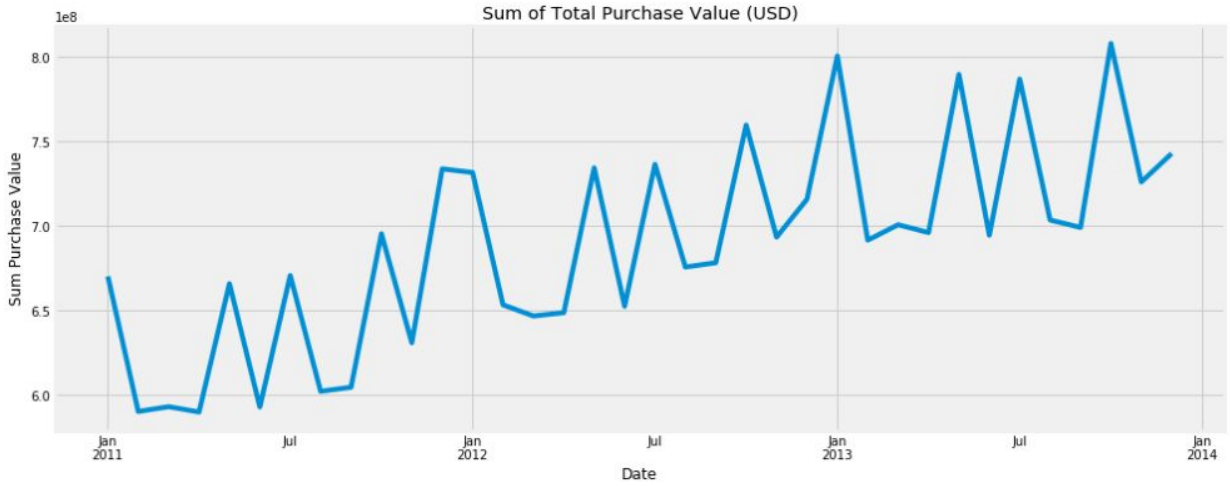
John Gilmore
Gfk Data Assignment
9/18/2019

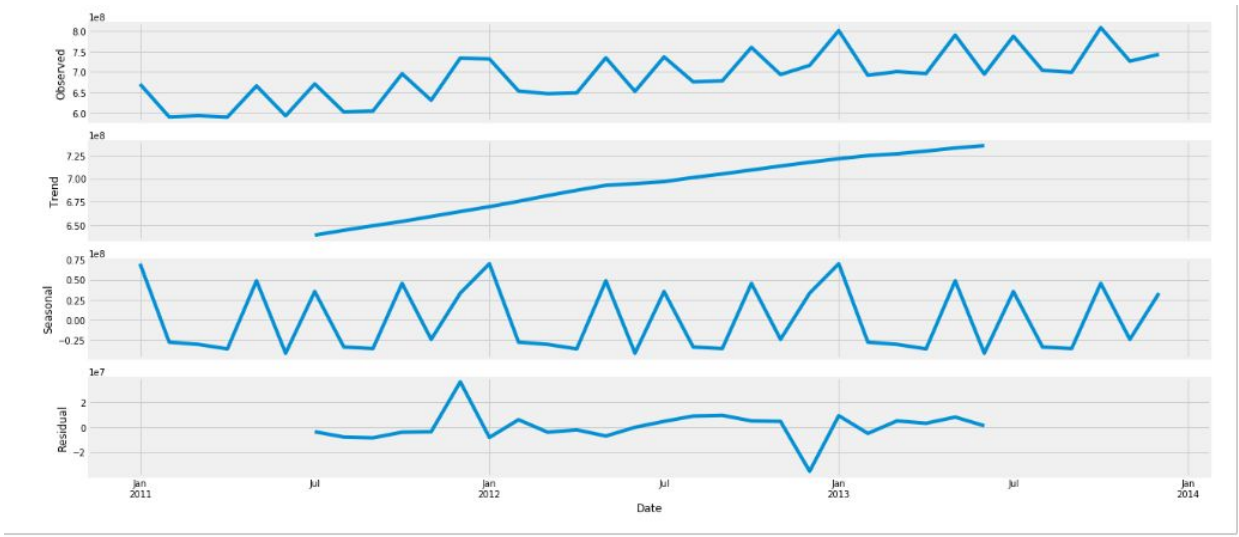
I conducted my analysis primarily in python and will format my write up largely by presenting visualizations accompanied by comments.

Trend Analysis in Sample Data:

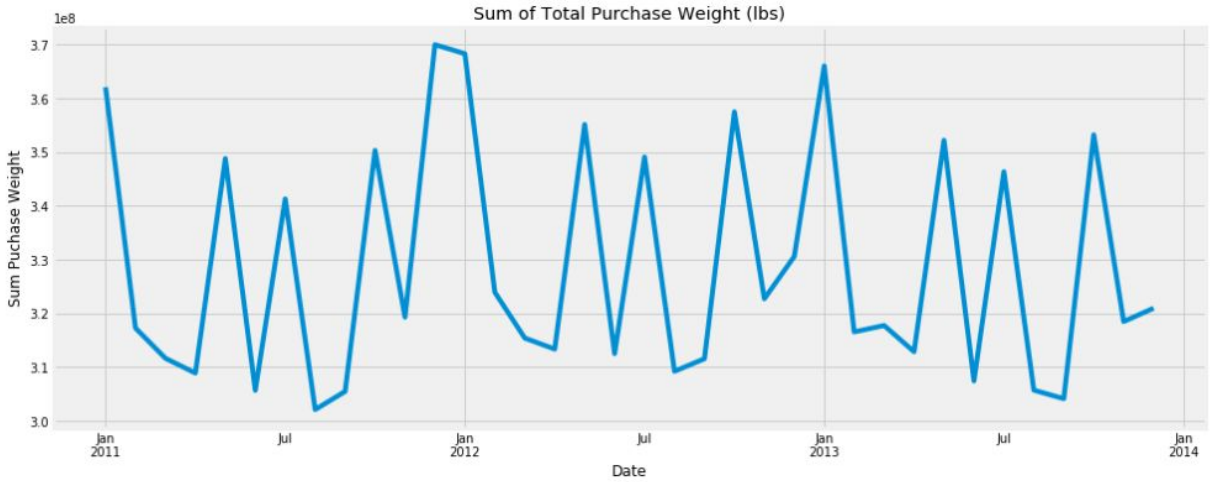
I began by examining the data and using python to visualize the data to better be able to spot trends in the data provided and establish variables of interest.



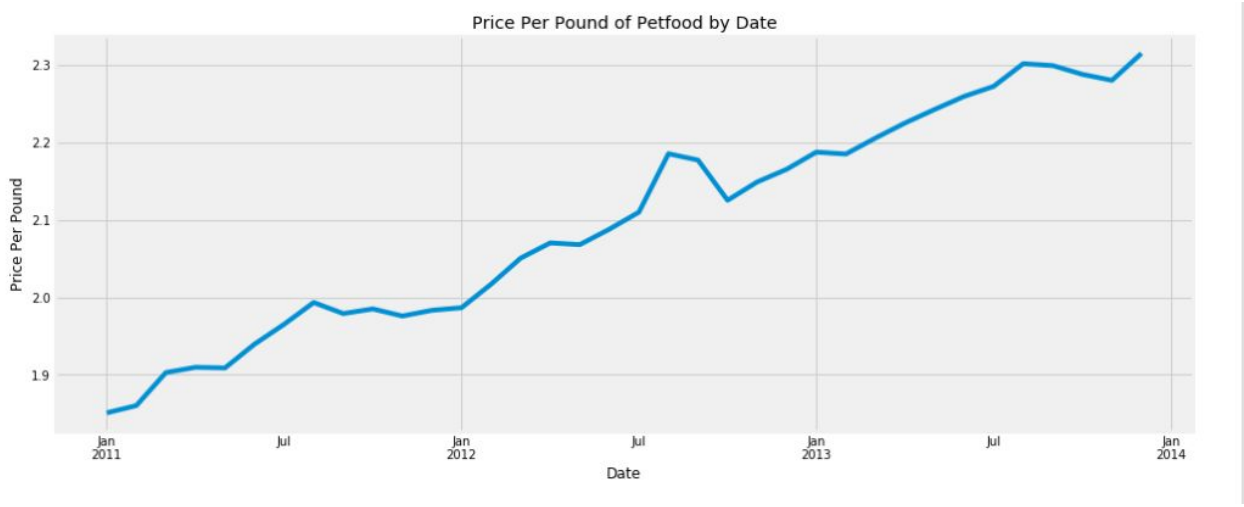
Total sales value is shown to have a high variance with repeating peaks and troughs. This could likely be due to dog food frequently being purchased in bulk and only requiring repurchase up to months later. However, that does not fully explain why people would be making their repurchases at the same time. The most notable spikes come in January, possibly due to a connection to the end of the Holiday season. More importantly, the data shows a steady increase in the total value of purchases increasing every year.



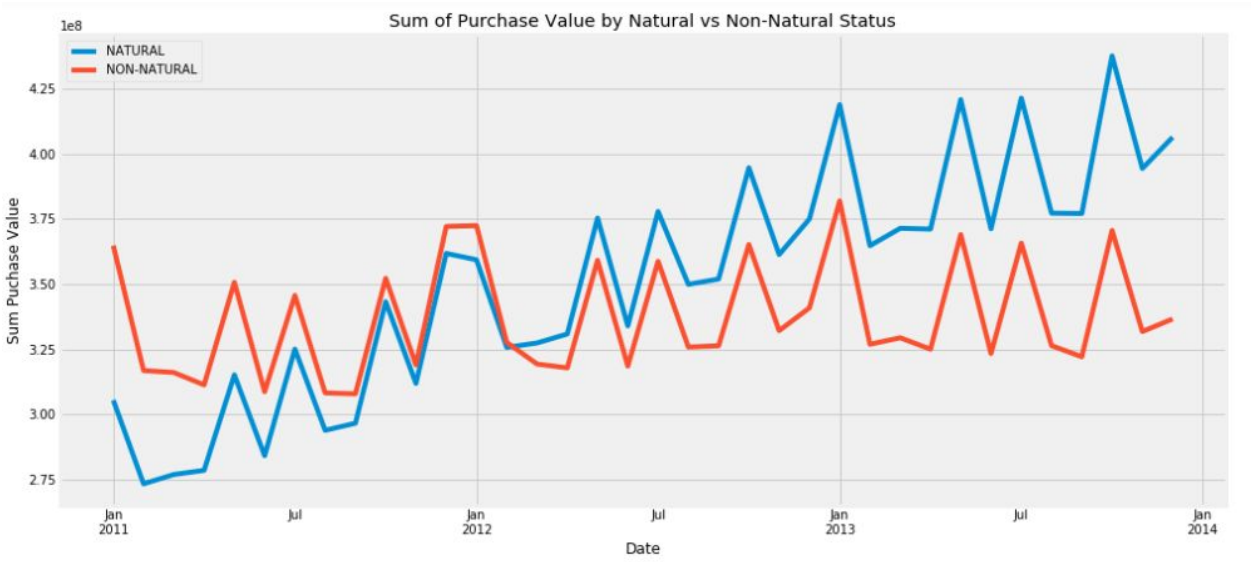
I also performed a time series decomposition to decompose my time series down to seasonality, noise, and trend. The most striking outcome of the decomposition is likely the strong upward trend shown in the second chart, which increases confidence that change we are seeing can not be attributed to randomness.



The sum of purchase weight is similarly variable and reflects comparable peaks and troughs to the value chart. However, unlike the value chart, there is not a clear continuing increase in the weight purchased suggesting the increase in value is due to an increase in the cost per pound of food.

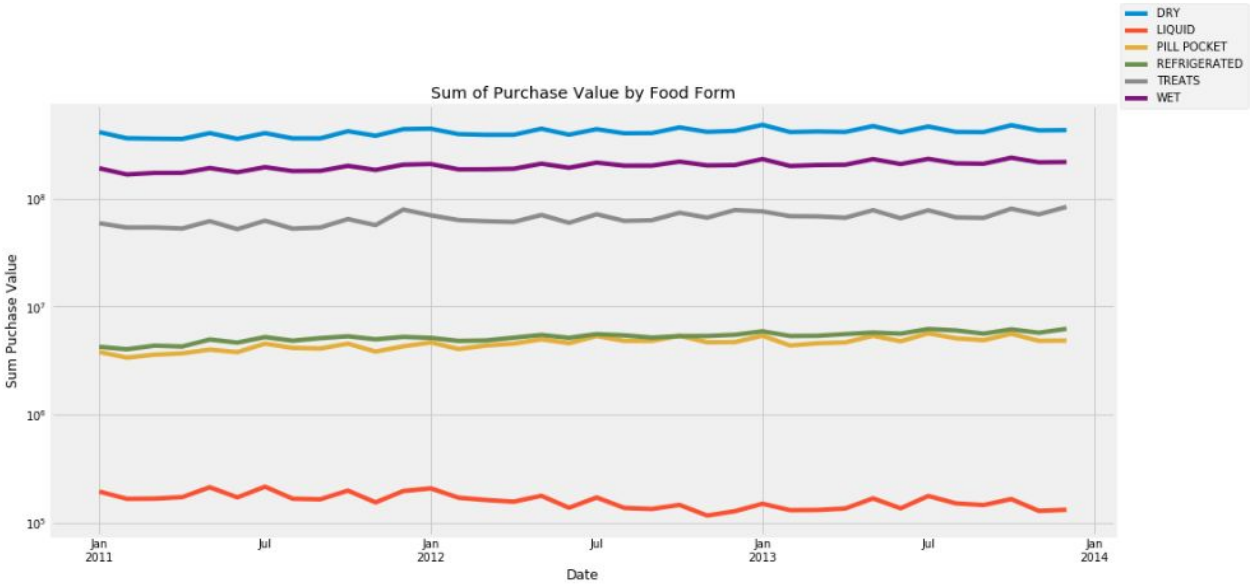


The increase in price per pound of food is confirmed by this chart showing the price per pound over time. There is a clear upward trajectory shown every year. This may be due to arbitrary price increases but may also be related to a shift towards pet food with higher quality ingredients.

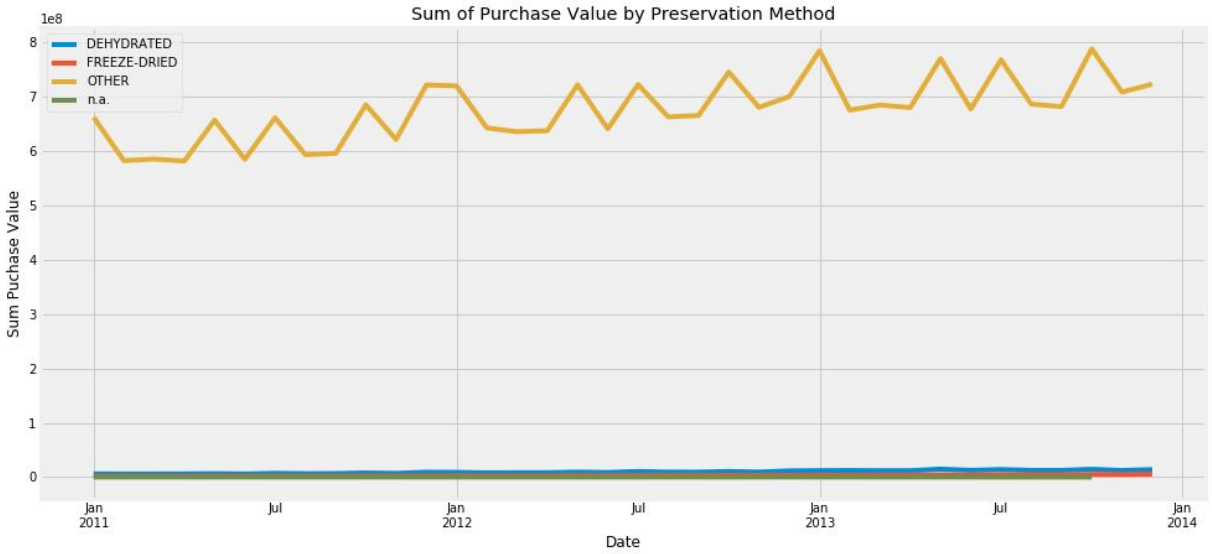


This chart comparing the total purchase value of Natural and Non-Natural pet food lends some credence to the hypothesis that pet food with higher quality ingredients is gaining in popularity. Non-Natural dog food is more popular at the start of the sample but is quickly eclipsed by the natural pet food, with the gap growing as time goes on.

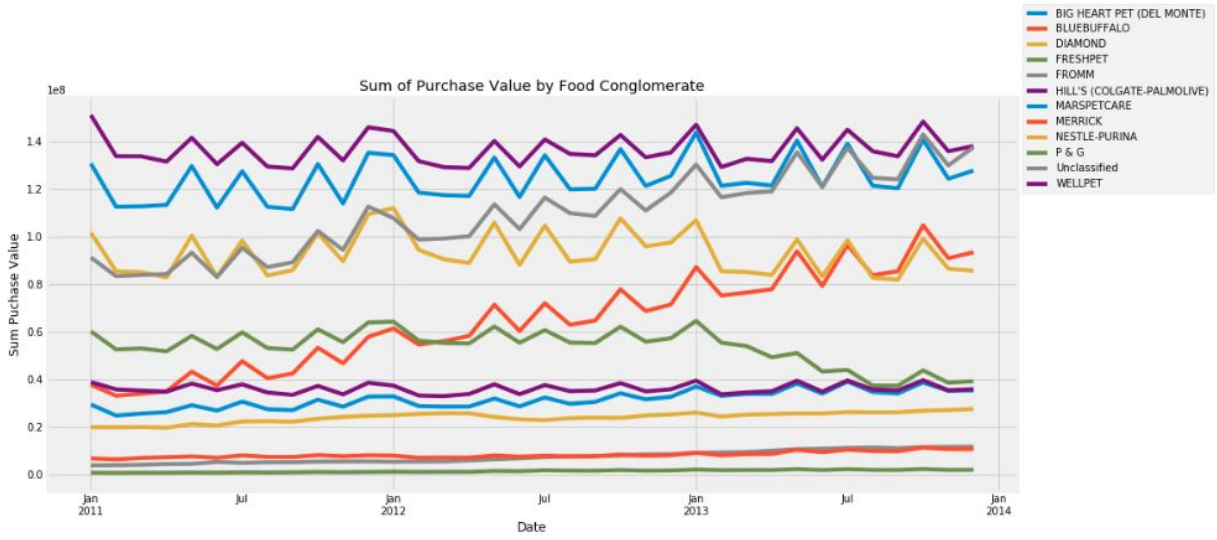
For the rest of my analysis, with a focus on forecasting 2014, I focused on predicting the total dollar value of the pet food market as included in the sample, and on the continued tradeoff between natural and non-natural pet food. I chose these particular variables because market value is likely to be the most desired information by important stakeholders and the natural pet food tradeoff seemed to be the variable experiencing the most change. However, before moving on to those investigations I did conduct analysis on the other provided variables.



Food form purchasing remained static. Variation can likely be attributed to random noise and purchasing remained largely the same by each product.



Purchasing by preservation method is comparably static. There is an increase in spending on the “other” preservation method but considering the extremely market share other already displayed relative to the other methods, this is not notable.



Given more time, the variable Conglomerate is the variable I would be most interested in analyzing further. The most notable changes over time are the decrease in popularity of Nestle-Purina and P & G and the increase in popularity of Fromm and Blue Buffalo. I believe it is likely that this would show correlation with the increasing popularity of natural options. Blue Buffalo and Fromm both emphasize natural ingredients and a transparent production process whereas P & G and Nestle-Purina pet foods have faced criticism for their ingredients and have mid to low review scores on dogfoodadvisor.com, a popular pet supply reviewing site.

Most of my efforts onward are focused on producing reliable forecasts for 2014. I chose to use an ARIMA (Autoregressive Integrated Moving Average) time series model in order to

predict what the total sales value would look like in 2014. After that made use of Facebook's Prophet package for my first time to do a comparative time-series analysis of natural and non-natural products and predict the trajectory of their tradeoff in 2014.

ARIMA Model:

I began my ARIMA model by using a grid search to find the optimal parameters for abest-fittingg model.

```
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(sum_val,
                                            order=param,
                                            seasonal_order=param_seasonal,
                                            enforce_stationarity=False,
                                            enforce_invertibility=False)
            results = mod.fit()

            print('ARIMA{}x{}12 - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

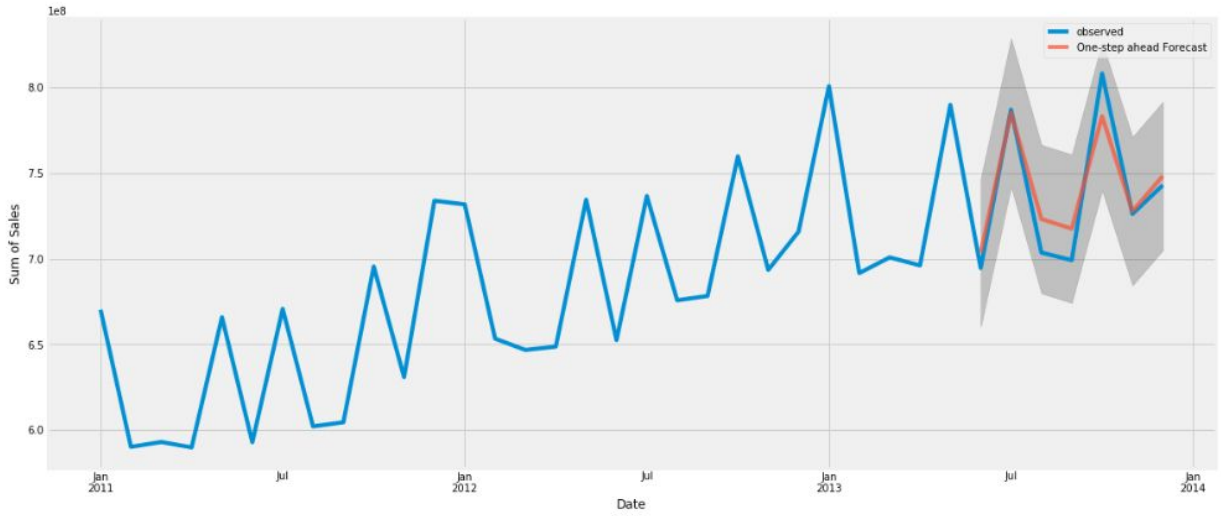
ARIMA(1, 1, 0)x(0, 0, 0, 12)12 - AIC:1311.1755693112536
ARIMA(1, 1, 0)x(0, 1, 0, 12)12 - AIC:810.5421764458929
ARIMA(1, 1, 0)x(1, 0, 0, 12)12 - AIC:833.5070094283749
ARIMA(1, 1, 0)x(1, 1, 0, 12)12 - AIC:365.6243214858375
ARIMA(1, 1, 1)x(0, 0, 0, 12)12 - AIC:1266.8168639357732
ARIMA(1, 1, 1)x(0, 1, 0, 12)12 - AIC:775.8104983358219
ARIMA(1, 1, 1)x(1, 0, 0, 12)12 - AIC:832.4559476440575
ARIMA(1, 1, 1)x(1, 1, 0, 12)12 - AIC:367.66861949931354

The italicized text above shows a snippet of the output and includes the parameters I used, being ARIMA(1, 1, 1)x(1, 0, 0, 12) because they had the lowest AIC of about 365.62

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.4889	1.005	-0.486	0.627	-2.459	1.481
ar.S.L12	0.0967	0.933	0.104	0.917	-1.732	1.925
sigma2	4.889e+14	6.74e-16	7.25e+29	0.000	4.89e+14	4.89e+14

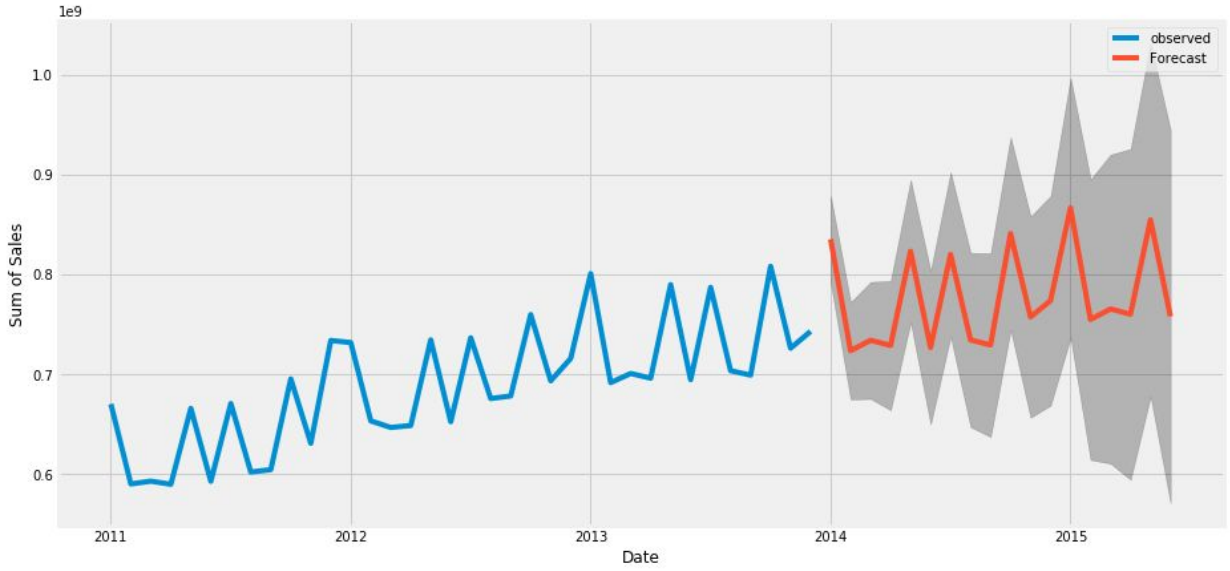
I performed a model diagnostic to ensure the quality of my model fit (shown above). The summary statistics are not completely ideal and if I were to continue this project I would likely want to lower the ar.S.L12 stat, but for the time being it is an acceptable fit.

Proceeding on, I wanted to test my predictive model against my data. I used all of the data up to 6/1/2013 as a training case and tested my model against the remainder of the data as seen below.



The forecast is a pretty good match. It does not capture the peaks and troughs quite as well as hoped for, but the sample size is small and it matches the trends well, so ultimately I was satisfied.

With the test completed, I made my forecast for the 2014 sum of sales value.

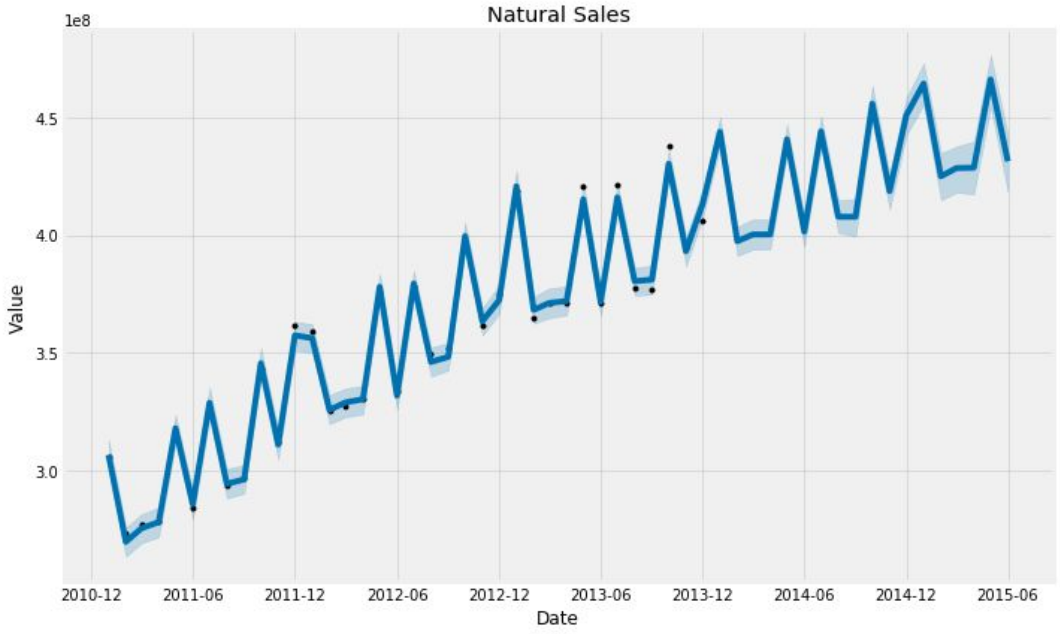


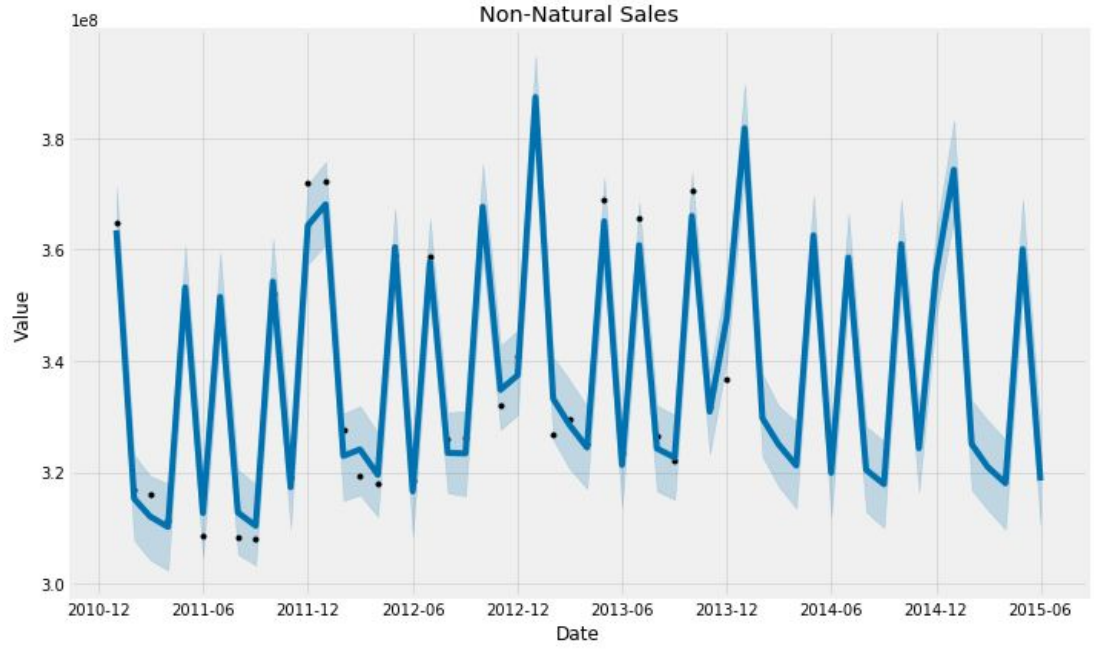
The forecast shows a continuation of the upward trend and mirrors the peaks and troughs of previous years. It is a particularly close comparison to 2013. Ultimately, this is what I expected and it seems consistent with the data I observed. Nothing in the data strikes me as a reason to expect a downturn in the pet food market and continuous growth shown over the previous three years should likely continue.

Prophet Analysis of Natural Pet Food Market:

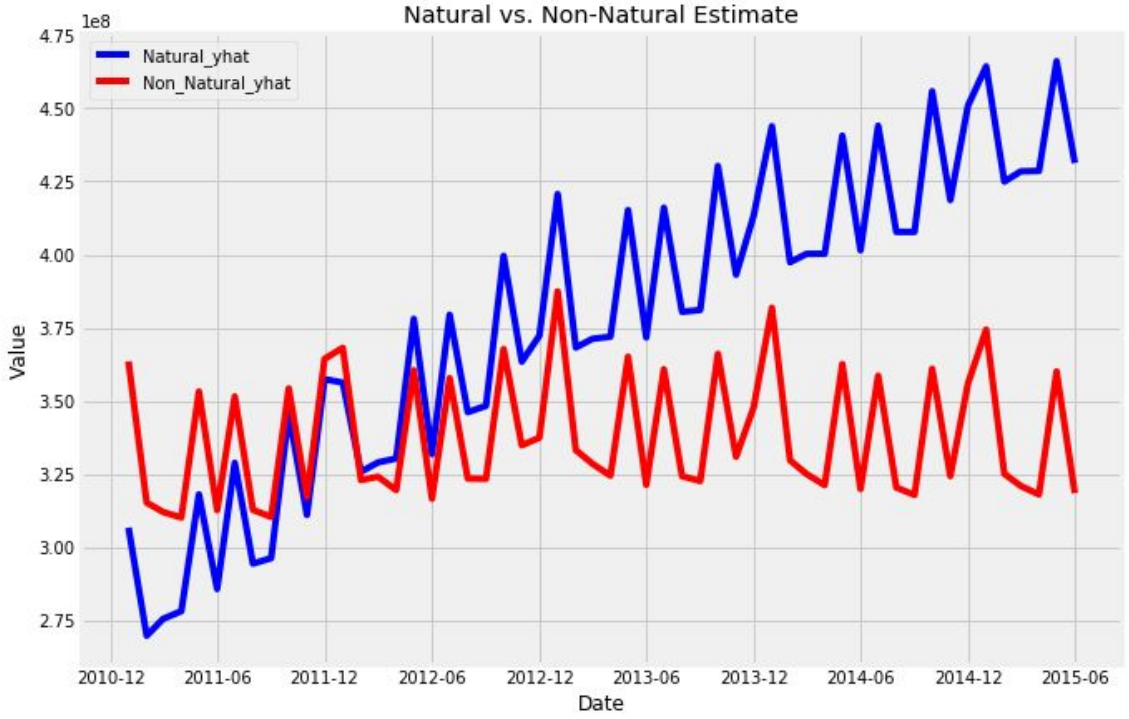
After completing my forecast for 2014, I wanted to do a more advanced analysis of the trends in the natural pet food market. In order to do so, I made use of Facebook Prophet which is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects (also I had never used it before and wanted to try it out).

Prophet benefits from requiring less manual tuning so you can get right into the forecasts. I began by forecasting the total sales value of natural and non-natural individually, as seen below.

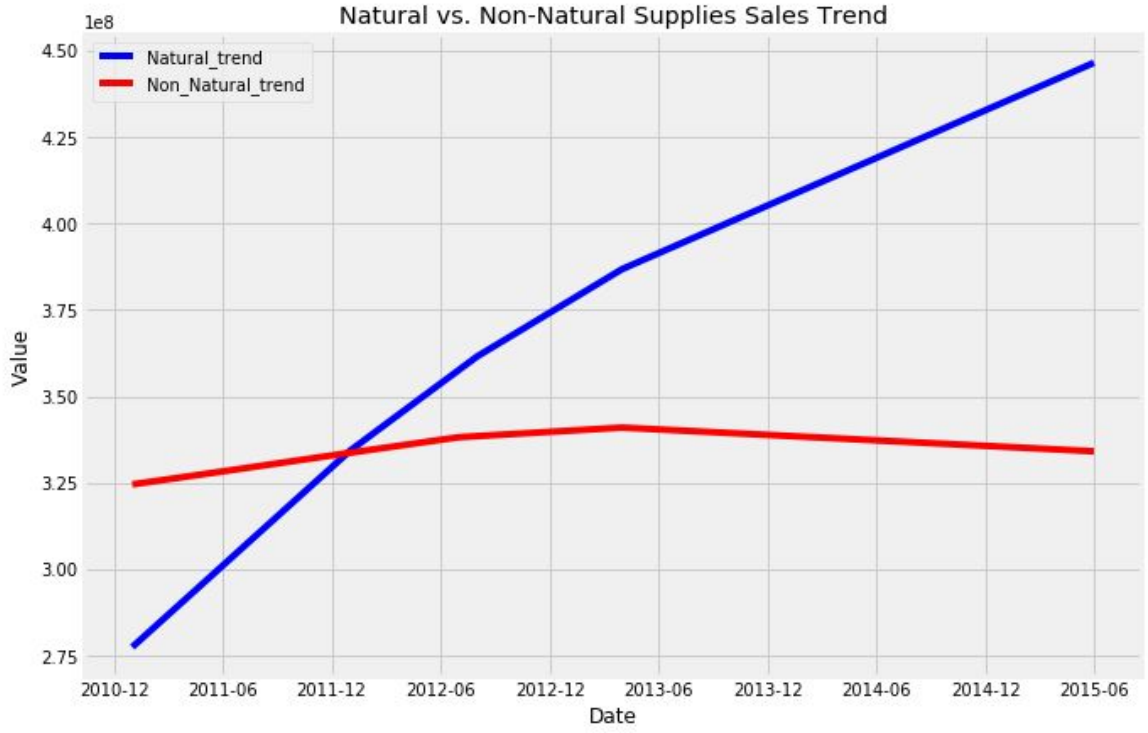




However, having the two forecasts together is better for comparison.



One can see that the expanding gap between natural and non-natural that I mentioned earlier should continue with the primary driver being an increasing popularity of natural pet food. Notably, non-natural pet food sales are actually staying fairly constant.



Finally, it is useful to isolate the forecasts to a trend line. The popularity of natural food is not increasing to quite the same degree as it was leading up to the start of 2013 but it is still clearly pulling away from non-natural food, which is now more notably showing decreasing sales totals since early 2013 that are projected to continue.

Conclusion:

After concluding my analysis by key conclusions are that the pet food market is showing growth in total sales value that should continue through 2014, the primary driver of this is an increase in popularity of more expensive dog food brands and a preference for natural food over non-natural food, which is experiencing a decline in sales value that I have forecasted to continue through 2014. Given my findings, I would expect that many less popular non-natural brands will make efforts to put out higher quality, natural products in an effort to access the increasing number of natural pet food preferring pet owners. These findings may also have implications for other industries. The increasing popularity of natural options may have some correlation with an increased interest in organic/natural food options for people as well as pets. Also, the natural preference could prove to be a good sign for the health and wellness industry, as people willing to spend extra on their pet's diet could also be willing to pay a premium for their diet and lifestyle with things like high-quality food, gym memberships, or fitness classes. The increase in the price per pound of pet food may also be indicative of a willingness to spend more in general on pets which could suggest increased interest in products like premium pet beds and accessories, pet boarding, or even something like dog walkers.