ANALYSIS OF START-UP COMPANIES

**Group 2**

Ashraf Adel 196280

Farah Aymen 194233

Aissatou Aida 197346

Table of Contents

[Introduction 2](#_Toc98820678)

[Datasets 2](#_Toc98820679)

[CSVs Used 2](#_Toc98820680)

[Features (Columns) 3](#_Toc98820681)

[Samples 3](#_Toc98820682)

[Acquisitions 4](#_Toc98820683)

[Additions 4](#_Toc98820684)

[Companies 4](#_Toc98820685)

[Investments 5](#_Toc98820686)

[Rounds 5](#_Toc98820687)

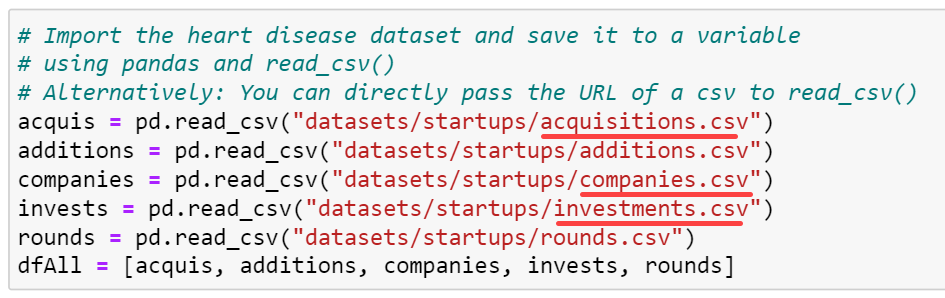
[Goals 5](#_Toc98820688)

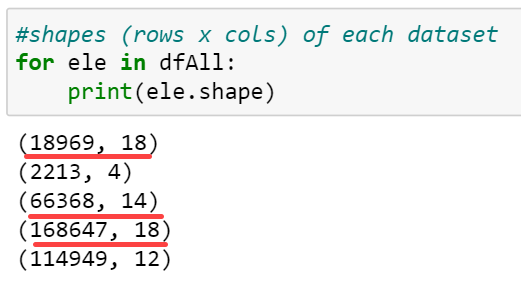
# Introduction

This project is planned to use a data set on the acquisitions and funding of start-ups. It is aimed to focus on the analysis of the dataset to deduce the reasons behind the success and failure of a start-up. The goal from the analysis is to deduce how much funding on average does a start-up usually require, Reasons behind their failure in case of dissolving or reasons behind their success. The dataset is based on 168647 start-up companies, while some of these start-ups were acquired by other companies, specifically 18969. Moreover, 168647 investments were made in those start-ups. The dataset consists of 5 files: acquisitions, additions, companies, investments and rounds. Each file will contribute to reaching those goals. First, the acquisitions file shows the companies that got acquired, acquired by whom and when, and how much it was acquired for. Second, the companies file shows all information about the start-up companies including their name, location, whether operating or not, and its funding information. Lastly, the investments file shows which company was invested in by which company, how much was paid for it, and the type and amount of funding.

# Datasets

## CSVs Used

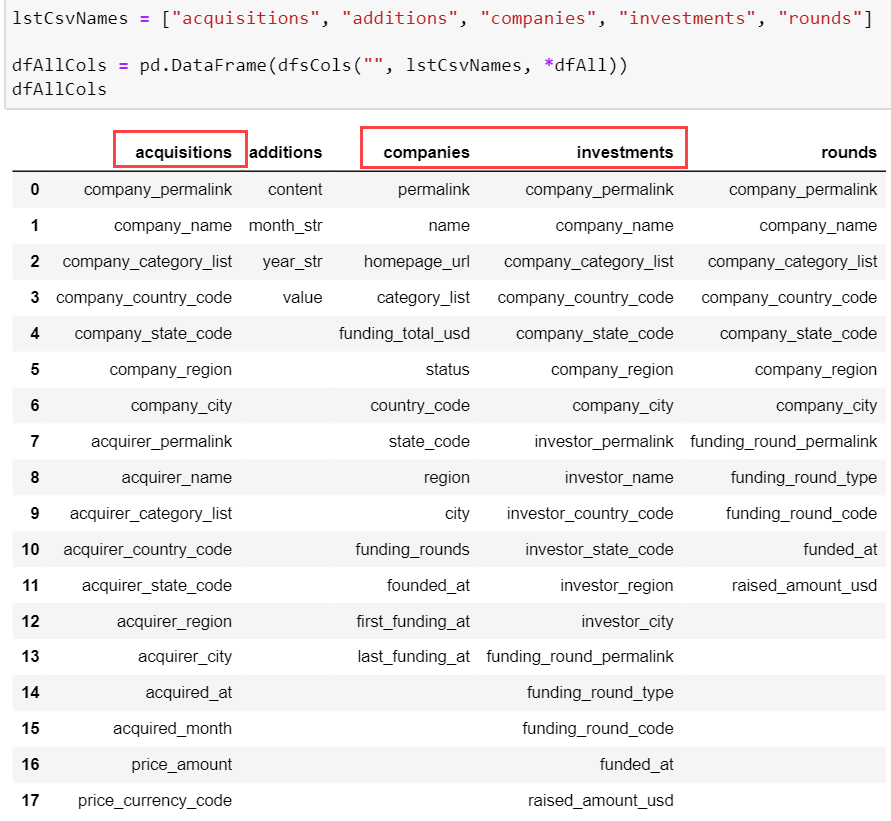




The analysis will mainly focus on the acquisitions, companies, and investments data sets while the others mostly contain redundant values that are already covered in the main datasets.

Python was used to represent samples of the data in a clear manner.

## Features (Columns)

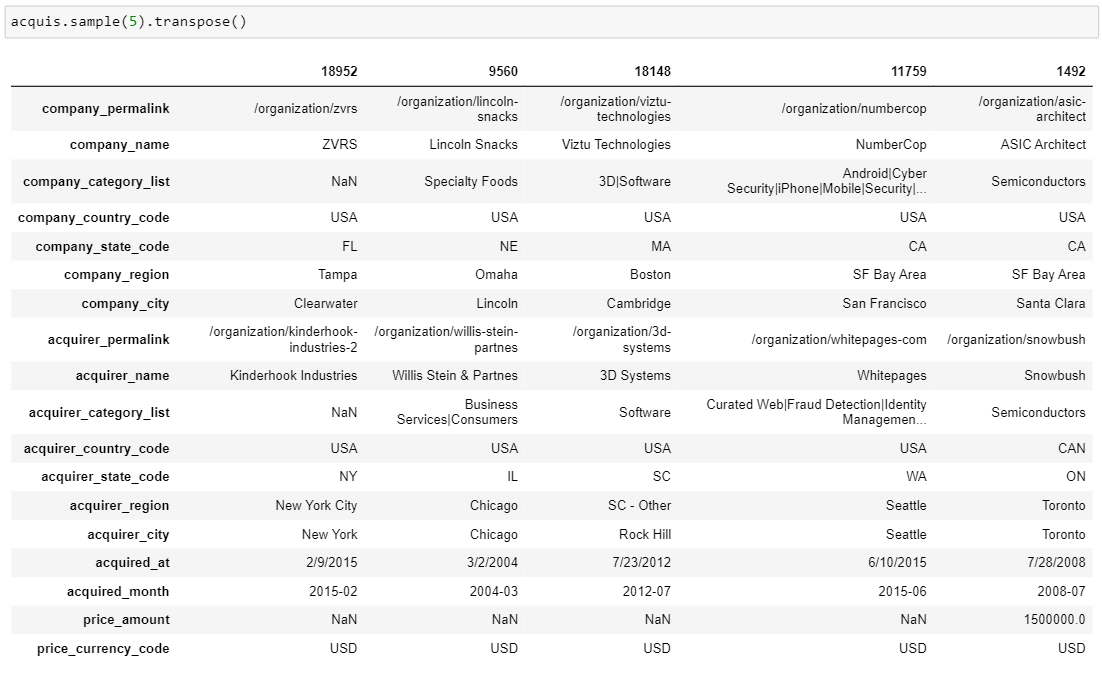


## Samples

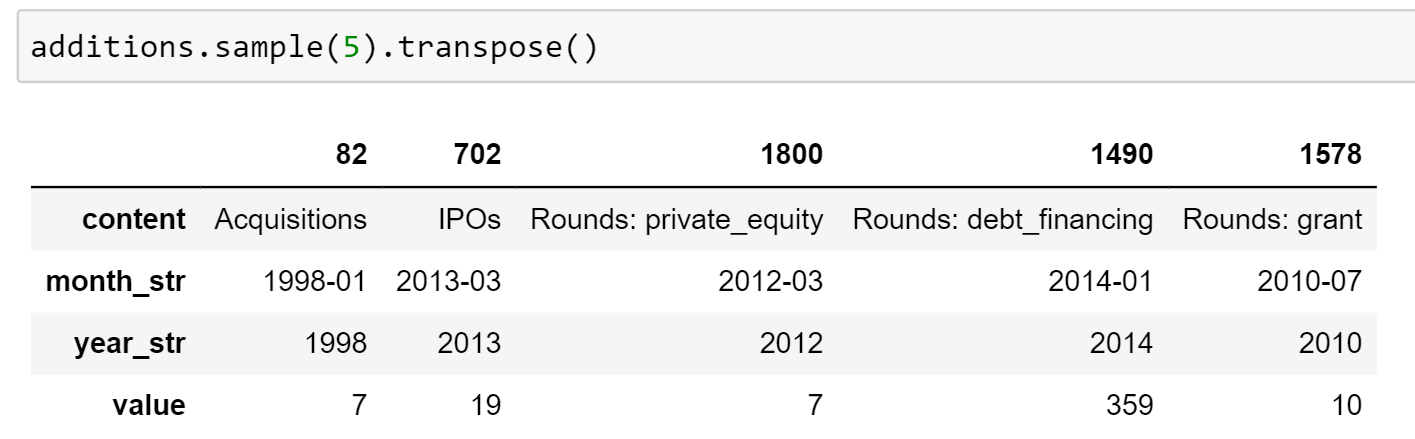
5 random samples from each dataset:

The Tables were transposed to make the features clearer.

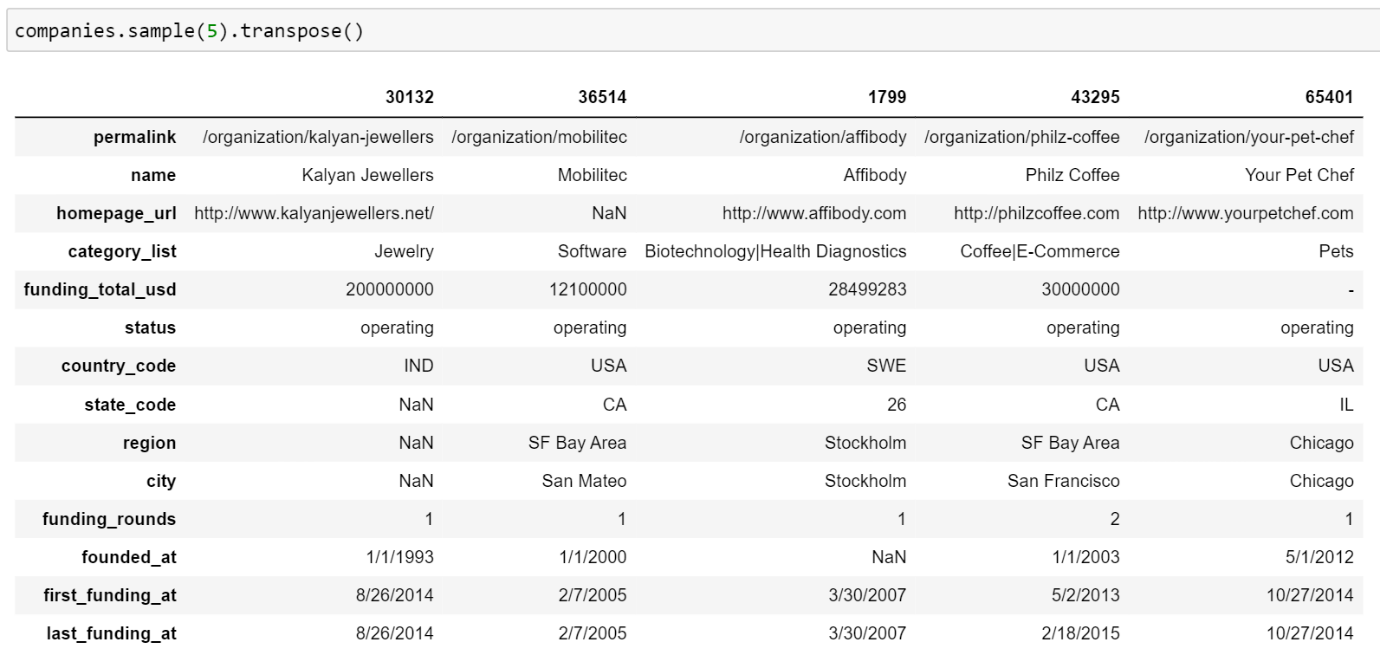
### Acquisitions



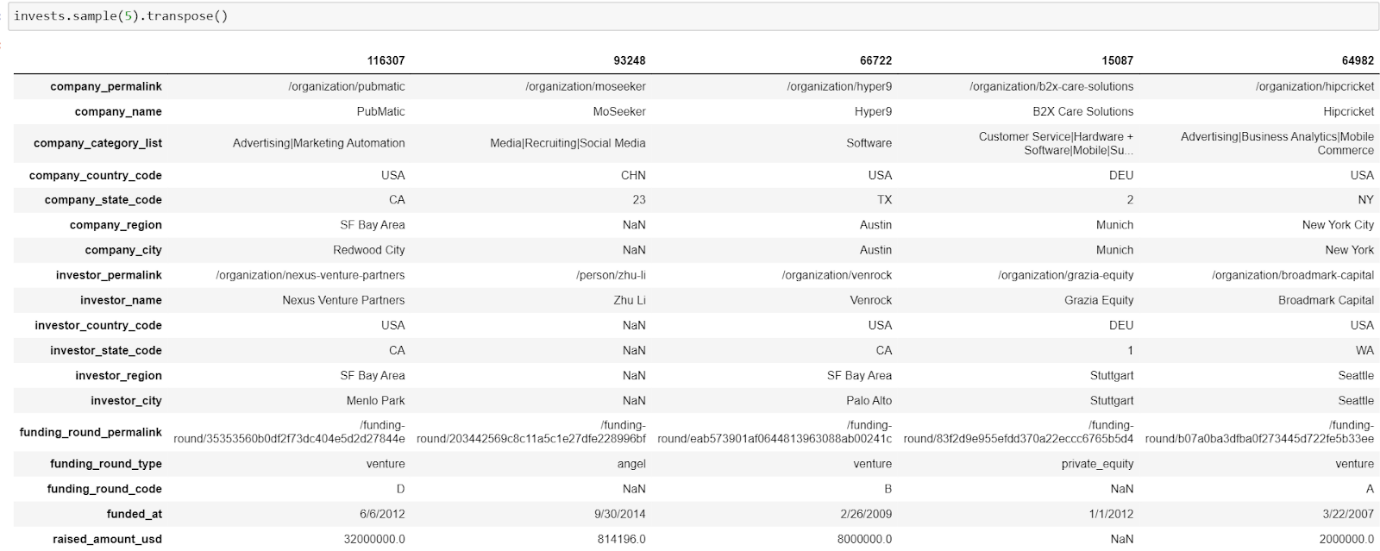
### Additions



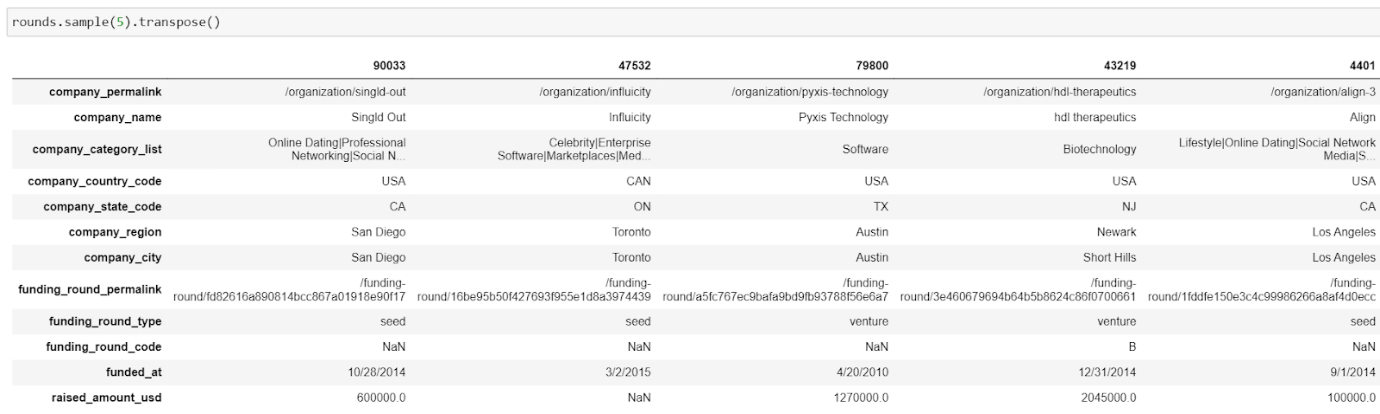
### Companies



### Investments



### Rounds



# Goals

* **Can the factors that affect a start-up’s growth be determined?**
  + - This can be determined by analysing a start-up’s funding from three angles: The total funding, the number of funding rounds, and the funding round type.
* **Do taxes play an important factor in deciding a start-up’s success?**
  + This can be determined by gathering information about each country’s tax revenues (found [here](https://stats.oecd.org/Index.aspx?DataSetCode=RS_GBL)) and correlating the results with the status of a start-up.
* **Is it possible to determine the suitable number of employees for each companies’ category lists**
  + This can be determined by gathering information about number of employees from each company’s permalinks and integrating that data in a new column in the dataset.
* **Can the number of technologies that are used by the companies affect the rate of its success or failure?**
  + - This can be determined by checking if these technologies used were outdated and / or faulty.

* **Which countries are most probable to have the most failed start-ups?**
  + - Correlating the country and a start-up failure, the analysis can show which regions are the hardest to kickstart a company there that may lead to its acquisition or dissolving. Then, it can be linked with the reasons behind this difficulty that makes the start-up fail. Difficulties can be presented in lack of funding, low funding frequency, or other external factors. Additionally, a comparison can be deduced by forming a ratio of the number of closed start-ups to total companies in that specific region.
* **Which category (field) attracts more funding?**
  + - Every start-up is built upon an idea or a category, and each category can attract interested investors. The data will be analysed to determine which companies under which category attract the most funding, and then it would show which category has the potential and is an interest to investors.