

SUBPRIME MORTGAGE CRISIS EDA

ABHISHEK YADAV

19b090001

Dept. of Mathematics, IIT BOMBAY

INTRODUCTION

In this project, we will focus on exploratory data analysis of banks stock prices. We will use pandas to directly read data from Yahoo Finance. The main objective is to analyze and visualize different features from the dataset to better understand the bank industry and how it behaves.

We will focus on bank stocks and see how they progressed throughout the financial crisis all the way to early 2020.

ABOUT DATASET

We will get stock information for the following banks:

- Bank of America (BAC)
- CitiGroup (C)
- Goldman Sachs (GS)
- JPMorgan Chase (JPM)
- Morgan Stanley (MS)
- Wells Fargo (WFC)

About the data:

- Stock data from Jan 1st 2006 to Jan 1st 2020.
- Six banks.
- 6 columns and 3523 rows.
- Source: Yahoo Finance

Feature Columns

- High: Is the highest price at which a stock traded during the course of the trading day.
- Low: Is the lowest price at which a stock traded during the course of the trading day.
- Open: Is the price at which a stock started trading when the opening bell rang.
- Close: Is the last price at which a stock trades during a regular trading session.
- Volume: Is the number of shares that changed hands during a given day.
- Adj Close: The adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions. Factors in corporate actions, such as stock splits, dividends, and rights offerings.

The following questions will be answered throughout the Project:

- What is the max Close price for each bank's stock throughout the time period?
- On what date did Citigroup stock reach its highest price?
- Did anything significant happen on 2009-01-20?
- Which stock would you classify as the riskiest over the entire time period?
- Which would you classify as the riskiest for the year 2015?
- What is the rolling 30 day average against the Close Price for Bank Of America's stock for the year 2008?

EDA AND CONCLUSIONS

RETURNS

- Created a new DataFrame called returns. This dataframe will contain the returns for each bank's stock.

Returns are typically defined by: $R(t) = (P(t) - P(t-1)) / P(t-1)$

Here, $P(t)$ is price of stock at time t .

- On analyzing the returns data for all 6 banks we observe that 4 of the banks share the same day for the worst drop.

```
In [96]: # Biggest single day Losses
         returns.idxmin()

Out[96]: BAC Return    2009-01-20
         C Return     2009-02-27
         GS Return    2009-01-20
         JPM Return    2009-01-20
         MS Return     2008-10-09
         WFC Return    2009-01-20
         dtype: datetime64[ns]
```

Did anything significant happen on 2009-01-20?

1. It can be seen that of the 6 banks, 4 had their worst performance on the same day.
2. President Obama took office on Jan. 20, 2009. ([More information here](#))
3. The [subprime mortgage crisis](#) also had a major part in the decline of prices.
4. Markets had little confidence in the economy and the future was uncertain.
5. The banking sector in general declining by 30%.
6. Bank of America Corporation (BAC) dropped 29%, and Citigroup Inc. (C) sank 20%.
7. The S&P 500 and the Nasdaq took similar hits on inauguration day, dropping 5.3% and 5.8%, respectively.

STANDARD DEVIATION OF RETURNS

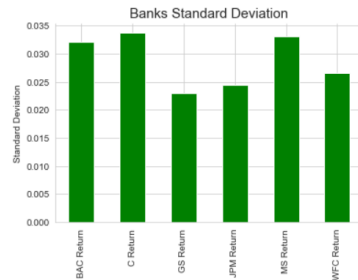
Standard Deviation: Is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

Standard deviation is the statistical measure of market volatility, measuring how widely prices are dispersed from the average price. If prices trade in a narrow trading range, the standard deviation will return a low value that indicates low volatility. Conversely, if prices swing wildly up and down, then standard deviation returns a high value that indicates high volatility.

Basically, standard deviation rises as prices become more volatile. As price action calms, standard deviation heads lower.

Which stock would you classify as the riskiest over the entire time period?

```
In [13]: # Standard Deviation
returns.std().plot(kind='bar', color='Green')
plt.ylabel('Standard Deviation')
plt.title('Banks Standard Deviation', fontsize=15)
sns.despine()
```



Looks like the two most riskiest stocks are Citigroup and Morgan Stanley.

Closing price vs Time

Let's create a line plot showing Close price for each bank for the entire index of time.

```
In [17]: # Option N°1
for tick in tickers:
    bank_stocks[tick]['Close'].plot(label=tick, figsize=(12,6))
plt.legend()
```

Out[17]: <matplotlib.legend.Legend at 0x1b3277718c8>



- Now we can clearly see the crash of Citigroup in 2008.
- Goldman Sachs also had a crash in 2008 but bounce back quite quickly after the recession.

Moving Average

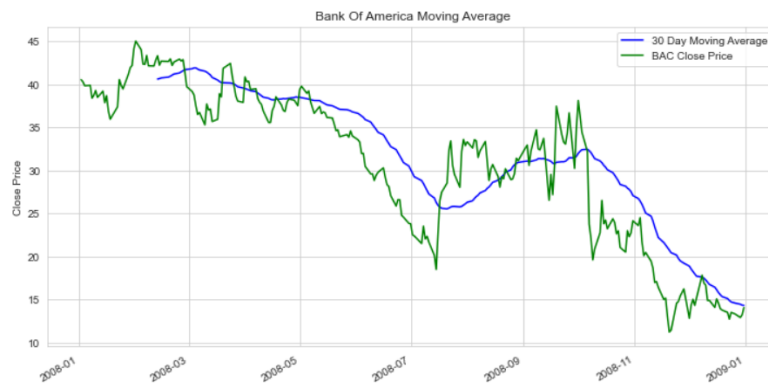
The [moving average \(MA\)](#) is a simple technical analysis tool that smooths out price data by creating a constantly updated average price. The average is taken over a specific period of time, like 10 days, 20 minutes, 30 weeks or any time period the trader chooses. The most common applications of moving averages are to identify trend direction and to determine support and resistance levels.

Let's analyze the moving averages for these stocks in the year 2008.

What is the rolling 30 day average against the Close Price for Bank Of America's stock for the year 2008?

```
In [111]: plt.figure(figsize=(12,6))
          BAC['Close'].loc['2008-01-01':'2008-12-31'].rolling(window=30).mean().plot(color='blue', label='30 Day Moving Average')
          BAC['Close'].loc['2008-01-01':'2008-12-31'].plot(color='green', label='BAC Close Price')
          plt.ylabel('Close Price')
          plt.xlabel('')
          plt.title('Bank Of America Moving Average')
          plt.legend()
```

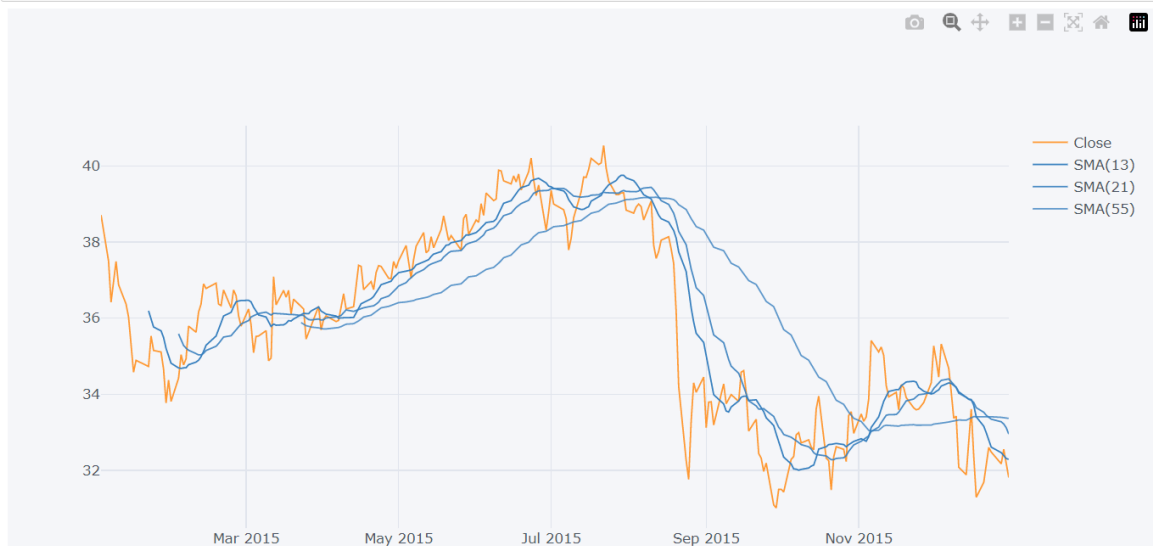
Out[111]: <matplotlib.legend.Legend at 0x1b32c7903c8>



Simple Moving Averages

A [simple moving average \(SMA\)](#) calculates the average of a selected range of prices, usually closing prices, by the number of periods in that range. Is a technical indicator that can aid in determining if an asset price will continue or if it will reverse a bull or bear trend.

```
In [23]: MS.loc['2015-01-01':'2016-01-01']['Close'].ta_plot(study='sma', periods=[13,21,55])
```



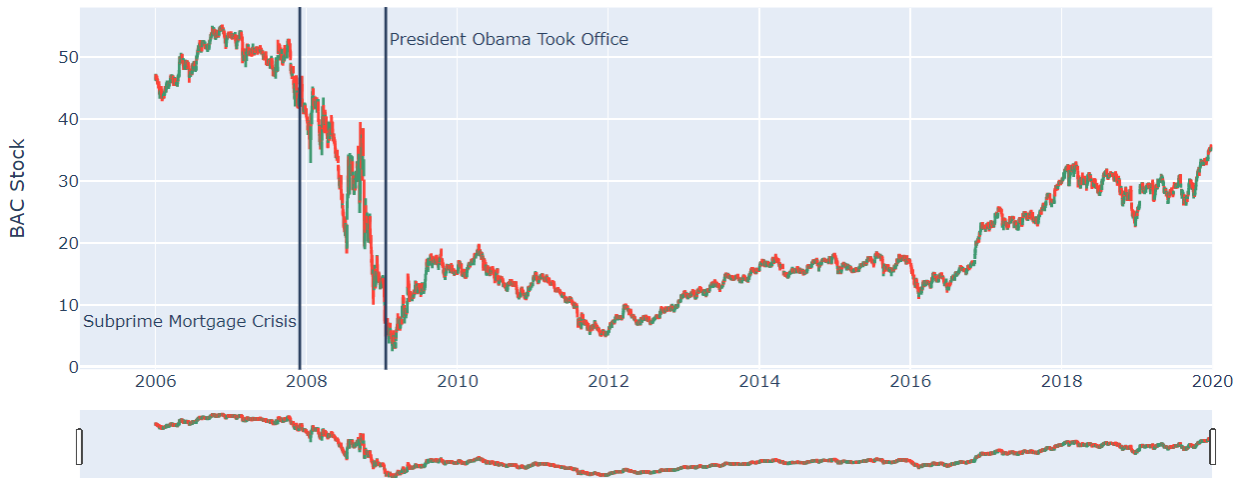
Simple Moving Averages plot of Morgan Stanley for the year 2015

Candlestick Charts

Documentation: [Candlestick Charts in Python](#)

The candlestick chart is a style of financial chart describing open, high, low and close for a given x coordinate (most likely time). The boxes represent the spread between the open and close values and the lines represent the spread between the low and high values. Sample points where the close value is higher (lower) than the open value are called increasing (decreasing). By default, increasing candles are drawn in green whereas decreasing are drawn in red.

Plot of Bank of America stock price in Candlestick format :

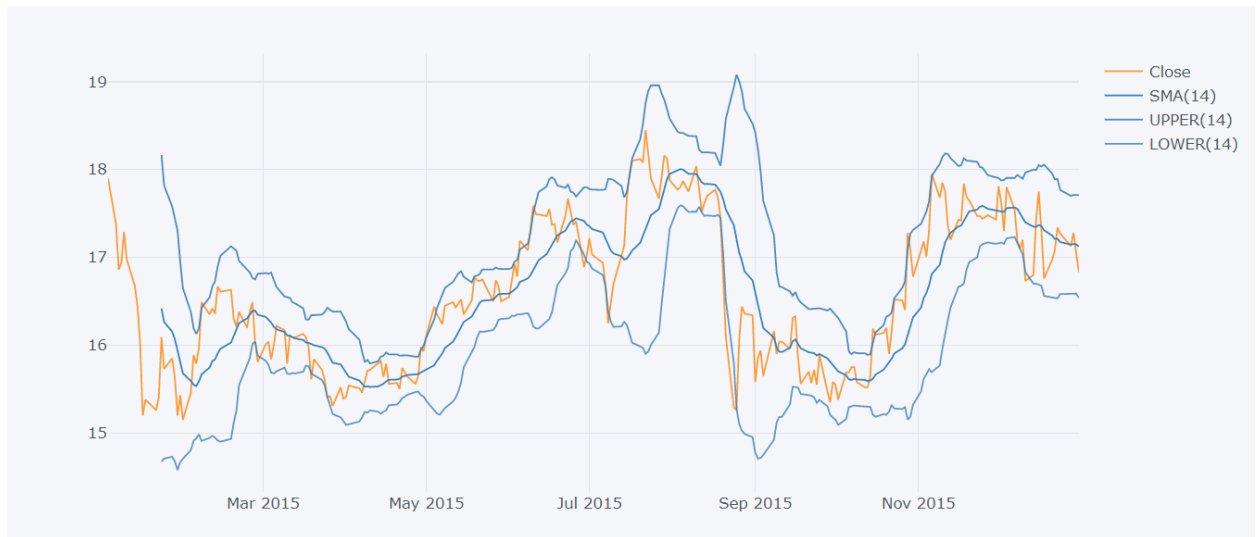


Bollinger Band

A [Bollinger Band](#) is a technical analysis tool defined by a set of trendlines plotted two standard deviations (positively and negatively) away from a simple moving average (SMA) of a security's price, but which can be adjusted to user preferences.

- When the bands tighten during a period of low volatility, it raises the likelihood of a sharp price move in either direction.
- When the bands separate by an unusual large amount, volatility increases and any existing trend may be ending.
- Prices have a tendency to bounce within the bands' envelope, touching one band then moving to the other band. You can use these swings to help identify potential profit targets.

Bollinger Band Plot for Bank of America for the year 2015



Bank Facet plot

Documentation: [Time Series and Date Axes in Python](#)

Facet plots, also known as trellis plots or small multiples, are figures made up of multiple subplots which have the same set of axes, where each subplot shows a subset of the data.

