Stock Market Analysis

3250 Group Report

By: Jayoti, Ali, Hunain, Hussam, Krupal

April 2020

Table of Contents

1. Introduction 2
2. Data Preparation 2
3. Analysis of AAPL/BB Stocks 3
4. Impact of Covid-19 on the Stock Market 5
5. Analysis of Dow Jones Industrial Average Stocks (30 stocks) 7
6. Yearly Return for Dow Jones Stocks 9
7. Stock Market Forecasting 11

Conclusions 13

Appendix 15

1. Introduction

Financial data is ubiquitous in the world and each day generates gigabytes of data that can be used for analysis and forecasting. For the term project we will be using financial market data for stocks to gain insight into historical trends that have taken place in the markets. First, we describe the data preparation procedure including the steps taken to extract financial data from the web-resource; then, as a bench mark, we perform time series data analysis on the Apple (AAPL) and Blackberry (BB) stocks to describe different trends in their stocks. Then, we conduct a statistical analysis of the 30 individual stocks comprising the Dow Jones Industrial Average (DJIA) and present the yearly investment return for 30 stocks used to calculate DJIA. Then, a recurrent neural network is developed to attempt to forecast the future stock price. Finally, the concluding remarks are presented in the conclusions.

2. Data Preparation

The Yahoo Finance Application Programming Interface (API) was used to acquire market data for over 3,000 stocks trading on various American stock exchanges dating back to January 1st, 1970. The data consists of the daily high, low, open, close and volume for these stocks. The data was of high quality with minimal to no errors and there was minimal preprocessing required to get the data ready for analysis. The procedure of getting date for data analysis included following steps:

1. Acquire a list of stock symbols for companies trading on the New York Stock Exchange (NYSE) and NASDAQ stock exchanges with a market capitalization over $1 billion and an average daily volume of 1 million shares (consisting of over 3,000 stocks).
2. Use the Yahoo Finance API to get historical financial market data in JSON format using the requests library. Loop through the list of stock symbols in step 1 to get the complete dataset.
3. Parse the JSON response for each stock and use pandas DataFrame to store the data in memory.
4. Save the data from DataFrame to csv file for future analysis.

Challenges we faced during the procurement of the data included finding a reliable data source that provided both current and historical data going back as far as possible on a wide range of stocks. We tried multiple APIs such as AlphaVantage, IEX and Quandl to procure the data however we ran into limitations such as maximum requests, limited historical data or a limited selection of stocks. We tried web scraping the data however the process was too time confusing and riddled with errors. We eventually found our way to the Yahoo Finance API which provided us with the data we required in an easy to parse format in the form of JSON response. A sample of the dataset can be found in the Appendix (Table A1). The code used to procure the data can be found in the file 'YahooFinanceAPI.ipynb'. Python libraries utilized to procure the data includes: requests, json, pandas, and time.

3. Analysis of AAPL/BB Stocks

Having the data prepared for all the stocks, in this section, we conducted a comprehensive analysis on the stock market historical data to shed light on the factors affecting the price. Two stocks of Apple (APPL) and Blackberry (BB) were chosen as a bench mark and we have followed their variations during time. However, similar analysis can be done for other stocks.

To perform a comparative analysis on the stock data of the APPL and BB, we used the AAPL.csv file which contains the apple stock data and BB.csv file which consists of the Blackberry data. We created another dataset named iPhoneData2.csv under which we included the iPhone model release dates. The reason is that the stock value of Apple skyrocketed after the release of a new iPhone every year. This trend will continue in the subsequent years. We merged datasets of Apple and iPhonedata into one, to have the information of the iPhone model along with the stock price in our time series, (see Table A2). As a result, a new dataset apple formed with the model column from the iphonedata dataset. For the Blackberry dataset, the models contributed very little in the actual stock price. That is why, we excluded appending the dataset with model data.

Fig. 1 shows the APPL vs BB stock's closing price variations for the time period 1985 to 2020. The Primary Y axis is fed with APPL data and the secondary Y axis is fed with BB data; while x axis shows dates. Fig. 1 shows that in early 2007, the BB stock has a closing stock value as high as \$133.55, whereas APPL stock sits at \$28 which is quite low as compared to BB. It is worth noting that in 2007 the release of Iphone resulted in an increase in the AAPL stock price. The subsequent demonstrates a clear upward trend in APPL stock, indicating a strong performance improvement of the company. Whereas BB goes through an upset, as their stock decreases into \$6 around 2013. At the same point of time, the AAPL sees a massive growth, reporting $100.3 as closing value.

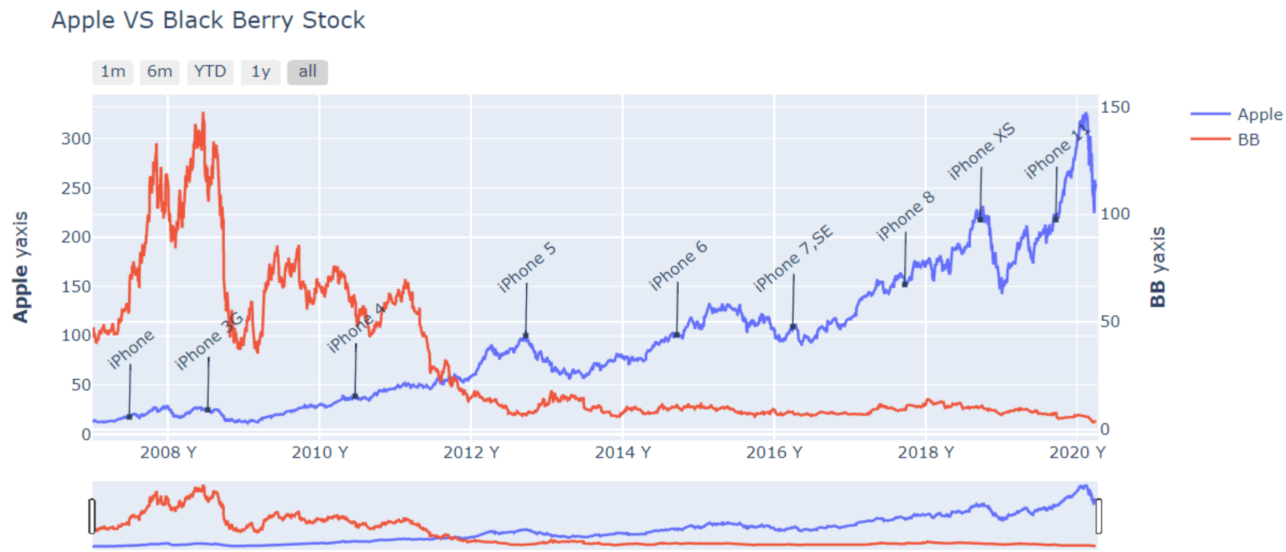

Figure 1. Time series data of APPL and BB stocks

After the first iPhone launch in 2007, the AAPL stock did not experience any significant growth, whereas BB stock increased significantly for a few months following the launch in early 2007. The subsequent release of apple models resulted in an overall constant rise in the AAPL stock. From 2010 to 2018, APPL had a trend of continuous increase, except for the launch of Iphone 5. However, BB stock showed a contrasting behaviour which experienced an initial phase of upward and downward values, only to plummet from $150 to $20 within the 2011-2012 financial year. BB never recovered from this phase, while Apple went on to show a 200% increase in its stock value from within 2012-2018.

From this data, we can say that though there is no evidence that the rise of APPL directly caused the drop in BB. However, the analysis shows that there is a market shift from Blackberry to apple. The year 2011 onwards follows the same trend, where AAPL and BB stocks continue to rise and drop respectively. Eventually, the BB fell down to $2.9 in 2020 from $147 in 2008. This indicates a significantly worsened performance. On the other hand, AAPL went up from $10 in 2006 to $327 in 2020, demonstrating a great performance.

In addition, Fig. 1 shows that the US-China Trade war in 2019 and COVID-19 pandemic in 2020 resulted in the biggest drops in Apple stock value since 2007. The effect of COVID-19 on the stock market, will be analyzed in the next section.

4. Impact of Covid-19 on the Stock Market

In this section, we will analyze the impact of Covid-19 on the stock market. The effects are presented for Canada and the US separately. The Toronto Stock Exchange is made up of over 1,500 companies. The S&P/TSX Composite Index is the benchmark Canadian index, representing roughly 70% of the total market capitalization on the Toronto Stock Exchange with about 250 stocks.

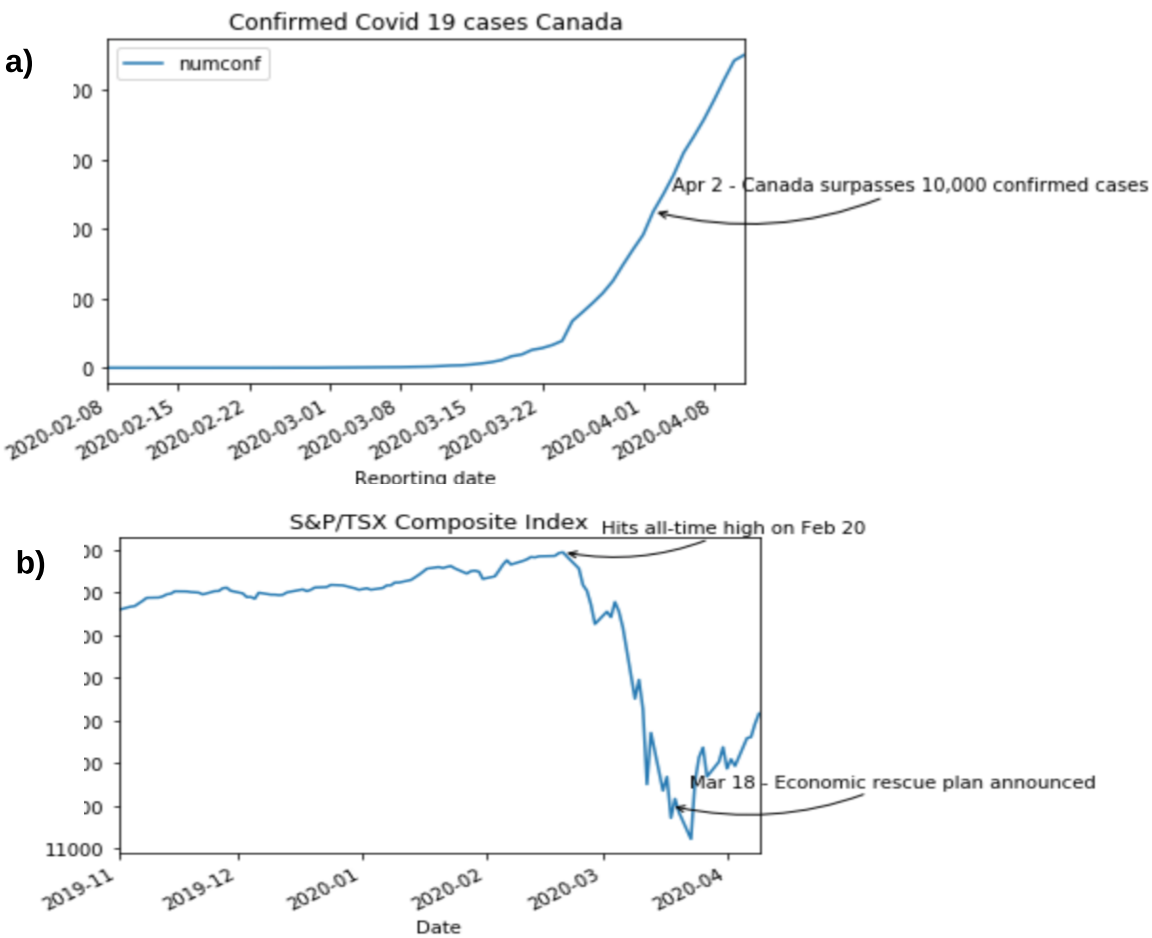

Figure 2 a) Cumulative COVID-19 confirmed cases in Canada, b) S&P/TSX Composite Index

Fig. 2 shows the cumulative numbers of COVID-19 confirmed cases along with the S&P/TSX Composite Index in Canada. The S&P/TSX composite index hit an all-time high of 17,944 on Feb

20th, 2020. By Mar 18th Canada had about 569 confirmed cases of COVID-19. By March 23rd 2020, the index had declined to 11,228, losing 37% of its value. After announcing an Economic rescue plan by government, market began to recover to some degree and ended in 14,166 on April 9.

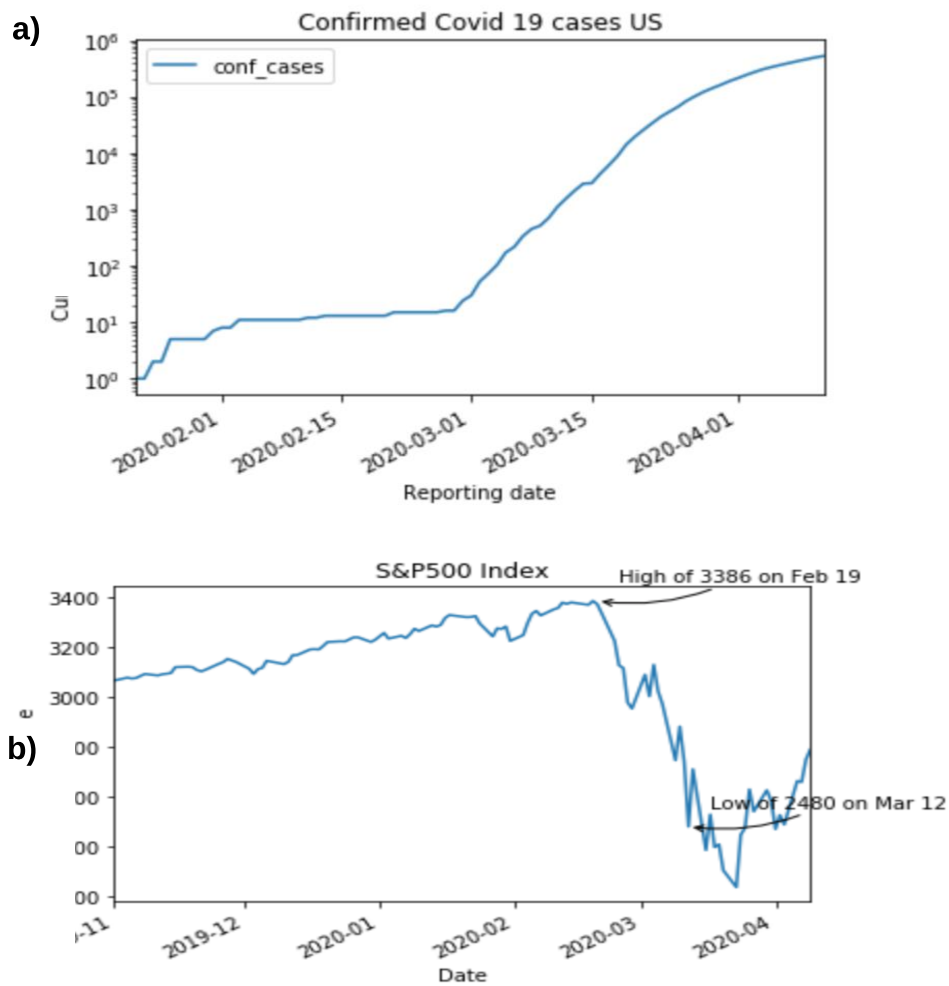

Figure 3 a) Cumulative COVID-19 confirmed cases in US, b) S&P/TSX Composite Index

In the case of the US, S&P 500 index was used to analyze the effects of COVID-19 on the stock market. The S&P 500, or simply the S&P, is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. Fig. 3 shows the cumulative numbers of COVID-19 confirmed cases along with the S&P 500 index. The number of COVID-19 cases in the US started to increase exponentially from the beginning of March, (see Fig. 7a). The S&P 500 reached a high of 3386 on Feb 19 2020, while declined to 2480 by 12th

March 2020, (see Fig 7b). The S&P 500 ended at 2789 as of April 9th resulting in a decline of -21% from a high reached on Feb 20th due to the COVID-19 pandemic.

5. Analysis of Dow Jones Industrial Average Stocks (30 stocks)

The Dow Jones Industrial Average (DJIA) is a price-weighted market index measuring the performance of 30 large companies listed on stock exchanges in the US. The companies represent every major sector of American commerce and are therefore a good barometer of the health of the economy.

We performed various statistical analysis on these 30 stocks to gain insights into the dataset related to trends, volatility, correlation and autocorrelation. The general trend in the data was upwards with large swings in the data marking the recessions that occurred during the 1970-2020 time period. This was to be expected as prices do tend to trend higher as time goes on. However, an autocorrelation analysis (see Figure 4) of the dataset revealed that there is a very weak correlation in the time series data no matter how long the lag. Meaning that past prices cannot be used to profitably predict future prices and that changes in the prices represent a random walk.

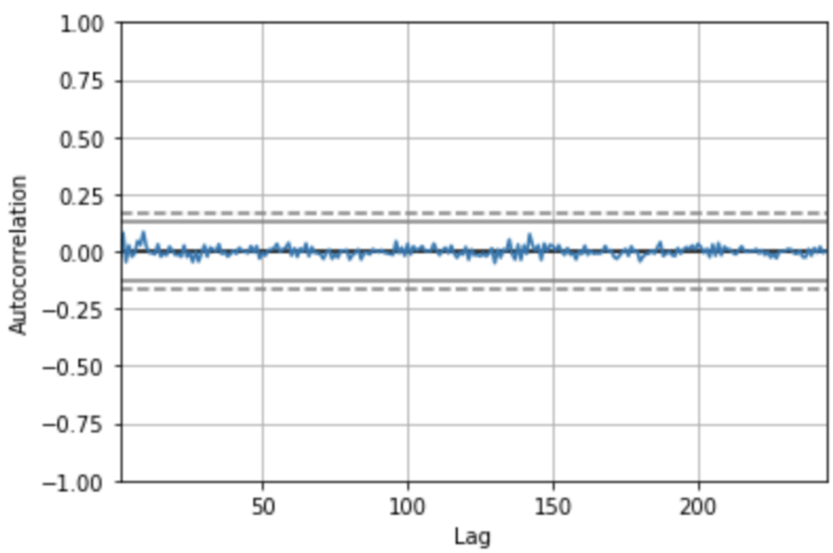

Figure 4 Autocorrelation plot of the DOW 30 stocks

One statistic to measure the volatility of stocks is the standard deviation of daily returns. This was calculated using the pct_change and std (standard deviation) method in the pandas DataFrame class for all 30 stocks. This analysis revealed that Apple (AAPL) was the most volatile stock in the DJIA with a standard deviation of 2.86% while Procter & Gamble (PG) was the least volatile stocks with a standard deviation of 1.39%, (see Fig. 5).

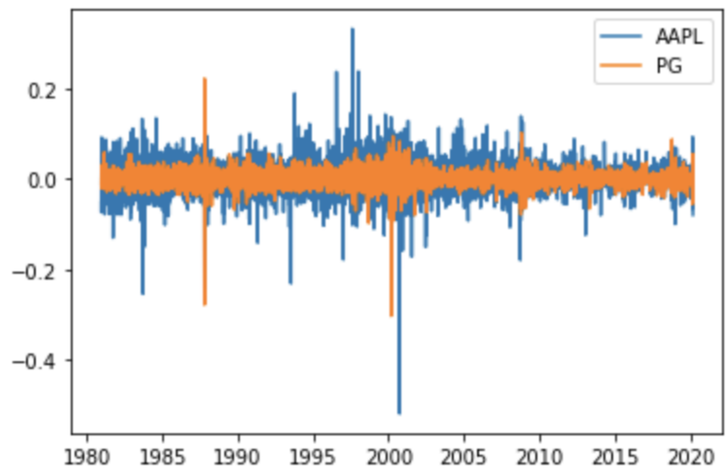

Figure 5 Daily percent change of AAPL and PG

A correlation analysis of the 30 stocks was calculated to gain insights into possible diversification benefits of investing in a variety of stocks while reducing risk. However, the analysis revealed that most stocks are positively correlated to each other and that there is little to be gained in diversifying across the 30 stocks. The three most and least correlated pairs are listed below.

3 most highly correlated pairs:

1. MSFT-V (0.9818)
2. TRV-DIS (0.9806)
3. HD-V (0.9793)

3 least correlated pairs:

1. DOW-MCD (-0.5842)
2. DOW-KO (-0.4150)
3. DOW-HD (-0.2631)

A heatmap chart (shown below in Figure 6) using seaborn helps to visualize the highly correlated nature of the individual components of the DJIA. The dark blue squares represent pairs of stocks with a strong positive correlation value while dark red squares represent the inverse.

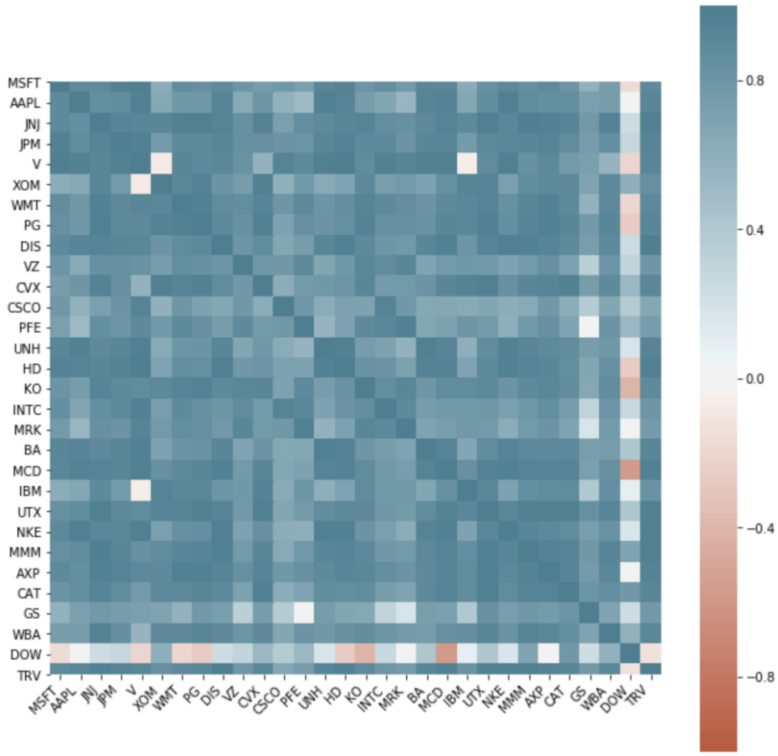

Figure 6 Heatmap showing correlation among the individual stocks of the DJIA

6. Yearly Return for Dow Jones Stocks

Yearly stock market returns provide a great way to see how much volatility and what return rates you can expect over time when investing in the stock market. Yearly stock market returns (% percent return) have been calculated for the period of calendar year 2015 through 2019, on a calendar-year basis for the Dow 30 stocks, (see Table A2). The Dow chemical, DOW symbol, data was not available. Table A3 is generated using the 'Closing' stock price and is sorted based on the highest return in 2019.

Table 1 demonstrates the top 5 rows of Table A2. It includes AAPL (Apple), MSFT (Microsoft), ETR (Entergy[1]), V (Visa), JPAM (JPMorgan Chase[2]) that showed the highest return among the Dow30 stocks in 2019. On the other hand, at the bottom, there are BA (Boeing), XOM

[1] Entergy Corporation is a Fortune 500 integrated energy company engaged primarily in electric power production and retail distribution operations in the Deep South of the United States.
[2] JPMorgan Chase & Co. is an American multinational investment bank and financial services.

(ExxonMobil), MMM (3M), PFE (Pfizer[3]), WBA (Walgreens Boots Alliance[4]); among them the last three showed negative return (loss) in 2019. Overall 2019 yearly return data showed that this year has been profitable for most stocks (89%) and has been a profitable year for the investors.

Table 1. Dow30 highest yearly returns stocks in 2019

STOCK	**2015**	**2016**	**2017**	**2018**	**2019**
AAPL	-2.3	10.6	47.29	-9.31	83.51
MSFT	19.65	15.4	36.98	16.8	57.2
ETR	-21.49	6.28	11.27	3.5	42.69
V	18.53	3.24	43.84	14.35	42.48
JPM	6.22	36.83	23.57	-10.3	40.11

Fig. 7 compares yearly returns for the top 5 stocks for the calendar years from 2015 to 2019 in a bar chart. Fig. 4 shows that the highest profitability is in 2019; In 2015, ETR (Entergy) showed more than 20% loss, while V (Visa) showed the highest profit (~20%). In 2016, all 5 top stocks have been profitable with JPM having the highest profit of 35%. Similar to 2016, in 2017, the percentage of return has increased for all the stocks, with AAPL and V demonstrating more than 40% profit. However, 2018 shows a slowdown in the Market, with AAPL and JPM showing yearly loss during this year. Finally, all stocks (Except V) showed the highest yearly return in 2019, confirming that 2019 has been the most profitable year among last years for the shareholders of these top companies. In addition, Microsoft (MSFT) has delivered at least ~20% profit to the shareholders, confirming it as a safe and profitable investment.

[3] Pfizer Inc. is an American multinational pharmaceutical corporation headquartered in New York City.

[4] Walgreens Boots Alliance, Inc. is an American holding company that owns a number of pharmaceutical manufacturing, wholesale, and distribution companies.

Figure 7 Dow30 Yearly Return for Top 5 Stocks from 2015 to 2019

7. Stock Market Forecasting

Time series forecasting in general and in particular stock market forecasting are difficult types of predictive modeling problems. It is being said that it is impossible to predict stock prices based on historic data due to the many factors affecting the market price. However, we give a shot to apply artificial intelligence to predict stock prices using historical data. A powerful type of artificial neural network designed to handle sequence dependence is called recurrent neural networks (RNNs). An RRN has connections that have loops, feedback, and memory. The memory allows this type of network to learn and generalize across sequences of inputs rather than individual patterns. In an RNN, we store the output activations from one or more of the layers of the network. Often these are hidden later activations. Then, the next time we feed an input example to the network, we include the previously-stored outputs as additional inputs. You can think of the additional inputs as being concatenated to the end of the "normal" inputs to the previous layer.

Now, even though RNNs are quite powerful, they suffer from the Vanishing gradient problem which means that they are good for storing memory 3-4 instances of past iterations but larger numbers of instances don't provide good results. Therefore, in practical applications, another version of RNN called Long Short Term Networks (LSTM) is used. The expression long short-term refers to the fact that LSTM is a model for the short-term memory which can last for a long period of time. An LSTM is well-suited to classify, process and predict time series given time lags of unknown size and duration between important events. For application of LSTM, there is no requirement for time series such as being stationary, etc. LSTMs were developed to deal with the exploding and vanishing gradient problem when training RNNs. The LSTM network was built using Keras library in Tensorflow. (Tensorflow version 2.1.0). The results are presented for APPL stock. The stock price from 2006 to 2018 is used as a training dataset and data from 2019 up today (2020-03-09) is used as a test dataset. We have chosen the 'Close' attribute for prices. Fig. 2 shows the training and test datasets.

Figure 8 Training and test dataset for AAPL stock

The LSTM architecture used a sequential deep learning model, having 4 LSTM layers and one output layer. Each LSTM layer had 50 neurons inputs, and 20% dropout to avoid overfitting. Each LSTM layer contains the following components: Forget Gate (a neural network (NN) with sigmoid), Candidate layer (a NN with Tanh), Input Gate (a NN with sigmoid), Output Gate (a NN with sigmoid), Hidden state (a vector), Memory state (a vector). For the detailed explanation of the LSTM see: https://medium.com/deep-math-machine-learning-ai/chapter-10-1-deepnlp-lstm-long-short-term-memory-networks-with-math-21477f8e4235).

Since LSTMs store long term memory state, we create a data structure with 60 time steps and 1 output, therefore, for each element of the training set, we have 60 previous training set elements. We need to predict a new point based on the actual last 60 points, the next point on 59 actual points and 1 forecasting, the next point on 58 actuals and 2 forecasting s, and so forth. Fig. 9 shows the

LSTM forecasting for APPL compared to the real stock data. The reason that we could get relatively good results is that we have created a test set of 60 actual data-points and predicted output 61, if we want to predict for example 50 forecasting s ahead of time, the results deviates a lot from the real data.

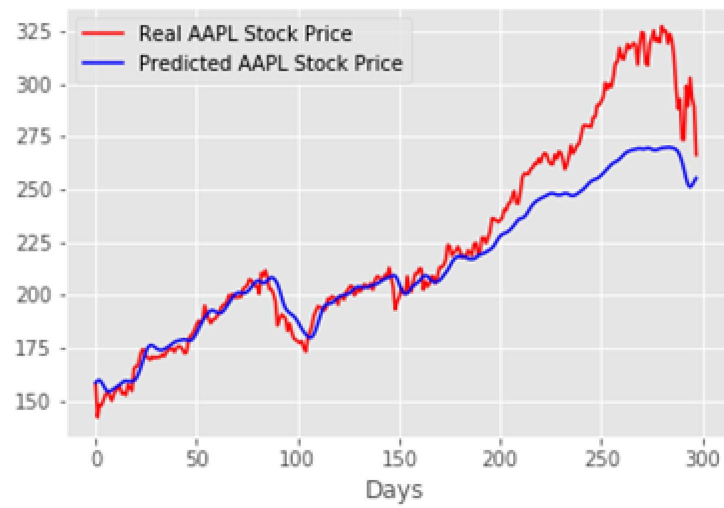

Figure 9 LSTM forecasting and real stock data for AAPL stock in 2019

Conclusions

In this project, we conducted data analysis on the stock market historical data obtained from Yahoo Finance API. AAPL and BB time series data were closely analyzed to find underlying trends behind their stock values. AAPL stock showed an increasing trend from 2007 up 2018, with the exception of iPhone 5 release in 2013. In contrary, BB stock after an initial rise around 2017, showed a decreasing trend for the same period. Our analysis revealed that the US-China Trade war in 2019 and COVID-19 pandemic in 2020 resulted in the biggest drops in Apple stock value since 2007. Overall, the growth of COVID-19 harmed the Canadian stock market severely in which S&P/TSX composite index, hit from an all-time high of 17,944 on Feb 20th to 11,228 by Mar 18th, losing 37% of its value. In the US, COVID-19 also harmed the stocked market. In the US, S&P 500 index declined 21% from a high of 3386 on Feb 19th to 2480 by March 12th. We performed various statistical analysis on the 30 stocks in Dow Jones Industrial to gain insights into the dataset related to trends, volatility, correlation and autocorrelation. Dow30 stocks showed a general upwards trend with large swings in the data marking the recessions that occurred during the 1970-

2020 period. However, autocorrelation analysis revealed a very weak correlation in times series data, no matter how long was the lag. A correlation analysis of various Dow30 stocks showed that most stocks are positively correlated to each other and that there is little to be gained in diversifying across the 30 stocks for the investment. Three most highly correlated stocks became: MSFT-V (0.9818), RV-DIS (0.9806), HD-V (0.9793). Yearly return for Dow30 stocks demonstrated that AAPL, MSFT, ETR, V, and JPAM had the highest yearly return percentage in 2019, which the AAPL having the highest of 83%. Among these top 5 stocks, MSFT provided ~20% yearly profit for the last 5 consecutive years. Finally, a recurrent neural network model, LSTM, was used to forecast the stock price. Although it is almost impossible to predict stock prices based on historic data due to the many factors affecting the market price, the LSTM model could predict the general trend within the AAPL stock. Overall, analyzing and evaluating historical stock data and employing the methods presented in this project, could provide the investors and traders with powerful tools to gain an edge in the markets by making informed decisions.

Appendix

Table A1. A sample of the dataset

Date	MSFT	AAPL	JNJ	JPM	V	XOM	WMT	PG	DIS	VZ	...	IBM
2017-01-03	62.580002	116.150002	115.839996	87.230003	79.500000	90.889999	68.660004	84.199997	106.080002	54.580002	...	167.190002
2017-01-04	62.299999	116.019997	115.650002	86.910004	80.150002	89.889999	69.059998	84.500000	107.440002	54.520000	...	169.259995
2017-01-05	62.299999	116.610001	116.860001	86.110001	81.089996	88.550003	69.209999	85.059998	107.379997	54.639999	...	168.699997
2017-01-06	62.840000	117.910004	116.300003	86.120003	82.209999	88.500000	68.260002	85.029999	108.980003	53.259998	...	169.529999
2017-01-09	62.639999	118.989998	116.279999	86.180000	81.750000	87.040001	68.709999	84.400002	108.360001	52.680000	...	167.649994

Table A2. A sample of Apple dataset with added Model

In [26]:

```
#Append iphone release date to dataset
Apple=pd.merge(Apple, iPhonrelease, how='outer', on='Date')
Apple.head(2)
```

Out[26]:

	Unnamed: 0	Date	Open	High	Low	Close	Adj Close	Volume	model
0	345479400	1980-12-12	0.513393	0.515625	0.513393	0.513393	0.406782	117258400.0	NaN
1	345738600	1980-12-15	0.488839	0.488839	0.486607	0.486607	0.385558	43971200.0	NaN

In [42]:

```
Apple.iloc[[5,2000, 4000, 6000, 8000, 8017, 8018]]
```

Out[42]:

	Unnamed: 0	Date	Open	High	Low	Close	Adj Close	Volume	model
5	346084200	1980-12-19	0.504464	0.506696	0.504464	0.504464	0.399707	12157600.0	NaN
2000	595089000	1988-11-09	1.366071	1.406250	1.357143	1.401786	1.122781	50430800.0	NaN
4000	844781400	1996-10-08	0.839286	0.866071	0.830357	0.830357	0.720784	47608400.0	NaN
6000	1095773400	2004-09-21	2.696429	2.776428	2.675714	2.715000	2.356730	96663000.0	NaN
8000	1346160600	2012-08-28	96.425713	96.585716	95.809998	96.400002	84.038406	66854200.0	NaN
8017	1348234200	2012-09-21	100.344284	100.724289	99.908569	100.012856	87.188019	142897300.0	iPhone 5
8018	1348493400	2012-09-24	98.122856	99.302856	97.571426	98.684288	86.029793	159941600.0	NaN

Table A3. Yearly %Returns for Dow30, sorted based on highest return in 2019

Stock	**2015**	**2016**	**2017**	**2018**	**2019**
AAPL	-2.3	10.6	47.29	-9.31	83.51
MSFT	19.65	15.4	36.98	16.8	57.2
ETR	-21.49	6.28	11.27	3.5	42.69
V	18.53	3.24	43.84	14.35	42.48

JPM	6.22	36.83	23.57	-10.3	40.11
PG	-11.63	8.41	9.35	0.58	38.14
NKE	34.29	-15.62	21.1	15.51	37.15
GS	-6.58	36.03	6.18	-36.23	34.08
DIS	14.4	2.11	1.59	-4.03	33.75
AXP	-25.01	10.92	32.32	-4.57	30.84
WMT	-29.28	13.15	44.77	-6.55	28.12
INTC	-3.93	8.77	26.28	-0.21	27.61
HD	28.28	3.04	41.3	-9.47	27.59
UNH	17.41	39.99	37.98	11.34	21.55
MRK	-7.61	13.49	-5.9	34.06	21.05
KO	3.2	-1.89	9.38	3.65	17.94
IBM	-15.09	22.63	-7.87	-26.72	17.41
CAT	-25.36	38.73	68.55	-20.01	17.33
TRV	7.63	11.05	12.21	-10.41	16.55
JNJ	-1.24	15.41	21.34	-8.59	14.09
MCD	27.33	4.73	44.71	1.35	12.56
CSCO	-1.09	15.6	26.36	10.06	11.22
VZ	-0.45	17.03	-2.11	3.25	9.84
CVX	-19.74	33.28	6.56	-14.84	8.68
BA	11.23	12.32	88.79	6.58	1.95
XOM	-15.18	17.11	-7.56	-19.83	0.29
MMM	-7.83	21.75	32.39	-19.64	-7.17
PFE	3.48	1.66	10.21	17.89	-9.09
WBA	12.95	1.48	-12.08	-9.21	-13.26