

Unicorn Companies Project

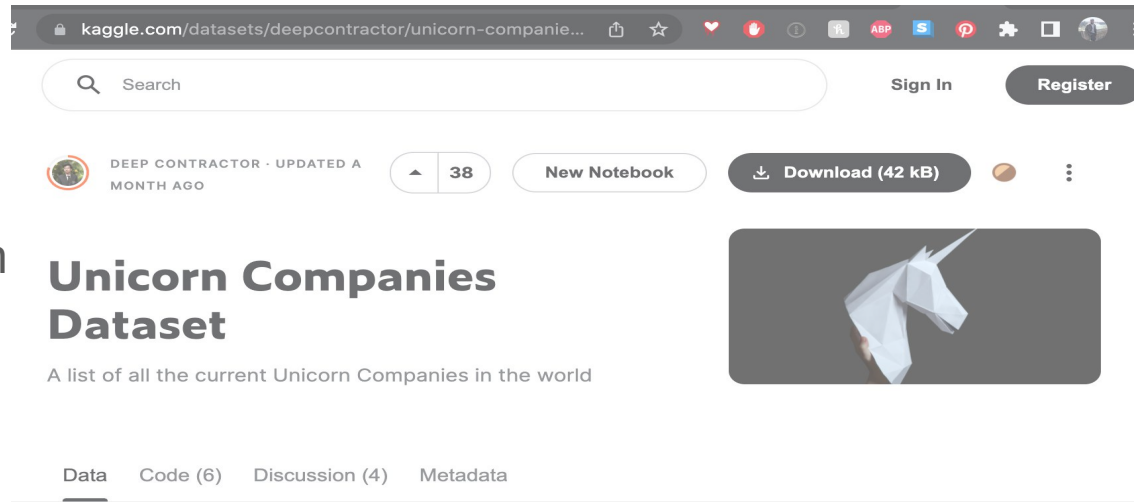


Project Overview

Reason the topic was selected - Unicorn companies are companies that have a valuation of at least 1B. The relativity of the dataset in the current economic climate and the unique projected value of these companies makes this project interesting. The dataset has mainly categorical data which will be used to show the current state of the companies under review.

Data Source

The dataset was sourced from the Kaggle website which is a renowned source for ready to use datasets



Data Exploration

We reviewed multiple possible data sets and landed on this one after checking the steps used to preprocess the data. As the project progresses the plan is to incorporate other datasets with similar characteristics and compare the trajecory of potential Unicorns.

The image displays three screenshots of the Kaggle website, illustrating the process of finding a suitable dataset for data exploration.

Top Screenshot: Shows the Kaggle homepage with the search bar and navigation links. The dataset "Indian startups - 2021" by Bharath Shivaiah is highlighted, showing it was updated 10 months ago and has 11 versions. The dataset description is "List of companies registered in India - 2021".

Middle Screenshot: Shows the dataset "unicorn companies around the world in 2021" by Priyanshu Chaudhary, updated 3 months ago. The dataset description is "The World's Unicorn Companies 2017". The dataset is available in Data, Code (6), Discussion (0), and Metadata formats.

Bottom Screenshot: Shows the dataset "Kickstarter Campaigns" by Yash Kantharia and 2 collaborators, updated 3 years ago (Version 1). The dataset description is "A dataset containing kickstarter crowdfunding campaign data". The dataset is available in Data, Code (5), Discussion (1), Activity, and Metadata formats.

Data Analysis

Using Pandas and Jupyter Notebook the original dataset Unicorn Companies.csv was used and evaluated for relevance of the data towards the objective of the project. Eventually, columns that provided little useful information were dropped and the remaining columns were used to extract the required data.

Preprocessing of dataset

```
In [3]: 1 unsupervised_df.drop(columns=["Financial Stage","Total Raised Ranges($)","Deal Terms"],inplace=True)
2 unsupervised_df.set_index('Company',inplace=True)
3 unsupervised_df.rename(columns= {'Transition Time':'Transition Time(Years)'},inplace = True)
4 unsupervised_df = unsupervised_df[unsupervised_df['Founded Year'] !=None]
5 # unsupervised_df.index.name = None
6 unsupervised_df.head()
7 # unsupervised_df.to_csv('Resources/Unicorn_Eval.csv')
```

Out[3]:

	Valuation (\$B)	Year Joined	Country	City	Industry	Select Investors	Founded Year	Investors Count	Total Raised(\$)	Transition Time(Years)
Company										
Bytedance	140.0	2017	China	Beijing	Artificial intelligence	Sequoia Capital China, SIG Asia Investments, S...	2012	28	7.440000e+09	5
SpaceX	100.3	2012	United States	Hawthorne	Other	Founders Fund, Draper Fisher Jurvetson, Rothen...	2002	29	6.874000e+09	10
Stripe	95.0	2014	United States	San Francisco	Fintech	Khosla Ventures, LowercaseCapital, capitalG	2010	39	2.901000e+09	4
Klarna	45.6	2011	Sweden	Stockholm	Fintech	Institutional Venture Partners, Sequoia Capita...	2005	56	3.472000e+09	6
Epic Games	42.0	2018	United States	Cary	Other	Tencent Holdings, KKR, Smash Ventures	1991	25	4.377000e+09	27