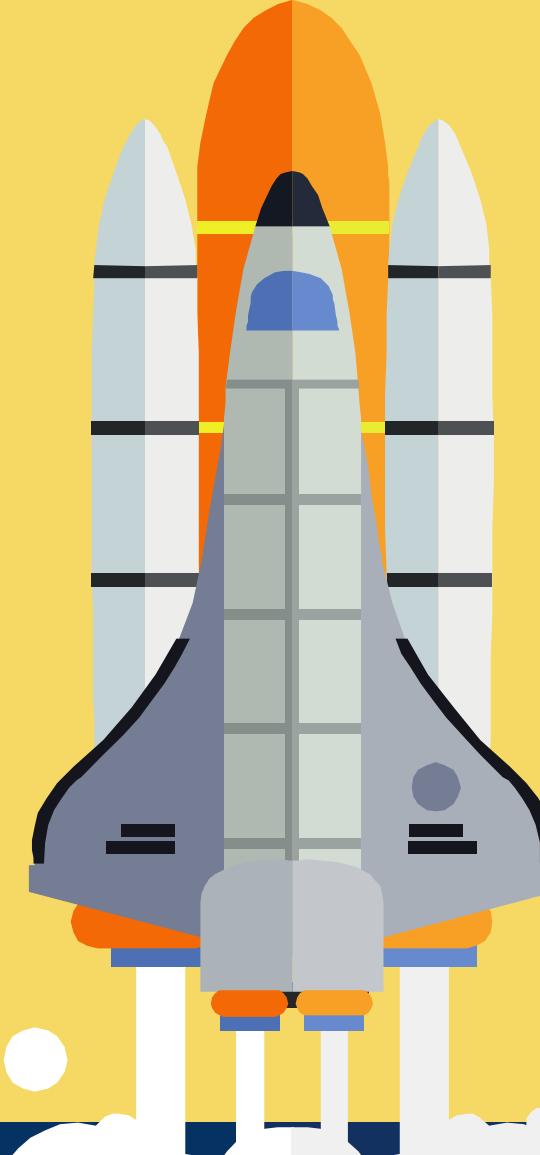


# How To Build The Perfect Startup

A Data Analysis of Venture Capital  
Funding

Sean Gallen & Jack Gallagher



University College Dublin  
An Coláiste Ollscoile, Baile Átha Cliath

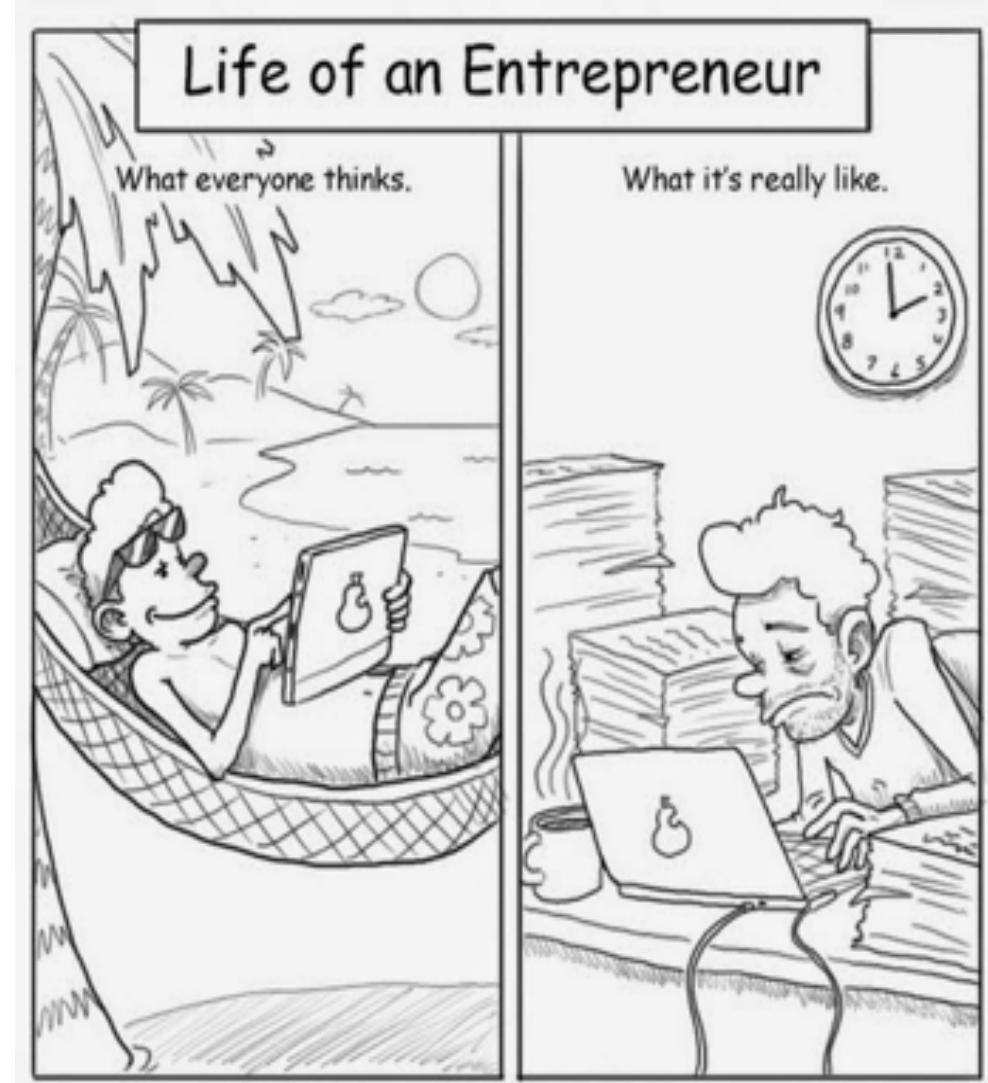
# Summary Aims and Objectives

- To merge the data we aim to use for our project, from multiple datasets into one single dataset.
- To analyze **Venture Capital (VC)** funding in order to learn more about how different attributes such as the founder, location, and industry affect how successful a company is in terms of funding.
- To help future startups understand the ideal location for their industry, and the best schools to hire from.



# Growing Love For Entrepreneurship

- Around 500,000 people become “entrepreneurs” every month according to entrepreneur.com
  - Although some don’t realise how much **hard work** is involved
- More and more people want to become entrepreneurs and it is beginning to be seen as a **viable career path**:
  - **Startup hype** in today’s society
  - Have a great idea
  - And of course \$\$\$



# VC Money

- The amount of money spent by VC funds is **HUGE!!**
  - Total of **\$164Billion** invested by VC funds globally in 2017
  - That is equivalent to \$34,359 for every person in Ireland, in ONE year alone
  - Irish tech companies alone raised \$1billion in funding in 2017
  - Median deal size globally is \$10.8Million



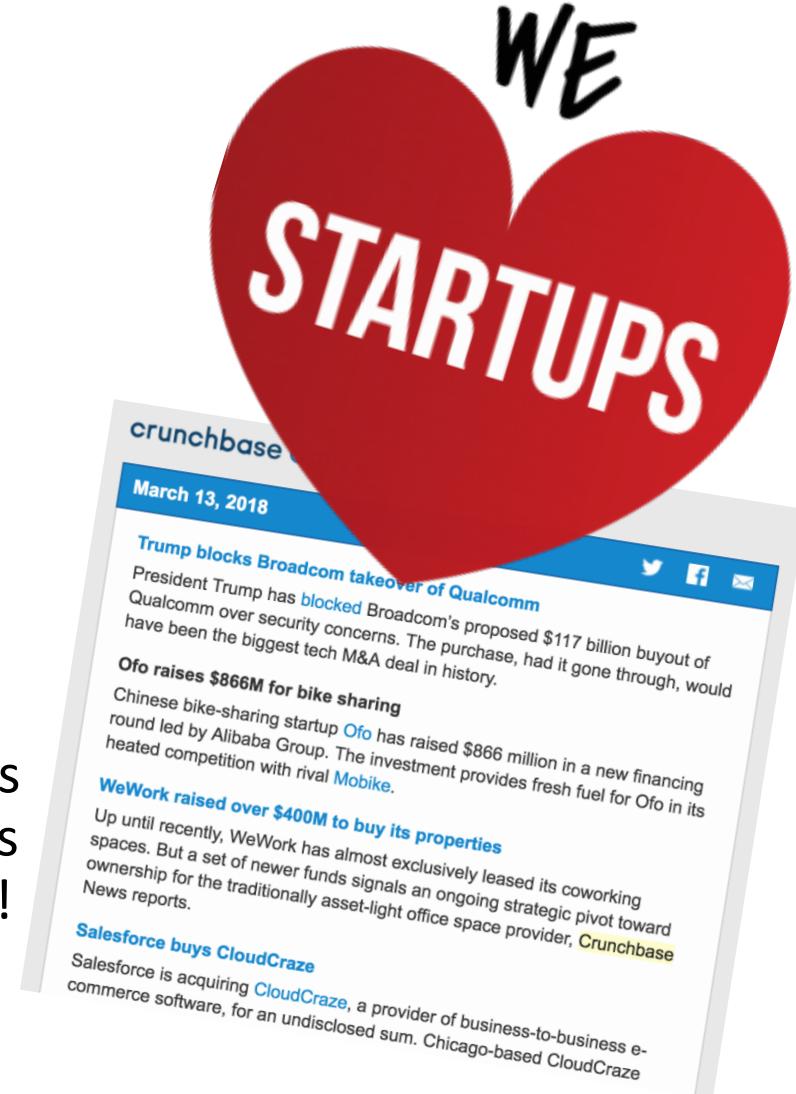
# Be the “other 1!!”

- 9 in every 10 startups FAIL!
  - Poor product
  - Ignore customers
  - Run out of cash
  - No business model
  - Lack of PASSION and focus
- BUT **someone** has to be the other 1!!



# Why we chose this dataset...

- We both have a huge interest in the world of startups.
- Such a broad topic, with a major link to the real world.
- We initially had thought of doing an Oscars movie speech data analysis as it was the first dataset which took our interest on kaggle.
  - Luckily we stumbled across the CrunchBase dataset while endlessly looking for large, and detailed datasets on startups
  - Unfortunately we then found out we had very limited access
  - Fortunately after a long week of emails, we got FULL access!



# Existing Research

- Current research on VC funding lies primarily around the following two areas:
  - How much funding companies, and industries have received in certain geographical areas
  - The growth of certain companies, and industries in terms of funding when compared with previous quarters
- There is a very fascinating gap in analysis when it comes to the correlation between the location of the company and the industry of that company, in terms of the amount of funding they achieve. (e.g. Tech company in San. Fran. vs Tech company in Boston).



# Primary Research Questions

- **RQ1.** Is there correlation between a company's industry and location, with the amount of funding they receive?
- **RQ2.** Does the founders education affect the amount of funding his company receives from Venture Capitalists?
  - Institution attended
  - Field of Study
  - Degree type
- **RQ3.** Do founders involved in startups in the past earn more VC in their current startups VS founders with no previous startup involvement?



# CrunchBase:



- Crunchbase offered all the data we needed and is a highly regarded startup information database online.
- Identified the data fields we required for our project:
  - Startup funding information e.g name of startup, location, industry
  - Data on founders e.g degree, institution, gender
- Applied for Research access to the Crunchbase API so we could extract the data we had identified for our project.



# CrunchBase:

- API allowed access to 9 datasets:
  - Organisations, people, funding, jobs etc.
- All the data that we required was available.
- Extra datasets included information on:
  - Company Acquisitions
  - IPOs (Initial Public Offering).

The image shows two overlapping screenshots of the Crunchbase platform. The top screenshot is the 'Crunchbase Data' API documentation page, showing sections for Documentation, Data Exports (Daily Excel Export, Daily CSV Export, Open Data Map, 2013 Snapshot, Complete Node List), REST API (Using the API, Release History), and API Entities Types (Address, Acquisition, Category, Deal, Funding, Person). The bottom screenshot is the main Crunchbase website homepage, featuring a search bar, a sidebar with links to Companies, People, Investors, Funding Rounds, Acquisitions, Schools, Events, My Searches, My Lists, Featured, Add New Profile, and Marketplace, and a main content area with sections for Latest insights and analysis, Activity this week, Introducing Crunchbase Marketplace, and Featured searches and lists.



University College Dublin  
An Coláiste Ollscoile, Baile Átha Cliath

# Ethical Considerations

- We obtained the data from CrunchBase under the conditions that:
  - It will be used for this project only
  - We will ensure that the data is not made publicly available and is secure
  - We will ensure we do not share the data with any third parties
  - The data will not be used commercially
- Data collected is already partially anonymized using hashing techniques
  - e.g. Institution name
  - Not necessary to anonymize any names of individuals

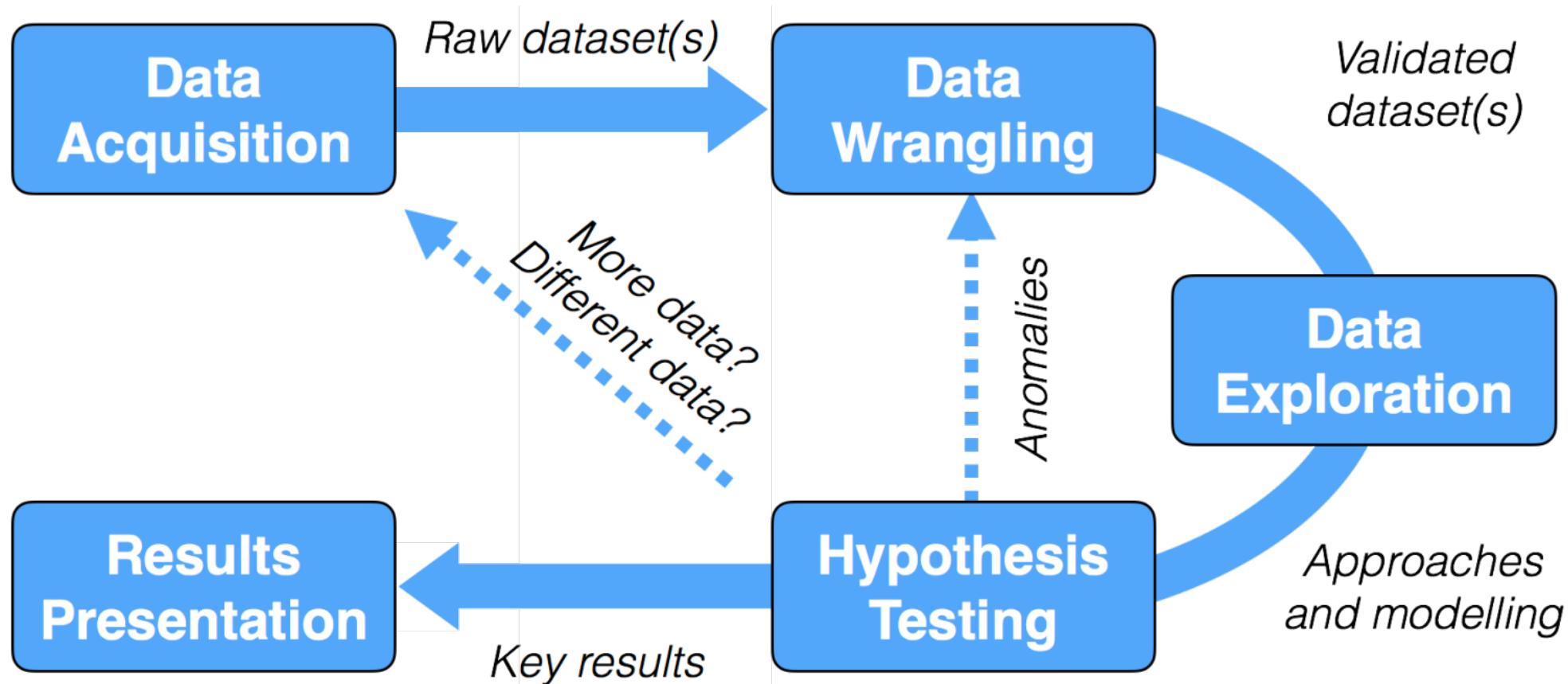


# Resulting Dataset

- Datasets that we collected are linked by the ‘id’ of the person and the ‘id’ of the company they are involved with.
- Data spans from 2007 - 2018.
  - Private dataset as discussed previously
- Important features:
  - Large dataset, 45,000+ rows of data on startups that have received funding
  - Includes fields detailing the startups information as well as the people involved in it



# Summary Approach



# Processing Stages

- Data Wrangling/Cleaning
  - Focus on identifying missing values specifically in the startups funding
  - Eliminating duplicate instances and further filtering of unnecessary columns
- Exploratory Analysis
  - Initially investigate how a founders education affect the amount of funding his company receives
  - Proceed to investigate the correlation between a company's industry and location on funding
  - Compare the funding people starting a company for the first time against people who have set up multiple companies
- Results
  - Use appropriate visualizations
  - Use tables to represent the insights we investigated



# Q&A SESSION



University College Dublin  
An Coláiste Ollscoile, Baile Átha Cliath