Bootcamp ETL Project

# Finding Data

We scouted several webpages to retrieve information for Amazon’s stock price, we saw that amazon shares trade at different stock exchanges in different countries.

We want to download the data for Amazon’s stock in different countries to try an analyze the spread of price between countries and exchange rates (FX).

For this we used two main sources of free financial information that have the information we seek available:

* Yahoo Finance
* Investing.com

Both webpages have the information available to download in CSV format (We got happy because simplicity is always better).

# Data Cleanup & Analysis

**Extract:**

**The extraction was made manually by downloading from the selected sources (Yahoo Finance and Investing.com) the CSV file from the selected stock, we repeated this process for the different stocks for amazon offered in different exchanges (US / NASDAQ, Switzerland / SIX, Mexico / BMV and Germany / Xetra).** The resulting information is given in 4 different currencies: USD, EUR, CHF y MXN

The final CSV files are 4, the data from the US was extracted from Yahoo Finance, the rest were extracted from investing.com

**Transform:**

1. Load the 4 CSV files extracted from both sources
2. Clean the information from each file
   1. Dates: We transformed the dates from string to date format
   2. Price: transform the price from string to float, including the price, open, high, low
   3. Volume: transform to float but also change the format because the volume column had a letter indicating the abbreviation of the number (K for thousands and M for millions)
   4. Change: transform the change day over day from the stock selected from string to float and remove % symbol.
   5. Exchange, Currency and Country: This were left the same as these values are string and were available in both datasets
3. Our database is now completed and ready to make the queries

**Load:**

We loaded the information to MongoDB, our database is a non-relational database and includes the price for the amazon stock traded at different exchanges (US, Mexico, Switzerland and Germany)

After the information was loaded in a database from MongoDB and then loaded it again and downloaded the final database in a CSV

# Conclusions

We chose this data because we were curious if investors from different countries are paying the same amount for the same amazon stocks in different countries, this because of the spread of FX Rates and also to spot if there is any additional difference in the price apart from the exchange rate (currency).

To analyze this data further we need to check the difference between the US stock vs each other stock (Mexico, Germany and Switzerland) to see the spread between the price and the implicit FX Rate, then compare it to the real exchange rates to see if the difference is explained with the exchange rate or there are other underlying factors that affect this price.