1. **List of group members, email addresses, and GitHub usernames**

Purvansh Jain - [purvansh@seas.upenn.edu](mailto:purvansh@seas.upenn.edu), Purvansh-Jain

Joseph Poirier — [jcpoir@seas.upenn.edu](mailto:jcpoir@seas.upenn.edu), jcpoir

Lisa Friedmann - [lisafrie@seas.upenn.edu](mailto:lisafrie@seas.upenn.edu), lisafrie

Rishi Patel - [rishi05@seas.upenn.edu](mailto:rishi05@seas.upenn.edu), rishipatel5799

1. **Description of application/website idea**

This multi-page web app will allow users to analyze startup and funding data from Crunchbase. There will be a homepage giving a quick overview of relevant summary statistics, such as number of companies and funding by year, and have hyperlinks to other pages. We may also have statistics on funding categorized by industry type and current operational status depicted graphically.

A second page will be a dashboard with a map displaying different parameters, like number of companies founded, number of financing rounds, dollar amount raised by state. We can also segregate the data based on US and non-US based companies.

A third page will give users the option to filter companies based on certain criteria, e.g. location, time of foundation, funding rounds, and show detailed characteristics of a selected company. Clicking on a state in the map on the dashboard page will bring users to this page and automatically apply the selected state as a filter.

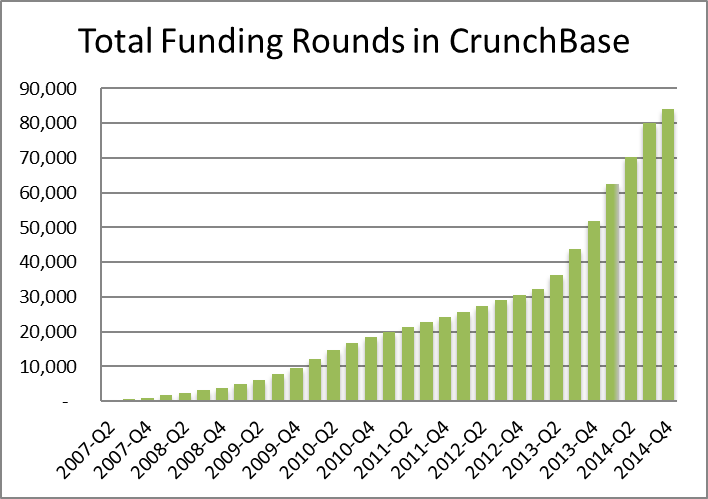
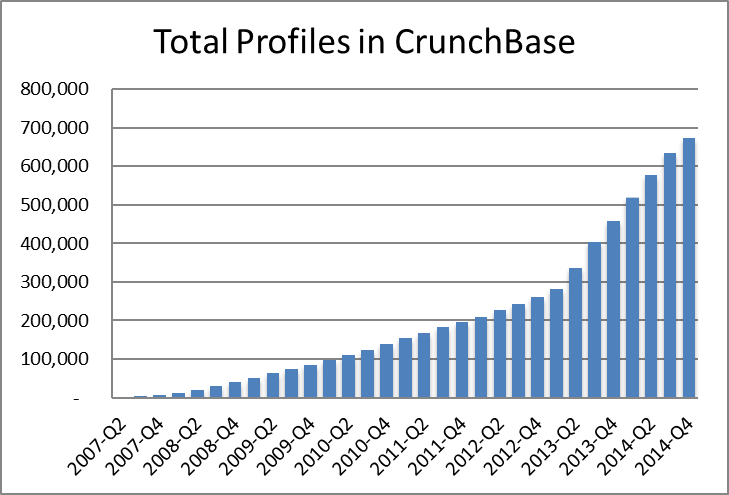
Whenever we click on a specific company, we will go to a fourth page showcasing statistics about the company chosen. It may have bar graphs about the different rounds of funding, and data specific to the company in question (location, founded date, industry type, current status and parent company if any)

1. **Description of the dataset:**

The dataset was found on [Tableau Sample Datasets](https://public.tableau.com/app/resources/sample-data?qt-overview_resources=1) and comes from the [Crunchbase](https://www.crunchbase.com/home) website.

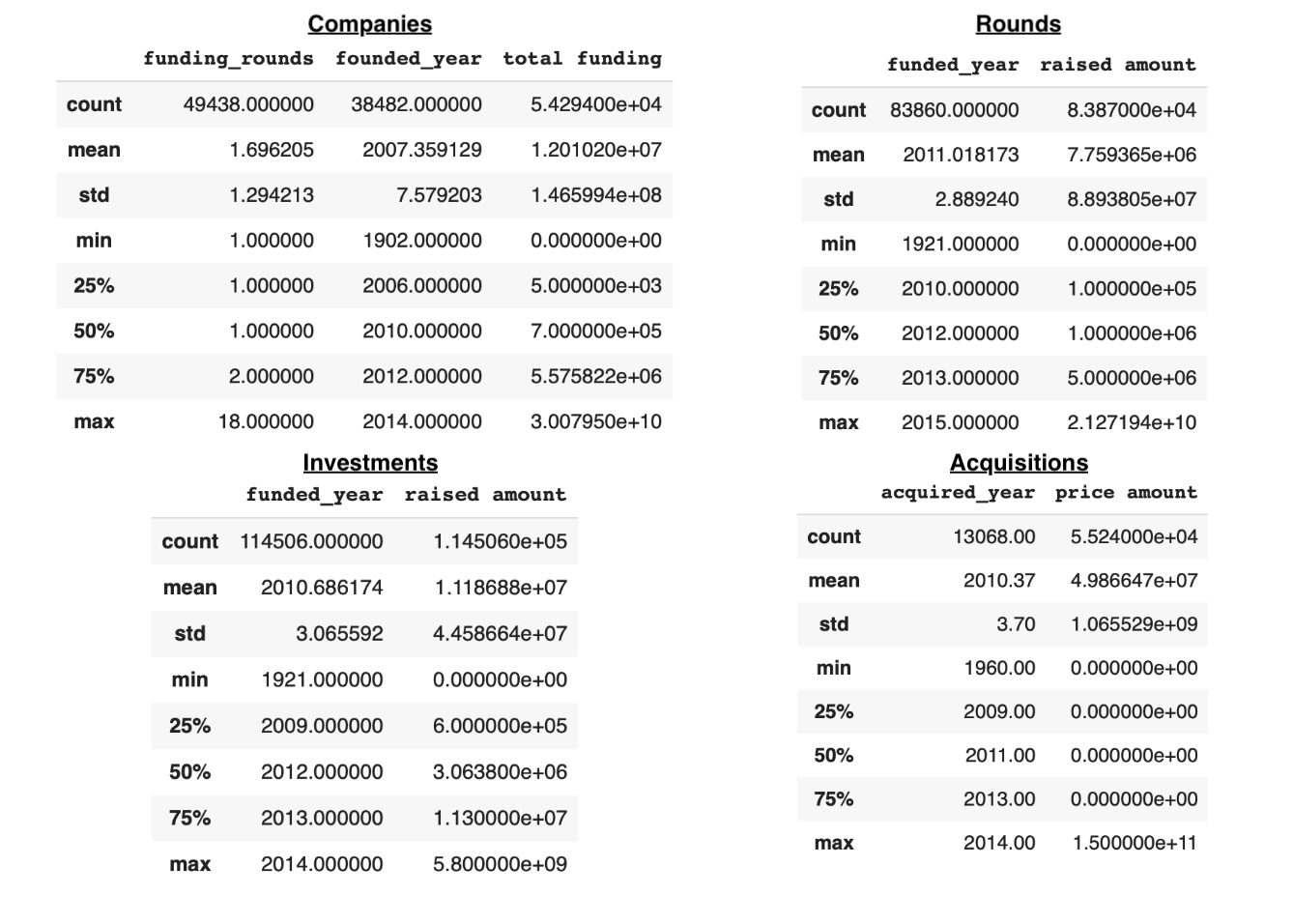
It contains information about companies with recorded fundraising 1960 and 2014. It has 4 separate tables:

* Companies: lists name, category, url, location, funding by company and other company characteristics
  + 49,439 rows, 18 columns
* Rounds: for each funding round for each company, lists type, date and amount
  + 83,871 rows, 16 columns
* Investments: lists information about the investors of each company, including name, location, round, date, amount invested
  + 114,507 rows, 24 columns
* Acquisitions: similar to Investments, this table lists information about acquisitions, including name of acquirer, location, date, price
  + 13,071 rows, 22 columns

Summary statistics:

Summary statistics of several attributes (e.g. report mean, standard deviation):

Calculated using Pandas ( for EDA)



1. **List of at least 5 queries (in natural language) you could write for your datasets. Some of these should require complex SQL (aggregations, subqueries, joins, etc.)**
2. Which companies have at least ‘x’ rounds of funding and the first round of funding is greater than $1,000,000 ?
3. What are the top 10 companies with the largest number of investors and what is their total funding?
4. The number of companies founded in a specific state/country
5. The funding round type, code and raised amount for a specific company
6. Show the number of companies funded in each year
7. Which companies were invested in by another company based in a specific state? (For example, which companies have had investors from California?)
8. Show the co-investors of co-investors in a certain company and which other companies they have been co-investors in
9. Which companies that have been acquired are from a market that is different from the market of the company that is acquiring them?