PREDICTING WITH PROPHET

FACEBOOK STOCK PRICE & BITCOIN PRICE FORECASTING AND THEIR RELATIONSHIP *ANALYSIS*

Waner Li 873464868 COMP4448

INSTRUCTOR: DR. POORAN NEGI

Table of Contents

Questions, Dataset, Motivation

Questions & Motivation

As we all note the increasing interest in cryptocurrency, people including me begin to wonder how could we make wiser investment decisions on bitcoin and stock market? What factor should we take into account to leverage our investment decisions? If we have extra money, should we make investment on bitcoin or stock market?

As a future data scientist, I am going to use Prophet which is an open sourcing package developed by Facebook (FB) to forecast FB stock price and bitcoin price and further analyze its relationship with FB stock price.

Background Information

Bitcoin is a form of electronic ash which a decentralized digital currency without a central bank or single administrator that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries.

Stocks are a type of security that gives stockholders a share of ownership in a company. The stock market works through a network of exchanges. Companies list shares of their stock on an exchange to raise money to grow their business and investors purchase those shares. Investors could make profit/loss out of selling their shares

Dataset

- Facebook historical stock price from 05/18/2012 to 03/27/2018
  This dataset is stored in the Stocker object as a data frame with 16 variables

	Date	Open	High	Low	Close	Volume	Ex-Dividend	Split Ratio	Adj. Open	Adj. High	Adj. Low	Adj. Close	Adj. Volume	ds	y	Daily Change
0	2012-05-18	42.05	45.00	38.00	38.2318	573576400.0	0.0	1.0	42.05	45.00	38.00	38.2318	573576400.0	2012-05-18	38.2318	-3.8182
1	2012-05-21	36.53	36.66	33.00	34.0300	168192700.0	0.0	1.0	36.53	36.66	33.00	34.0300	168192700.0	2012-05-21	34.0300	-2.5000
2	2012-05-22	32.61	33.59	30.94	31.0000	101786600.0	0.0	1.0	32.61	33.59	30.94	31.0000	101786600.0	2012-05-22	31.0000	-1.6100
3	2012-05-23	31.37	32.50	31.36	32.0000	73600000.0	0.0	1.0	31.37	32.50	31.36	32.0000	73600000.0	2012-05-23	32.0000	0.6300
4	2012-05-24	32.95	33.21	31.77	33.0300	50237200.0	0.0	1.0	32.95	33.21	31.77	33.0300	50237200.0	2012-05-24	33.0300	0.0800

- Bitcoin historical price from 01/17/2014 to 06/02/2019
  This dataset is downloaded from Quandl and loaded as a data frame with 7 variables and Date as index column.

	Open	High	Low	Close	Volume (BTC)	Volume (Currency)	Weighted Price
Date							
2014-01-07	874.67040	892.06753	810.00000	810.00000	15.622378	13151.472844	841.835522
2014-01-08	810.00000	899.84281	788.00000	824.98287	19.182756	16097.329584	839.156269
2014-01-09	825.56345	870.00000	807.42084	841.86934	8.158335	6784.249982	831.572913
2014-01-10	839.99000	857.34056	817.00000	857.33056	8.024510	6780.220188	844.938794
2014-01-11	858.20000	918.05471	857.16554	899.84105	18.748285	16698.566929	890.671709

Data Cleaning

Data Types and Nulls/NAs

We need to correct data types and clean nulls to process further analyze properly.

Facebook data types

```
FB_history.dtypes
```

```
Date              datetime64[ns]
Open                     float64
High                     float64
Low                      float64
Close                    float64
Volume                   float64
Ex-Dividend              float64
Split Ratio              float64
Adj. Open                float64
Adj. High                float64
Adj. Low                 float64
Adj. Close               float64
Adj. Volume              float64
ds                datetime64[ns]
y                        float64
Daily Change             float64
dtype: object
```

Facebook Nulls/NAs

```
FB_history.isna().any()
```

```
Date              False
Open              False
High              False
Low               False
Close             False
Volume            False
Ex-Dividend       False
Split Ratio       False
Adj. Open         False
Adj. High         False
Adj. Low          False
Adj. Close        False
Adj. Volume       False
ds                False
y                 False
Daily Change      False
dtype: bool
```

Bitcoin data types

```
bit.dtypes
```

```
Open                   float64
High                   float64
Low                    float64
Close                  float64
Volume (BTC)           float64
Volume (Currency)      float64
Weighted Price         float64
dtype: object
```

Bitcoin Nulls/NAs

```
bit.isna().any()
```

```
Open                  False
High                  False
Low                   False
Close                 False
Volume (BTC)          False
Volume (Currency)     False
Weighted Price        False
dtype: bool
```

We can see that the data types of both datasets are correct and ready to process. Luckily, the datasets I am using for this project don't have any missing values so no further data cleansing is necessary.

For this project, I only use "Date" "Adj. Close" "Daily Change" variables in Facebook dataset and "Date" "Weighted Price" variables in Bitcoin dataset. The following statistics help us better understand prices for both Facebook stock and Bitcoin.

```
FB_history['Adj. Close'].describe()
```

```
count    1472.000000
mean       89.482903
std        48.536888
min        17.729000
25%        49.167250
50%        80.902500
75%       123.902500
max       193.090000
Name: Adj. Close, dtype: float64
```

```python
bit['Weighted Price'].describe()
```

```
count     1973.000000
mean      2729.844882
std       3447.188387
min          0.000000
25%        402.917501
50%        691.393590
75%       4296.631535
max      19135.469160
Name: Weighted Price, dtype: float64
```

Modeling activity and Evaluation

Hypothesis and Independent Variables

H_0: There is no relationship between Facebook stock price and Bitcoin price;

H_1: There is relationship between Facebook stock price and Bitcoin price.

Independent Variables:

- Date
- Bitcoin price % difference calculated by “Weighted Price” and “Prior Price”
- Facebook price % difference calculated by “Adj. Close” and “Prior Price”

```
bit2 = bit_df[['Weighted Price']]
bit2['Prior Price'] = bit2['Weighted Price'].shift(1)
bit2['Bitcoin Diff Pct'] = ((bit2['Weighted Price'] - bit2['Prior Price'])/bit2['Prior Price'])
bit2.tail()
```

Date	Weighted Price	Prior Price	Bitcoin Diff Pct
2019-05-29	8625.220359	8668.153528	-0.004953
2019-05-30	8541.545609	8625.220359	-0.009701
2019-05-31	8331.197122	8541.545609	-0.024627
2019-06-01	8540.000449	8331.197122	0.025063
2019-06-02	8695.634771	8540.000449	0.018224

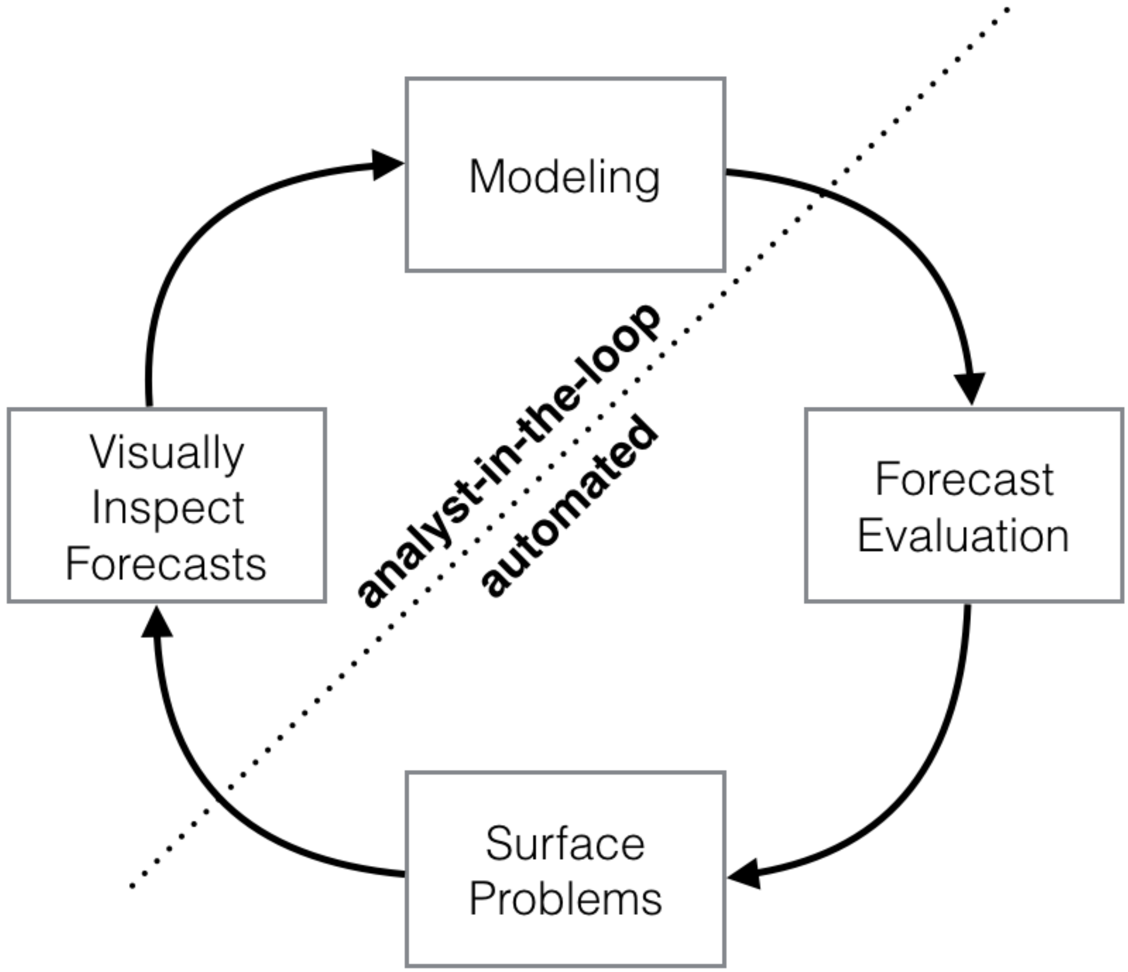
```
FB2 = FB_history
FB2.index = FB_history['Date']
FB2 = FB2[['Adj. Close']]
FB2['Prior Price'] = FB2['Adj. Close'].shift(1)
FB2['FB Diff Pct'] = ((FB2['Adj. Close'] - FB2['Prior Price'])/FB2['Prior Price'])
FB2.head()
```

Date	Adj. Close	Prior Price	FB Diff Pct
2012-05-18	38.2318	NaN	NaN
2012-05-21	34.0300	38.2318	-0.109903
2012-05-22	31.0000	34.0300	-0.089039
2012-05-23	32.0000	31.0000	0.032258
2012-05-24	33.0300	32.0000	0.032188

Model Selection and Evaluation Activities

All the analysis and prediction for Stocker is done using additive models with the [Prophet package developed by Facebook](). Additive models consider a time series an additive combination of an overall trend along with seasonal patterns on different time scales such as daily, weekly, or monthly.

We can evaluate Prophet perform by comparing observations with predications to see if real data could fall into it confidence interval width.



Model parameter/coefficient interpretation and their error analysis

Prophet utilizes the additive regression model y(t) comprising the following components:

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

where:

- Trend $g(t)$ models non-periodic changes;

- Seasonality $s(t)$ represents periodic changes;

- Holidays component $h(t)$ contributes information about holidays and events.

- $\epsilon(t)$ represents information that was not reflected in the model. Usually it is modeled as normally distributed noise.

Analysis, Visualization and Conclusion

Facebook Stock Price Overall Trend

It’s really important for investors to see the overall trend of price of your target stock. If you are a risk-averse investor, you should avoid those stocks with dramatic drop and rise. We can tell from the following graph that Facebook stock perform well in the past 6 years.

```
1  FB.plot_stock()
```

```
Maximum Adj. Close = 193.09 on 2018-02-01 00:00:00.
Minimum Adj. Close = 17.73 on 2012-09-04 00:00:00.
Current Adj. Close = 152.19 on 2018-03-27 00:00:00.
```

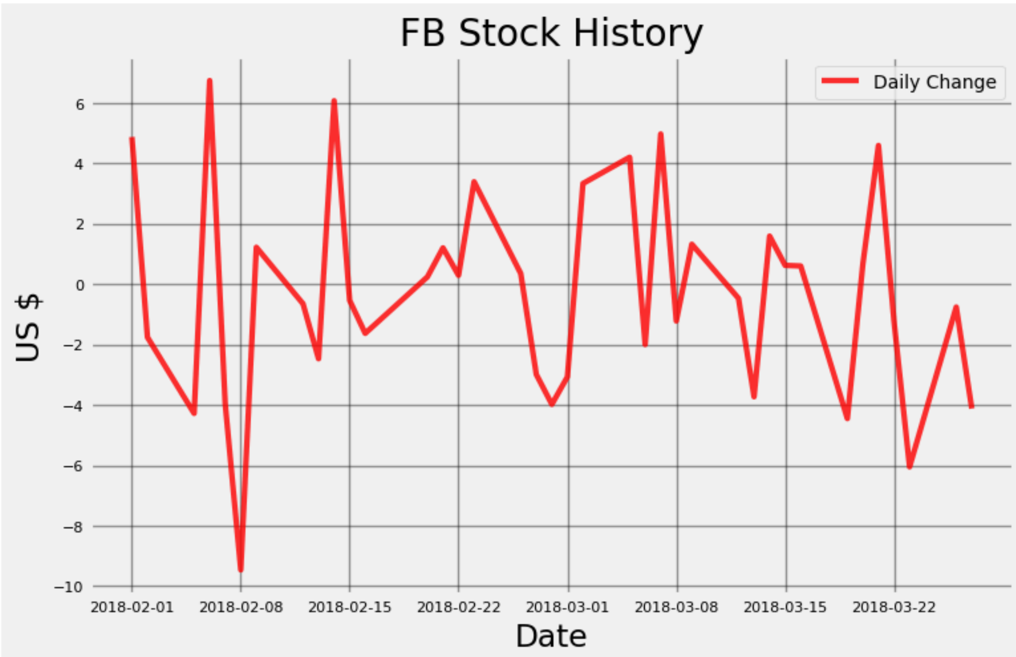
Facebook-Cambridge Analytica Data Scandal

The Facebook-Cambridge Analytica data scandal was a major political scandal which was erupted in March 2018 by a whistleblower. It was revealed that Cambridge Analytica had harvested the personal data of millions of people's Facebook profiles without their consent and used if for political advertising purpose.

I narrowed down the timeline to 02/01/2018 to 03/27/2018 to analyze how Facebook-Cambridge Analytica data scandal affected Facebook stock price. We can see it was knocked off around $4/share on March 17, 2018 but bumped up the very next day.

```
FB.plot_stock(start_date = "2018-02-01", end_date = "2018-03-27",
              stats=['Daily Change'])
```

```
Maximum Daily Change = 6.74 on 2018-02-06 00:00:00.
Minimum Daily Change = -9.46 on 2018-02-08 00:00:00.
Current Daily Change = -4.12 on 2018-03-27 00:00:00.
```

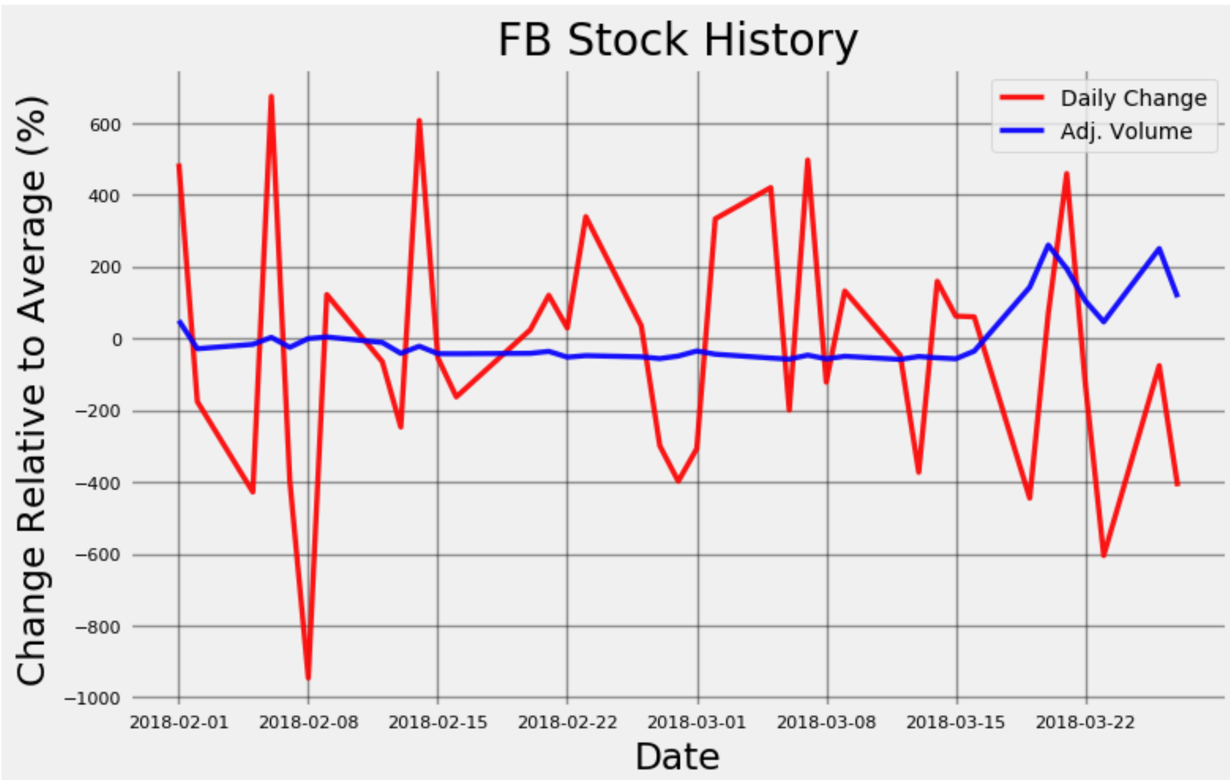


Change Relative to Average % graph also shows that this scandal didn't affect Facebook stock price much in the long-term. The fluctuate is more like a normal float of price when it compares to other timeframes.

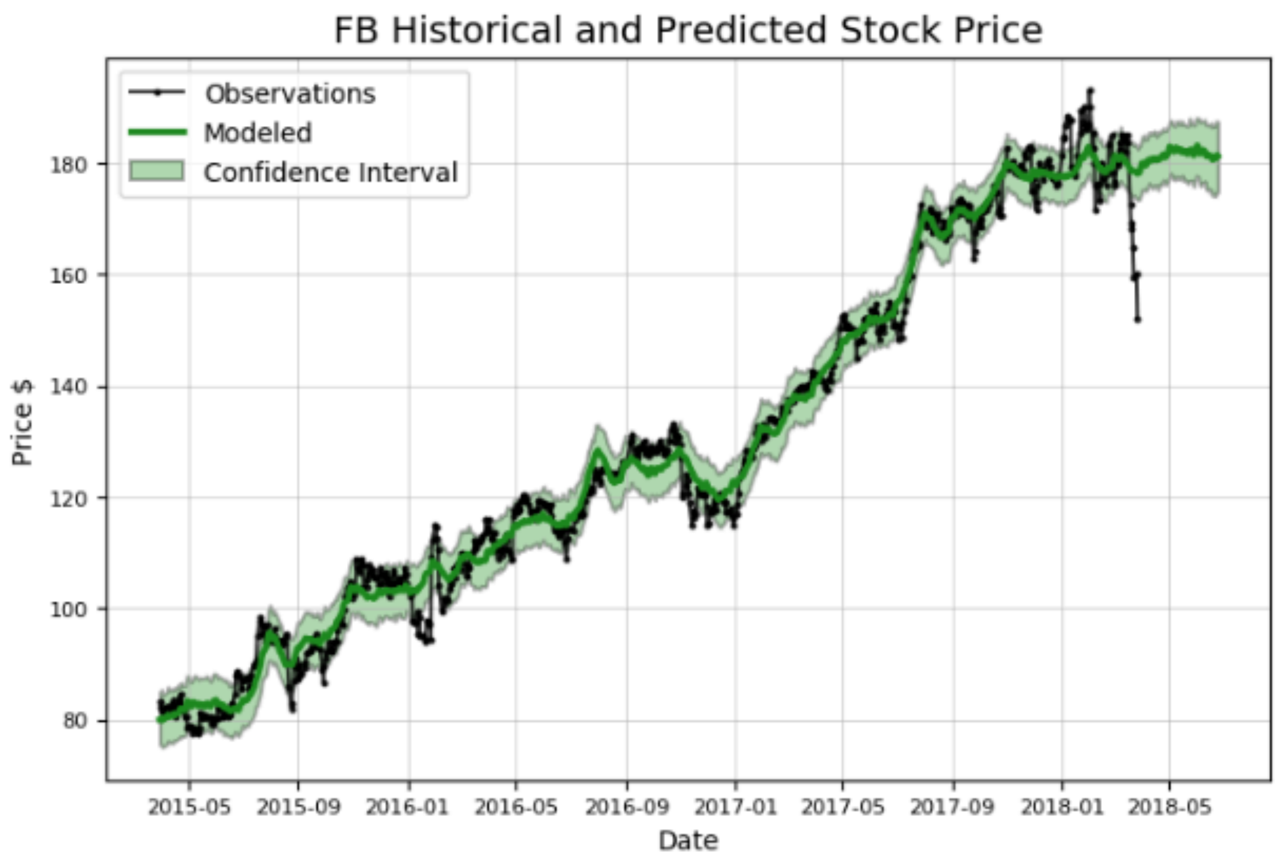
```
FB.plot_stock(start_date = "2018-02-01", end_date = "2018-03-27",
              stats=['Daily Change', 'Adj. Volume'], plot_type='pct')
```

```
Maximum Daily Change = 6.74 on 2018-02-06 00:00:00.
Minimum Daily Change = -9.46 on 2018-02-08 00:00:00.
Current Daily Change = -4.12 on 2018-03-27 00:00:00.

Maximum Adj. Volume = 128925534.00 on 2018-03-20 00:00:00.
Minimum Adj. Volume = 14873538.00 on 2018-03-12 00:00:00.
Current Adj. Volume = 76787884.00 on 2018-03-27 00:00:00.
```



Facebook Stock Price Forecast

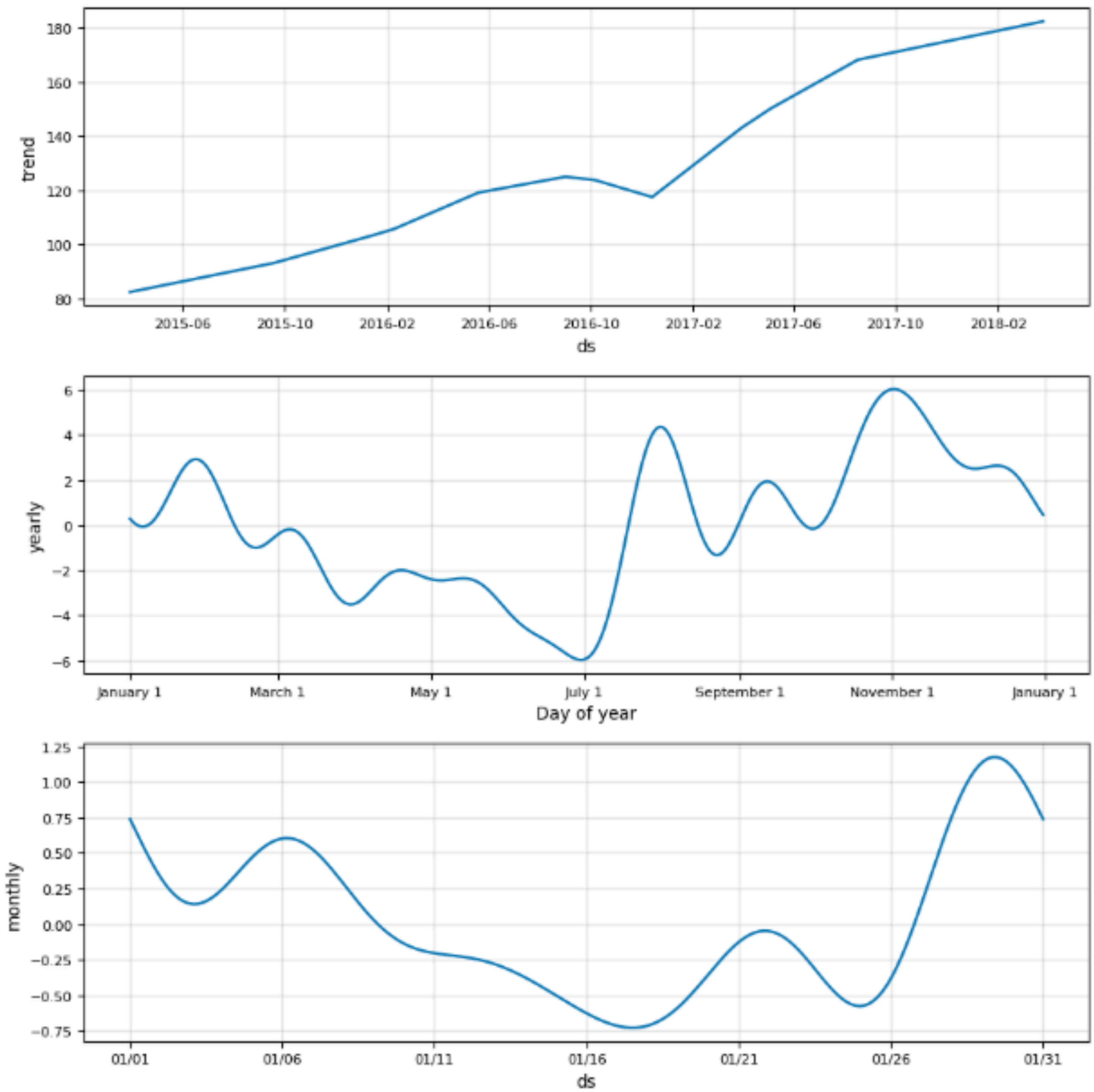


The green line, predication, contains a 80% confidence interval. This represents the model's uncertainty in the forecast. We can tell from the graph the FB stock price will most likely increase steadily in 90 days.

We can see our model failed the market since the observations in March to April 2018 are out of confidence interval. Please keep in mind that Prophet is an additive model which means it will perform better at capturing general movements over a long period rather than daily fluctuations. It will not be a good decision using Prophet as a guidance to play daily market.

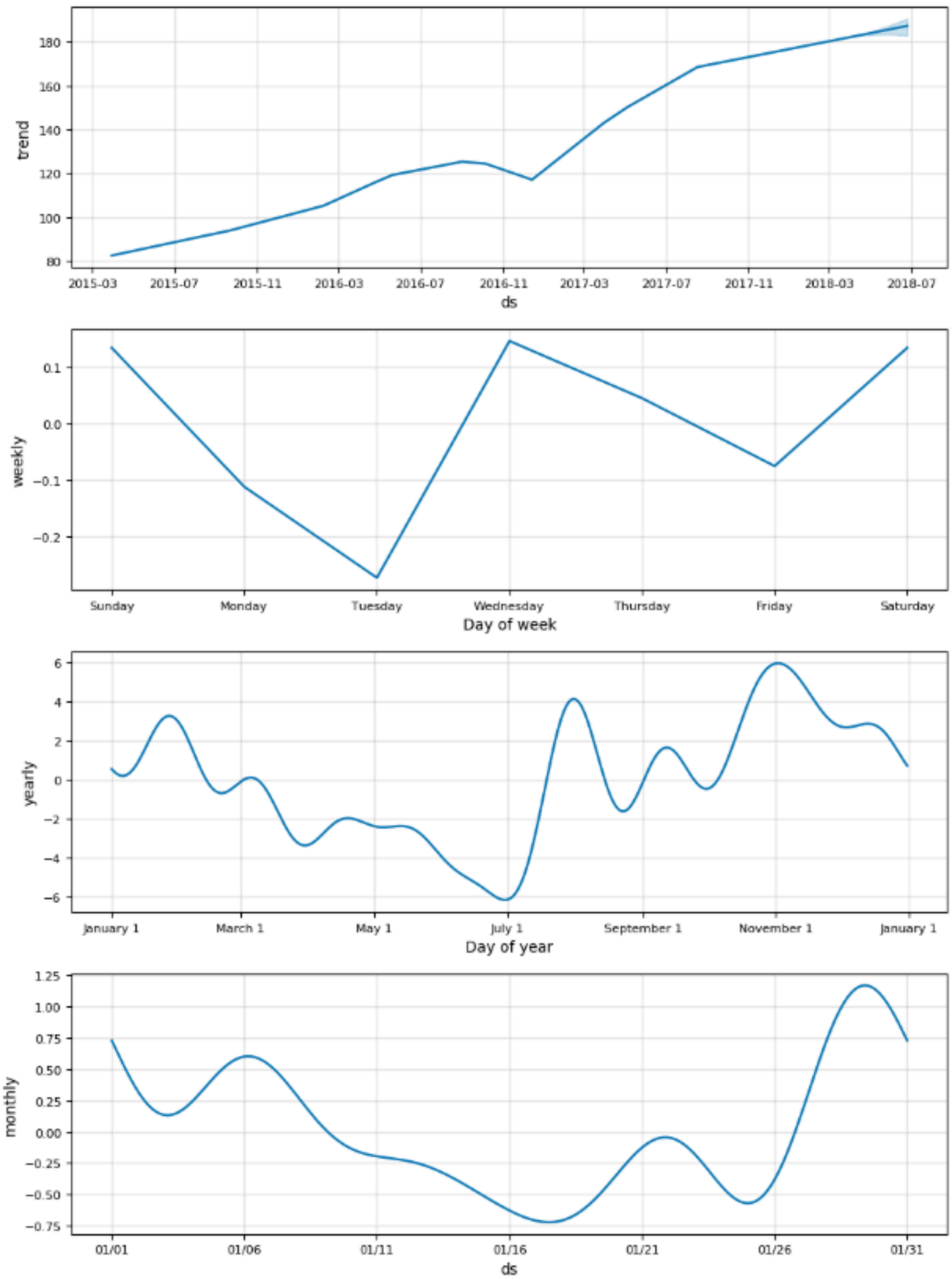
Facebook Stock Price Seasonality Analysis

```
# Variables assigned from previous method call
model.plot_components(model_data)
plt.show()
```



The overall trend is clearly in the upwards direction over the past three year and over the course of a year it appears to be bottomed out in late June and early July. As the time scale decreases, the patterns grow more noisy. The monthly pattern appears to be slightly random, and I would not advise you to make investment during 4nd-6th and 26th-29th of the month.

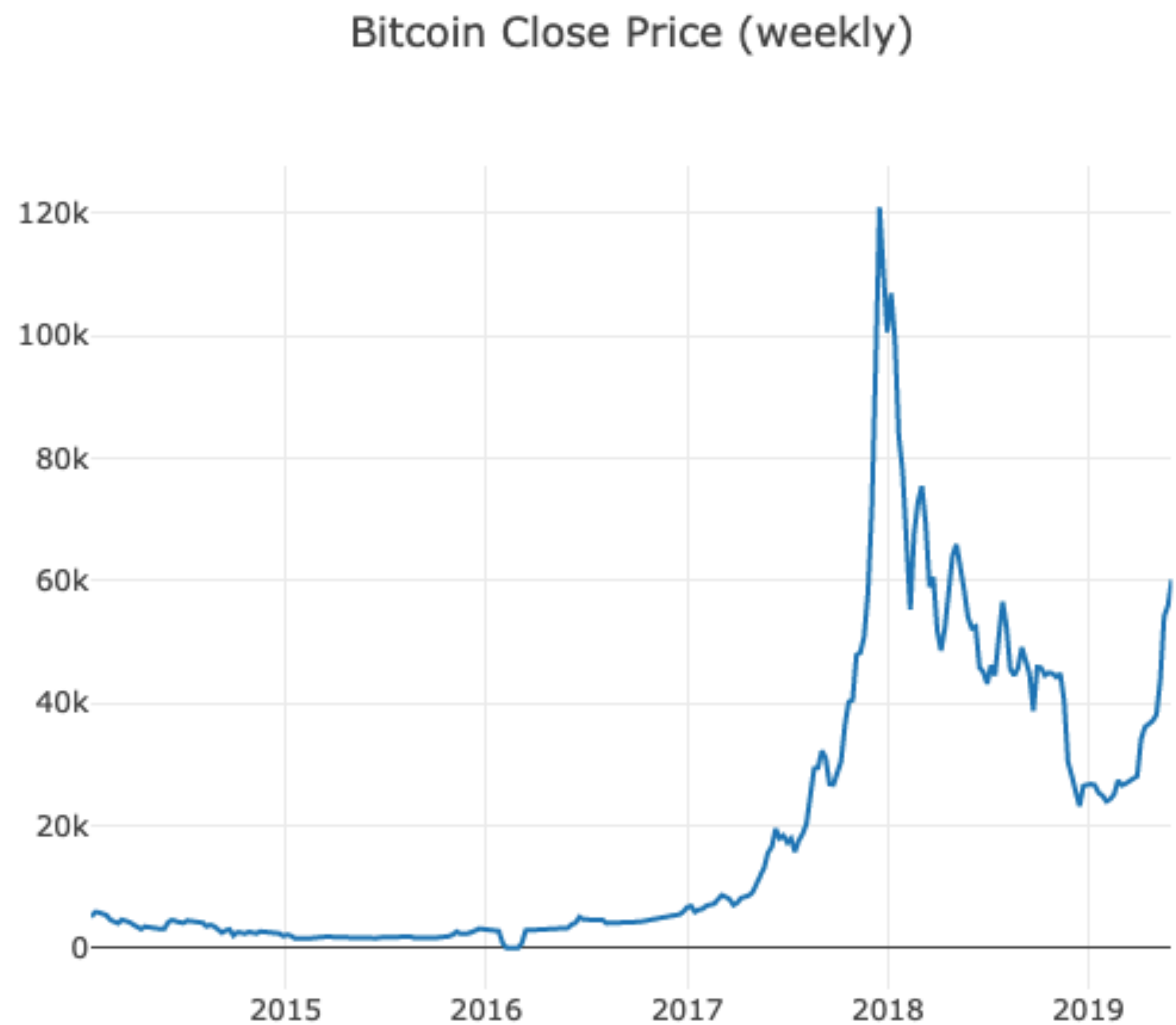
```
FB.weekly_seasonality=True
model.plot_components(model_data)
plt.show()
```



We can ignore the weekends because market is closed during weekends. We can see a pattern that price drops on Tuesday and Friday which means it might be profitable to buy FB stock on Tuesday or Friday and sell on Wednesday. Since the we can see week seasonality affects stock price, I will keep it to do further prediction.

Bitcoin Price Overall Trend

Since bitcoin dataset is over 5 years, it is hard to get anything meaningful from this line graph. To reduce the noise, I resampled the dataset to weekly bins.



Bitcoin Price Forecast

To measure the quality of my forecast, I split bitcoin price dataset into train part and test parts, then pass the parameters to Prophet and train the dataset by invoking its fit method.

```python
from fbprophet import Prophet
import logging
logging.getLogger().setLevel(logging.ERROR)
```

```python
bit = bit.reset_index()
bit.columns = ['ds', 'y']
```

```python
prediction_size = 30
train = bit[:-prediction_size]
train.tail(n=3)
```

	ds	y
1940	2019-05-01	5309.527703
1941	2019-05-02	5373.034059
1942	2019-05-03	5641.844757

```python
pro = Prophet()
pro.fit(train)
```

```
<fbprophet.forecaster.Prophet at 0x1c2596c160>
```

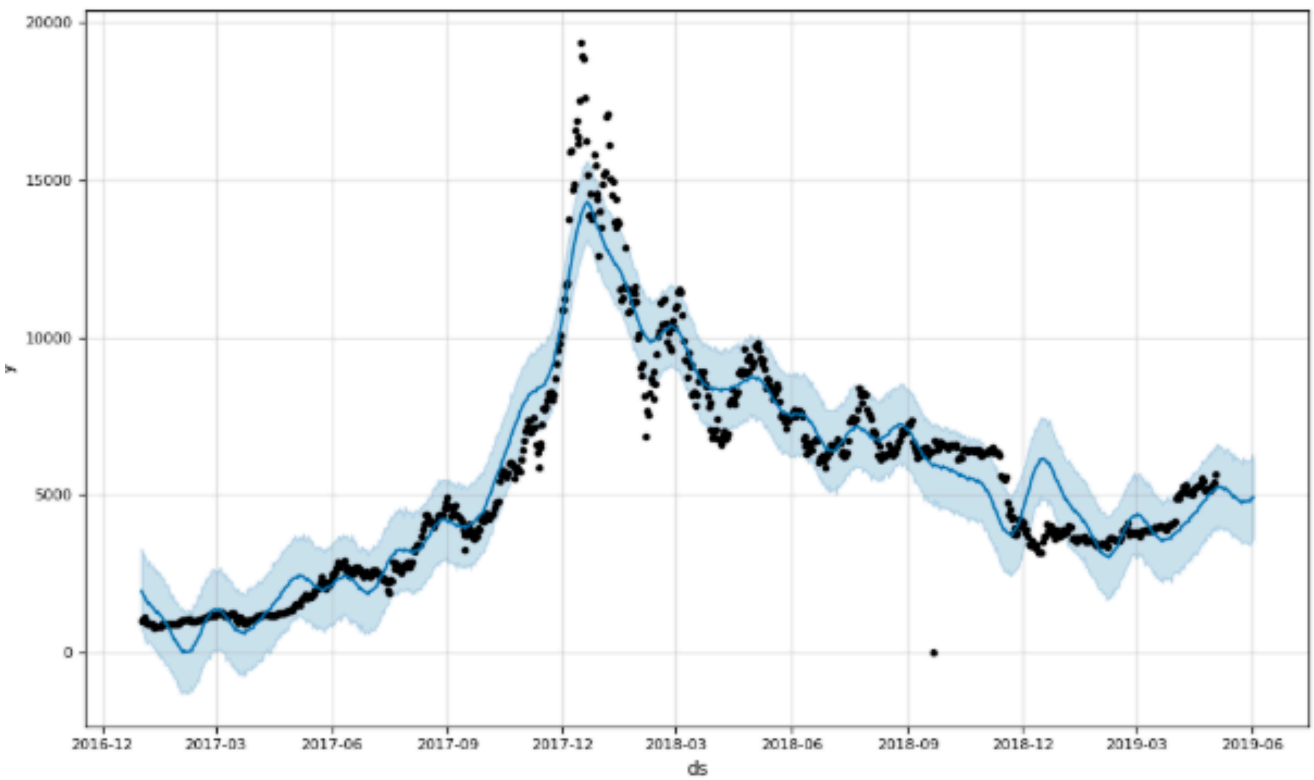
```python
future = pro.make_future_dataframe(periods=prediction_size)
future.tail(n=3)
```

	ds
1970	2019-05-31
1971	2019-06-01
1972	2019-06-02

```python
forecast = pro.predict(future)
forecast.head(n=3)
```

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive
0	2017-01-01	1425.530445	711.954403	3284.539484	1425.530445	1425.530445	524.836393	
1	2017-01-02	1431.748659	694.757121	3169.843290	1431.748659	1431.748659	446.057778	
2	2017-01-03	1437.966873	544.304853	3117.204929	1437.966873	1437.966873	371.752611	

Prophet returned lots of columns including tread and seasonality components, I only need to use yhat column since it forecast itself.

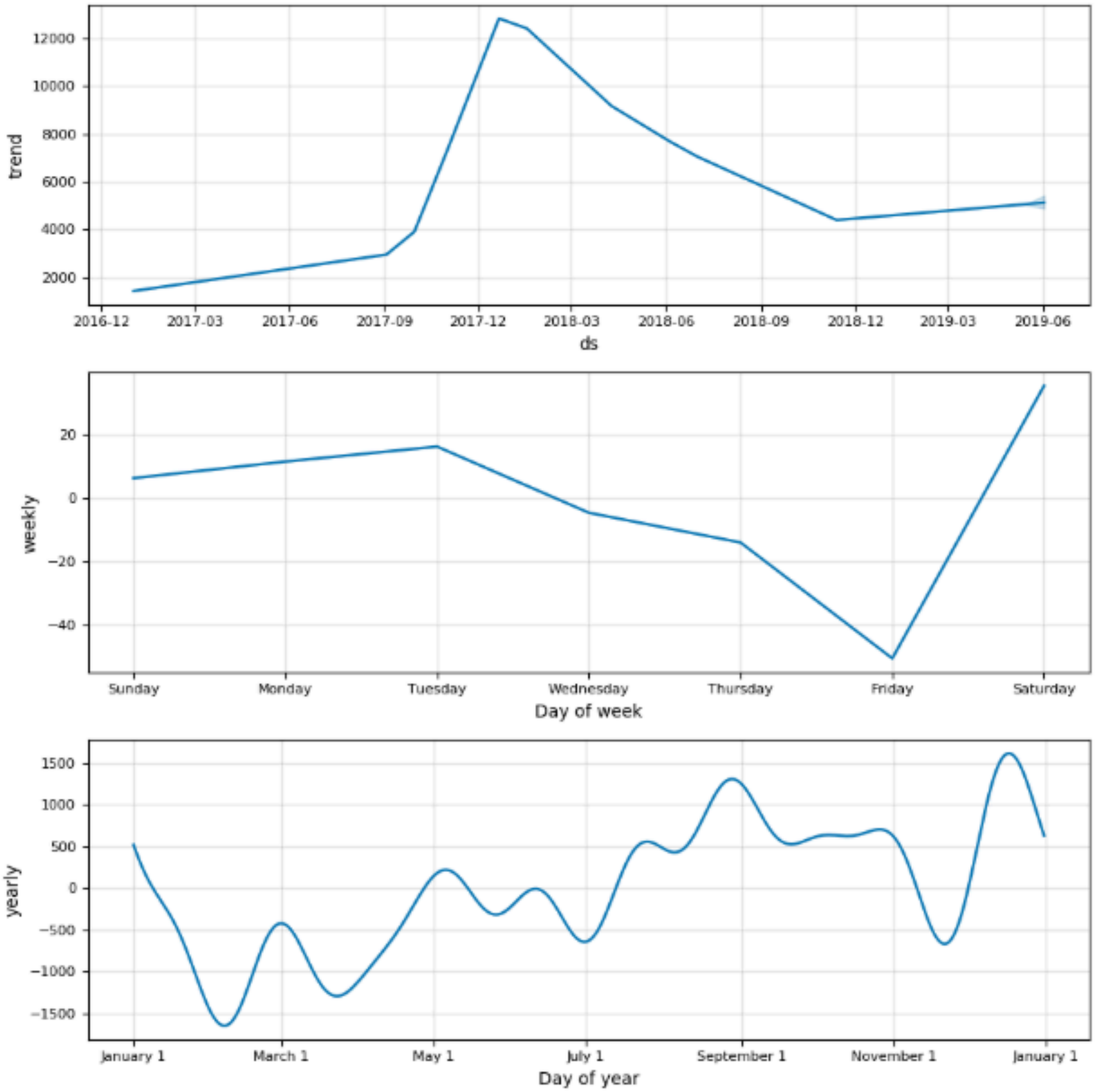


This graph is not informative and the only definitive conclusion that we can draw here is that the model treated many of the data points as outliers as bitcoin price was ridiculous high all of sudden around early 2018.

Bitcoin Price Seasonality Analysis

Let's dive into yearly and weekly seasonality. We can tell from yearly and weekly seasonality that it's better to invest on bitcoin on Friday in February and April.

```
1  pro.plot_components(forecast)
```



Relationship between Facebook Stock Price and Bitcoin Price Analysis

After calculating Price Difference% for both dataset, we need to inner join two datasets per dates

```
bit2 = bit_df[['Weighted Price']]
bit2['Prior Price'] = bit2['Weighted Price'].shift(1)
bit2['Bitcoin Diff Pct'] = ((bit2['Weighted Price'] - bit2['Prior Price'])/bit2['Prior Price'])
bit2.tail()
```

Date	Weighted Price	Prior Price	Bitcoin Diff Pct
2019-05-29	8625.220359	8668.153528	-0.004953
2019-05-30	8541.545609	8625.220359	-0.009701
2019-05-31	8331.197122	8541.545609	-0.024627
2019-06-01	8540.000449	8331.197122	0.025063
2019-06-02	8695.634771	8540.000449	0.018224

```
FB2 = FB_history
FB2.index = FB_history['Date']
FB2 = FB2[['Adj. Close']]
FB2['Prior Price'] = FB2['Adj. Close'].shift(1)
FB2['FB Diff Pct'] = ((FB2['Adj. Close'] - FB2['Prior Price'])/FB2['Prior Price'])
FB2.head()
```

Date	Adj. Close	Prior Price	FB Diff Pct
2012-05-18	38.2318	NaN	NaN
2012-05-21	34.0300	38.2318	-0.109903
2012-05-22	31.0000	34.0300	-0.089039
2012-05-23	32.0000	31.0000	0.032258
2012-05-24	33.0300	32.0000	0.032188

Clean joined dataset

```
merge = bit2.merge(FB2,left_index=True, right_index=True)
merge = merge[['Bitcoin Diff Pct','FB Diff Pct']]
merge.isna().any()
```

```
Bitcoin Diff Pct     True
FB Diff Pct         False
dtype: bool
```

```
np.count_nonzero(merge.isnull())
```

```
0
```

```
merge = merge.dropna(subset=['Bitcoin Diff Pct'])
merge.isna().any()
```

```
Bitcoin Diff Pct    False
FB Diff Pct         False
dtype: bool
```

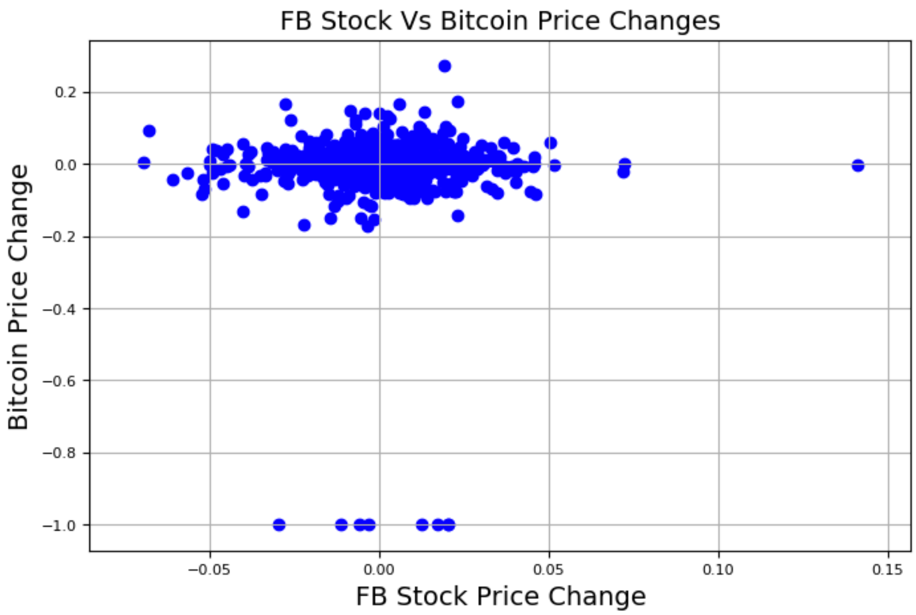
```
merge.count()
```

```
Bitcoin Diff Pct    1035
FB Diff Pct         1035
dtype: int64
```

Since there are only 27 missing values in “Bitcoin Diff Pct” column, I remove them from further analysis. Again, check if drop NAs performs successfully.

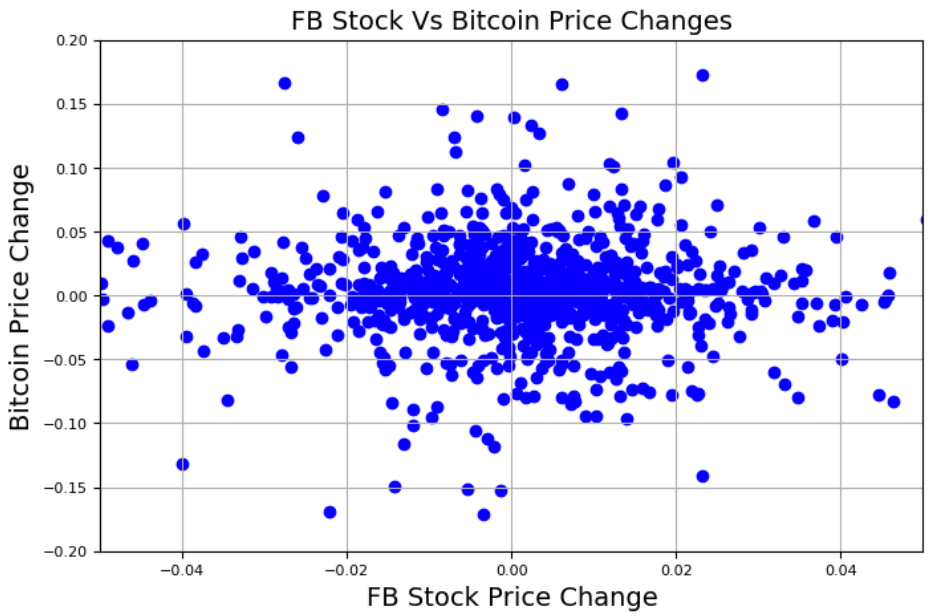
Scatter Plot Bitcoin Price Change % and Facebook Stock Price Change %

```
axes = plt.gca()
plt.scatter(merge['FB Diff Pct'], merge['Bitcoin Diff Pct'], color='blue')
plt.title('FB Stock Vs Bitcoin Price Changes', fontsize=14)
plt.xlabel('FB Stock Price Change', fontsize=14)
plt.ylabel('Bitcoin Price Change', fontsize=14)
plt.plot(np.unique(merge['FB Diff Pct']), np.poly1d(np.polyfit(merge['FB Diff Pct'], merge['Bitcoin Diff Pct'], 1))
         (np.unique(merge['FB Diff Pct'])))
plt.grid(True)
plt.show()
```



We can see some outliers in our scatter plot. Since they won't affect much on our result, we could zoon in to take a closer look at their relationship by reset axis range.

```python
# Remove outliners
axes = plt.gca()
axes.set_xlim([-0.05,0.05])
axes.set_ylim([-0.2,0.2])
plt.scatter(merge['FB Diff Pct'], merge['Bitcoin Diff Pct'], color='blue')
plt.title('FB Stock Vs Bitcoin Price Changes', fontsize=14)
plt.xlabel('FB Stock Price Change', fontsize=14)
plt.ylabel('Bitcoin Price Change', fontsize=14)
plt.plot(np.unique(merge['FB Diff Pct']), np.poly1d(np.polyfit(merge['FB Diff Pct'], merge['Bitcoin Diff Pct'], 1))
plt.grid(True)
plt.show()
```



We can see from the scatter plot that there is no significant relationship between FB stock price changes and Bitcoin Price Changes.

Disclaimer

This project and the information contained herein is not intended to be a source of advice or credit analysis with respect to the material presented, and the information and/or documents contained in this report do not constitute investment advice.

Reference

1. Prophet: forecasting at scale: https://research.fb.com/prophet-forecasting-at-scale/

2. Facebook-Cambridge Analytica Data Scandal:
https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

3. Prophet repository: https://github.com/facebook/prophet

4. Open Machine Learning Course. Topic 9. Part 2. Predicting the future with Facebook Prophet

https://medium.com/open-machine-learning-course/open-machine-learning-course-topic-9-part-3-predicting-the-future-with-facebook-prophet-3f3af145cdc