinalProjectUdacityJamesSykesMay72017/MapofStores?:e_mbed=y&:display_count=yes)

Map of Stores



## 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.)

*Overall, I chose the boosted model to predict the best store format for the new stores. The reason why is because it is a classification problem that is non-binary in a data rich environment and it had the highest F1 score of my models.*

*I also created a decision tree and a forest model and I found that the accuracy of the F1 for the boosted was 85.4% in comparison to the 82.5% of the Forest Model and Decision Tree.*

Fit and error measures					
Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
Boosted_Model	0.8235	0.8543	0.8000	0.6667	1.0000
Forest_Model	0.8235	0.8251	0.7500	0.8000	0.8750
Decision_Tree_41	0.8235	0.8251	0.7500	0.8000	0.8750

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

Store	Final Score
S0086	3
S0087	2
S0088	3
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?

*I ultimately went with ETS(M,N,M). Even though my ETS model had a higher AIC of 1283 compared to my ARIMA(1,0,0)(1,1,0) that had an AIC of 880. The overall, statistics favored ETS.*

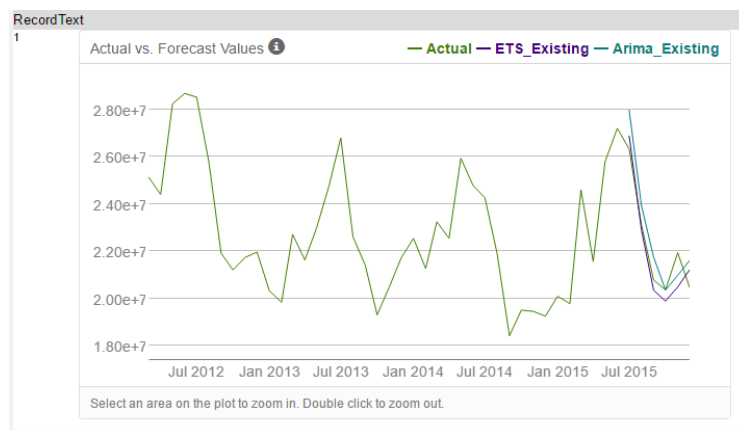
*In the chart below, you'll see that my mean error, root mean squared error, and Mean absolute scaled error for ETS are all lower than the Arima for when I had a holdout period of 6.*

#### Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_Existing	210494.4	760267.3	649540.8	1.0288	2.9678	0.3822	NA
Arima_Existing	-604232.3	1050239.2	928412	-2.6156	4.0942	0.5463	NA

*I*

*The chart below is only for the existing stores.*



2. Please provide a Tableau Dashboard (saved as a Tableau Public file) that includes a table and a plot of the three monthly forecasts; one for existing, one for new, and one for all stores. Please name the tab in the Tableau file "Task 3".

[https://public.tableau.com/views/FinalProjectUdacityJamesSykesMay182017/Dashboard1?:embed=y&:display\\_count=yes](https://public.tableau.com/views/FinalProjectUdacityJamesSykesMay182017/Dashboard1?:embed=y&:display_count=yes)

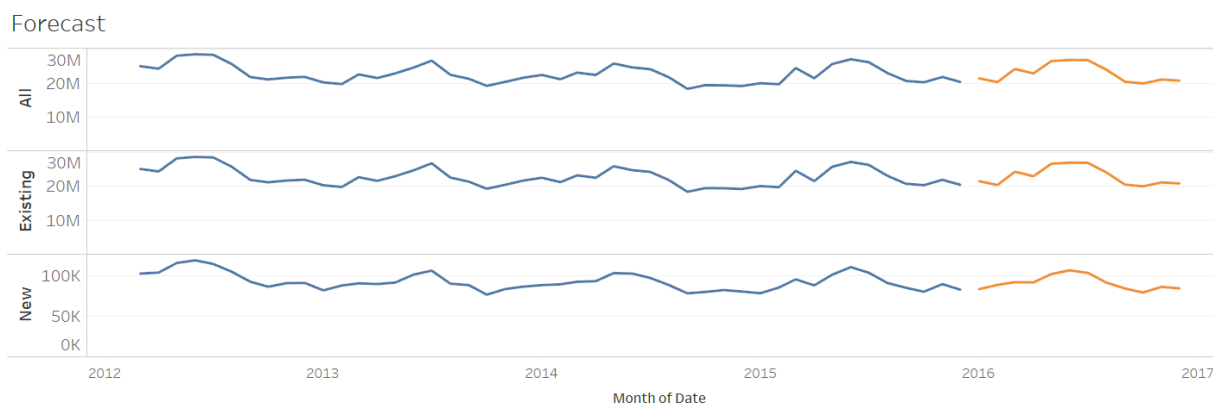


Table														☑
Predicted Actual	Year of Date		Date											
			January	February	March	April	May	June	July	August	September	October	November	December
	2012	New			102,580	103,777	115,563	118,892	114,432	104,819	92,558	86,341	90,799	91,005
		New & Existing			25,254,105	24,509,825	28,365,102	28,810,256	28,650,140	25,898,340	22,008,200	21,289,903	21,826,598	22,053,986
		Sum of Produce Existing			25,151,526	24,406,048	28,249,539	28,691,364	28,535,707	25,793,521	21,915,642	21,203,563	21,736,159	21,962,977
	2013	New	81,859	87,919	90,486	89,706	91,455	101,357	106,216	90,176	88,323	76,603	83,505	86,486
		New & Existing	20,404,543	19,917,540	22,807,556	21,715,092	23,091,607	24,856,763	26,909,321	22,690,393	21,489,589	19,373,181	20,573,278	21,802,193
		Sum of Produce Existing	20,322,684	19,829,621	22,717,070	21,625,385	23,000,152	24,755,406	26,803,106	22,600,217	21,401,266	19,296,578	20,489,773	21,715,707
	2014	New	88,301	89,359	92,461	93,215	103,177	102,557	97,155	88,490	78,228	79,934	82,246	80,419
		New & Existing	22,632,759	21,351,773	23,339,629	22,635,203	26,046,224	24,884,735	24,360,272	21,968,479	18,485,491	15,757,506	15,927,000	19,320,804
		Sum of Produce Existing	22,544,458	21,262,413	23,247,169	22,541,988	25,943,047	24,782,178	24,263,118	21,879,989	18,407,264	14,957,572	14,444,753	19,240,385
	2015	New	78,370	85,357	95,466	88,035	101,239	110,526	103,635	90,995	85,110	80,205	89,553	82,680
		New & Existing	20,166,900	19,857,690	24,703,873	21,647,764	25,893,313	27,322,990	26,442,112	23,221,622	20,859,526	18,404,186	22,026,460	20,545,580
		Sum of Produce Existing	20,088,529	19,772,333	24,608,407	21,559,729	25,792,075	27,212,464	26,338,477	23,130,627	20,774,416	20,359,981	21,936,307	20,462,899
Predicted	2016	New	83,372	88,641	91,943	91,707	102,005	106,543	103,378	91,443	84,132	79,226	86,205	84,336
		New & Existing	21,623,308	20,502,412	24,417,896	23,085,173	26,793,956	27,096,507	27,052,009	24,183,022	20,607,624	20,090,975	21,263,641	20,940,135
		Sum of Produce Existing	21,539,936	20,413,771	24,325,953	22,993,466	26,691,951	26,989,964	26,948,631	24,091,579	20,523,492	20,011,749	21,177,435	20,855,795

### 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.