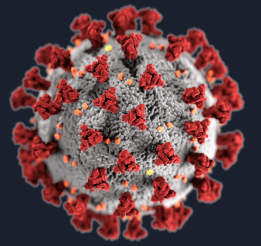
Pandemic vs. Stock Market

By: Analytics Anonymous

Justin Moon, Davis Page, Alberto Castellanos, Bhavin Sanghavi, and Tameka Kuar

Motivation & Summary Slide



The Big Question:
How did the pandemic (COVID-19) affect certain industries?

Other Questions:
Did quarantine affect the consumer industry more than the tech industry?
Has the tech industry done better because its ability to be anywhere and everywhere without physical presence?

Assumption:
We felt that the Technology Sector would outperform when compared to other sectors during the pandemic.

Findings:
Although we were unable to determine any sort of causation, we believe that our data shows that industries like the Tech and Real Estate industries have done much better than that of the Healthcare industry.

Questions & Data

We decided to use the dates January 1, 2020 - June 30, 2020 as the basis for the pandemic period. Our data assumes that stock percentage increase and decreases were affected by COVID-19.

We needed stock data for the industries we are comparing. We found stock data using a source (AlphaVantage) and downloaded csv files from 100 different companies. We used a sample of the top 20 companies (by market capitalization) to represent each of the five sectors.

- Technology
- Real Estate
- Consumer Staples
- Healthcare
- Financials

Data Cleanup & Exploration

First, we had to concatenate all of our csvs into one combined csv.
We also decided to only use a 5 year period to use as our historical data so we cut data from 2014 and below.
Next, we needed to add tickers to the csv because the original csv did not have a ticker column for each stock.

Discoveries:
Explain “glob” and how it was used to combine csvs

Problems:
API Key limit to 5 per minute so we used csvs

- 20 companies x 5 Sectors x 12 months x 6 years (7200 calls)
- 7200 / 5 = 1440 mins (24 hours)

APPLE STOCK DATA

```
timestamp,open,high,low,close,volume,Ticker,Sector
2020-07-02,365.12,370.47,363.64,364.11,55454924,AAPL,Information Technology
2020-06-30,317.75,372.38,317.21,364.8,810900890,AAPL,Information Technology
2020-05-29,286.25,324.24,285.85,317.94,701660022,AAPL,Information Technology
2020-04-30,246.5,294.53,236.9,293.8,816530808,AAPL,Information Technology
2020-03-31,282.28,304.0,212.61,254.29,1570331732,AAPL,Information Technology
2020-02-28,304.3,327.22,256.37,273.36,755223231,AAPL,Information Technology
2020-01-31,296.24,327.85,292.75,309.51,734044103,AAPL,Information Technology
2019-12-31,267.27,293.97,256.29,293.65,598871365,AAPL,Information Technology
2019-11-29,249.54,268.0,249.16,267.25,448922253,AAPL,Information Technology
2019-10-31,225.07,249.75,215.132,248.76,621478768,AAPL,Information Technology
2019-09-30,206.43,226.42,204.22,223.97,547408488,AAPL,Information Technology
2019-08-30,213.9,218.03,192.58,208.74,683515746,AAPL,Information Technology
2019-07-31,203.17,221.37,198.41,213.04,473957094,AAPL,Information Technology
2019-06-28,175.6,201.57,170.27,197.92,515218768,AAPL,Information Technology
2019-05-31,209.88,215.31,174.99,175.07,739456573,AAPL,Information Technology
2019-04-30,191.64,208.48,188.38,200.67,506117812,AAPL,Information Technology
2019-03-29,174.28,197.69,169.5,189.95,650981384,AAPL,Information Technology
2019-02-28,166.96,175.87,165.93,173.15,472540723,AAPL,Information Technology
2019-01-31,154.89,169.0,142.0,166.44,828099179,AAPL,Information Technology
2018-12-31,184.46,184.94,146.59,157.74,898917007,AAPL,Information Technology
2018-11-30,219.05,222.36,170.26,178.58,961321947,AAPL,Information Technology
2018-10-31,227.95,233.47,206.09,218.86,789748068,AAPL,Information Technology
2018-09-28,228.41,229.67,215.3,225.74,678972040,AAPL,Information Technology
2018-08-31,199.13,228.87,197.31,227.63,700318837,AAPL,Information Technology
2018-07-31,183.82,195.96,183.42,190.29,393843881,AAPL,Information Technology
2018-06-29,187.9912,194.2,180.73,185.11,527624365,AAPL,Information Technology
2018-05-31,166.4102,190.37,165.27,186.87,620976206,AAPL,Information Technology
2018-04-30,167.88,178.9365,160.63,165.26,666360147,AAPL,Information Technology
2018-03-29,178.54,183.5,164.94,167.78,701387082,AAPL,Information Technology
2018-02-28,167.165,180.615,150.24,178.12,888378184,AAPL,Information Technology
2018-01-31,170.16,180.1,164.7,167.43,639245534,AAPL,Information Technology
2017-12-29,169.95,177.2,166.46,169.23,518560008,AAPL,Information Technology
2017-11-30,169.87,176.24,165.28,171.85,581876496,AAPL,Information Technology
2017-10-31,154.26,169.6499,152.46,169.04,496135305,AAPL,Information Technology
2017-09-29,164.8,164.94,149.16,154.12,669594016,AAPL,Information Technology
2017-08-31,149.1,164.52,148.41,164.0,638221161,AAPL,Information Technology
2017-07-31,144.88,153.99,142.41,148.73,411377229,AAPL,Information Technology
```

COMBINING CSV's

In [1]:
```
import pandas as pd
import csv
import glob
import os
os.chdir('Resources')
```

In [2]:
```
extension = 'csv'
ticker_files = [x for x in glob.glob('*.{}'.format(extension))]
```

In [3]:
```
#combine all files in the list
combined_csv = pd.concat([pd.read_csv(f) for f in ticker_files ])
#export to csv
combined_csv.to_csv( "combined_csv.csv", index=False, encoding='utf-8-sig')

combined_csv
```

SORTING AND LOCATING 2015-2019

In [7]:
```
sorted_df = file_df.sort_values(by=['timestamp'])
sorted_df.head()
```

Out[7]:

	timestamp	open	high	low	close	volume	Ticker	Sector
11173	2000-07-31	85.44	91.94	78.12	81.81	21970800	UNH	Health Care
13820	2000-07-31	101.30	101.40	90.25	93.06	67023900	JNJ	Health Care
959	2000-08-31	20.50	23.69	20.00	23.13	17514100	TMO	Health Care
18859	2000-08-31	75.25	111.60	74.88	108.00	13343200	GILD	Health Care
5857	2000-08-31	99.00	129.50	97.00	129.50	53400600	GS	Financials

In [8]:
```
#start_drop_dates = sorted_df
cleaned_df = sorted_df.loc[(sorted_df['timestamp']>'2014-11-31')]
predict_df = cleaned_df.loc[(cleaned_df['timestamp']<'2020-01-01')]
print(cleaned_df)
predict_df.tail()
```

Adding Tickers to the Data

In [2]:

```
# List with all the ticker names
ticker = ['WMT', 'PG', 'KO', 'PEP', 'COST', 'PM', 'MO', 'MDLZ', 'EL', 'CL',
          'KMB', 'KHC','WBA', 'GIS', 'MNST', 'STZ', 'BF-B', 'SYY', 'CLX', 'KR',
          'BRK.B', 'JPM', 'BAC', 'WFC', 'C', 'BLK', 'SPGI', 'AXP', 'MS', 'GS',
          'CME', 'CB', 'USB', 'MMC', 'MCO', 'TFC', 'ICE', 'PGR', 'PNC', 'AON',
          'JNJ', 'UNH', 'MRK', 'PFE', 'ABT', 'LLY', 'ABBV', 'TMO', 'AMGN', 'BMY',
          'MDT', 'DHR', 'GILD', 'CVS', 'VRTX', 'CI', 'REGN', 'SYK', 'ANTM', 'ISRG',
          'AMT', 'CCI', 'PLD', 'EQIX', 'DLR', 'PSA', 'SBAC', 'SPG', 'AVB', 'EQR',
          'WELL', 'O', 'ARE', 'WY','CBRE', 'ESS', 'PEAK', 'BXP', 'VTR', 'MAA',
          'AAPL', 'MSFT', 'V', 'MA', 'INTC', 'NVDA', 'ADBE', 'PYPL', 'CSCO', 'ORCL',
          'CRM', 'ACN', 'AVGO', 'TXN', 'IBM', 'QCOM', 'FIS', 'NOW', 'INTU', 'FISV']

# List of ticker with the specified sector
consumer_staples = ['WMT', 'PG', 'KO', 'PEP', 'COST', 'PM', 'MO', 'MDLZ', 'EL', 'CL',
                    'KMB', 'KHC', 'WBA', 'GIS', 'MNST', 'STZ', 'BF-B', 'SYY', 'CLX', 'KR']

finance = ['BRK.B', 'JPM', 'BAC', 'WFC', 'C', 'BLK', 'SPGI', 'AXP', 'MS', 'GS',
           'CME', 'CB','USB', 'MMC', 'MCO', 'TFC', 'ICE', 'PGR', 'PNC', 'AON']

health_care = ['JNJ', 'UNH', 'MRK', 'PFE', 'ABT', 'LLY', 'ABBV', 'TMO', 'AMGN', 'BMY',
               'MDT', 'DHR', 'GILD', 'CVS', 'VRTX', 'CI', 'REGN', 'SYK', 'ANTM', 'ISRG']

real_estate = ['AMT', 'CCI', 'PLD', 'EQIX', 'DLR', 'PSA', 'SBAC', 'SPG', 'AVB', 'EQR',
               'WELL', 'O', 'ARE', 'WY','CBRE', 'ESS', 'PEAK', 'BXP', 'VTR', 'MAA',]

technology = ['AAPL', 'MSFT', 'V', 'MA', 'INTC', 'NVDA', 'ADBE', 'PYPL', 'CSCO', 'ORCL',
              'CRM', 'ACN', 'AVGO', 'TXN', 'IBM', 'QCOM', 'FIS', 'NOW', 'INTU', 'FISV']


# Loop to add tickers to the corresponding csv
for name in range(len(ticker)):

    csv_input = pd.read_csv(f'Resources/monthly_{ticker[name]}.csv')
    csv_input['Ticker'] = ticker[name]
    csv_input.to_csv(f'Resources/monthly_{ticker[name]}.csv', index=False)
```

```
# Loops to add sector names to corresponding csv
for name in range(len(consumer_staples)):

    csv_input = pd.read_csv(f'Resources/monthly_{consumer_staples[name]}.csv')
    csv_input['Sector'] = 'Consumer Staples'
    csv_input.to_csv(f'Resources/monthly_{consumer_staples[name]}.csv', index=False)

for name in range(len(finance)):

    csv_input = pd.read_csv(f'Resources/monthly_{finance[name]}.csv')
    csv_input['Sector'] = 'Financials'
    csv_input.to_csv(f'Resources/monthly_{finance[name]}.csv', index=False)

for name in range(len(health_care)):

    csv_input = pd.read_csv(f'Resources/monthly_{health_care[name]}.csv')
    csv_input['Sector'] = 'Health Care'
    csv_input.to_csv(f'Resources/monthly_{health_care[name]}.csv', index=False)

for name in range(len(real_estate)):

    csv_input = pd.read_csv(f'Resources/monthly_{real_estate[name]}.csv')
    csv_input['Sector'] = 'Real Estate'
    csv_input.to_csv(f'Resources/monthly_{real_estate[name]}.csv', index=False)

for name in range(len(technology)):

    csv_input = pd.read_csv(f'Resources/monthly_{technology[name]}.csv')
    csv_input['Sector'] = 'Information Technology'
    csv_input.to_csv(f'Resources/monthly_{technology[name]}.csv', index=False)
```

Data Analysis

Step 1: Find the quarterly percent change for each stock in a sector (not % formatted)

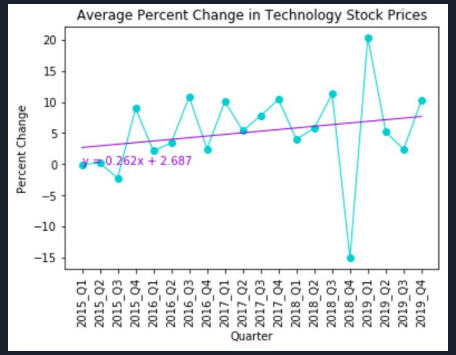
	Ticker	2015_Q1	2015_Q2	2015_Q3	2015_Q4	2016_Q1	2016_Q2	2016_Q3	2016_Q4	2017_Q1	...	2018_Q1	2018_Q2	2018_Q3	2018_Q4
0	AAPL	0.127288	0.007996	-0.120590	-0.045694	0.035436	-0.122855	0.182531	0.024502	0.240373	...	-0.008568	0.103290	0.219491	-0.301232
1	ABBV	-0.105440	0.147762	-0.190207	0.088770	-0.035787	0.083859	0.018737	-0.007135	0.040562	...	-0.021301	-0.021130	0.020831	-0.025270
2	ABT	0.029098	0.059357	-0.180522	0.116609	-0.068582	-0.060244	0.075808	-0.091747	0.156209	...	0.049939	0.017857	0.202820	-0.014040
3	ACN	0.049043	0.032981	0.015292	0.063505	0.104306	-0.018284	0.078383	-0.041254	0.023478	...	0.002678	0.065733	0.040406	-0.171504
4	ADBE	0.017056	0.095618	0.014936	0.142544	-0.001490	0.021215	0.133104	-0.051502	0.264012	...	0.233052	0.128332	0.107215	-0.161919
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
92	WBA	0.111286	-0.002834	-0.015869	0.024729	-0.010745	-0.011515	-0.031824	0.026544	0.003504	...	-0.098458	-0.083321	0.214696	-0.062689
93	WELL	0.022334	-0.151629	0.031845	0.004578	0.019256	0.098500	-0.018380	-0.104855	0.058120	...	-0.146464	0.151755	0.026001	0.079136
94	WFC	-0.007661	0.033824	-0.086949	0.058617	-0.110375	-0.021299	-0.064441	0.244580	0.009980	...	-0.136146	0.057813	-0.051948	-0.123288
95	WMT	-0.042268	-0.137629	-0.085859	-0.054596	0.117292	0.066141	-0.012325	-0.041597	0.042824	...	-0.099038	-0.037316	0.096439	-0.008093
96	WY	-0.076344	-0.049774	-0.132063	0.096562	0.033356	-0.039057	0.072892	-0.057921	0.129279	...	-0.007374	0.041714	-0.114920	-0.322591

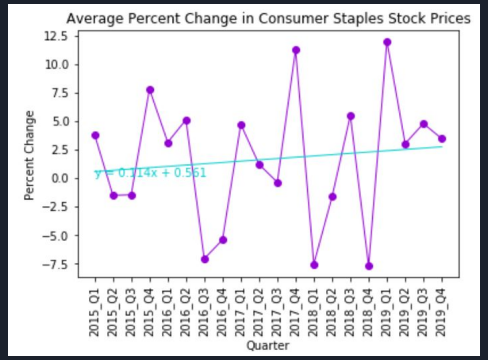
Step 2: Calculate the average percent change per sector

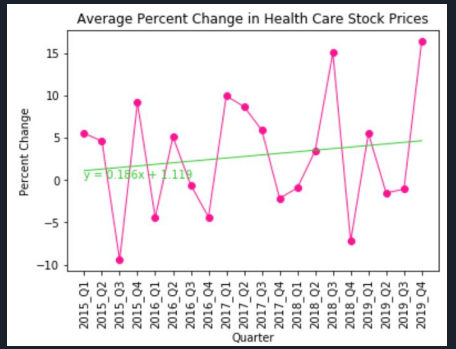
Information Technology Sector Quarterly Average Percent Change

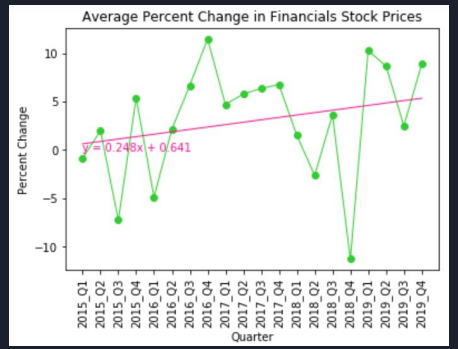
	Quarter	Percent Change
0	2015_Q1	-0.0388%
1	2015_Q2	0.1938%
2	2015_Q3	-2.2266%
3	2015_Q4	8.9643%
4	2016_Q1	2.1432%
5	2016_Q2	3.4437%
6	2016_Q3	10.8199%
7	2016_Q4	2.4793%
8	2017_Q1	10.0641%
9	2017_Q2	5.3750%
10	2017_Q3	7.7881%
11	2017_Q4	10.4001%
12	2018_Q1	4.0167%
13	2018_Q2	5.7831%
14	2018_Q3	11.2682%
15	2018_Q4	-15.0720%
16	2019_Q1	20.3657%
17	2019_Q2	5.1983%
18	2019_Q3	2.3812%
19	2019_Q4	10.2363%

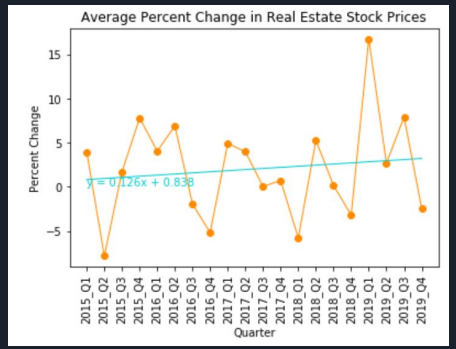
Step 3: Plot points on a line/scatter plot and perform a linear regression to predict 2020 Data



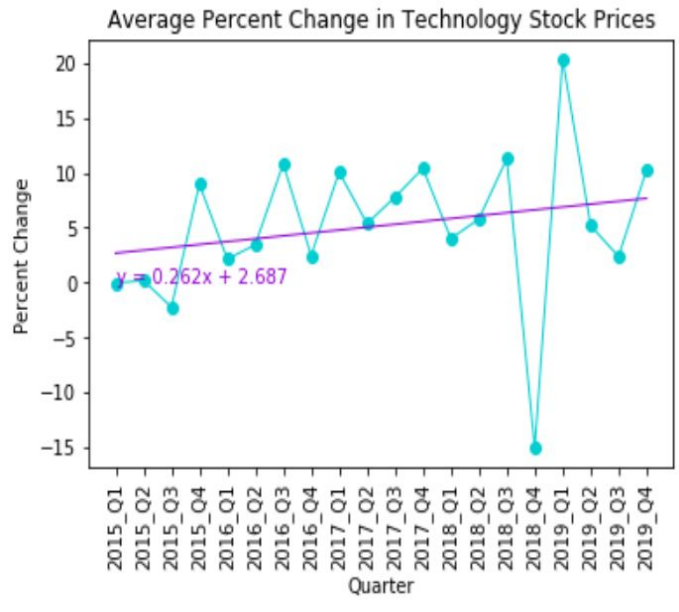


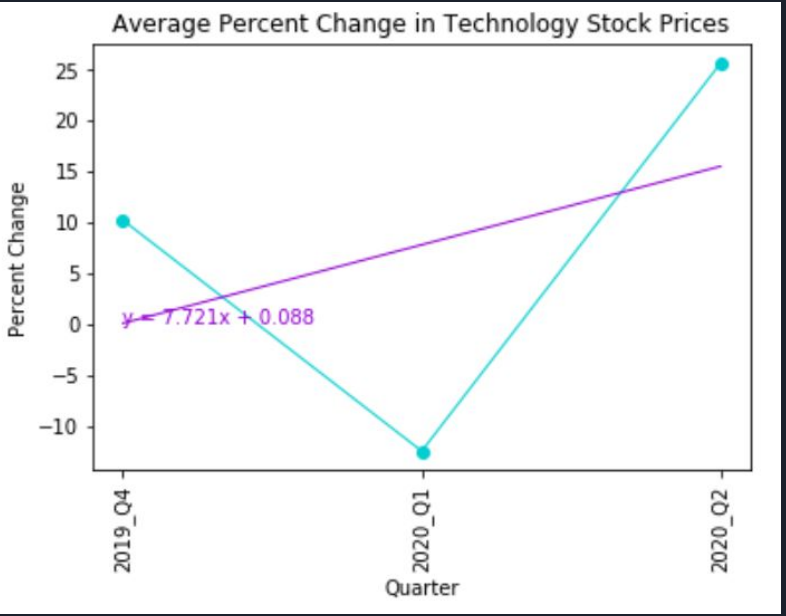




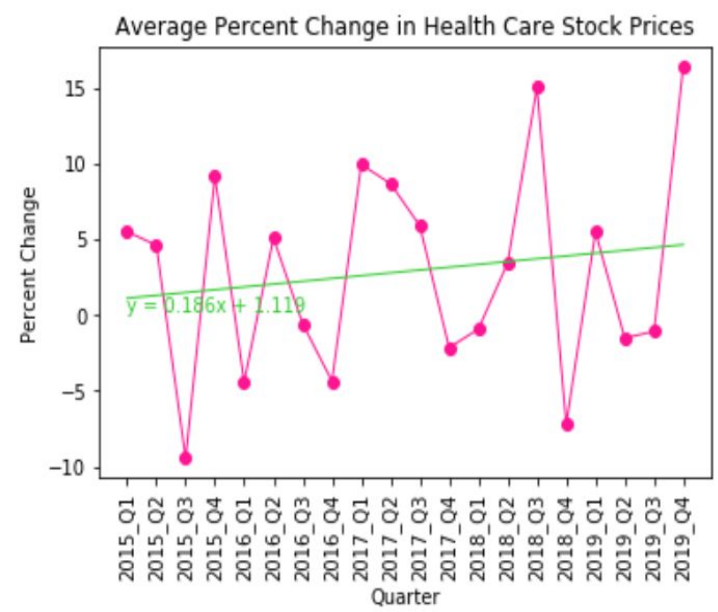


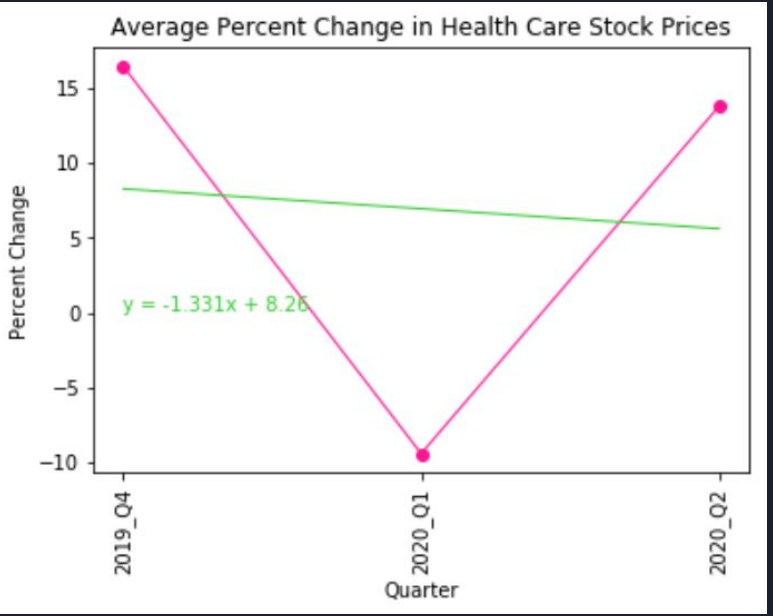
Step 4: Compare Predicted Data with Actual Data



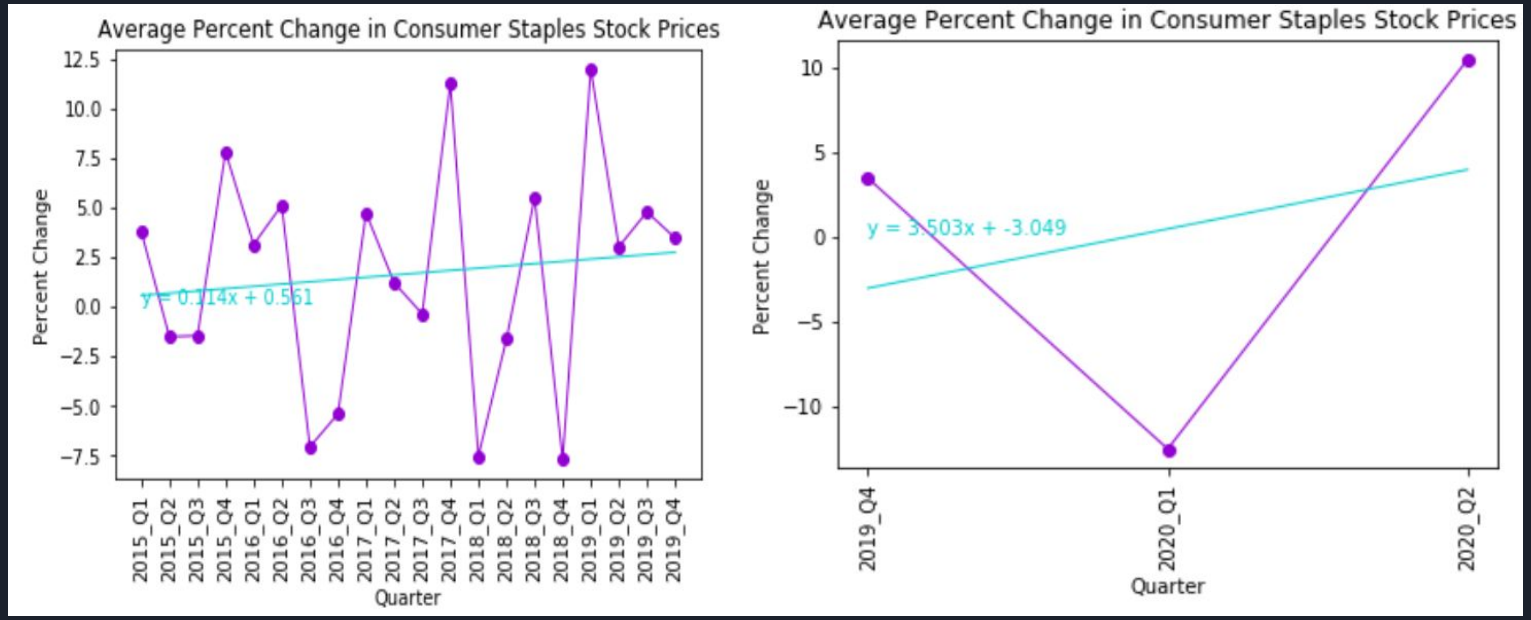


Step 4: Compare Predicted Data with Actual Data

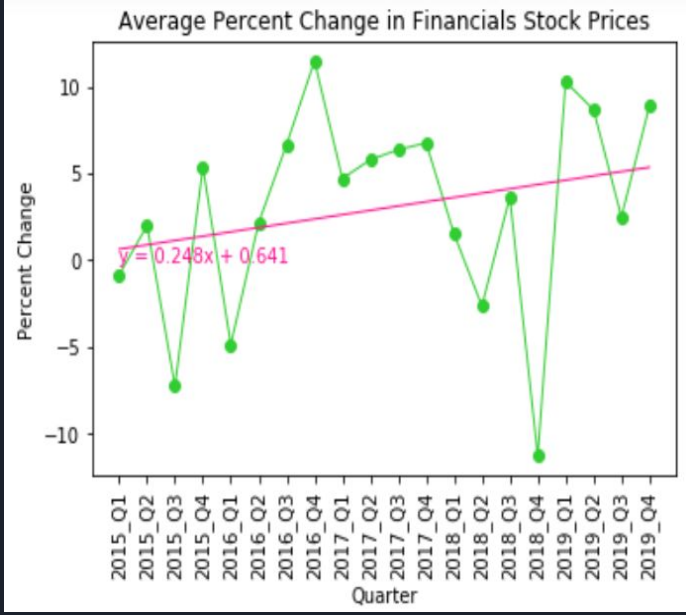


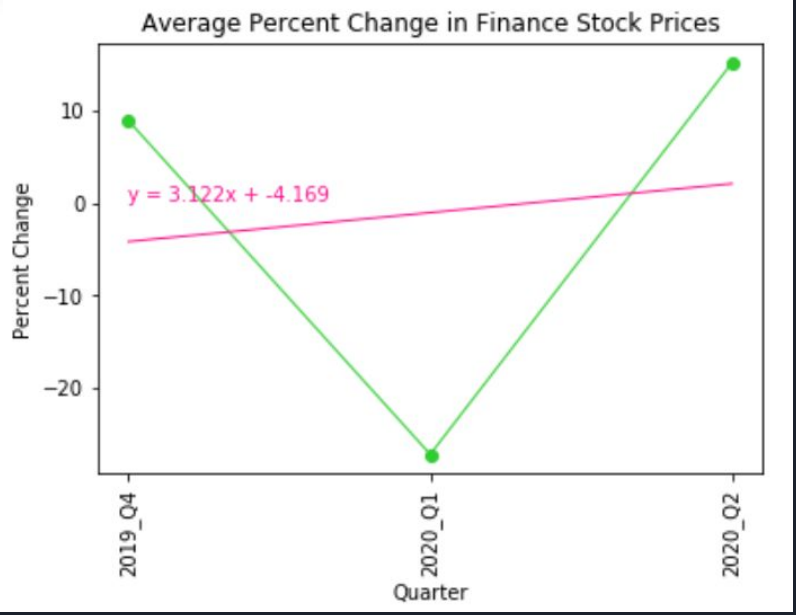


Step 4: Compare Predicted Data with Actual Data

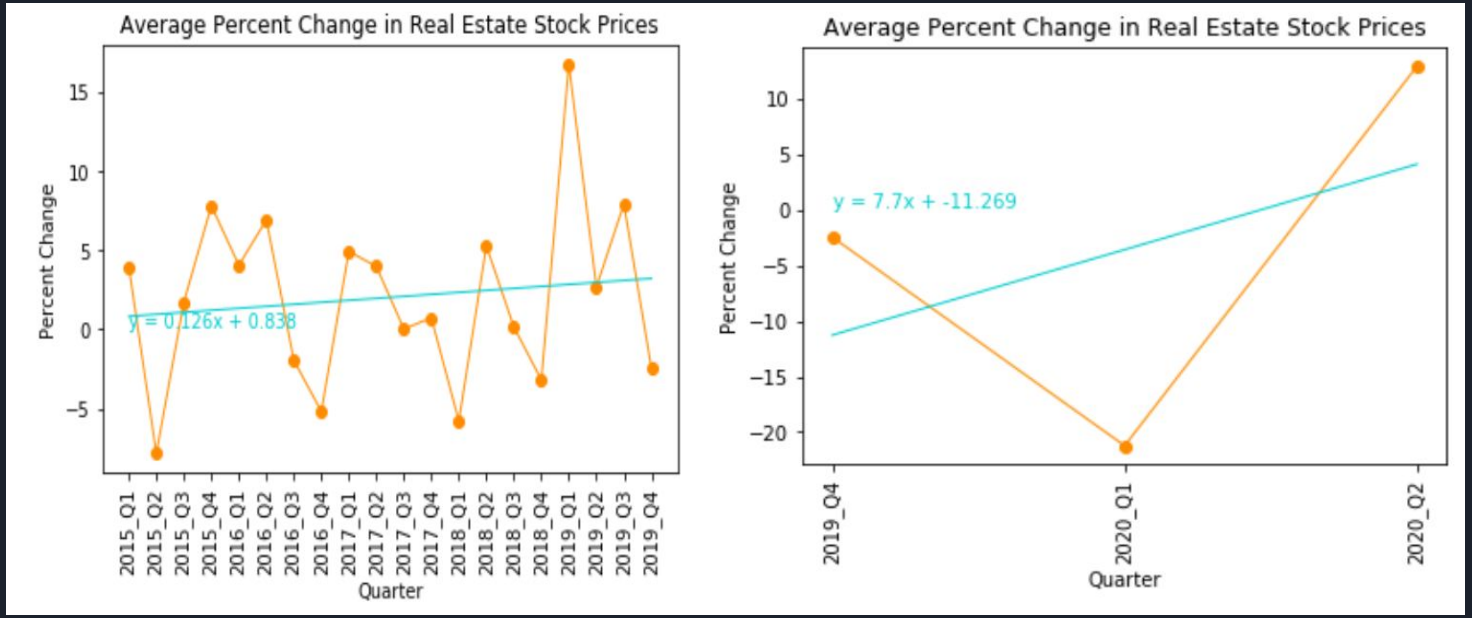


Step 4: Compare Predicted Data with Actual Data





Step 4: Compare Predicted Data with Actual Data



Discussion

Predicted Average Growth ROC:
Technology: 0.30
Real Estate: 0.10
Consumer Staples: 0.10
Healthcare: 0.20
Finance: 0.20

Actual Average Growth ROC:
Technology: 7.7
Real Estate: 7.7
Consumer Staples: 3.5
Healthcare: -1.3
Finance: 3.1

We expected the Technology sector to be much higher than average predicted growth because we felt that these companies would not be impacted as hard because of they are not required to have a physical presence. In fact, the pandemic would possibly increase the performance of such an industry.

However, we were surprised to find that the Real Estate industry to have increased over the pandemic higher than its average predicted growth.

Discussion

V-Shape Recovery

All of the sectors made a short recovery that resembles a V-shape even if the slope was negative such as Healthcare.

Post Mortem

Difficulties:
Incorrectly presuming that stock data prices should be averaged when comparing sectors

- Some stocks have higher prices and lower prices so we must compare evenly by looking at percent change

If we had more time:
Go deeper into tech companies and companies that had online presence and compare their data before, now, and after.
Has stimulus checks affected company stocks?
Instead of using 2020 as a general timeframe for the pandemic, using real data such as confirmed cases to compare against stock data.

Questions?

Sources

Stock Data:

https://www.alphavantage.co/

Determine Top 20 Stocks per Sector:

https://www.barchart.com/