

SHINY APP REPORT

The Dataset

For this project, I decided to compile the historical data from 80 different companies from January 2008 till the end of September 2019. When creating the list of the 80 companies I made sure to record the industry, the sector, the country, and the stock exchange in which they are listed. Based on this list, I then downloaded the historical data for all these companies individually from Yahoo! Finance, meaning that I collected 80 different CSV files.

In order to design a clean dataset with categorical variables, my job has been to add columns to each of these files. On each one of them, I added the country, the ticker, the stock exchange, the sector, and the industry thanks to my previous list.

All my CSVs were complete, which meant I could now merge them together. To do so I used an R command:

```
##{r}
setwd('/Users/adriencoucaud/Desktop/Finance Datasets')
getwd()
filenames <- list.files(path = "/Users/adriencoucaud/Desktop/Finance Datasets", full.names=TRUE)

All <- lapply(filenames,function(i){
  read.csv(i, header=TRUE, sep=";")
})

df <- do.call(rbind.data.frame, All)

write.table(df, 'Full_dataset.csv', row.names=FALSE, sep=";", dec = ".")
##
```

With my dataset ready, I now had to specify where were my categorical variables (factor) as well as formatting the date.

By this time I already started to work on my shiny dashboard and building a prototype. However, later I played around my dataset on Tableau and discovered that I should add one more variable: Return.

I used Excel and built a formula that calculates the daily return based on the Adjusted Close $[(n-1 - n)/n]$ but only if the former row belongs to the same company, avoiding to calculate a return between two different companies.

Apple Inc.	NasdaqGS	24/09/2019	217.679993	-0.004755
Apple Inc.	NasdaqGS	25/09/2019	221.029999	0.01538959
Apple Inc.	NasdaqGS	26/09/2019	219.889999	-0.0051577
Apple Inc.	NasdaqGS	27/09/2019	218.820007	-0.004866
Credit Agricole S.A.	Euronext	02/01/2008	12.194887	0
Credit Agricole S.A.	Euronext	03/01/2008	12.17353	-0.0017513
Credit Agricole S.A.	Euronext	04/01/2008	11.938276	-0.019325
Credit Agricole S.A.	Euronext	07/01/2008	11.820648	-0.009853

This variable was key because it allowed me to perform more operations, find other insights and calculate the volatility (STDV[Return]).

With my dataset_2.0, I started to work on new ideas of graphs for my dashboard as I was now able to tell a new story.

The Dashboard

Purpose

Since I started this class, I knew I would want to tell a story related to the world of Finance.

Why?

Because this industry has always triggered my curiosity and interest. I also took the opportunity given by this class to add a Finance project to my portfolio which will be useful if I ever want to apply for jobs in banks or asset managers.

It seemed mandatory to select a relevant timeframe which is why the story begins in 2008 and ends in Sept 2019, somehow the end of the chaos generated by the **US-China trade war**. The sectors and companies chosen are giants from the past or from today and it will be interesting to dig down and study what happened throughout the time.

One crucial information to remember is that the dataset is not a loyal representation of the economical horizon of both this time period, the countries selected or even the sectors.

At the core of this limit is **the survivor bias**: even if we see stocks performing poorly we only see the best stock which survived until now. The data and graphs for the Financial Services Sector would have surely been different if they took into account the Lehman brothers' crash in 2008.

It must be clear that only companies that I liked or found interesting to analyze made it to the final dataset and they are **not a representative sample**.

So what is the purpose? - Once again, having fun!

Prior to the start, I had a list of hypothesis and questions to prepare the investigation:

- How badly did the Financial Sector crash in 2008?
- The Financial Sector never recovered from the Subprime crisis.
- The biggest growth came from the Technology Sector.
- The American market is the most volatile because they are willing to take more risks.
- Which company generated the best return?
- Which Stock exchange choose if you are risk-averse?
- What was the impact of the 2008 crisis?
- What was the impact of the trade war?

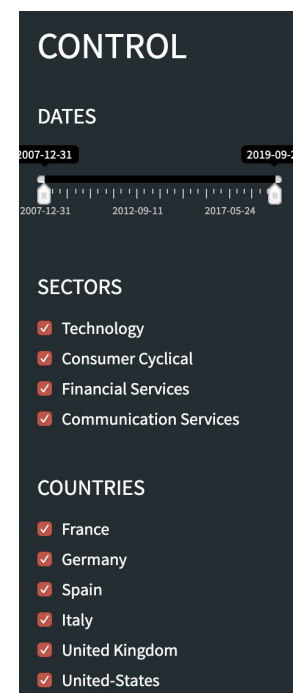
The dashboard should be designed as a tool allowing users to answer questions, and find insights. We shall now review the features of the dashboard!

Features

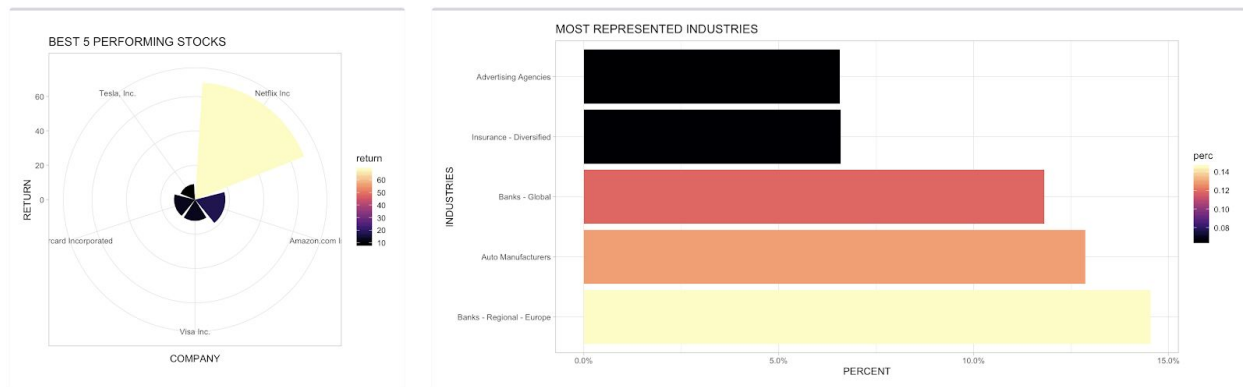
To simplify everything and make the app as intuitive as it can be, I built a single control panel where the user can maneuver and change the dataset to perform different analyses. Having such a feature allows the user to go more and more in-depth filtering the sectors and/or the countries to have a smaller sample analyzed.

To keep it simple again the dashboard has two tabs each performing a different analysis. Launching the app the user arrives on the "General" tab which will give him an overview of the dataset (according to the filters activated).

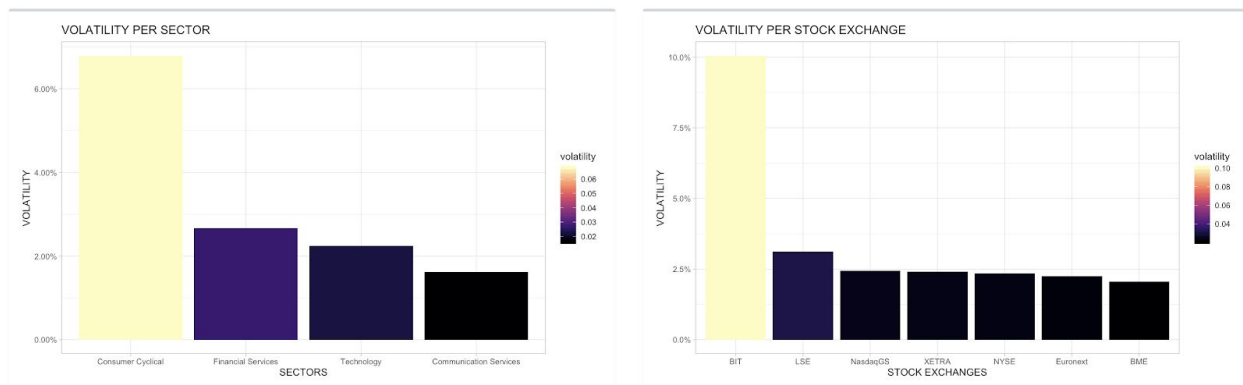
For instance, the top of the tab is comprised of 3 value boxes telling how many companies are analyzed with the current filter, which company performed the best and the worst during the time frame selected.



The first graphs show two TOP 5, first is the five best stocks based on return and then the most represented industries in the sample. This way you can see if maybe the best-performing stocks belong to the most represented industries, or you can have a glimpse of your sample population by reviewing the industries analyzed by the dashboard.



The second layer of graphs provides information concerning the volatility. The user can analyze how the volatility evolved in percent.



- For the graph on the left, the user should play with the time slider and the country filter to see how the sectors' volatility is impacted if you remove the most volatile country of the dataset (Italy).
- In the second one, the user should play with the time slider and the sector filter to see how maybe one Sector being highly volatile influenced the overall volatility of a stock Exchange. The investigation could find out "how to make the Nasdaq less volatile and less risky?"

As its name states it, the first tab is for general information and for the user to understand how the economy was, based on its filters: the big performers, the population composition, and the riskier countries or sectors.

After reviewing the general tab, the user might want to have a deeper analysis of a company's performance. The second tab is designed to offer a focus on the best performing stock identified in the first tab, giving you more details on why or how this stock performed well compared to the others.

The tab has three value boxes reminding the user which company is analyzed and showing what were the best and worst daily return this stock knew during the time frame selected.



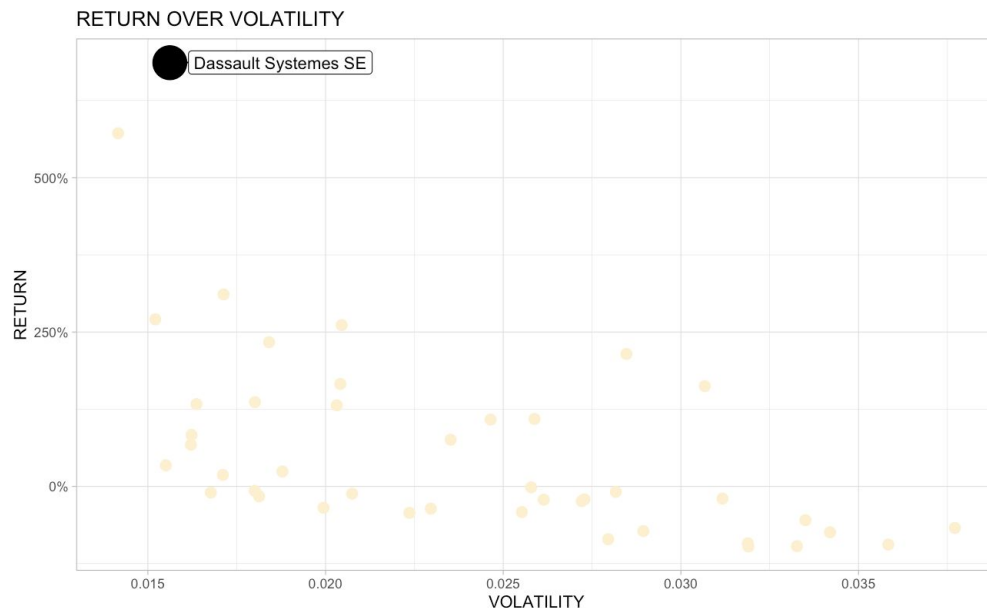
Then come, according to me, the most interesting graphs of the dashboard, giving a detailed vision of the stock performance both individually and also compared to the other companies of the sample.

First, the user can see the price evolution of the stock and understand visually why it obtain the best return performance among all the stocks. Every point on the graph symbolizes the Daily Adjusted Close.



Moving to the next graph which shows the best-performing stock compared to the others. What is interesting in the second graph is what is being measured: Return and Volatility.

It allows the user to see if a stock had a high return with low or high risk, what is interesting is to question if you would have been ok to bet on such a risky stock or if you would have stepped back and never take the bet. In the case of Dassault Systèmes, the stock was not really volatile compared to the others, so less risky.



Finally, to better understand the concept of volatility, the user has the opportunity to review the standard distribution of all the daily returns performed by the stock.

The flatter the normal curve is, the more volatile the stock is, and reverse.

To summarize, the whole dashboard is fully interactive and responds to the user's inputs on the control panel, enabling the user to investigate and perform an analysis.

For the story, someone even said to me that I built a "mini Bloomberg Terminal", which made me laugh but also flattered me to the highest point.

What to do next?

MANY things should be done on this app, but I have a list of the main improvements I wish I had been able to do, or that I will do later in the year.

- At the moment my Dates are filtered as an interval, which forced me to perform a product of returns instead of my original $\frac{r[2]-r[1]}{t[1]}$. My issue was that doing $\frac{r[2]-r[1]}{t[1]}$, I needed a vector with two time inputs while my volatility needed an interval. Another point is that the vector of time caused a lot of trouble to my visualizations as if the slider was not positioned on a day when the market was open then some graphs became all empty.
- On my second tab, when I state the best and worst daily return, I wish I had been able to plot the day it happened.
- I would like to make changes in the font, but I faced a lot of troubles with that and just abandoned.
- I would like to make the dashboard responsive, as for now, it will only be nice on a desktop. Please do not open the mobile version.
- The overall look of my dashboard can be improved.
- Have a simpler data collection using (**quantmod package (via getSymbols())**). It's **super easy**) enabling me to have more companies and thus a more representative sample.

I am missing a lot of improvements and updates ideas I thought about during the building process of my dashboard but I will surely keep working on it as it is the most advanced project I have on my data science portfolio at the moment.

[USE THE APP HERE](#)

ADRIEN COUCAUD