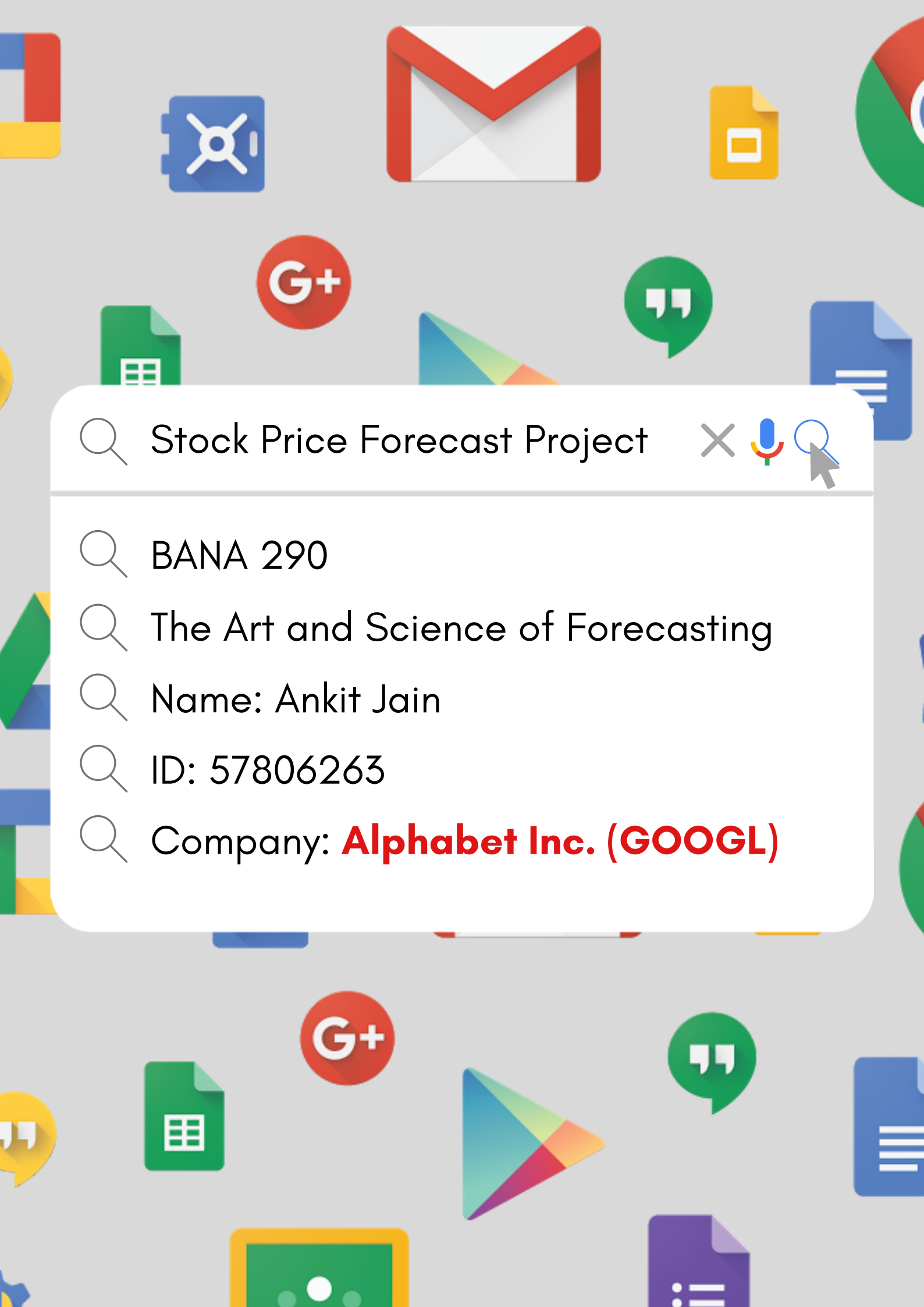
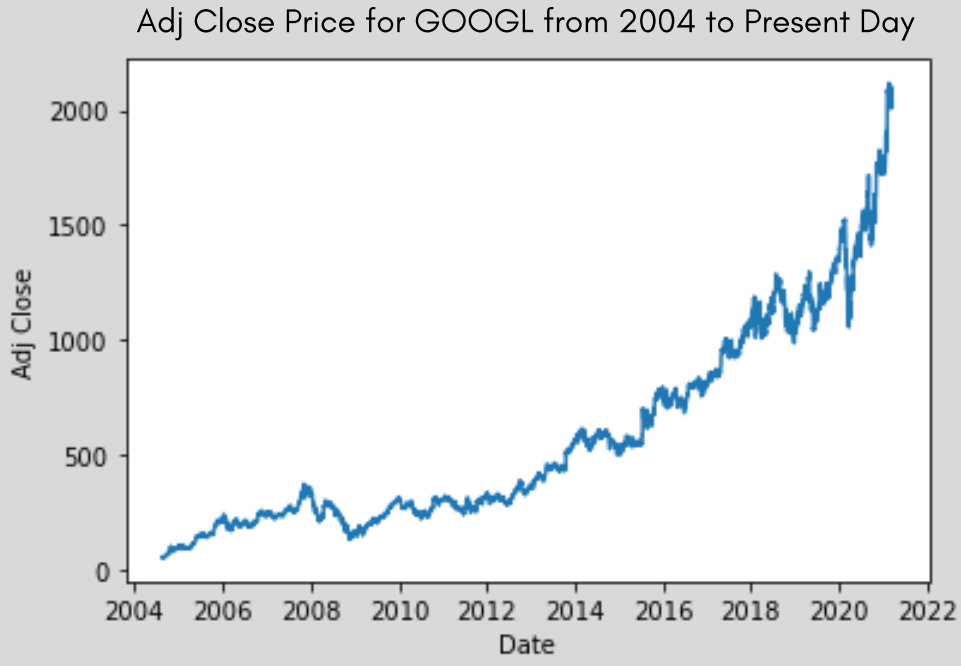

Stock Price Forecast Project
BANA 290
The Art and Science of Forecasting
Name: Ankit Jain
ID: 57806263
Company: Alphabet Inc. (GOOGL)

About the Company

Alphabet Inc. is a multinational company based in Mountain View, California, which was created due to the restructuring of Google in October 2015. It is the parent company of Google and its many subsidiaries.

Alphabet

Alphabet has two classes of shares: GOOGL and GOOG. The main difference between these two stock ticker symbols is that GOOG shares do not have voting rights, whereas GOOGL Shares do. GOOGL is the more commonly associated stock ticker symbol for Alphabet and is the Class A Share for the company.



Industry: Computer and Technology

Competitors: Apple (AAPL), Microsoft (MSFT), Facebook (FB), Taiwan Semiconductor Manufacturing (TSM), NVIDIA (NVDA)

Data Sources and Methods

Price History for the GOOGL Stock was obtained through the Pandas Python DataReader Tool from Yahoo Finance from October 4th, 2004 to the present day.

	Date	High	Low	Open	Close	Volume	Adj Close
0	2004-08-19	52.082081	48.028027	50.050049	50.220219	44659096.0	50.220219
1	2004-08-20	54.594597	50.300301	50.555557	54.209209	22834343.0	54.209209
2	2004-08-23	56.796799	54.579578	55.430431	54.754753	18256126.0	54.754753
3	2004-08-24	55.855858	51.836838	55.675674	52.487488	15247337.0	52.487488
4	2004-08-25	54.054054	51.991993	52.532532	53.053055	9188602.0	53.053055
...	...	...	...	...	...	...	...
4167	2021-03-10	2061.399902	2019.329956	2061.399902	2036.189941	1353100.0	2036.189941
4168	2021-03-11	2111.270020	2056.449951	2058.219971	2100.540039	1384200.0	2100.540039
4169	2021-03-12	2077.610107	2032.420044	2076.409912	2050.000000	1690000.0	2050.000000
4170	2021-03-15	2054.989990	2027.790039	2044.979980	2054.439941	1308400.0	2054.439941
4171	2021-03-16	2112.989990	2059.290039	2065.989990	2083.889893	1526368.0	2083.889893

4172 rows × 7 columns

I used the DataReader to also get information about three major competitors of the GOOGL Stock which were **AAPL (Apple), MSFT (Microsoft), and FB (Facebook)**.

I also got information about different economic variables associated with the US, as well as Alphabet Inc. from the **FRED Website** (https://fred.stlouisfed.org/)

For formulating forecasts, I used: **Python, Alteryx, and Excel.**
Types of Models Used in Ensemble: **ARIMA, ETS, Linear Regression, Support Vector Regression**

The goal of the project is to forecast the Adj. Close Price for GOOGL on Dec 31st, 2021

Overall Process

I based my thought process to predict the price around the idea that December 31st is the end of the period in terms of different time metrics in a year, i.e. that it is the last day of the year, last day of the last quarter of the year, last day of the last month of the year, and coincidentally, it is also the last day the stock market will open in 2021. So I used a lot of the end-of-the-period values to formulate my forecasts. For each of the models, wherever possible, I tried to graph out the forecasts to see how realistic the forecast is, and for the ensemble, I weighted each of my models based on the Mean Absolute Percentage Error (MAPE).

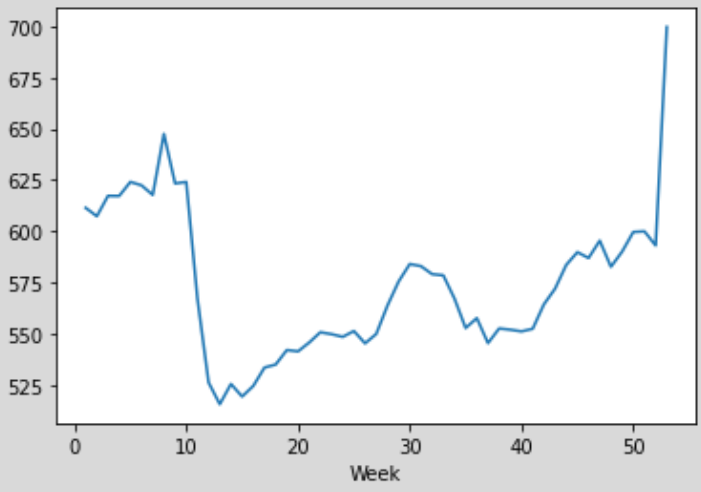
For all models, I did an 80-20 percent split for creating the training and testing sets.

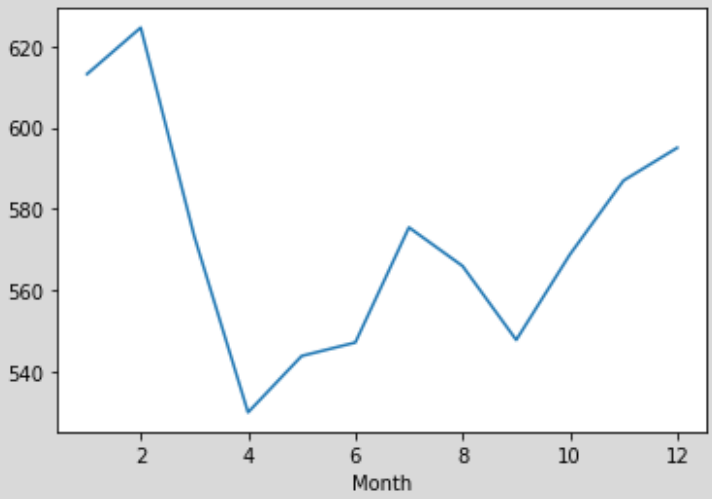
Process: ARIMA Models (Python)

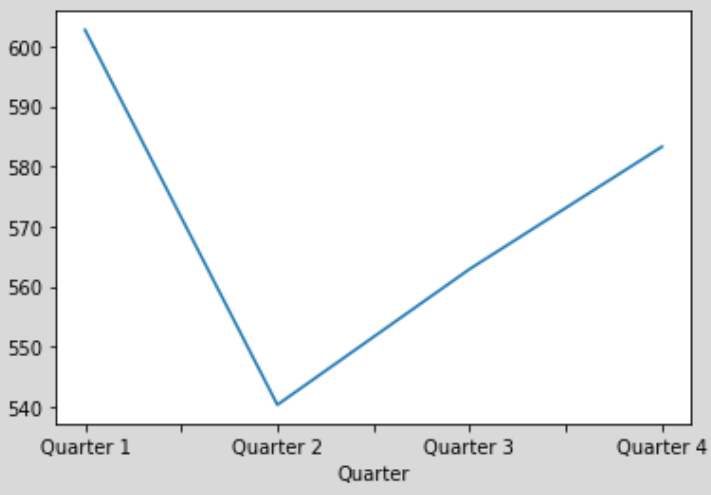
I created 4 ARIMA Models on Python, based on different levels of granularity:

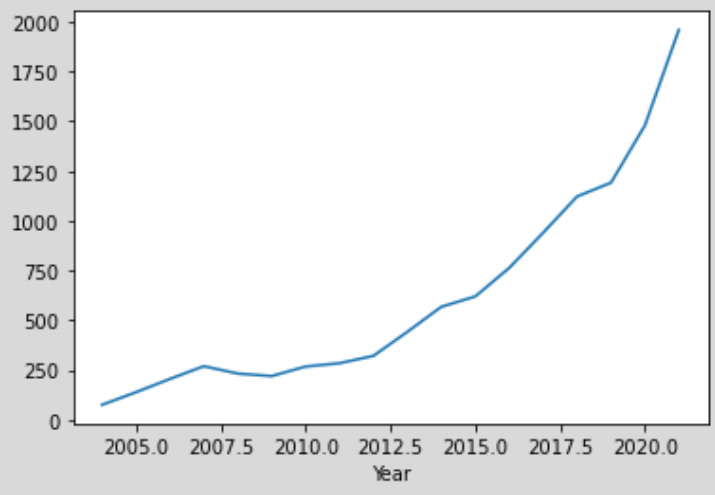
- Weekly
- Monthly
- Quarterly
- Annually

I did different levels of granularity because when I graphed each of them other than the daily prices to notice trends from 2004 to 2021 present, it seems that there was seemingly often a dip in the price around the month of April, which was confirmed when I explored the dataset. In order to account for the different levels of seasonality at each point, I changed around the PDQ order for the ARIMA Models and used the ones which provided the lowest MAPE, as well as a realistic result.









Process: ARIMA Models (Python)

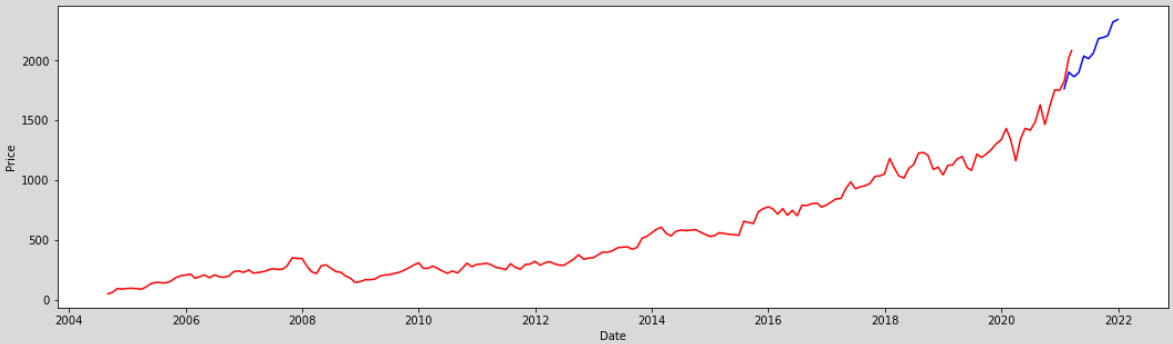
I got 4 different results from my forecasts with the corresponding MAPEs:

ARIMA (Python) - Annual
End of Year Forecast: $2,253.56
MAPE: 0.205049338

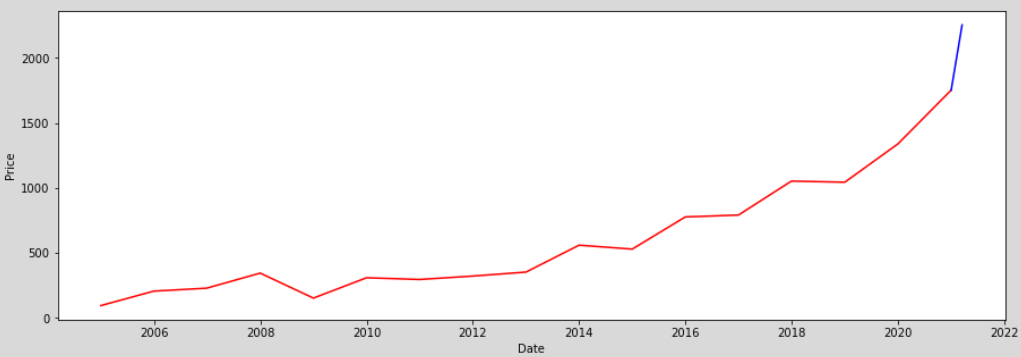
ARIMA (Python) - Quarterly
End of Year Forecast: $2,191.66
MAPE: 0.152041316

ARIMA (Python) - Monthly
End of Year Forecast: $2,342.87
MAPE: 0.039636115

ARIMA (Python) - Weekly
End of Year Forecast: $2,373.25
MAPE: 0.072101207



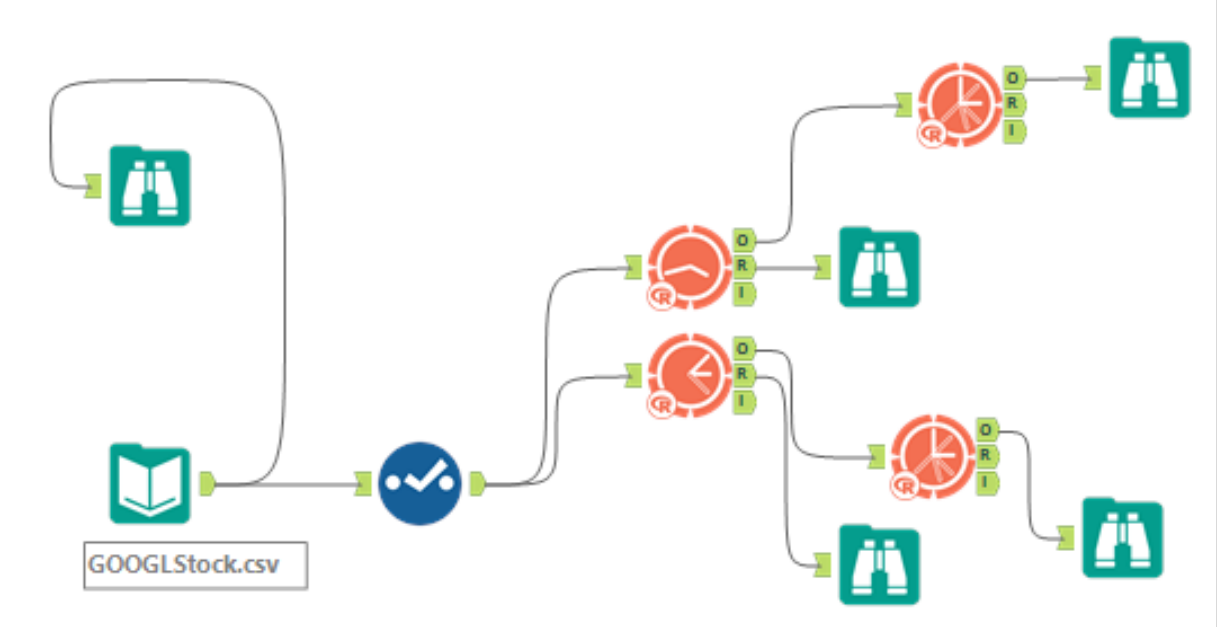
Monthly Forecast



Annual Forecast

Process: ARIMA and ETS Model (Alteryx)

After seeing my results for ARIMA on Python, I wanted to attempt running two-time series forecast models on daily data, while considering seasonality, and hence I chose to use Alteryx for that. For both models, I changed the different parameters and created the following workflow:



These were the forecasts I obtained:

ARIMA (Alteryx) - Daily

End of Year Forecast: $2,156.43

MAPE: 0.92927489

ETS (Alteryx) - Daily

End of Year Forecast: $2,297.67

MAPE: 0.92849588

Process: ETS Models (Excel)

Similar to Python, I ran an ETS Forecast Sheet Model using Excel and based on Annual, Quarterly and Weekly Granularity to get three more forecasts, but in this case I didn't edit any of the formula metrics offered by Excel.

These were the forecasts I obtained:

ETS (Excel) - Annual

End of Year Forecast: $2,070.63

MAPE: 0.145981156

ETS (Excel) - Monthly

End of Year Forecast: $2,478.88

MAPE: 0.153589774

ETS (Excel) - Quarterly

End of Year Forecast: $2,147.42

MAPE: 0.100830551

Process: Support Vector Regression Model (Python)

SVR Models are useful because they recongize patterns of time series datasets and incorporates them to create a regression.With this model, I created a moving average predictor and used those values from 207 days before (as 207 business days remaining in this year) in SVR Regressor with an extremely low gamma value to be able to predict the forecast 207 days from today. I did this by shifting the Adj Close column by 207 and then trained the model to predict the value for every 207 days after.

Model Forecast

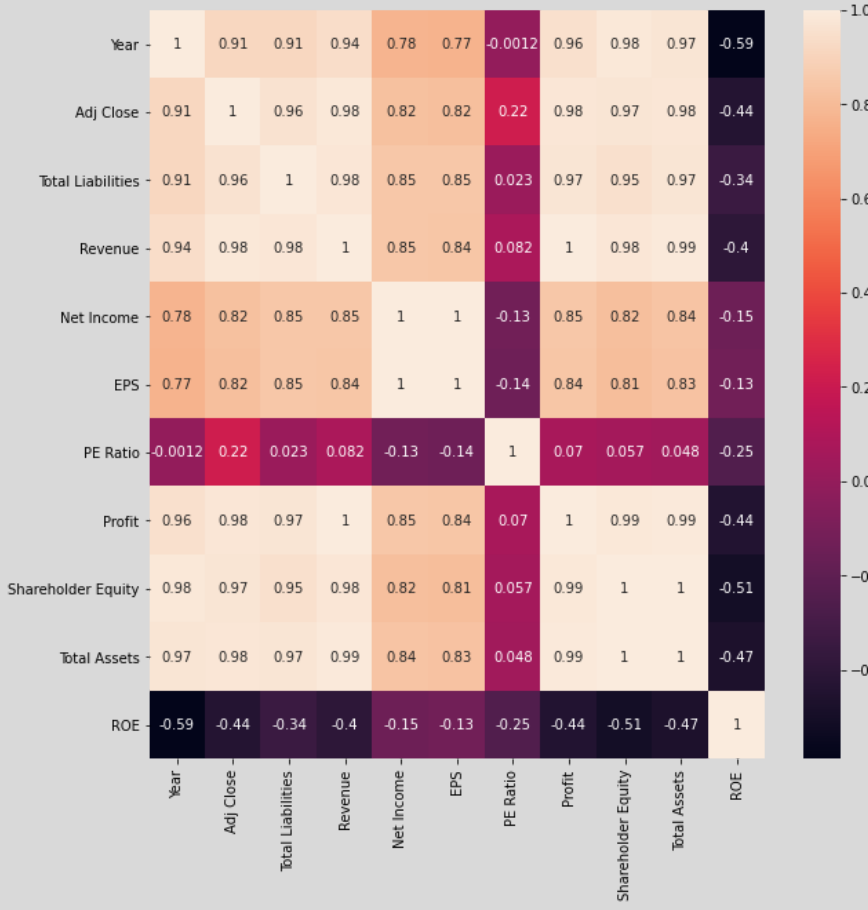
End of Year Forecast: $2355.19

MAPE: 0.168883126

SVR Model Forecast

Process: Linear Regression - 1 (Python)

Upon researching different variables which could possibly have an effect on the stock price of a company on the FRED Website, the data available was mainly quarterly, so in the model, I was predicting end-of-quarter prices. From the FRED Website, I got variables such as Quarter Net Income, P/E Ratio, Total Assets, Total Liabilities, Revenue, Profit, EPS, Return on Equity (ROE), etc, and created a heatmap showing correlation.



I put them all in a linear regression model to see which variables would be significant in predicting the stock price, and then I noticed there was a high degree of multicollinearity, so I combined some of the variables. I created a Share Price Variable (EPS * PE Ratio), Equity Return (Shareholders Equity * ROE).

The significant variables were in predicting the end of quarter stock price were: **Net Income, Profit, SharePrice, EquityReturn**

Process: Linear Regression - 1 (Python)

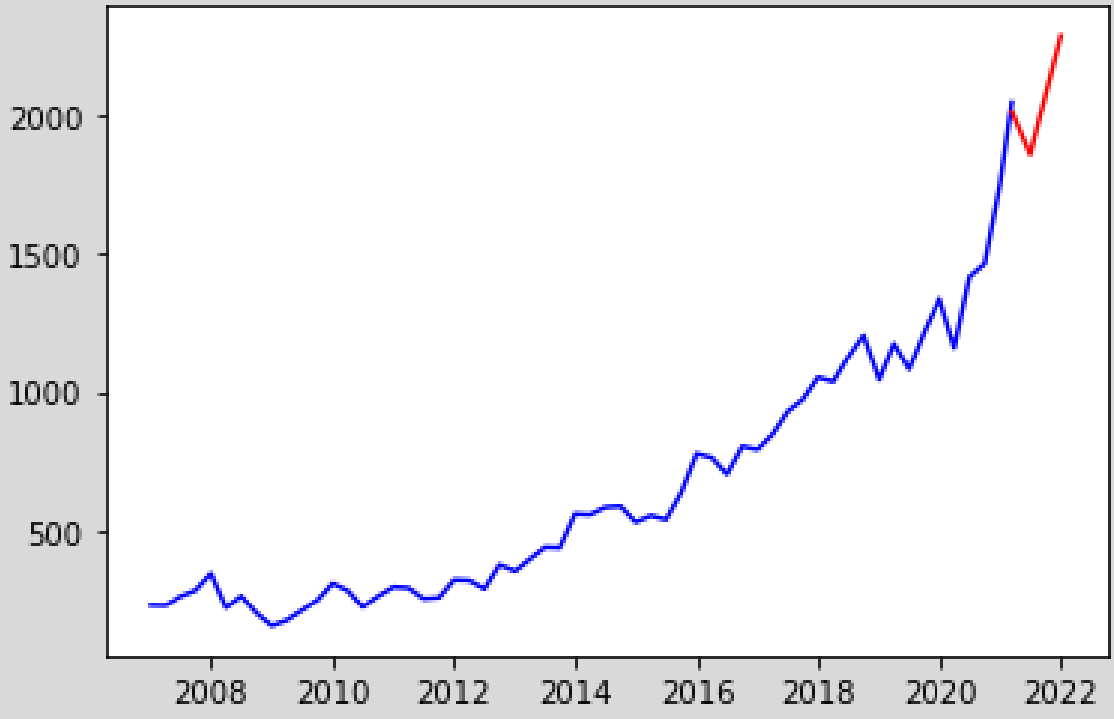
Regression Summary

	coef	std err	t	P>|t|	[0.025	0.975]
Net Income	-0.1033	0.016	-6.423	0.000	-0.136	-0.071
Profit	0.0271	0.007	4.012	0.000	0.013	0.041
SharePrice	2.0564	0.277	7.417	0.000	1.496	2.616
EquityReturn	0.0246	0.005	4.660	0.000	0.014	0.035

Model Forecast

End of Year Forecast: $2290.415

MAPE: 0.072389659



Forecast Visualization

Process: Linear Regression - 2 (Python)

In order to account for the market situation, to the previous regression model, I obtained values for GDP, Employment and Competitor Volumes and Prices. Since Volume represents how much a stock was traded from open to close, I believed that it would be a good indicator as to how a certain stock is related to its competitors. To remove multicollinearity and reduce the immense values of these competitor volumes, I multiplied the log values of the three competitors (Apple, Facebook, Microsoft) and created a new variable called CompVolume.

The significant variables were in predicting the end of quarter stock price for this new model were: **Net Income, Profit, SharePrice, EquityReturn, GDP, CompVolume**

Process: Linear Regression - 2 (Python)

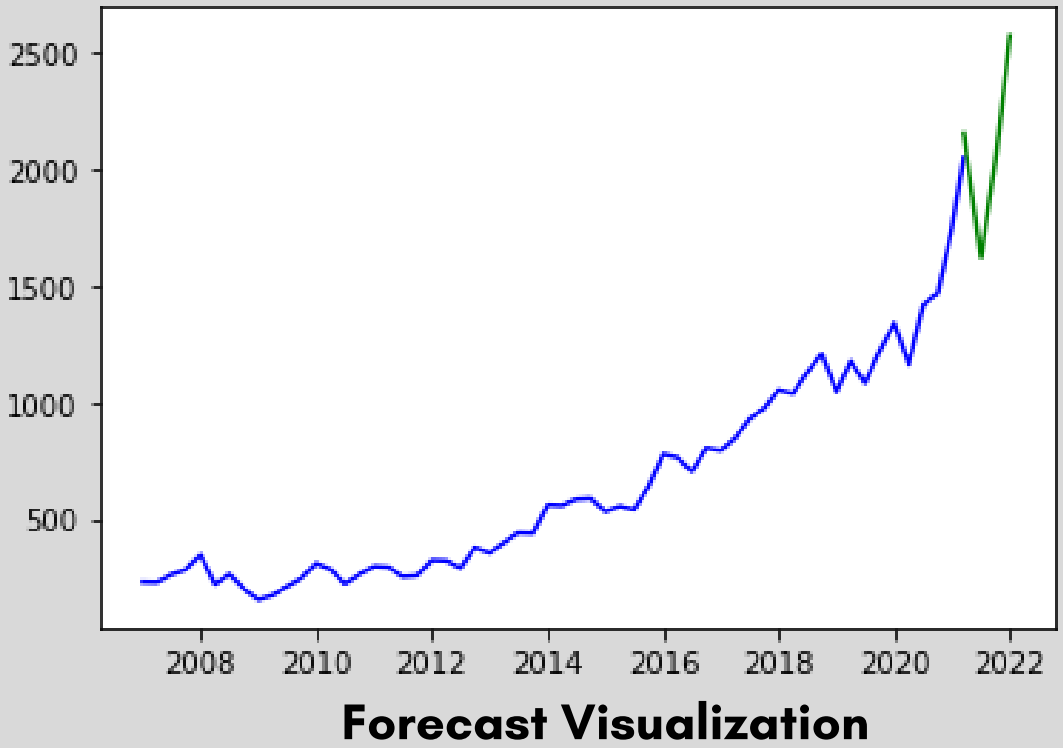
Regression Summary

	coef	std err	t	P>|t|	[0.025	0.975]
Net Income	-0.1185	0.018	-6.405	0.000	-0.157	-0.080
Profit	0.0656	0.007	9.711	0.000	0.052	0.080
SharePrice	1.6695	0.263	6.354	0.000	1.123	2.216
EquityReturn	0.0197	0.005	4.153	0.000	0.010	0.030
GDP	-0.0359	0.011	-3.387	0.003	-0.058	-0.014
CompVolume	0.0700	0.024	2.875	0.009	0.019	0.121

Model Forecast

End of Year Forecast: $2,572.20

MAPE: 0.044363559



Forecast Visualization

Ensemble Model

After creating all these models and obtaining forecasts, in order to obtain an ensemble forecast, I combined them on basis of weighted average, weighed by the MAPE given by the respective models. I did so on Excel.

ENSEMBLE MODEL (STOCKS FORECASTING PROJECT) COMPANY: ALPHABET INC. (GOOGL)					
No.	**Model**	**Forecast for Dec 31st, 2021**	**MAPE**	**1 - MAPE**	**Weight**
1	ARIMA (Alteryx) - Daily	$2,156.43	0.92927489	0.07072511	0.007864992
2	ARIMA (Python) - Annual	$2,253.56	0.205049338	0.794950662	0.088402554
3	ARIMA (Python) - Monthly	$2,342.87	0.039636115	0.960363885	0.106797345
4	ARIMA (Python) - Quarterly	$2,191.66	0.152041316	0.847958684	0.094297315
5	ARIMA (Python) - Weekly	$2,373.25	0.072101207	0.927898793	0.103187062
6	ETS (Alteryx) - Daily	$2,297.67	0.92849588	0.07150412	0.007951622
7	ETS (Excel) - Annual	$2,070.63	0.145981156	0.854018844	0.094971236
8	ETS (Excel) - Monthly	$2,478.88	0.153589774	0.846410226	0.094125119
9	ETS (Excel) - Quarterly	$2,147.42	0.100830551	0.899169449	0.099992213
10	Support Vector Regression Model (Python) - Moving	$2,355.20	0.163851804	0.836148196	0.092983929
11	Linear Regression 1 (Python) - Quarterly	$2,290.42	0.072389659	0.927610341	0.103154984
12	Linear Regression 2 (Python) - Quarterly	$2,572.20	0.044363559	0.955636441	0.106271629

Ensemble Model: Forecast

End of Year Forecast: $2,309.54

Growth: 10.83%

Future Improvements

- I was working on creating a Long Short Term Memory (LSTM) Model, but due to it being a very new topic to me, I didnt include it in my forecast because I am still in the process learning about it, so in the future, after gaining more knowledge, I would want to include that in my ensemble as well.
- For my regression at the moment, it only includes few competitors of Alphabet Inc., in the future I hope to incorporate more competitors to give a more accurate result, along with more defining economic metrics and variables.