**Data Cleaning and Integration**

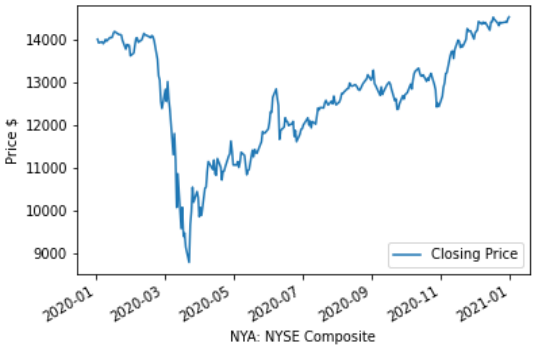
The dataset used to analyze the legal operating businesses New York City was obtained from NYC Open Data. This data set features businesses and individuals holding a Department Of Consumer Affairs (DCA) license so that they may legally operate in New York City. The data is formatted as a comma separated value (csv) file. This dataset as well as the other datasets used for economic analysis was read into DataFrames and parsed with the Python Pandas library. The dataset contained records dating back to 2001. For this analysis, we focused on 2019 and 2020. The two separate subsets were created; one for each year of interest. There were many rows that had empty values. The *License Expiration Date* and *License Creation Date* columns had numerous formatting issues. For these columns, we used Pandas to drop rows that had invalid values.

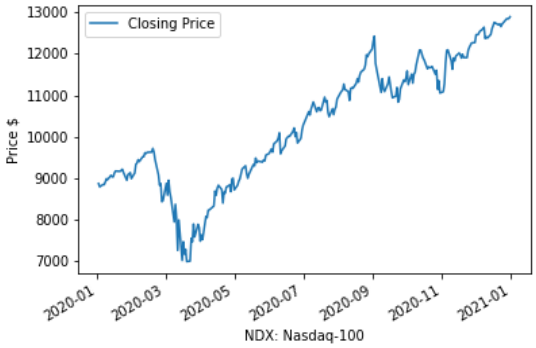
The dataset to analyze the S&P 500 was initially downloaded from *polygon.io*. I wanted historical market data for other indices and ETFs but *polygon.io* did not have the data. I ended up retrieving all the financial market data for my analysis from *investing.com*. I chose *investing.cim* because it was free and it provided a very friendly and easy to use user interface. The datasets for all indices and ETFs span the entire year of 2020. The data was formatted as csv. We used Pandas to read the csv data into DataFrames. The closing prices had to be modified to remove all commas.

**Data Analysis and Findings**

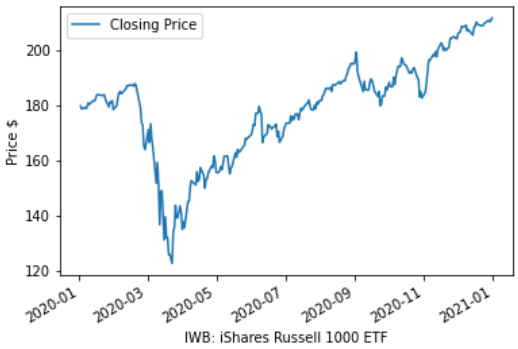
**<u>Stock Market</u>**

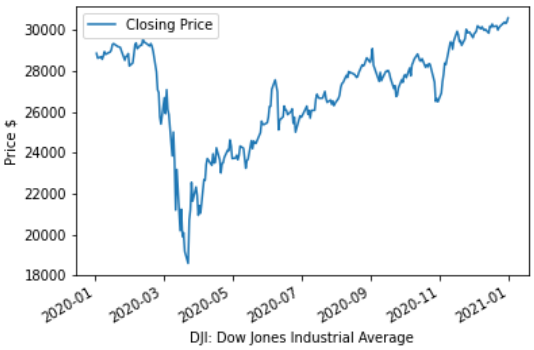
The U.S stock markets have several tradeable financial products that track the performance of the specific industries and market sectors. The Standard and Poor's 500, or simply the S&P 500, is a free-float, weighted measurement stock market index of 500 of the largest companies listed on stock exchanges in the United States. It is one of the most commonly followed equity indices. In this analysis, we also looked at data for NYSE Composite, Nasdaq-100, Russell 2000 ETF, Russell 1000 ETF, and the Dow Jones Industrial average. In these charts we observe that between a significant dip between the months of March and April 2020. In the months following the crash, we observe a period of recovery until the original prices are surpassed.

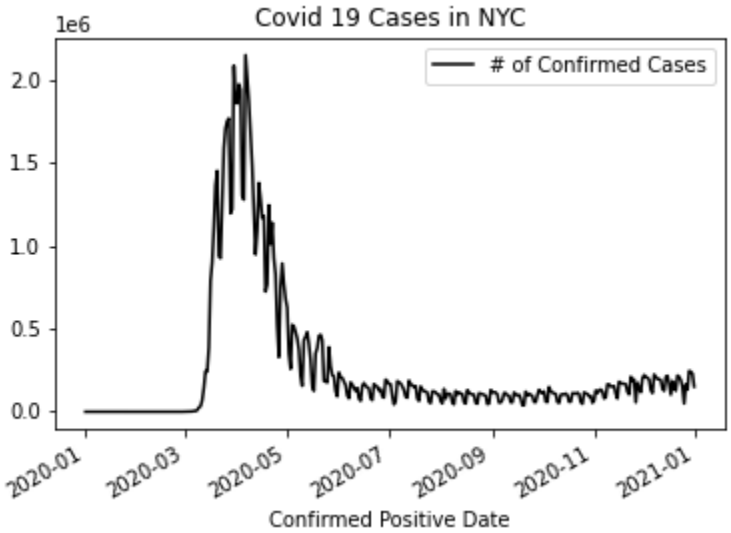




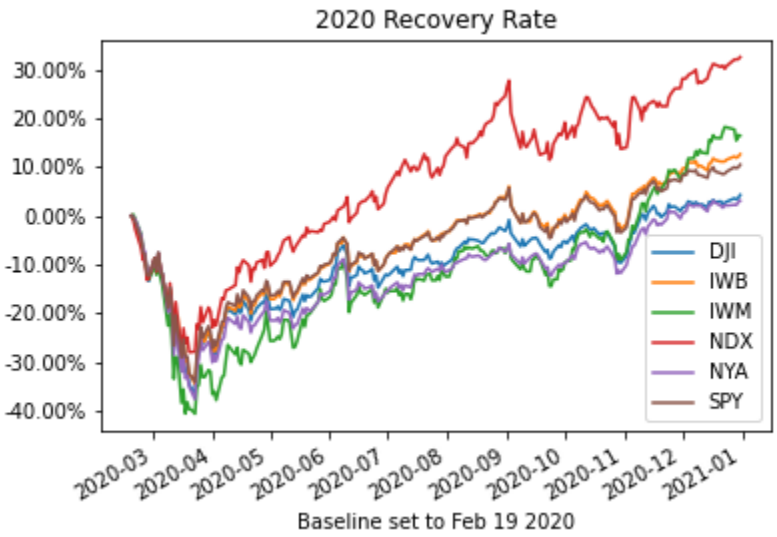


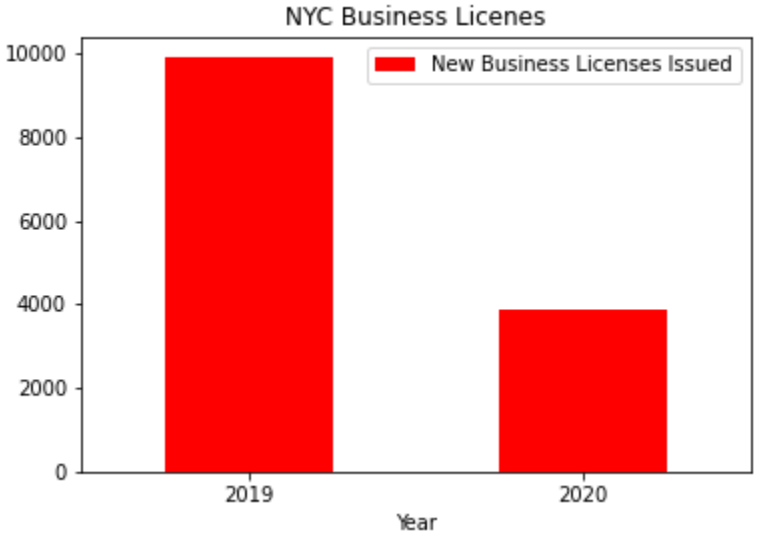

From the plot titled *Covid 19 Cases in NYC*, we see that there was a significant spike between the months of March and April of 2020. The spike went from almost zero cases to 1 million in a matter of days.
Our findings show that there is a strong relationship to the number of confirmed positive COVID19 cases that could be used as a historical tool to predict stock market performance in future pandemics.



In the chart titled *2020 Recovery Rate*, we see the rate at which each index and ETF recovered. Feb 19 2020 was selected as the baseline price. The percentages show that the Nasdaq-100 had the best recovery performance. The outstanding performance is mainly due to the Nasdaq-100 heavily focusing on technology and media industries.




For many small businesses, the pandemic quickly resulted in revenue decline.
The revenue troubles appear to be more concentrated in some industries than others. Our analysis shows that there was a 60% drop in the number of new business licenses that were issued by New York City Department of Consumer Affairs.



In this analysis we the top 10 industries that were affected by the Covid19 pandemic shutdown. We see that the home improvement was impacted the most. These 10 industries amounted to 60% of the total decline in new business licenses issued in 2020.