W200 - Project - Summer 2019
Financial markets data and Recession Predictors
Praveen Joseph, Sameed Musvee, Jade Chia-Chun Hou, Wenqi Liu

Contents

Theory of recessions and credit cycles:

Since the beginning of capitalism, markets follow a sequence of ups and downs - from the Tulip mania of the 17th century to the Great Financial Crisis (GFC) of the 21st century - periods of growth have always been followed by times of recession. The "cyclical" behavior of the capitalist economy shows periods (roughly 5-7 years) of economic growth followed by times of recession.

When talking about the sources of economic growth there is a very wide range of variables that could be considered. Factors like capital, technology, labor, investments, interest rates and their relationships all play a crucial role in the economy. These factors are intrinsically captures by financial time-series data: macroeconomic varies, stock prices, commodity prices and credit and sentiment data. We will use financial markets data to analyse and make statistically significant inferences specifically in the context of various recessionary periods of US economic history.

Exploring recessions:

For the purpose of this project, we will focus our attention on a few specific recessionary period to better understand the behaviour of financial market data and see how different variables can be used to predict financial recessions. Specifically we will focus our attention on 3 recessionary periods in US history:

- **1980-1982 recession** - January 1980 - November 1982
- **Early 2000s recession** - March 2001 - November 2001
- **Great Recession** - December 2007 - June 2009

Data Exploration:

Source:
All data used in the analysis was sourced from Bloomberg, which is a proprietary platform for extensive and reliable financial markets data. Bloomberg collects and aggregates data from various sources including, U.S. Bureau of Economic Analysis (BEA), US Department of Labor (DoL), National Bureau of Economic Research(NBER) and Federal Reserve(Fed) which collect and publish fundamental macroeconomic data for the US economy, including GDP, Inflation, unemployment, consumer Credit etc.

Dataset
The data set consists of 4 distinct set of financial time-series data
1. Macroeconomic Data (econ_data)
2. Global Stocks Data (stock_data)
3. Commodity price and Index Data (commodity_data)
4. US Credit and Sentiment Data (credit_data)

Macroeconomic data (econ data) is a set of macroeconomic variables, many of which are published quarterly and a primary determinants for understanding the health of the economy. This data is also called fundamental data because of its explanatory power of economic effects and plays a key role in policy decision making and assessing the health of an economy.

	Date	Fed_Rate	GDP_YOY	GDP_USD	Inflation	Unemployment	Consumer_Credit_USD	Consumer_Credit_YOY	FED_Recession_Index
0	2019-06-28	2.38	4.0	21337.90	1.6	3.63	17.086	5.2	12.51
1	2019-03-29	2.41	4.6	21098.83	1.9	3.87	10.982	4.9	10.84
2	2018-12-31	2.27	4.9	20897.80	1.9	3.80	11.997	4.7	11.46
3	2018-09-28	1.95	5.8	20749.75	2.3	3.80	14.311	4.9	10.33
4	2018-06-29	1.82	6.0	20510.18	2.9	3.90	8.875	4.2	9.87

Credit and Leading indicators data includes credit growth and change in the economy and measures of confidence and expectation among the various players in the economy and it gives a sense of how different participants in the economy (individuals, households, businesses, corporates, govt, etc). Credit is also the lifeline of the economy and is a key determinant of stock market and housing prices.

	Date	Experian_Cons_Credit	Umich_Cons_Sent	Umich_Cons_Expect	Citi_econ_surprise	Consumer_Credit_USD)	Consumer_Credit_YOY	FED_Recession_Index
0	2019-06-30	NaN	98.2	89.3	-68.3	NaN	NaN	12.51
1	2019-05-31	0.83	100.0	93.5	-32.3	17.086	5.2	11.12
2	2019-04-30	0.88	97.2	87.4	-58.7	17.457	5.3	11.21
3	2019-03-31	0.92	98.4	88.8	-35.9	10.982	4.9	10.84
4	2019-02-28	0.92	93.8	84.4	-29.8	15.500	4.9	9.14

Stock prices reflect the value of financial assets in the economy and are a strong indicator of the overall health of corporations in the economy. Specifically stock prices tell us how companies in specific sectors of the economy are performing.

	Date	S&P_Index	MSCI_World	EURO_STOXX	MSCI_APAC	MSCI_EM	Nasdaq_Index	Bitcoin	Citi_econ_surprise	USD_Index
0	2019-06-01	2941.76	2178.35	3473.69	160.05	1054.86	8006.244	12212.70	-68.3	96.130
1	2019-05-01	2752.06	2046.25	3280.43	152.36	998.00	7453.148	8503.38	-32.3	97.750
2	2019-04-01	2945.83	2178.67	3514.62	162.36	1079.24	8095.388	5237.87	-58.7	97.479
3	2019-03-01	2834.40	2107.74	3351.71	159.81	1058.13	7729.321	4072.16	-35.9	97.284
4	2019-02-01	2784.49	2085.84	3298.26	158.69	1050.95	7532.532	3802.53	-29.8	96.157

Commodity prices reflect prices of actual commodities traded in the market and signal the level of demand and supply in the real economy. Commodity prices give us an understanding of the underlying transactions in the economy that drive the engine of global productivity.

	Date	Gold	Silver	Oil	NatGas	Agriculture	Soybean	Corn	Wheat
0	2019-06-28	1409.55	15.3150	488.916	15.83163	354.7374	3579.369	74.3491	74.2323
1	2019-05-31	1305.58	14.5942	448.499	16.76547	353.5957	3504.901	75.9928	71.2631
2	2019-04-30	1283.53	14.9525	535.766	17.79443	322.1981	3403.133	64.3832	60.6207
3	2019-03-29	1292.38	15.1183	503.268	18.64069	333.5600	3567.618	64.6971	65.1003
4	2019-02-28	1313.32	15.6120	480.793	19.72403	338.9276	3665.369	67.1522	65.2221

Data transformations:

The data transformation is done in 5 stages to clean, streamline and organize the data to prepare for the data analysis. This process also make the data more consistent across different dataframes and allows us to access the data. Steps in the data transformation process include:

1. Read & clean the data
2. Rename the column (to have more pythonic titles) and organize the data
3. Creating bins for the date variable ordering them all to the first day of the month to enable merging the data between different dataframes
4. Setting date as the index and visualizing the data to understand economic trends
5. Create a new column to define a period of recession vs no recession (apply groupby on the data)

Data Cleaning:

The data is read from an excel file with multiple tabs. This data was originally downloaded from the Bloomberg data repository using a custom SQL query built for the platform. Each of the 4 tabs are specially setup to download data for each of the 4 dataset described above with the correct time series labels, characteristics (% change vs. price index levels) and frequencies (monthly vs. quarterly).

The data is read into 4 separate dataframes in a jupyter notebook and then each data set is cleaned to check for missing variables. Missing data points are excluded to ensure that the sample data correctly represents the economic effects. Filling missing data with '0' or any standard fill is incorrect treatment for financial data as it might impair the consistency of statistical information.

Data Merging:

Econ data is released quarterly, whereas stock prices, commodity prices and credit data are recorded with monthly, daily, and sometimes at miili-second frequencies. This can lead to data challenges when we analyse GDP data with stock prices or commodity prices to understand cross-sectional effects.

In order to overcome this challenge we bin the data by aligning the dates across quarterly and monthly data (Jupyter notebook has detailed explanations of this) and this allows us to seamlessly merge data across the 4 data sets. In cases where we merge data that is produced quarterly with monthly time series, we allow a static backfill to ensure consistency of data capture.

Section 1: Macroeconomic Data

Macroeconomics is the study of the behavior of the economy as a whole. While microeconomics looks at single factors that affect individual decisions, macroeconomics studies general economic factors. Macroeconomics is a complicated discipline, with many factors that influence it. These factors are analyzed with various economic indicators that tell us about the overall health of the economy.
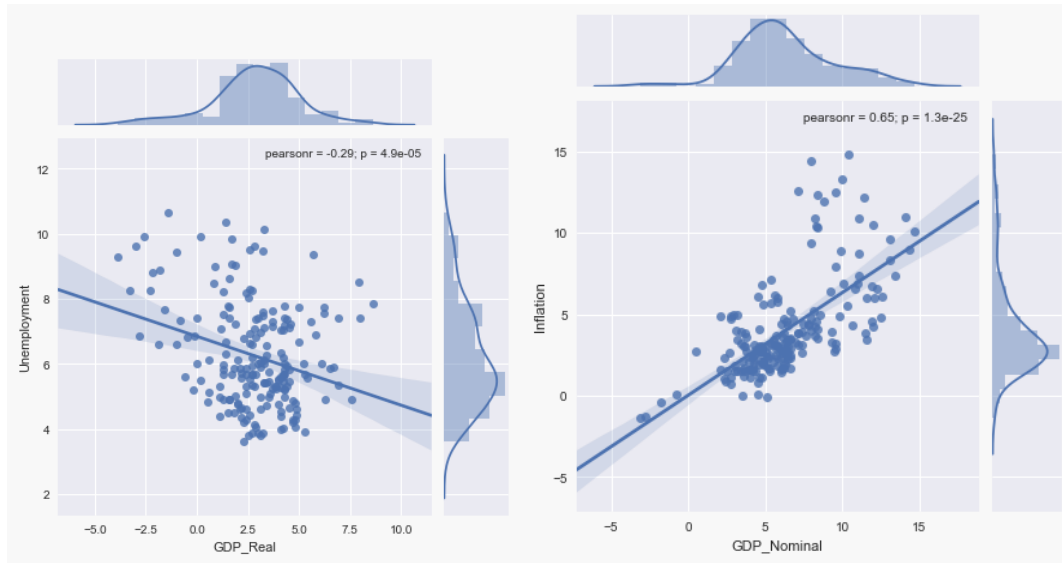
Gross Domestic Product (GDP) is one of the most important concepts in macroeconomics, it refers to the total amount of goods and services a country produces, commonly known as the GDP. This figure is like a snapshot of the economy at a certain point in time.

The Unemployment Rate tells us about the health of the labor force are individuals' ability to find work. The regression plot on the LHS (low p-value and negative correlation) shows that when the economy witnesses growth from period to period (indicated in the GDP growth rate), unemployment levels tend to be low. This is because with rising (real) GDP levels, output is higher and, hence, more laborers are needed to keep up with the greater levels of production.

Inflation is the third main factor macroeconomists look at which represents the rate at which prices rise. If nominal GDP is higher than real GDP, we can assume the prices of goods and services has been rising. The regression plot on the RHS shows the strong correlation between nominal GDP and Inflation with a strong positive relationship.

Lattice plot of GDP, Unemployment AND Inflation
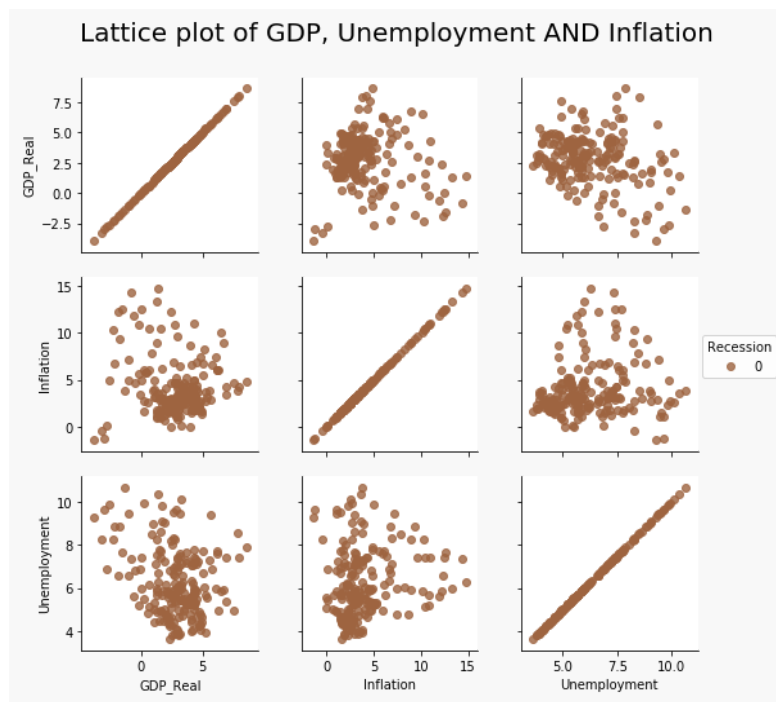
The lattice plot above captures the distributional information between the 3 core variables during recessionary periods (GDP < 0) and normal state of the economy. The strongest link is between GDP_Real and unemployment and between GDP and Inflation both showing opposing effects.

Section 2: Credit and Recession Data
### Are Economic Indexes correlated with Credit in recessions?

Credit is the exchange of buying power today for the promise to pay it back in the future with interest. In short time periods (5-7 years), credit plays a huge role in driving the economic growth as the increasing amount of credit makes it possible to drive consumption levels higher than income levels, fostering the expansion of the economy until it reaches its maximum level. The ability for municipalities and consumers in the US to purchase items using credit overall is a good thing. Loans fund the construction of bridges and railroads, allow consumers to purchase houses and finance college educations. This is turn leads to greater economic development, for when there is little credit provided there is usually little development. Too little credit/debt growth can also create economic problems coming in the form of foregone opportunities.

However, at a certain point the cycle inverts, and consumption and the economy slows down. This creates a problem with credit i.e. when there is too much of it and an inability to pay it back. When lenders are unable to recoup prior loans, this reduces future credit growth and even leads to a sharp decline in credit availability. The great recession of 2008-2009 is the most recent example of this phenomenon, mortgage defaults in the U.S. in 2007/2008 led to a contraction in credit availability in the economy, one of the leading causes of the great recession.

In our data, we looked at our dataset to see which credit metrics have the highest correlation to gdp growth and which appeared to be predictive of prior recessionary periods. In other words, can a change in credit be predictive i.e. was there usually a decline in credit prior to the other recessions (a period in which economic growth contracts for two periods). As apparent from the above heat map, GDP real has the highest correlation to leading indicators, and Consumer Credit YoY. We examine both these variables below.
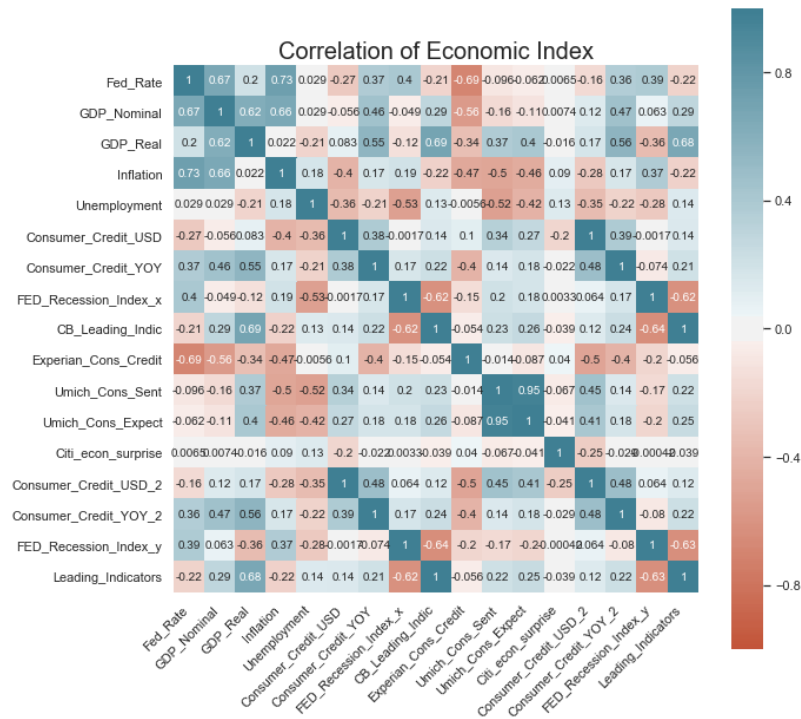
The conference board leading indicators index is a composite index of 10 leading economic indicators including economic, interest rates and credit metrics. This data had the highest correlation to GDP over a long period and appears to be predictive of recessionary periods. For example, the leading indicators index appears to dip down prior to all three known recessionary periods (1980s, 200s and 2008-2009) we are aware of in our dataset.
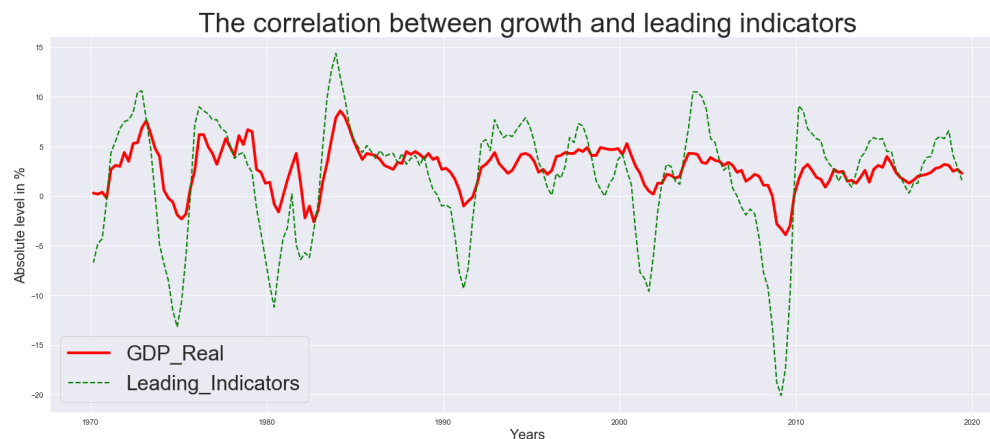
We then looked at consumer credit data. This data also had a very strong relationship to GDP over a long period. However, the data appears to be predictive of only certain recessionary periods. For example, credit data declines prior to the 1980s recessions and the 2007-2009 financial crisis. However, there was no decline in credit observed prior to the 2000s financial recession. While this recession was milder with a GDP decline from peak to trough of 0.3%, this recession was not "predicted" by a credit decline.

The correlation between growth and credit
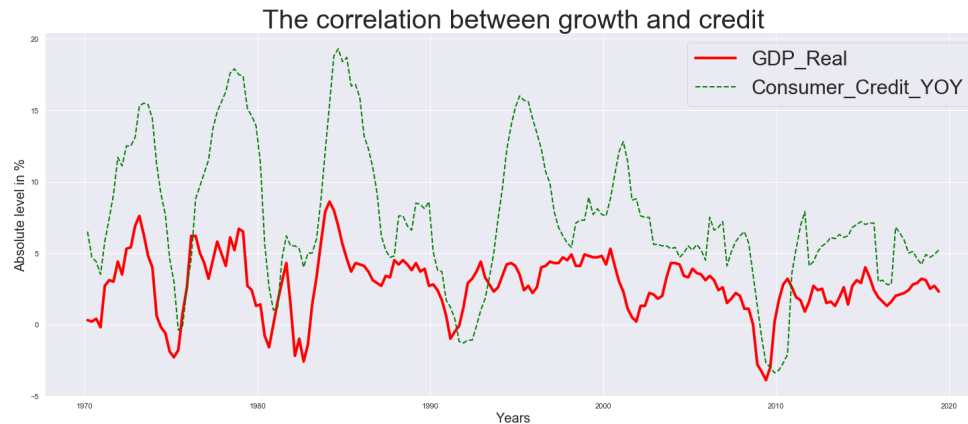
Deep diving into the 2008 crash, this is exactly what happened, an unprecedented monetary and credit expansion. However just as the credit growth starts to decline in 2006 we see a corresponding decline in GDP growth culminating in the financial crisis in 2009. The recovery of economic growth from the financial crisis also appears to occur just a credit growth begins to pick up again underscoring the importance of credit to economic growth particularly during the recent financial crisis.

The correlation between growth and credit before, during and just after the financial crisis
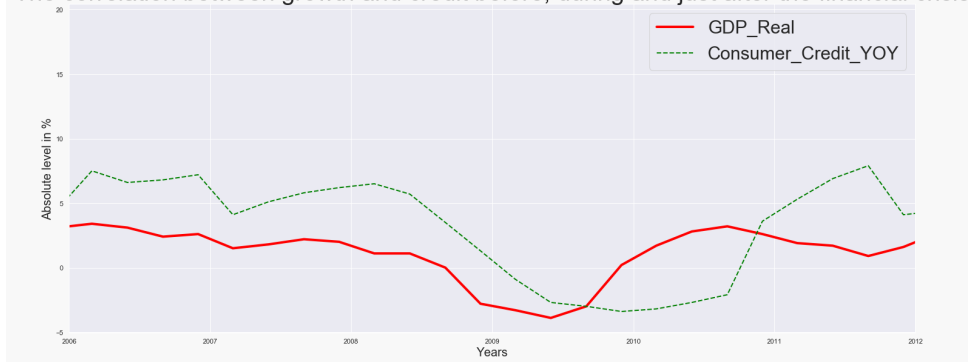
Section 3: Stock Market and Recession
## Are Economic Indexes correlated with Equity Indexes in recessions?

In the section, we focus on the trends of Nasdaq index where many dynamic technology companies (Apple, Dell, Microsoft, Cisco, Intel) are listed. Financial experts and economists carefully monitor changes in the Nasdaq composition index because they feel it gives a good indication of the present and future state of the US economy.

Our goal is to predict the recession before it's coming. We are curious which variables are related to the Nasdaq. Our heatmap shows the Spearman correlation of each variable, it points out that S&P 500 TR Index, MSCI World Stocks, EURO STOXX 50, MSCI APAC Stocks and bitcoin are more related to the Nasdaq, on the other hand, Citigroup Economic Surprise Index and US Dollar Index are less related.

Spearman Correlation of variables:
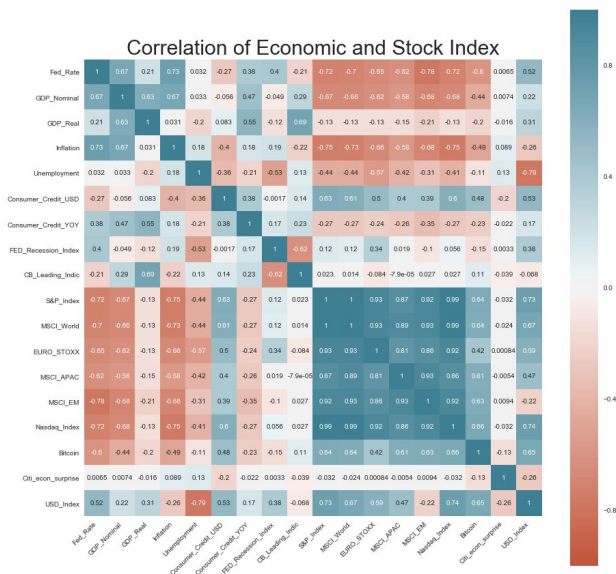
Our heatmap leads us whether these highly correlated variables could predict the trend of Nasdaq. Could we use these variables to forecast the Nasdaq index, including the recession? We apply random forest regression to characterize the trends of the Nasdaq, using highly correlated variables, including S&P 500 TR Index, MSCI World Stocks, EURO STOXX 50, MSCI APAC Stocks and bitcoin to train our machine learning model. Further investigating, we wanted to know which variables are more important to the Nasdaq index. We found these 4 indexes are more important for predicting the Nasdaq, including S&P 500 TR Index, MSCI World Stocks, EURO STOXX 50, MSCI EM Index. Our plot below shows the random forest regression
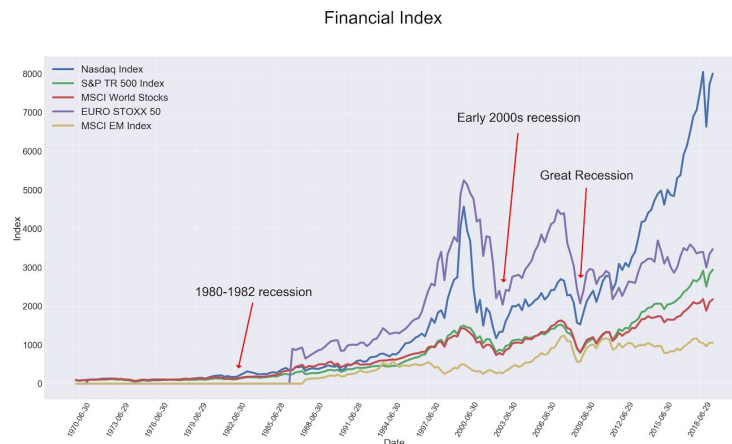
Predict Nasdaq from S&P 500 Index, MSCI World Stocks, EURO STOXX 50, MSCI EM Index

Random Forest Regression for Nasdaq
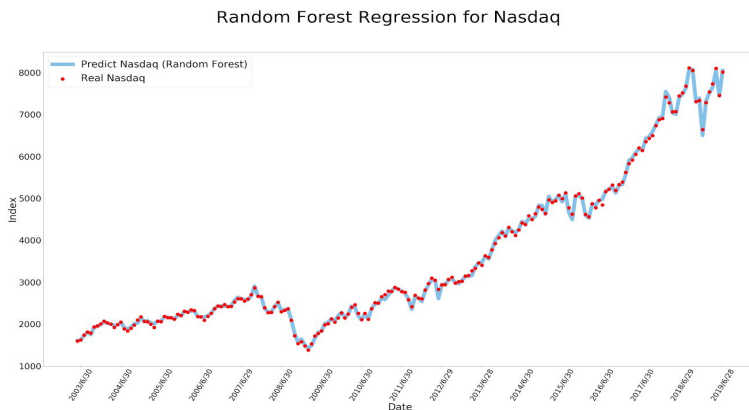
Section 4: Commodity Prices and Recession
Are Economic Indexes correlated with Commodities in recessions?

In section, historical data of commodity prices are collected, including, gold, silver, oil, natural gas, soybean, corn and wheat. In order to adjust the price, indices of them are used. From the heatmap of correlation of Economic Index and Commodity, we can see that GDP and Inflation is correlated with gold price index and WTI oil price. Similar correlation analysis is conducted between Stock Index and Gold and Oil prices as well, and MSCI World is highly correlated with Gold and Oil prices.

Correlation of Economic Index and Commodity
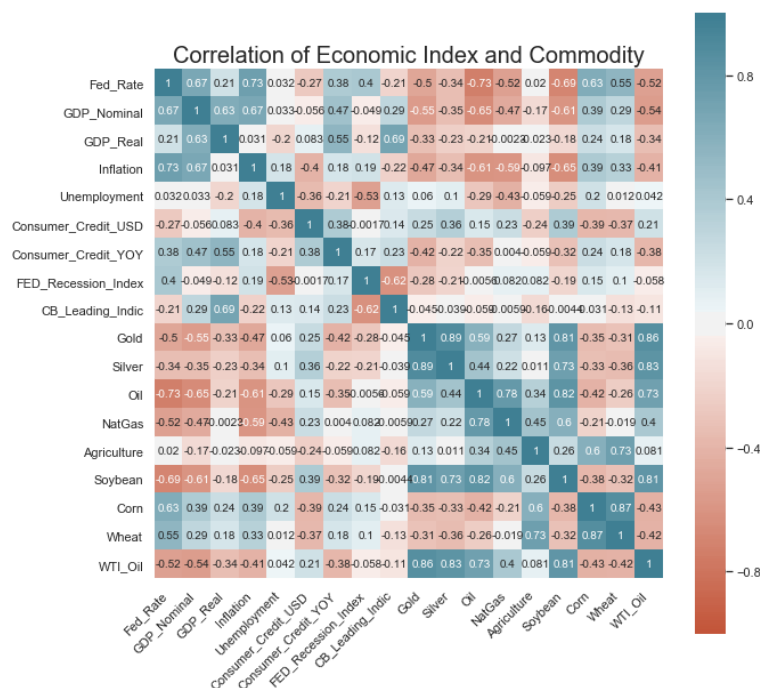
Gold is used as a counter-market securities tool during the recession time. That means people buy gold during the recession to avoid losing money. In general, gold price moves up during the recession. From char of gold below, we can see that from 1980 to 1981, there is a spike of gold price. Since the recession in 1980 is due to oil crisis, so there is a spike of inflation too. eIn contrast, the recession in 2008, when the gold price moves up, inflation goes down. The reason is that, people lost faith in the US economy at that time, so they buy gold to keep the value of their investment while inflation is low.

The correlation between Inflation and Gold Price Index
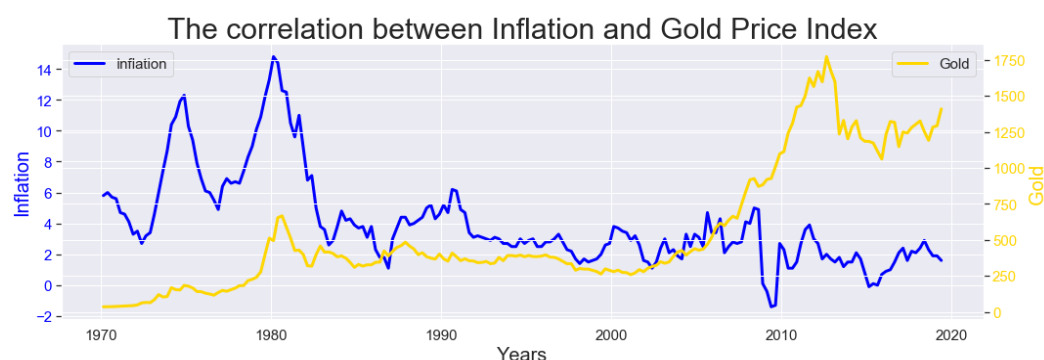
From the chart of inflation and WTI oil price, we can see that oil price moves up in 1980, when there is an oil crisis. Inflation moves up too. The reason is similar as the gold price trend moving. That recession is due to inflation, so the prices of commodities increase a lot, as well as inflation. In the last recession in 2008. We can see the oil price moves up very high, and inflation moves up just a bit. That is due to the same reason as explained in gold section, the whole economy is down, and that makes the oil price drops sharply in 2009. Demand of oil reduce a lot, so the oil price dropped.

The correlation between Inflation and WTI Oil Price
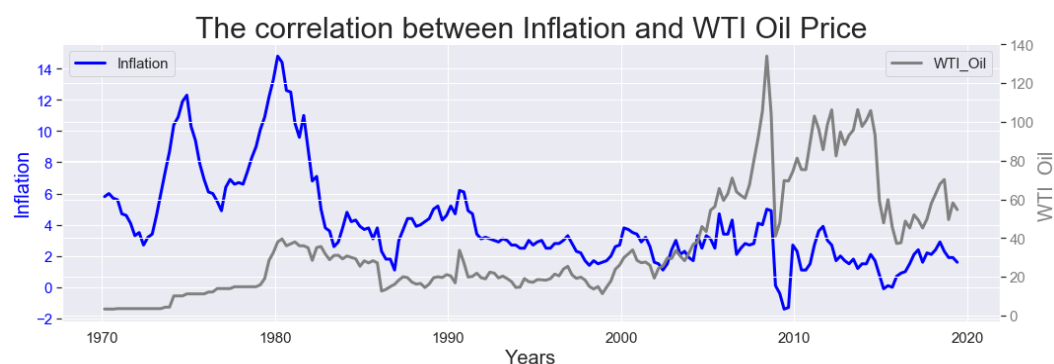
Conclusion

What we've attempted is a holistic analysis of financial markets data. Comparing economic data with financial market data we make statistical inferences which leads to a fundamental understanding of the economy as a whole. The visualization of data provided useful insights regarding causal relationships between variables. Overall, Python is a powerful tool for analyzing large financial datasets and panda dataframes are helpful in analyzing, exploring and visualizing time series data.