Combining Predictive Techniques

REVIEW | HISTORY

Meets Specifications

Excellent! You do well. you resolve all observations everything seems to be perfect 💯.
✅ Task1 The map looks great. map legends for color and sales are there. although all circles are the same size you need to revise that.👏🏼
✅ you are supposed to select ETS over ARIMA against the holdout sample.👏🏼
Your answers are readable, concise, and supported with clear visualization.
I enjoy reviewing your submissions thanks. Keep learning.
Don't forget to give review feedback ⭐⭐⭐⭐⭐

Additional resources

🔗Dimensions and Measures, Blue and Green
🔗 THE DIFFERENCE BETWEEN SEGMENTATION AND CLUSTERING
🔗 decision-tree-vs-random-forest-vs-boosted-trees-explained
🔗 ARIMA vs ETS
🔗tableau help

Overall

✓	**The write up is written clearly, in complete sentences, and without major typos.**
	Awesome: The write-up is written clearly, in complete sentences, and without major typos.
✓	**Several visualizations are included. All visualizations are clearly labeled and help answer the related questions.**
	✅ Task1 The map looks great. It has legends. Color is used to show the clusters and size is used to show total sales. ✅ Task3 visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts

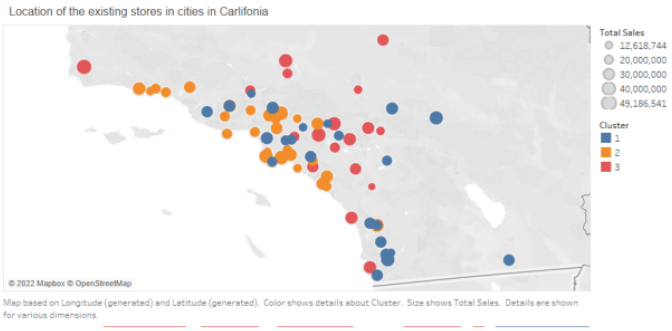
Task 1

✓	**Accurately identifies the correct number of formats and provides justification using the Adjusted Rand and CH indices.**
	✅ Awesome yes, cluster 3 is optimal . ✅ you use K-means as the clustering method. ✅ provide justification using the Adjusted Rand and CH indices according to the rubric. CH (Calinski-Harabasz) indices and the box-whisker plots in the Rand indices show how tight the indices for each data point are within each other
✓	**Identifies the correct number of stores that fall into each store format.**
	✅ Awesome every cluster has the right number of stores assigned to it.25-35-25
✓	**Provides one observation about the differences among clusters, and uses the results of the clusters to provide justification.**
	✅ Awesome you, provide much correct observation suggestions: that clusters differ from each other, they have very close average distance, max distance, and separation values but you may say the max distance is the most distinctive. We could add a more specific description by pointing towards what type of goods the clusters are oriented more We can also include a table or some visualization to support the comparison between the clusters You can use the table from the "R" output end of the "K-Centroids Cluster Analysis" tool. For example, a higher

value in our case means that the cluster is oriented towards selling more of that type of goods. For example, cluster one sells a lot of
General Merchandise compared to the other two cluster

✓ **Includes a map that shows the location of the stores, uses color to show cluster, and size to show total sales. A legend is used for both color and size.**

✅ Awesome the map looks great.
✅ It has Color legend. Color is used to show the clusters
✅ It has size legend.size is used to show total sales.



Task 2

✓ **States the type of classification model used and adequately justifies the choice using at least one model comparison method.**

✅ you should use a 20% validation set and random seed = 3.
✅ You compare a decision tree, random forest, and boosted model
✅ You should defend you choice by comparing the overall accuracy against the validation set
you can choose either Boosted or Forest Model,. you choose Forest model which is great.

✓ **Includes a table that correctly identifies the format for each of the 10 new stores.**

✅ Awesome the stores are correctly segmented - great job.
I can see cluster 1 , cluster 2 ,and cluster 3 great

Task 3

✓ **Compares and identifies the best ETS or ARIMA model to use for forecasting. Justifies the decision by showing the plot and shows forecast error measurements against the holdout sample.**

✅ Awesome great job you selecting ETS over ARIMA against the holdout sample.
✅ You jusify your answer for choosing ETS

- should either present graphs showing the actual vs. forecasted values
- or the forecast error measurements against the holdout sample.
  ✅should choose either an ETS(M,N,A) or ETS(M,N,M) model over any other ETS model by showing the decomposition plot. you should either present graphs showing the actual vs.
  forecasted values or the forecast error measurements against the holdout sample.
  decomposition plot we can see
- that there is quite a bit of seasonality , so we should apply it multiplicatively.(M)
- From the plot, we can also see that the trend turns up at the end, so trend should not be applied,(N)
- it appears that the remainder is changed in magnitude,so we should apply it multiplicatively.(M)

✓ **A table with the correct 12 month forecasts for existing and new stores is provided. A visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts is provided.**

✅ Awesome the forecasts for the new stores are in the acceptable range.
✅ Awesome the forecasts for existing stores are in the acceptable range.
✅ Great job! visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.

historical data contains from 2012 to 2015
You can do that by I guess you generate sum_produce for all data by year and month this will be actual data or historical
then add in the same CSV file the data you get for new with type new and existing for type existing

DOWNLOAD PROJECT

RETURN TO PATH