UDACITY

PROJECT

Combining Predictive Techniques

A part of the Business Analyst Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Requires Changes

1 SPECIFICATION REQUIRES CHANGES

Really great job so far ! Your work needs some further improvements though. I'm sure your next submission will be the ONE.
Keep up the good work !!!

Overall

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.

Task 1

Accurately identifies the correct number of formats and provides justification using the Adjusted Rand and CH indices.

Awesome: Well done using the Adjusted Rand and Calinski-Harabasz indices to find the best number of clusters. 👍

Identifies the correct number of stores that fall into each store format.

Awesome: Your classification of stores is also accurate. The clustering model does a good job.

Provides one observation about the differences among clusters, and uses the results of the clusters to provide justification.

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: Great job providing the map here along with its legend and title

Task 2

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

Awesome: The boosted model is indeed the best choice, due to the higher F1 score.

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

Task 3

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.

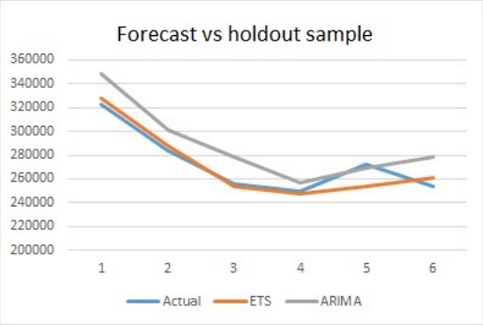
Awesome: The forecasts for both new and existing stores are accurate though. 👍

Identifies the best ETS or ARIMA model to use, and justifies the decision by showing forecast error measurements against the holdout sample.

Required:

You should have found ETS(M,N,M) to be optimal here. We have to use a holdout sample when comparing the models. It will give a more reliable comparison of the models.
Here is the procedure for finding the optimal model between ARIMA and ETS
The optimal ETS model to use for each forecast is ETS(M,N,M). First, analyze the Time Series plot in order to decide which is the best ETS model to use. You'll notice that the Error component always increases and decreases (M), the Trend component is missing (N), and finally, the Seasonal component decreases slightly every year (M).
You should also build a seasonal ARIMA model (0,1,1)(0,1,0)12. The reason for that ARIMA model is that (1) It takes one seasonal differentiation and one first difference to make the data stationary (d=1, D=1); (2) the autocorrelation at lag-1 is negative (q=1) and (3) there is no significant autocorrelation at the seasonal lags to include a seasonal term.
You'll then compare both models against the holdout sample (year 2015) and based on the error measurements as well as the forecasted values vs. actual values, you'll be able to determine the optimal model between the two.

Here is an example of a six period holdout sample. We can see that ETS is forecasting the holdout sample more accurately.



RESUBMIT

DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

RETURN TO PATH

Rate this review