Hyperiondev

TASK

Exploratory Data Analysis on the Automobile Data Set

Visit our website

Introduction

The dataset in question is that of Crunchbase. A financial site focused on start-up's in early and late stages of funding. I was immediately drawn to this dataset due to the fact that I would also like to get into business and start my own company one day. What intrigued me was knowing that the questions I've always had would be answered, but most of all put into perspective. I am well aware of the US funding pool, but I would like to know what makes it tick and possibly why? Besides Silicon Valley are there any other places that provide similar opportunity? Are there other countries catching up? How many businesses have been successful and so forth.

DATA CLEANING

The data was generally clean except for the missing values and zero's. What needs to be bared in mind is that this is not data that can be manipulated. It is factual data based on actual companies. If there is one round of funding, then that's all there was. Perhaps the company shut down or was acquired. Sometimes the company is liquidated and investors rename the company and take it in a completely new direction. So, no filling in the blanks!

MISSING DATA

I handled the missing data as follows, I remove the rows that had missing values such as "Nan" and removed the columns I would not be using. Which brought down the amount of data by a significant amount. But I was still left with 20 000+ rows, so it was enough to work with.

DATA STORIES AND VISUALIZATIONS

The first question I wanted to know was regarding status, who is still operating and who is not. Most of all who has been acquired. I used a pie graph to represent this and to my surprise. 95% of businesses started are still in business. The start-up culture is more than just a passing phase. Younger people starting sustainable companies has become the norm. With an unequal distribution of 85% operating, 10% acquired and 5% closed. It's safe to say business founders today are much more equipped than those of the past.

My next graph displays the number of start-ups started from 2000 to 2014 (since that's the time period of the dataset). In the beginning, there is a small dip due to the dotcom crash of the year

2000. The dotcom crash resulted in one of the biggest market crashes in American history. A steady incline in years to come where it reaches its peak in 2011. There is a rapid decline soon after. This does not mean anything bad. In fact, it's a good thing. Most companies during this time were successful so, most people in the industry were employed. Bear in mind as of this analysis. The US has an unemployment rate of 3.9%, now take into consideration of that small percentage, the elderly make up more than half of that. The reason I mention the US is because that's where most start-ups are, which you will see later.

As seen in the next graph which picks up where I left off. The US dwarf's all other countries in regard to the amount of businesses founded per country. Great Britain comes in second but is a meagre 5% of the US at best. Canada, China and Germany come follow up. Which are all quite close.

Moving on to the next visualisation. I wanted to know which market dominated the start up space. Software was at number 1 unsurprisingly, but biotechnology is close as second making up about 80% of the software. While Mobile is less than half of biotechnology, curated web and enterprise software are almost equal and so are the rest of the top 20.

I did further digging into the US, and wanted to know which states produced the most start-up's! Not surprising is the fact that California is at the top due to the fact that Silicon Valley is in San Francisco. Which is in California. What was surprising is that New York, in second place was a mere quarter of start-ups produced. The reason why is because New York was always known to be the financial hub of the world due to Wall Street. I guess there's not much room with investment banks such as Goldman Sachs and JP Morgan Chase running the show, as they have been for the past few decades. It's a massive shift in economic power from the East coast to the West. Although this is not a first for San Francisco. The gold rush occurred there from 1848 – 1855. Many people from all walks of life came to the city in search of fortune and frankly that time has come right back around again as many founders leave their home countries to seek funding from San Francisco. Land of opportunity.

Towards the end I wanted to put everything into perspective. So I used a library called Plotly which can produce interactive graphs. Every dot represents a company in the dataset and the necessary details of said company are available. Lastly I took one last look at the US and wanted to know the funding status of all the companies in the country. Verizon is at the very top with no other company coming close. So, software might be the most in terms of volume of companies but Telecoms is the ultimate fund raiser!

My take away from the dataset are as follows:

- Most start-ups survive and become great companies
- The most start-ups were founded in 2011
- USA trumps all others in sheer volume regarding the founding of start-ups
- Software and Biotech are the leading markets
- California has the most founded companies
- Telecoms Verizon is the highest fund raising start-up in the world

THIS REPORT WAS WRITTEN BY: EVAN TAYLOR