

The background of the slide is a light gray gradient. It is decorated with several realistic water droplets of various sizes, some at the top and some at the bottom. In the center, there is a faint, circular logo that appears to be a stylized 'T' or a similar emblem.

Data Analysis on startup's Acquisition Status

**SUBMITTED BY | ANKIT DADARWALA | DATA ANALYSIS INTERN AT
TECHNOCOLABS SOFTWARES**

Introduction

Problem Statement:

“THE SHARK TANK INDIA” BY SONY the TV show introduce different business ideas by peoples or group to INVESTORS to invest in their business to grow.

In our project logic as above in that base on crunchbase.com platform industries like Administrative, AI, Agriculture, Biotechnologies, E-commerce, Food, Gamming etc.. Are funded by investors like Brand capital, Franklin templeton, NEDFI venture capital, Elevation capital etc..

Thousands of companies are emerging around the world each year. Among them, some are merged and acquired (M&A), or go to public (IPO), while others may vanish and disappear.

- ✓ What makes this difference and leads to the different endings for each company?
- ✓ How to predict the success of companies?



GOAL

The objective of the project is to predict whether a startup which is currently Operating, IPO, Acquired, or closed.

WHAT IS ACQUISITIONS?

An acquisition is when one company purchases most or all of another company's shares to gain control of that company. Purchasing more than 50% of a target firm's stock and other assets allows the acquirer to make decisions about the newly acquired assets without the approval of the company's other shareholders. Acquisitions, which are very common in business, may occur with the target company's approval, or in spite of its disapproval. With approval, there is often a no-shop clause during the process.

WHAT IS IPO?

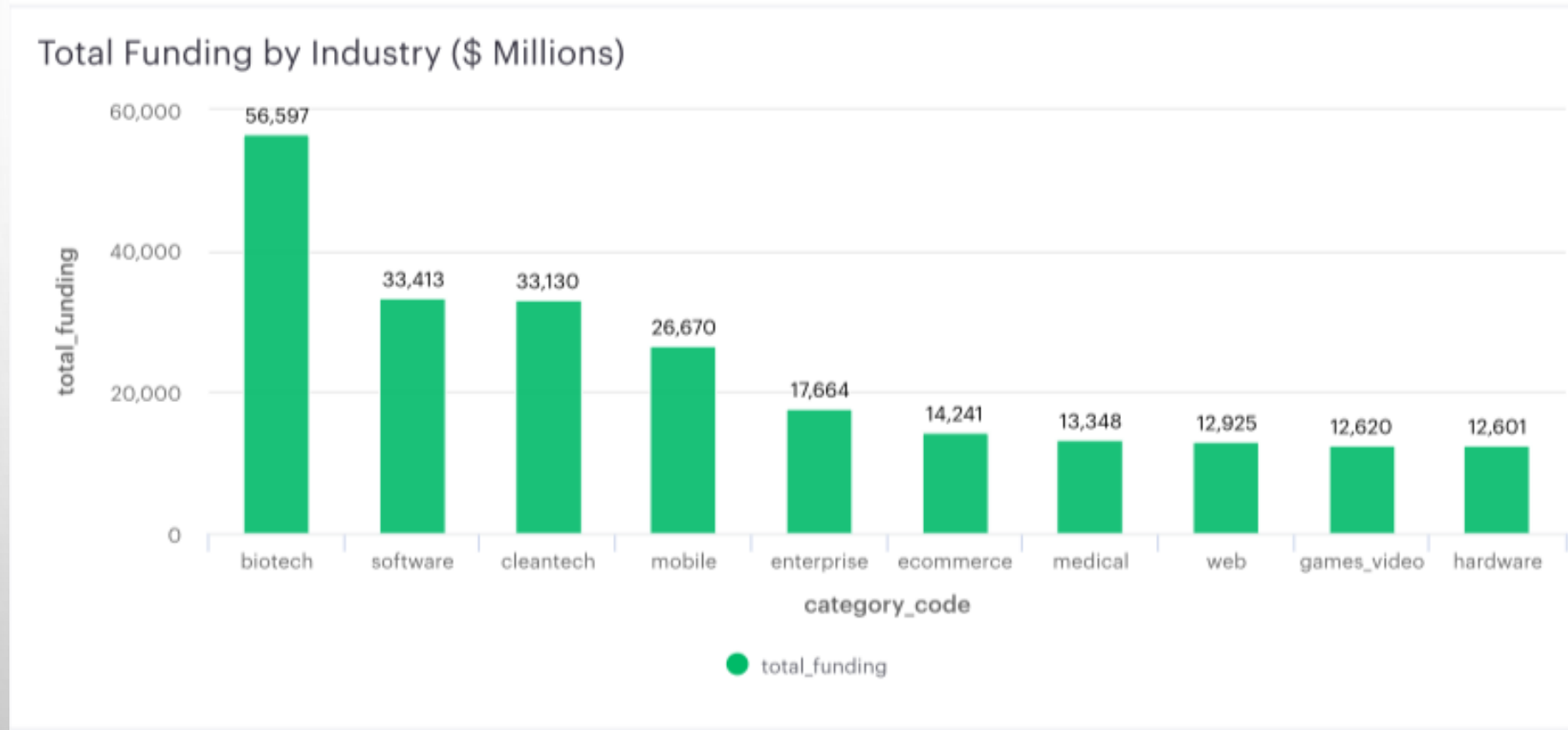
An initial public offering (IPO) refers to the process of offering shares of a private corporation to the public in a new stock issuance. An IPO allows a company to raise capital from public investors. The transition from a private to a public company can be an important time for private investors to fully realize gains from their investment as it typically includes a share premium for current private investors. Meanwhile, it also allows public investors to participate in the offering.

DATASET

The dataset we are using is extracted from Crunchbase Data Export containing 60K+ companies' information updated to December 2015. There were four data files, named "company", "investments", "rounds" and "acquisition". The "company" file contains most comprehensive information of the companies, while other files contain more detailed information regarding the investment operations. Thus, we chose the file "company" as the base and extracted meaningful features from other files to add into it.



GRAPH SOURCE FROM TOWARDS DATASCIENCE



- ✓ IN GRAPH SHOWING BAR PLOT OF DIFFERENT CATEGORIES OF COMPANIES ON X-AXIS AND THEIR TOTAL FUNDS PROVIDING ON Y-AXIS
- ✓ IN GRAPH SHOW THAT THE BIOTECH INDUSTRIES ARE MORE FUNDED
- ✓ FUNDED STARTS FROM 2007-01-01 FROM THE DATA

DATASET FEATURES

1. Name: company's name
2. Homepage_url: the website of the company
3. Category_list: the industry category the company belongs to, including up to four subcategory divisions
4. Funding_total_usd: the total amount of funding in all rounds of investments
5. Status: the operation status of the company (0 = closed or operating, 1 = ipo or acquired)
6. Country_code: the country of company's headquarter
7. State_code: the state of company's headquarter
8. Region: the region of company's headquarter
9. City: the city of company's headquarter
10. Funding_rounds: total number of funding rounds
11. Founded_at: the date company founded (in string format '2007-01-01').
12. First_funding_at: the first time the company raised money (in string format '2008-03-19')
13. Last_funding_at: the last time the company raised money (in string format '2008-03-19').

MODEL IMPLEMENTATION

- LOGISTIC REGRESSION :→ Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is dichotomous in nature. Dichotomous means there are only two possible classes. For example, it can be used for cancer detection problems. It computes the probability of an event occurrence.

SIGMOID FUNCTION APPLY ON LINEAR REGRESSION:

$$p = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p)}}$$

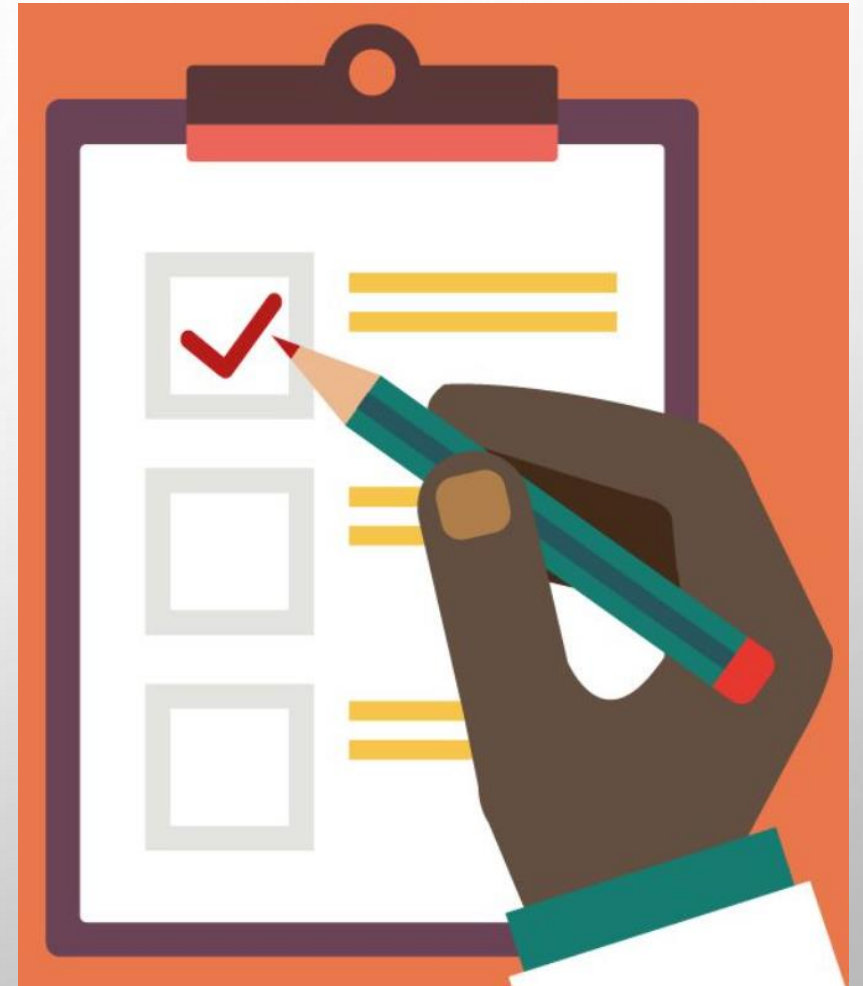
- K-NEAREST NEIGHBORS :→ K-nearest neighbour is one of the simplest machine learning algorithms based on supervised learning technique. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

WORK ON EUCLIDEAN DISTANCE FORMULA $(D) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$.

- RANDOM FOREST CLASSIFIER :→ Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

Project Goals

- ANALYTICAL PROBLEM FRAMING
 - EXPLORATORY DATA ANALYSIS (EDA)
 - VISUALIZATIONS
- DATA PRE-PROCESSING ON TRAIN AND TEST DATASETS
- MODEL/S DEVELOPMENT AND EVALUATION
- PERFORMING HYPER PARAMETER TUNING, SAVING THE BEST MODEL AND PREDICTING THE TARGET
- CONCLUSION AND FUTURE WORK DISCUSSION





THANK YOU