Projecting Coffee Prices in Kenya

Riley Wong

Project Components

This project consisted of two main components: scraping coffee pricing data from PDFs for each month between 2014 and 2018, and forecasting coffee prices for the 2018-2019 session using a predictive model.

Data Scraping

The PDF reports for coffee pricing data come from the Nairobi Coffee Exchange. To convert these PDFs into usable data, I used the Tabula Python library. For half of the PDF files, I had to manually extract and clean the parsed data.

Forecasting

To forecast future coffee prices for the 2018-2019 session, I used the [Prophet](#) library, as it works well on time series data that contains seasonality effects.

Results

For the following charts, the x-axis is time and the y-axis is average coffee price for that month. The middle and left side of the graph containing a scatterplot with black points shows the historical data. The right side of the graph, without the accompanying black points, represents the forecast for future coffee types.

For all three coffee types, we can see that there seasonal trends occuring. AA and AB grade coffee prices appear to be trending upward, while C grade coffee prices appear to be trending downward.

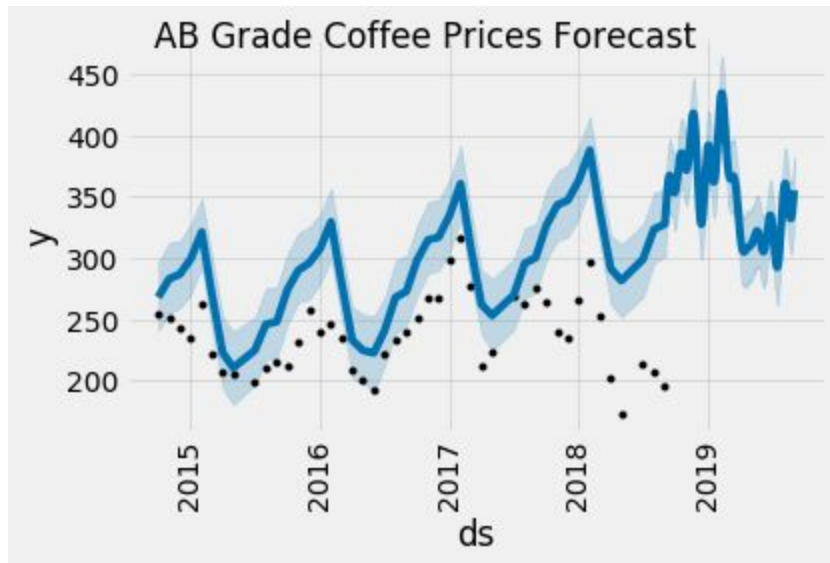For all three coffee types, it seems that coffee prices peak in February.

AA Grade Coffee
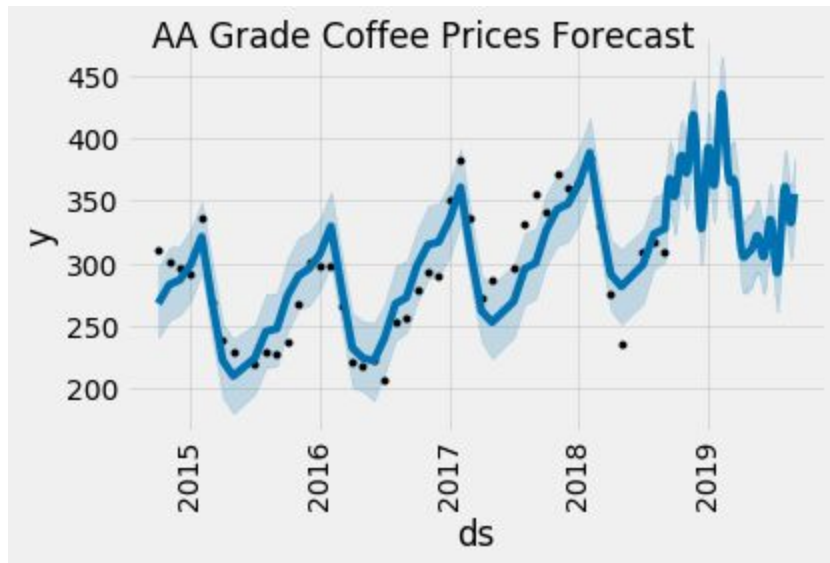
Predicted Coffee Prices

Oct 2018	374.95
Nov 2018	397.96
Dec 2018	353.84
Jan 2019	380.02
Feb 2019	428.38
March 2019	365.40
April 2019	308.62
May 2019	312.04
June 2019	316.14
July 2019	325.59
Aug 2019	352.31
Sept 2019	355.41

AB Grade Coffee

AB Grade Coffee Prices Forecast

Predicted Coffee Prices

Oct 2018	291.75
Nov 2018	271.29
Dec 2018	49.80
Jan 2019	184.46
Feb 2019	381.15
March 2019	248.53
April 2019	156.03
May 2019	159.21
June 2019	99.08
July 2019	190.94
Aug 2019	207.54
Sept 2019	196.72

C Grade Coffee
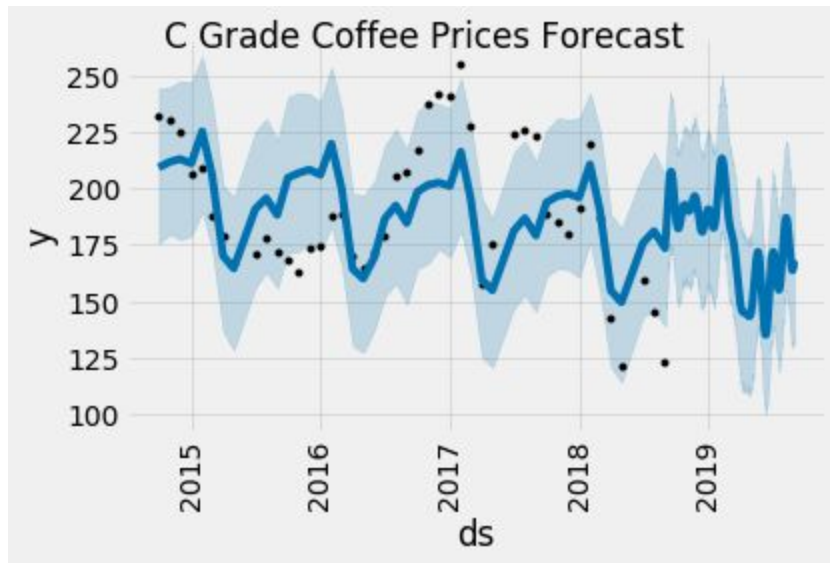
Predicted Coffee Prices

Oct 2018	187.72
Nov 2018	192.98
Dec 2018	185.62
Jan 2019	188.61
Feb 2019	203.94
March 2019	171.09
April 2019	145.10
May 2019	160.80
June 2019	147.75
July 2019	163.78
Aug 2019	175.99
Sept 2019	167.91

Constraints and Possible Extensions

The main constraint for this project was time.

Give the time constraint, I ended up not using any other data besides month, average monthly price, and coffee grade. I also only used one method.

Give more time, we could try a variety of methods, and compare them. Using other time-series predictive methods, such as simple averages, moving averages, Holt's, or ARIMA, and comparing these methods with a train/test split and mean squared error metric, could give us a fuller picture for these forecasts. We could also compare these methods against other machine learning methods, such as random forest, gradient boosted trees, or just a simple linear regression.

We could also incorporate all available data, or incorporate other data. The PDFs also included data on the bags bought, weight bought, min and max prices, and value in USD. These are all features that could potentially be useful in modeling. We could also do a bit of feature engineering such that for each month, we also include the price difference since the last month, and the price from the last month.