# CAPSTONE PROJECT - 1 REPORT ON

# Data Analysis on Bank Stocks

**CONTENTS**

|  |  |
| --- | --- |
| **S.No.** | **Topic** |
|  | Chapter 1: Introduction   * Abstract * Project description   Objective of the study |
|  | Chapter 2: Tools and Skills Used |
|  | Chapter 3: Source Code and Output Snapshots |
|  | Chapter 3: Conclusion |
|  | Chapter 5: Future Scope |

**CHAPTER 1: INTRODUCTION**

**ABSTRACT**

* The Great Recession was the sharp decline in economic activity during the late 2000s. It is considered the most significant downturn since the [Great Depression](https://www.investopedia.com/terms/g/great_depression.asp). The term Great Recession applies to both the U.S. recession, officially lasting from December 2007 to June 2009, and the ensuing global recession in 2009.
* As name itself suggest, in this project we will be mainly focusing on how certain bank stocks behaved, what their trend was and how they progressed throughout the financial crisis all the way to early 2016.
* In this project I have done Exploratory Data Analysis and incurred visualization through seaborn and matplotlib and done manipulation with data using pandas

## PROJECT DESCRIPTION

* In this project I’m going to work on the data cleaning and exploratory data analysis. The visualization is done on bank stocks progressed throughout the financial crisis all the way to early 2016.
* Visualizations are done using seaborn and matplotlib. This data has been gathered from Yahoo finance using ***pandas-datareader.***

### This project is divided into 2 parts:

* **Data Gathering and preparing it for the visualization part:**
  + The dataset was not readily available so I have to extract data from the internet specifically from yahoo finance using pandas datareader.
  + The dataset was a bit messy as it was a Multi Index dataframe so I did some manipulation to make it more accessible for EDA and visualizations.

### Draw the necessary plot for the visualization:

* + Now as the data is structured and cleaned, we can now jump onto the plotting part. In this part we have to plot different graphs like line chart, Bar plot, pair plot etc.

# CHAPTER 2:

**TOOLS AND SKILLS USED**

## TOOLS AND SKILLS

### TOOLS AND PLATFORM:

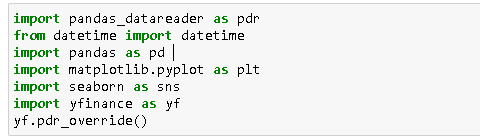
* + Anaconda IDE- (python distribution) anaconda is a free and open source distribution of the python and R programming language for data science and machine learning related application (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment.
  + Jupyter Notebook- the Jupyter notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualization and narrativetext.
  + Python 3.8.3:
* Seaborn – it is a python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
* Numpy – it is the fundamental package for scientific computing with python. It contains among other things:
  + A powerful N-dimensional array object.
  + Tools for integrating C/C++ and Fortran code
  + Useful linear algebra, Fourier transforms and random number capabilities.
* Matplotlib – it is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
* Pandas – library written for python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulation numerical tables and time series.

# CHAPTER 3:

**SOURCE CODE AND OUTPUT SNAPSHOTS**

## SOURCE CODE AND OUTPUT SNAPSHOTS

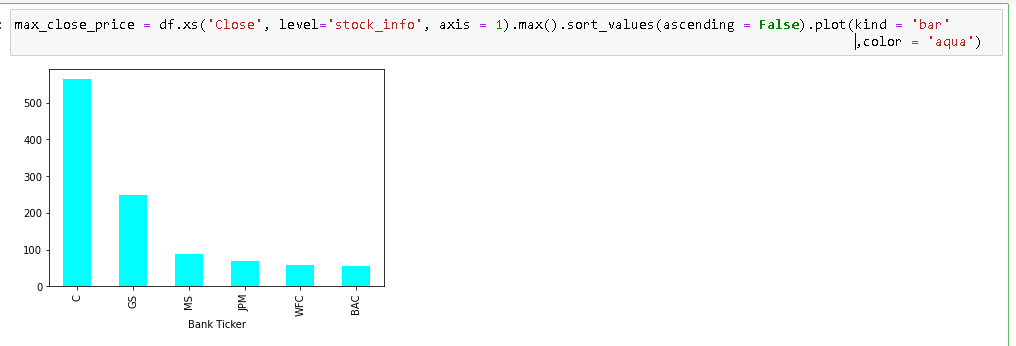
* IMPORTING LIBRARIES



* IMPORTING DATASET AND GLIMPSE OF FIRST 3 ROWS



### What is the max Close price for each bank's stock throughout the time period?



### Create a new empty DataFrame called returns. This dataframe will contain the returns for each bank's stock. Returns are typically defined by:

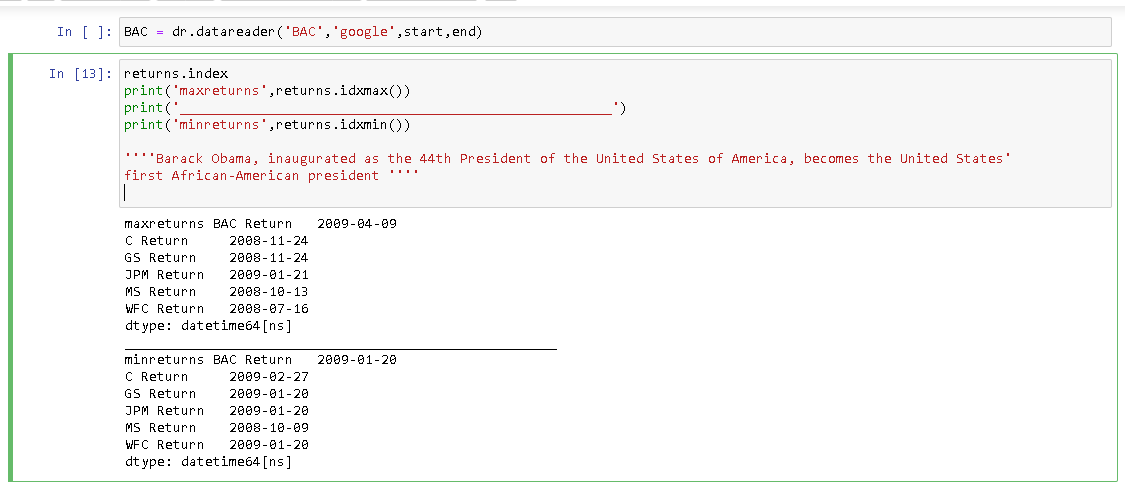
### C:\Users\Anshal\Desktop\sadas.PNG

### Create a pair plot using seaborn of the returns dataframe. What stock stands out to you? Can you figure out why?[¶](http://localhost:8888/notebooks/Documents/Stock%20data%20Analysis.ipynb#d.)-Create--a--pairplot--using--seaborn--of--the--returns--dataframe.--What--stock--stands--out--to--you?--Can--you--figure--out--why?)

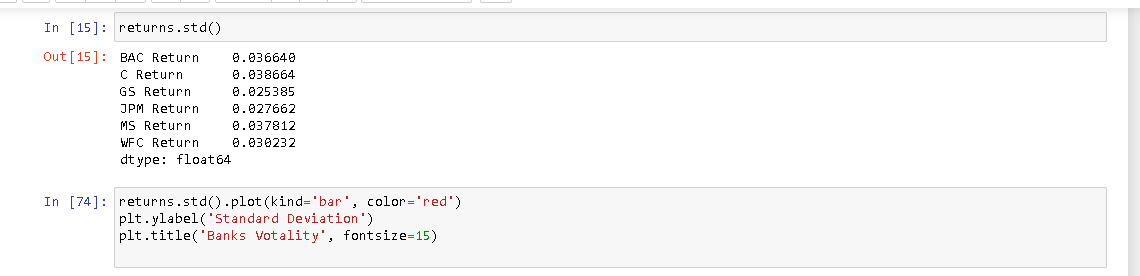
### C:\Users\Anshal\Desktop\sfdfds.PNG

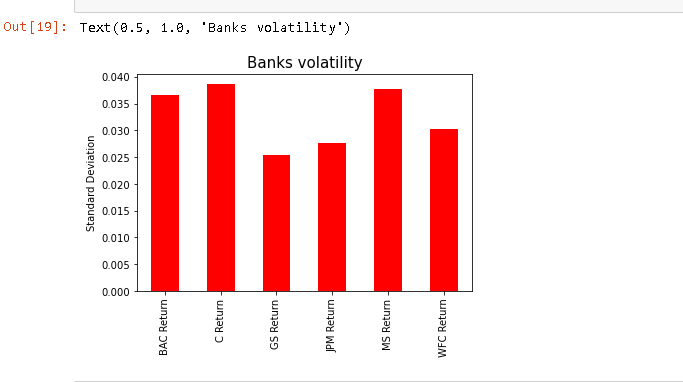
### C:\Users\Anshal\Desktop\gdfd.PNG

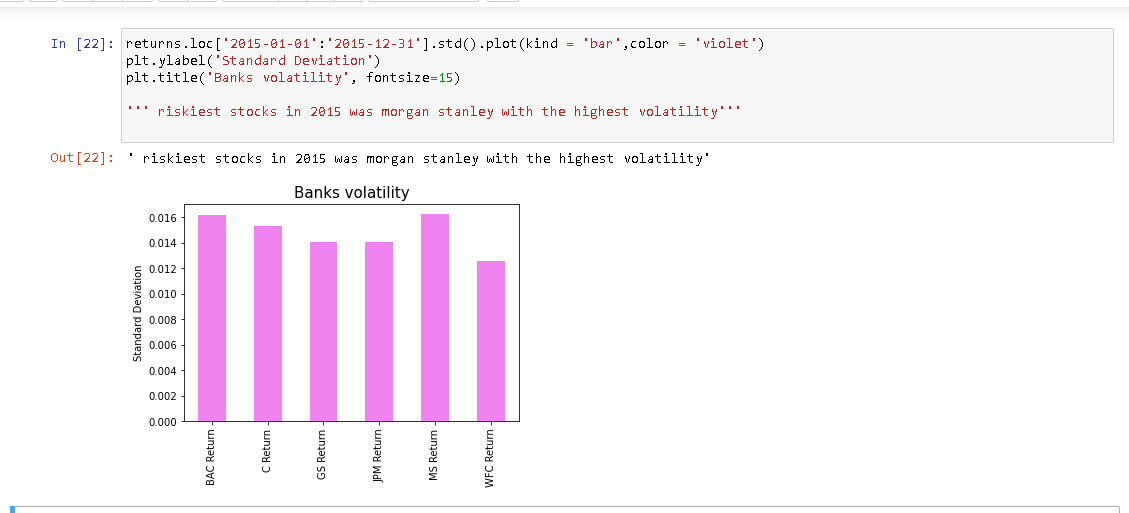
### Using these returns DataFrame, figure out on what dates each bank stock had the best and worst single day returns. You should notice that 4 of the banks share the same day for the worst drop, did anything significant happen that day?



### Take a look at the standard deviation of the returns, which stock would you classify as the riskiest over the entire time period? Which would you classify as the riskiest for the year 2015?







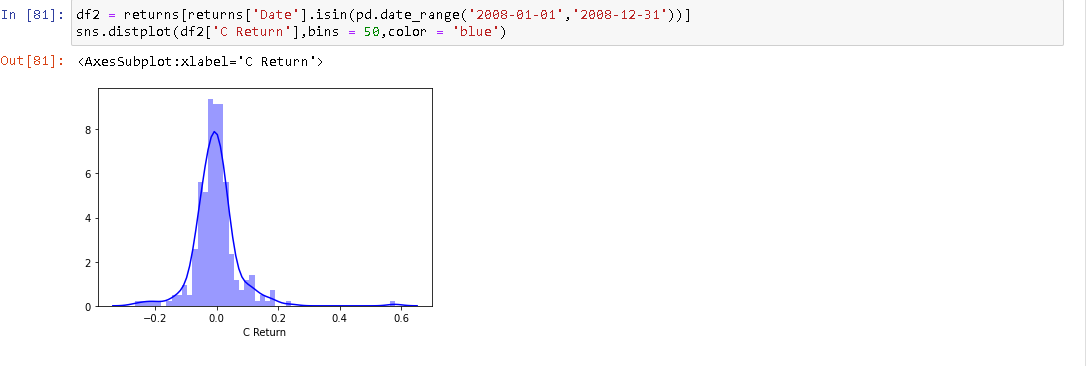
### Create a distplot using seaborn of the 2015 returns for Morgan Stanley

### C:\Users\Anshal\Desktop\dsa.PNG

### Create a distplot using seaborn of the 2015 returns for citi group-

### C:\Users\Anshal\Desktop\sda.PNG

* Create a ditsplot using seaborn for the 2008 returns of citi group



### Create a line plot showing Close price for each bank for the entire index of time. (Hint: Try using a for loop, or use .xs to get a cross section of the data.)

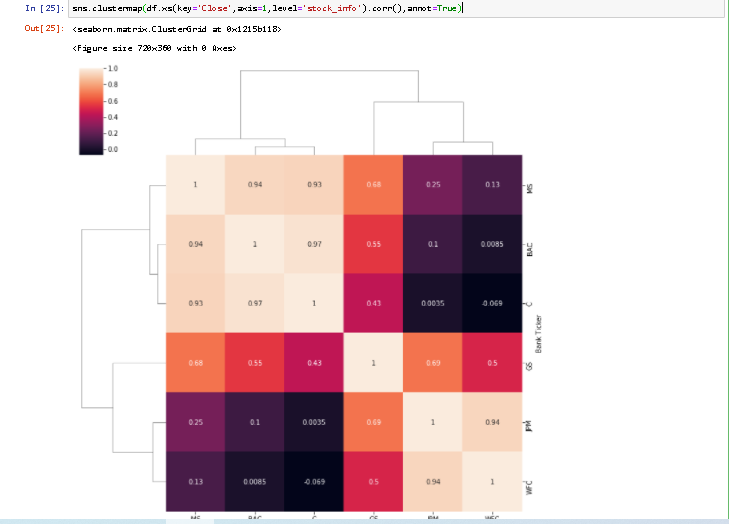
### C:\Users\Anshal\Desktop\dasdasdasdasd.PNG

### Create a heatmap of the correlation between the stocks Close Price

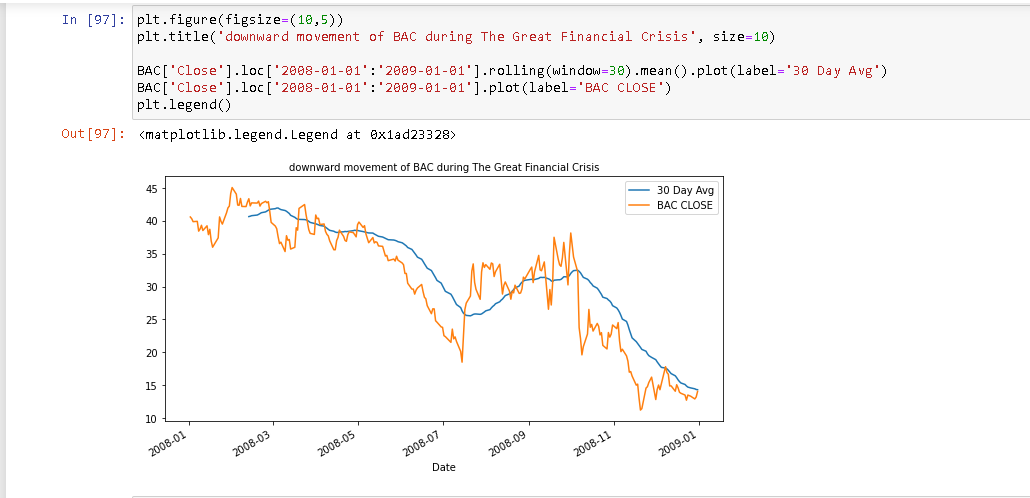
### C:\Users\Anshal\Desktop\dasdasd.PNG

### Use seaborn's clustermap to cluster the correlations together

### C:\Users\Anshal\Desktop\dasdasdasrf.PNG



* Plot the rolling 30 day average against the Close Price for Bank of America's stock for the year 2008.



# CHAPTER 4: CONCLUSION

## CONCLUSION

The results of the analysis helped us to understand how the Great Recession affected the stocks at that time. While working on this project learned many new things like how to handle multi index data and how gather data from the internet etc.

The findings were pretty amazing, as all the stocks belongs to banking sector they should very strong relationship among them and if someone owned all these stocks he/she were the worst affected people. The best stocks among all of the others was Goldman sachs as it was able to stand and was able to maintain its price well among its competitors’ before and after the great recession.

# CHAPTER 5: FUTURE SCOPE

## FUTURE SCOPE

With more sophisticated data analysis we can incur answers to much more business question or I should say stock market related question.

In US nowadays almost 60% of trading is done by bots , we can implement sophisticated machine learning and deep learning algorithms to predict future price of a stock and if we are able to implement them in a correct way then the results from these algorithms will be much more accurate than humans.