# Data Visualization – Assignment 1

## ***Project Description***

1. Describe the chosen topic of interest and explain why this topic is interesting/important.

Our topic of interest is **Financial/Capital Markets**.

This topic is interesting because of the sheer amount of available data and the existing connection between numerical figures, sociological behaviors, and the state of our economies at any given time.

1. Provide the name and a description of the dataset, highlighting its relevance to the chosen topic. Also, briefly describe how the data was collected.

The dataset we collected is a snapshot of most important macroeconomic indicators throughout the past few years, alongside with some predominant stocks’ key markets information and fundamental corporate data.

**Macroeconomic indicators selected**: Inflation rate, Policy Rate, Unemployment Rate, Change in Money Supply, VIX, …

**Stocks selected**: Apple (AAPL), Meta (META), J.P. Morgan (JPM), Nike (NIKE), Nvidia (NVDA), & Alibaba (BABA).

**Markets indicators**: Price, Volume & potentially option data (depending on availability).   
Also will be included some categorical data (sector of a stock, for e.g. Technology, Consumer-Discretionary, etc…)

**Fundamental factors:** Sales, Production, Free Cash Flow, Interest Expense, Net Income…

**Other assets** : Gold, Copper, 2Y US Bond Rates, 10Y US Bond Rates

All data will be gathered from the **Bloomberg Terminal** & **Yahoo Finance!**

The period of analysis has not been decided yet, but most likely from 1 year before Covid to today.

1. What are the attributes in the dataset? What are the types of each variable? Is there any missing or outlier data? Do you need to transform the data before the analysis?

Most variables are numerical data.

Some categorical ones.

Some transformations will make them categorical, for example, whether the Federal Reserve in the U.S. has increased or not its policy rate.

Many transformations will be required.

Ghali will take care of the feature engineering process.

1. Include the URL link or reference to access the dataset:

Yahoo Finance! : Python package.

Bloomberg Terminal : Special access.

## ***Exploration***

Four sub-topics I have thought about:

1. Dynamic Factor Correlation Analysis : how does the correlation between different asset classes evolve over time? This will be better visualized under a graph network.   
   The idea is to select, say, one of our stocks’ price return (i.e. change in the stock price over, say, 1 month), and try to correlate it with changes in other variables, in this case: macroeconomic data, other asset classes’ returns (Gold, Copper, some government bond rates, etc…). I will provide you with all the needed data.   
   The edges of the graph should be the correlation force, and the size of the node should be the return size.
2. Time Series Clustering of Financial Markets.   
   In this topic, I essentially try to find technical market sub-structures and categorize them. The idea is to identify market topologies at any time and any time frame, and see whether, varying lag factors, a ‘fractal’ analysis can be performed.
3. Sector-analysis :   
   Essentially, visualize the proportion of stocks in the S&P500 (an index where the 500 most valuable US companies (The Standard and Poor's 500, or simply the S&P 500, is a stock market index tracking the stock performance of 500 of the largest companies listed on stock exchanges in the United States) that are up or down under a sector-based analysis.   
   The idea here is to say, over the past month, investing in the Pharmaceutical sector was more profitable than the Industrial sector.
4. Basic stock-return analysis:   
   Out of our selected stocks, create a basic equally-weighted portfolio and visualize the return.   
   The idea is to show, should you have invested $1,000 last month, how much you would have this month, and show the proportion of return per stock.   
   A pie chart can be a good idea here.