# Forecasting Retail Gasoline Prices

*Nicole Leong-Lee*

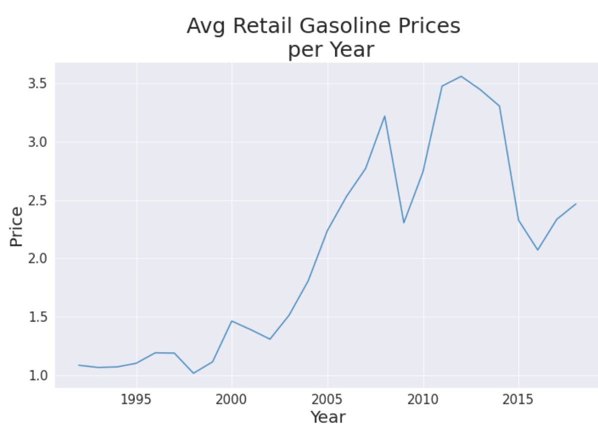
---

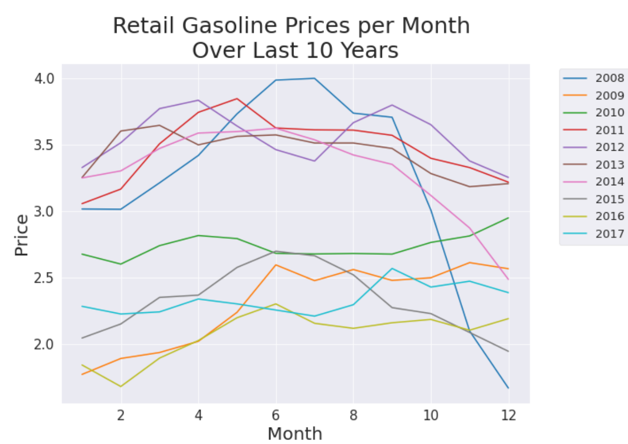
The purpose of this analysis was to explore and forecast U.S. retail gasoline prices from 1992 to 2018 using Facebook Prophet (FB Prophet), a tool used to forecast time series data. All work (exploratory data analysis and modelling) was done in a Jupyter Notebook with supporting Python modules that execute visualizations and modelling. Additional packages like Numpy, Pandas, and SkLearn were used.

**Analysis:**
The Excel file was reformatted to remove extra rows and restructure the data into an acceptable output for analysis. There were no missing values.

The file contained monthly data on U.S. retail gasoline prices from January 1992 to January 2018. Based on the average retail gasoline prices per year, there were significant declines in price from 2008 to 2009 and 2014 to 2016.

Due to these 2 drastic changes, an assessment on monthly trends over the last 10 years was performed to check for seasonality. There was no monthly seasonality of gasoline prices found. Interestingly, there was a steeper drop in prices from July 2008 to December 2008, while there was a gradual decline from June 2014 to December 2015.




**Modelling:**
The last year of the dataset was used as the test set, whereas the remainder of the data was assigned as the training set. The dataset was separated in this manner as more data was needed to train the model and produce outputs that were in line with the general trend.

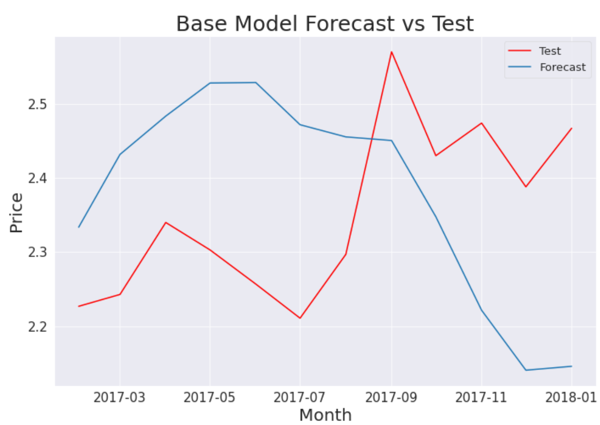
A base FB Prophet model (with default hyperparameters) was created as a basis of comparison for future models. The Root Mean Squared Error (RMSE) of the test set for the base model was 0.21. RMSE, instead of Mean Absolute Error, was deemed a more appropriate metric to evaluate the model. It gives higher weight to larger errors, penalizing outliers. Larger errors in this analysis are unwanted.

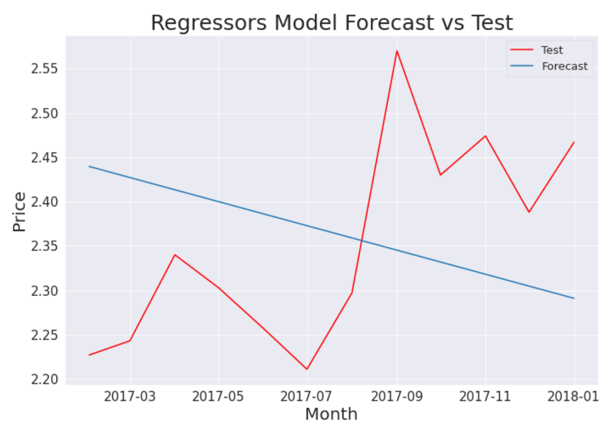
For example, an error in the positive or negative direction of gasoline prices would have a significant effect on consumers and/or gasoline companies. Higher prices would result in lower gasoline consumption, incentivizing people to drive less, carpool, or use public transportation. Conversely, lower prices would cause gasoline companies to be less profitable, regardless of how much was sold at the lower price, due to costs of the business.

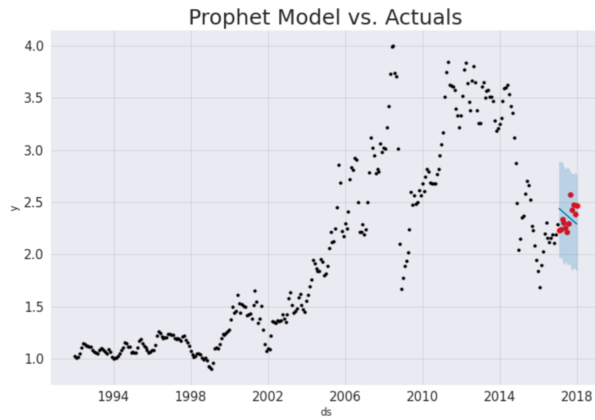
To improve the model, additional regressors were added. 2008 to 2009 and 2014 to 2016 were important historical years for gasoline prices. The Financial Crisis of 2008 caused a great recession in the U.S. and the Oil Glut of 2014 led to a plunge in oil prices after developing economies were demanding much less of it compared to prior years.

Regressors were added to take these 2 extraordinary events into account. In addition, given the Coronavirus Pandemic and Russia-Saudi Arabia Oil Price War this past year, these seemingly one-time "crisis" events appear to be more common and were incorporated into the model to account for similar future occurrences.

The final model used the same hyperparameters as the first, but with the yearly seasonality hyperparameter set to False due to the inclusion of additional regressors. The RMSE of the test set for this model improved by 29%, decreasing to 0.15. As shown below, the final model does a better job of generalizing the test set compared to the base model. The uncertainty levels (portion highlighted in blue on the last graph) of the final model also capture the true values of the test set (red points on the last graph).






**Running Code and Analysis:**
To run the code and analysis, simply download the "forecasting_take_home_nicole_leong-lee.ipynb" file and "src" folder from Github (https://github.com/nicleong27/forecasting_take_home_nicole_leong-lee.git), and run all cells in the Python Notebook.