DSC520: STATISTICS FOR DATA SCIENCE
BELLEVUE UNIVERSITY

FINAL PROJECT

A Closer Look at Common Types of Investments: Bank Stocks versus Growth Stocks

By: Tai Ngo
Date: 2/27/2020

INTRODUCTION

In our modern world, there are countless ways that we can use and spend our money; however, the best way that undisputedly grows our money legally over a life-time is investments. When it comes to personal investments, it becomes another world with endless possibilities. The purpose of this project does not aim to display a strategy or sell a way to invest, but relies on actual data to gain educational knowledge on common-type investments. After reading through this paper, the readers should be able to tell which type of given investments will produce the highest and lowest amount of return on investment over a fixed period of time. Other factors such as investment costs and duration will also be taken into account. The datasets for this project consist of many CSV files, which include the bank stocks and the growth stocks. For the bank stocks, the datasets consist of four of the most popular banks in the US: Bank of America, JP Morgan, Goldman Sachs and Wells Fargo. The dataset for the growth stocks consists of Adobe, Salesforce, Shopify and RingCentral.

PROBLEM STATEMENT

The problem of this project is to figure out a definitive answer to the question between a more-stable investment such as bank stocks and a more-risky investment such as growth stocks, what is a better option in a fixed amount of time such as three months? This project will answer this question by using real data and visually show the audience by using boxplots.

METHODOLOGY

The problem is addressed by using the following datasets, the links to the datasets are attached below:

- Bank Stocks - https://www.kaggle.com/rohan8594/stock-data
- Growth Stocks - https://www.kaggle.com/fjgonzalez/growth-technology-stocks

After the CSV files from each dataset are imported, the data is cleaned and prepared for the analysis. Basic data visualization techniques such as plotting are used to represent the data. The calculations of the mean and the percent difference are performed to compare the stocks. The boxplots are used to show the mean of the high values and low values of each stock.

For the bank stocks, I have chosen to use the datasets for four popular banks: Bank of America, JP Morgan, Goldman Sachs and Wells Fargo. For the growth stocks, I have chosen to use the datasets of Adobe, Salesforce, Shopify and RingCentral.

INTERESTING INSIGHTS

1. It may be a common belief that defensive stocks such as bank stocks have a low rate of return on investment, but what exactly is the rate of return on investment for bank stocks? Across all four popular bank stocks (Bank of America, JP Morgan, Goldman Sachs, Wells Fargo), the percent differences between the average of the high values and the low values are 65.9%, 61.2%, 63.0% and 59.4%, respectively. Although the percent differences for these bank stock are very close, they definitely show that their value can go up a lot from the financial crisis in 2008 to early 2016. These are also the expected raise if investors decided to buy them during the crisis. In my personal opinions, these percent difference or percent increase is very low compared to the S&P500; however, if we factor in the dividend and how safely they are, then they may be a good investment for older people.

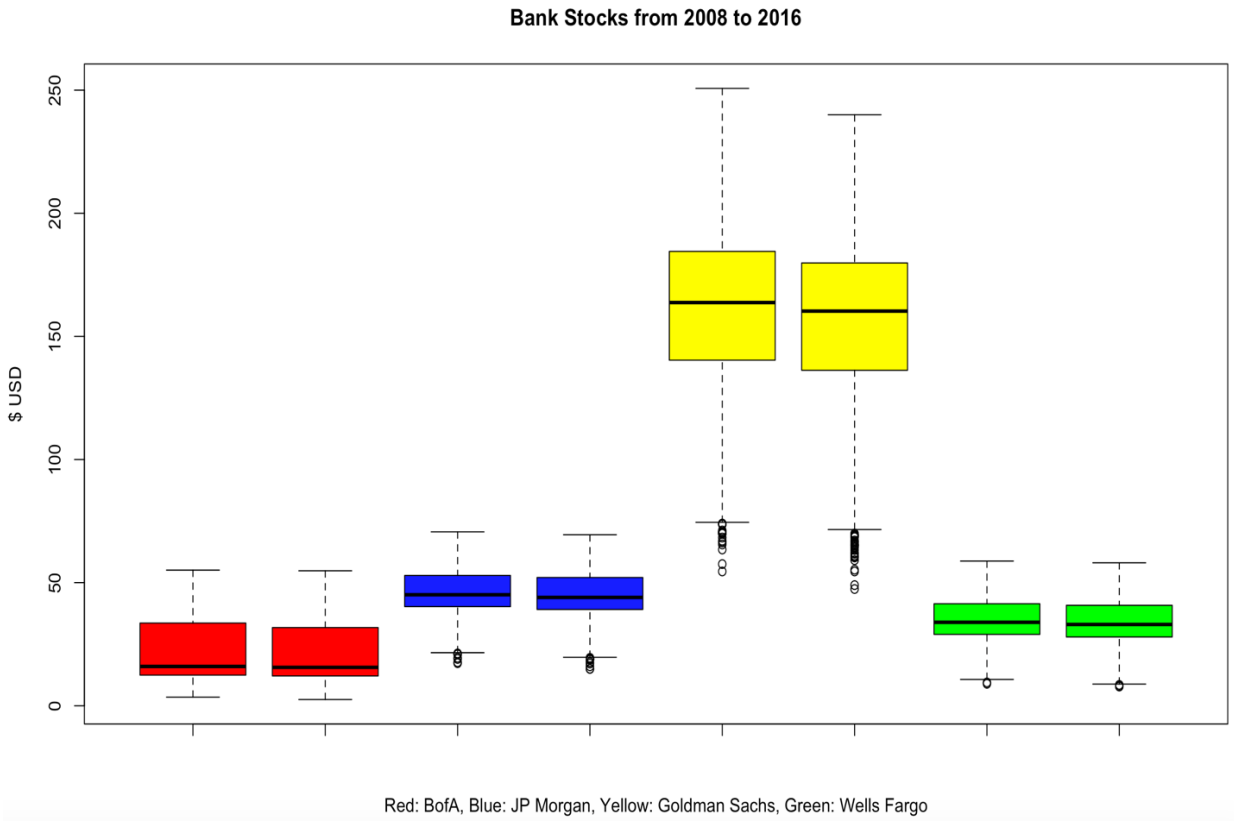

Figure 1. The boxplot of the bank stocks. These boxes represent the range of the high values on the left and the low values on the right.

2. As shown in Figure 1, the highest price and the lowest price fluctuates very little on a daily basis for each bank stock. This suggests that as we buy a bank stock, we can expect its value to move very slowly unless there is a special event such as earning report or abnormal event such as a Covid-19 outbreak.
3. Although the price of the yellow box (Goldman Sachs) shows a much bigger range and higher price than the other three stocks, its characteristics are the same. It has the same level of price fluctuation

on a daily basis and the percent increase from the year of 2008 to 2016. The only difference is its price tag is more expensive.

4. When it comes to growth stocks, people should expect to see wild rides in their portfolio. The percent differences between the average of the high values and the low values are 47.6%, 55.0%, 87.7% and 78.9% for Adobe, Salesforce, Shopify and RingCentral, respectively. These are the percent changes in less than three months. On a side note, during the period of data collection for these stocks, the market was heading into the trade war, which is the main reason why such a big difference occurs.

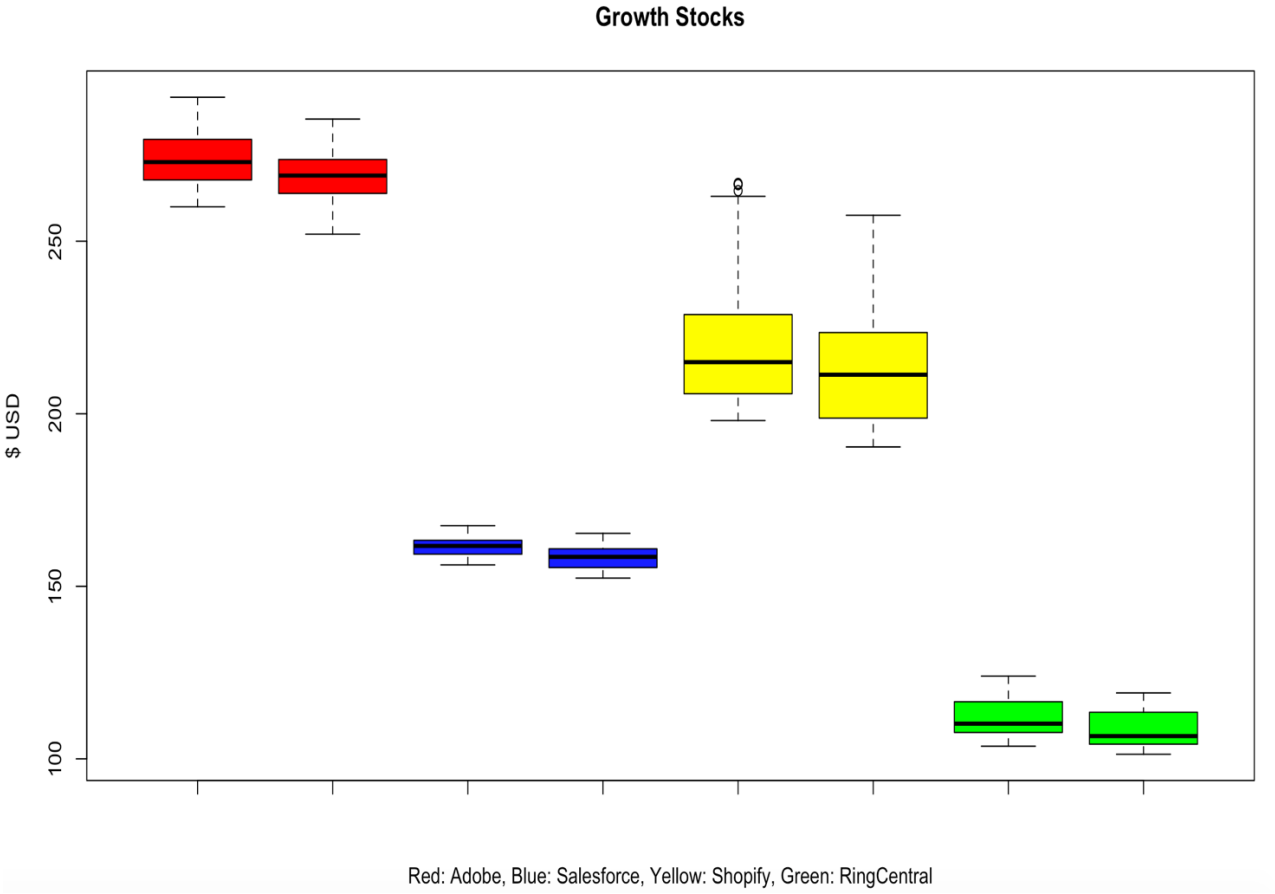

Figure 2. The boxplot of the growth stocks. These boxes represent the range of the high values (box on the left) and low values (box on the right).

5. When it comes to choosing a growth stock for our portfolio, it is a lot more difficult than choosing a bank stock. As displayed in Figure 2, Shopify has much better raise and higher fluctuations than the rest. The second best is the Adobe stock, but the price of this stock is also a lot higher.
6. Understanding the nature of the growth stocks: high price fluctuation and high variation between growth stocks, it suggests that we need to spend more resources (time, research) into finding a good stock. In addition, due to the high volatility, it may be a wise decision to keep these stocks for the short term and be ready for any incoming downfall as its value can drop significantly.
7. As the growth stocks face a lot of volatility in price, it is more suited for aggressive investors such as young investors who want to grow their portfolio more quickly and are willing to take more risks.

PROJECT IMPROVEMENTS

Although the datasets do not cover the entire market, its four representatives for each type of investment show sufficient results that can be drawn and provide enough information to discuss about their natures, strengths, weaknesses and characteristic. With this knowledge, investors have more resources to make better financial decisions in regards to their investments. To improve the project or further validate the results, more data of the stocks should be used. Deeper technical analysis should be performed to validate the results. If possible, machine learning models can be applied to predict the trends. If we continue to dig more into the data, we should be able to find the best bank stock as well as the best growth stock although these kinds of information only hold true for a very short period of time.