Final Report - Group Ky and Peter (Yen)

Topic: 7 technology companies (tech giants) and their acquisition history and how it affects their stock price value/performance throughout each acquisition year. In the end we only use 1 tech giant: Google and its stock price.

**Extract: Source data**

Source 1 - Tech Giants Acquisition Companies Data - (Kaggle)

Link: <https://www.kaggle.com/anshumoudgil/apple-google-facebook-via-graph-patterns/data>

The data containts Google, Microsoft, IBM, Yahoo, Apple, Facebook and Twitter. The acquisition.csv (Excel file) has the acquisition date, company that got bought, and parent company bought (acquired) it.

Source 2 - Google Stock Price Data - (Yahoo)

Link: <https://finance.yahoo.com/quote/GOOG/history?p=GOOG&.tsrc=fin-srch>

We found that yahoo finance contains stock price data since Google’s IPO 8/19/2004 - 11/11/2019.

**Transform:**

**Import two CSV file (acquisition.csv and GOOG csv) from Resources folder into Pandas with file path.**

**Turn both data into Pandas dataframe.**

**Basic finding is that: each of “ParentCompany” has acquired how many (“count”) company in the past**

Google 215

Microsoft 210

IBM 162

Yahoo 114

Apple 95

Facebook 67

Twitter 53

Name: ParentCompany, dtype: int64

**Filter by ParentCompany: Google** (focus on Google instead of the other 6 companies)

Convert three columns that indicate the “date” into one Date column by import datetime and timedelta.

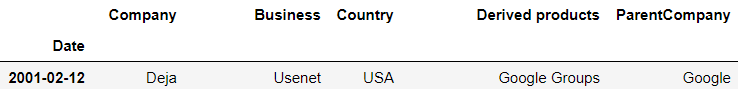
Create a date\_str\_list and use For Loop to striptime the acquisition\_year, mont, and date into date format such as 2001-02-12 (the first acquisition by Google, company name Deja). Create new data frames and keep the columns we want and drop the columes we don’t want. We called the first dataframe **google\_acquisitions\_df.**

We want the first and second data set to be able to join together, since we are interested in the date that a Company was acquired, versus Google stock price of that date, 30 days after, 60 days after, to make an observation on the acquisition behavior affecting the general confidence of the stock price.

We called the second dataframe **google\_df**.

We run into two issue:

First, we are yet to figure out how to add 30 days on top of the acquisition date.



Second, it turns out we also need to have the date align, but first dataframe has 204 date, while the second google stock price dataframe has 3838 date that we need to somehow mix them together.

We were able to drop the index column and replace it with our Date.

Our goal was to set **acquisition date** and **30 days after acquisition date** and **60 days after acquisition date** by datetime and timedelta.

**Load: Destination data**

We choose to Load the data into a SQL database Postgres by launching pgAdmin4. We want to merge and join both dataframe from Pandas, and SQL is more convenient in joining two separate tables together. We run into another issue of Loading all the data to SQL, the tables in SQL are created, and connection is established, but SQL seems to reject the format of our Pandas dataframe, not sure why.

We are able to query the table’s column name such as acquired\_date, acquired\_company…

