

ECE 143 Project (Group 11): Analysis of Startup Success

Presenters:

1. Kevin Pham (Presenter) (A16136224)
2. Mikhael Alex (A15763534)
3. Sathwik Madhula (A59019455)
4. Surya Pilla (A59015821)
5. Yizheng (A53282437)



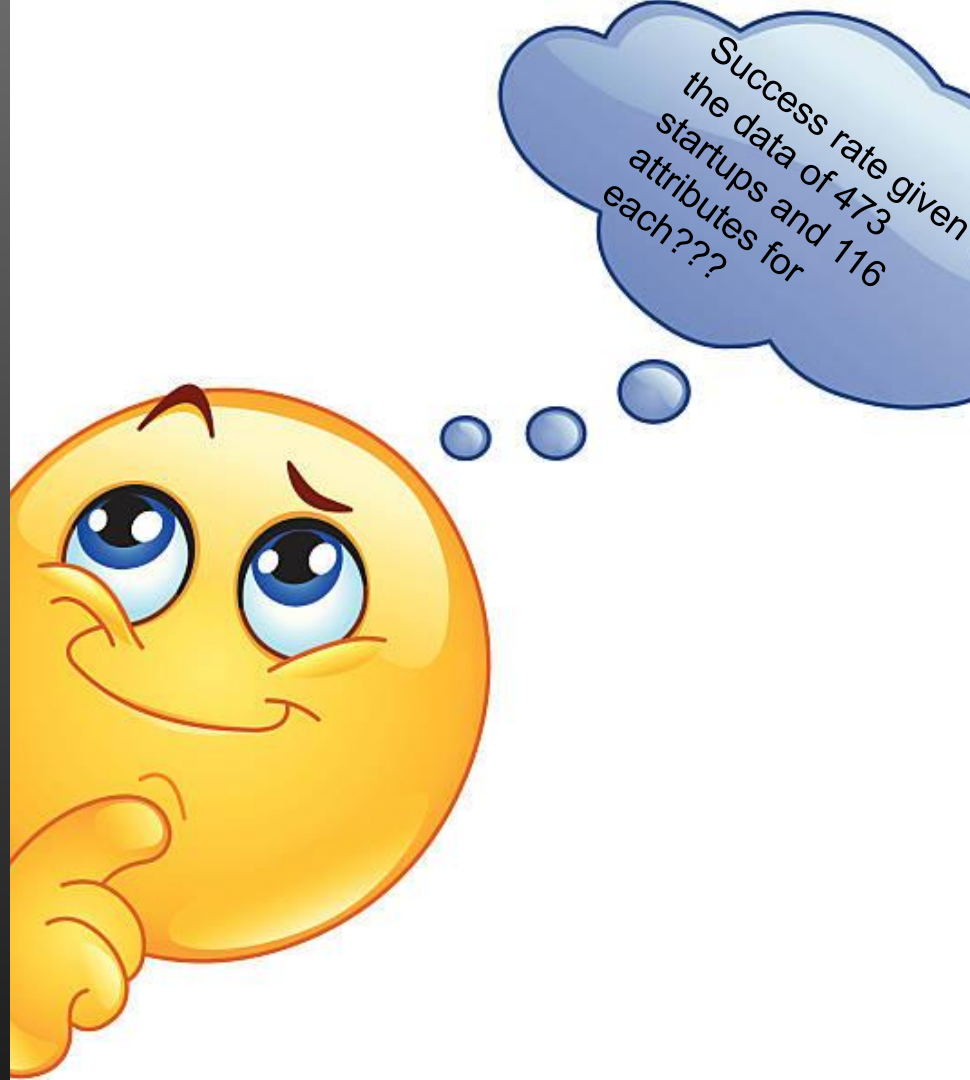
Flow of the presentation

1. Overview of data
2. Understanding the data/ geographical spread
3. Data cleaning/ wrangling
4. Data visualization
5. Summary

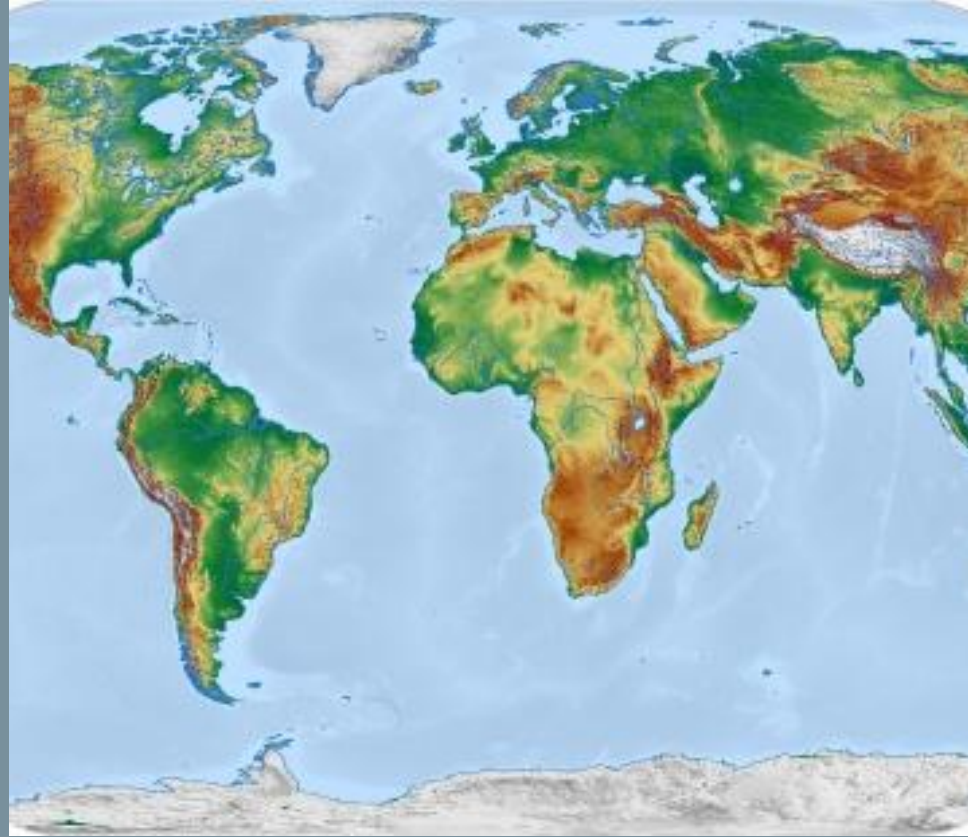
Overview of data

Overview of Data

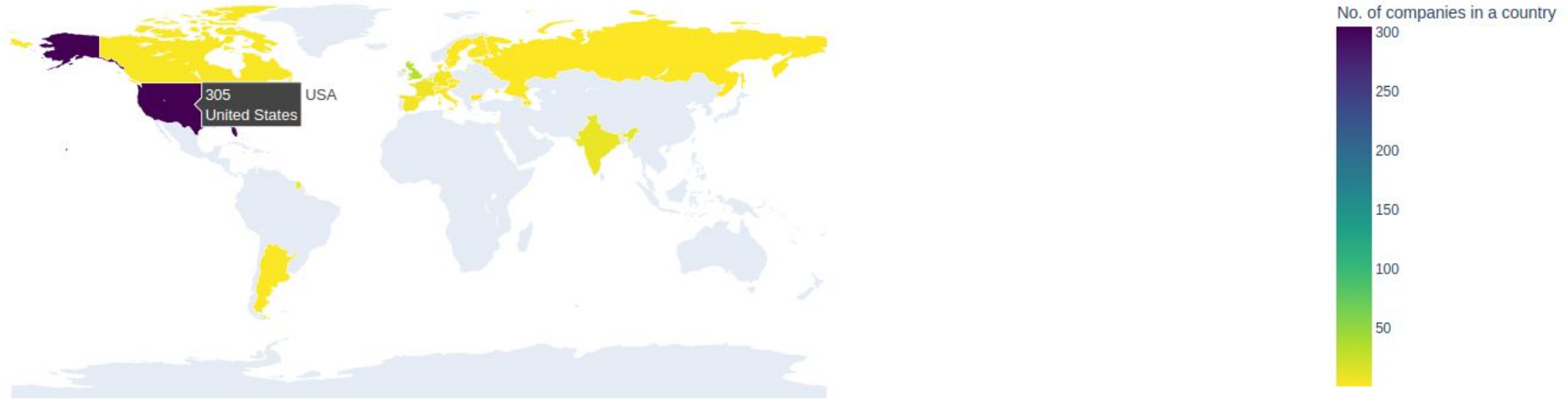
1. Dataset:
<https://www.kaggle.com/datasets/ajaygorkar/startup-analysis>
2. The common trends of the data help in figuring out success rate of a startup



Geographical spread

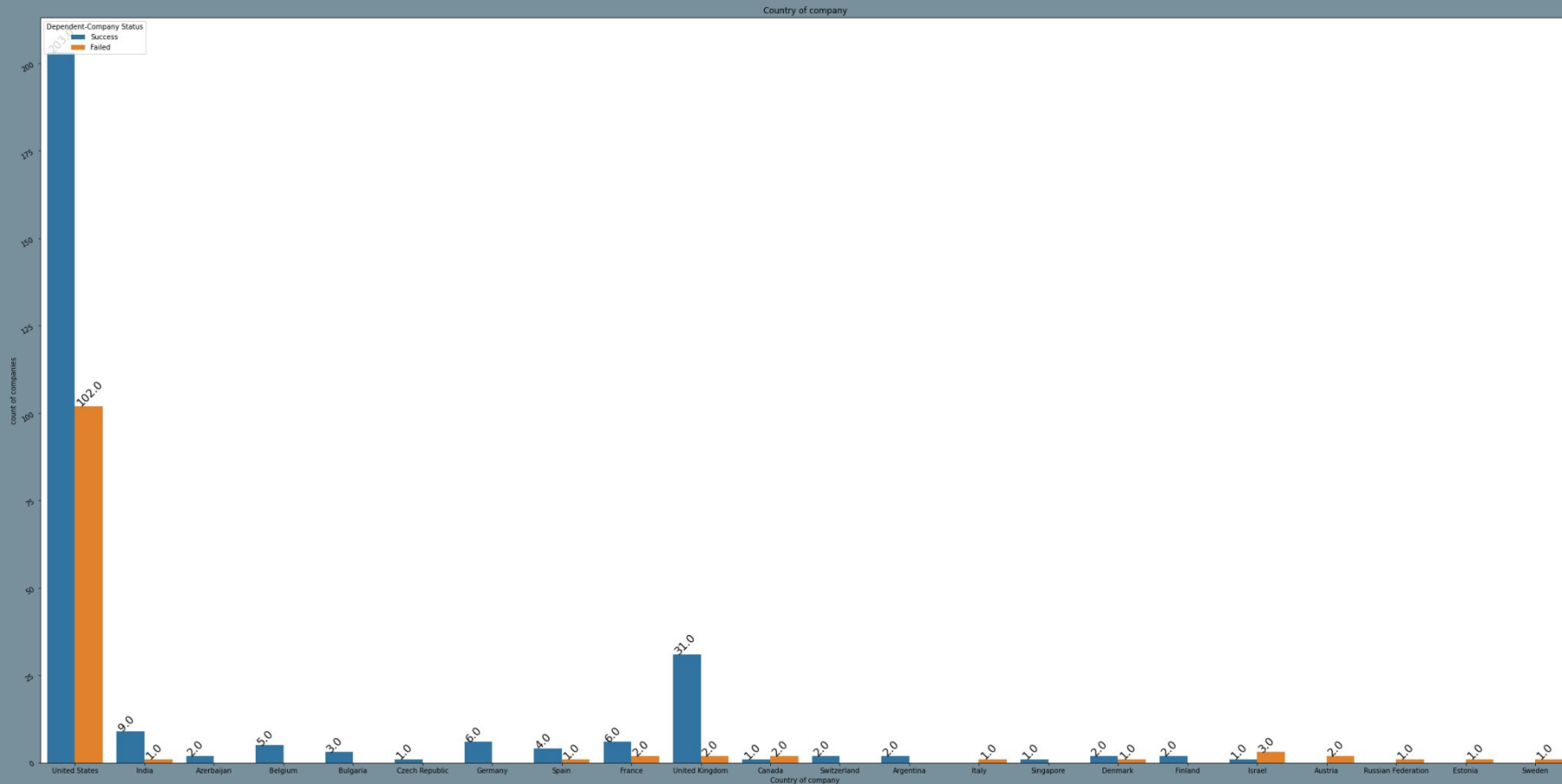


Number of Companies in different Countries

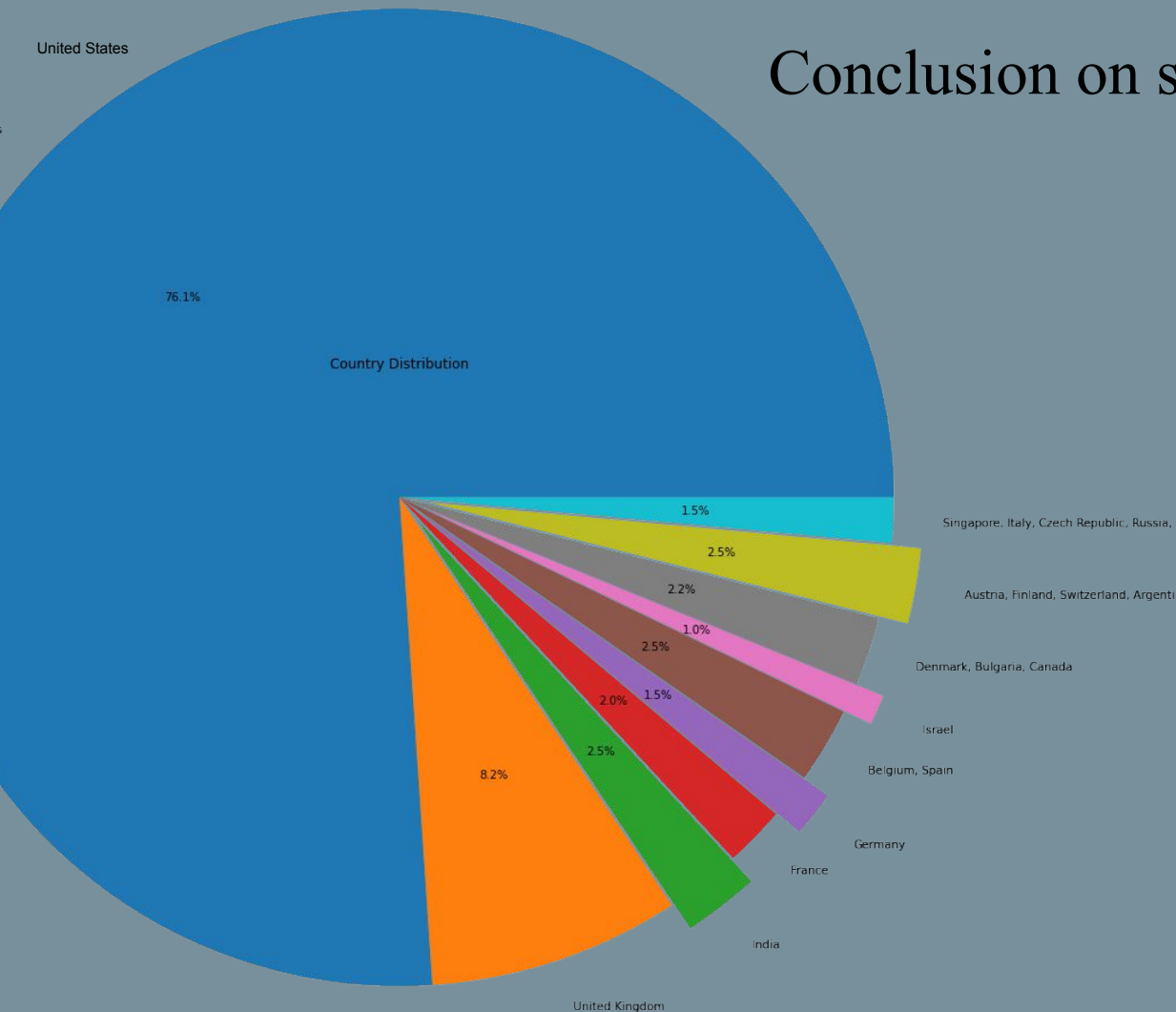


Important note: As per data available majority of the startup companies belong to U.S.

Number of Successful and Failed Companies by Country



Conclusion on startup geography



- The data contains more than 75% of U.S. based startups.
- Although, there is 100% success rate as evident from the previous slide for countries like South America it only has 2 startups.
- Country/ geographical location cannot be used as a valid attribute that could estimate the success of a startup based on this dataset.

Data Cleaning/ Wrangling

Fill Missing Data

	Null Values	% Missing Values
Exposure across the globe	85	18.008475
Number of of Research publications	84	17.796610
Time to market service or product	83	17.584746
Company awards	85	18.008475
Survival through recession, based on existence of the company through recession times	98	20.762712

Categorical columns has “No Info” value, replace with mode.

	Null Values	% Missing Values
google page rank of company website	154	32.627119
Number of Direct competitors	80	16.949153
Employees per year of company existence	128	27.118644
Last round of funding received (in millionUSD)	167	35.381356
Avg time to investment - average across all rounds, measured from previous investment	98	20.762712

Numerical columns has “No Info” & “unknown amount” value, replace with mean.

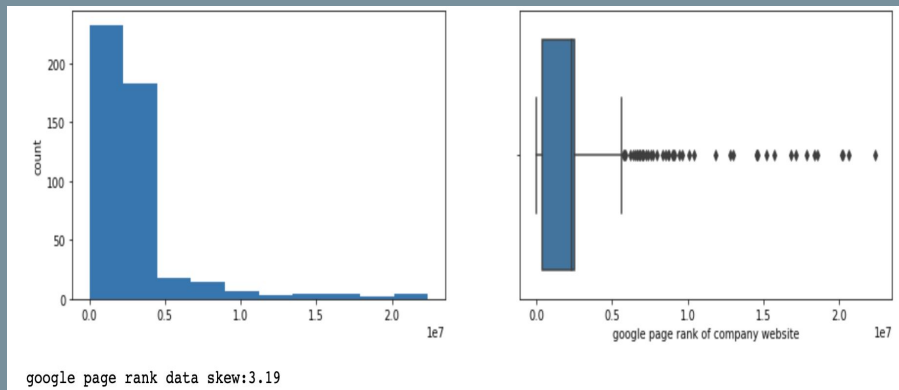
Last Funding Amount	160	33.898305
Employee Count	166	35.169492
Employees count MoM change	205	43.432203
Number of Co-founders	0	0.000000
Number of of advisors	0	0.000000

Stats

	count	unique	top	freq
Exposure across the globe	472	2	Yes	331
Number of of Research publications	472	3	None	334
Time to market service or product	472	3	Low	336
Company awards	472	2	No	396
Survival through recession, based on existence of the company through recession times	472	3	Not Applicable	366

	count	mean	std	min	25%	50%	75%	max
google page rank of company website	472.000000	2518862.836478	3393685.673554	483.000000	380600.250000	2348954.000000	2518862.836478	22391670.000000
Number of Direct competitors	472.000000	2.257653	4.055082	0.000000	0.000000	1.500000	2.257653	33.000000
Employees per year of company existence	472.000000	18.436076	56.977528	0.000000	5.000000	12.665000	18.436076	833.300000
Last round of funding received (in milionUSD)	472.000000	5.866330	6.894420	0.010000	1.315288	5.866330	5.866330	62.500000
Avg time to investment - average across all rounds, measured from previous investment	472.000000	10.563414	12.369628	1.000000	4.000000	10.233333	12.000000	156.000000

Numerical Data Skew & Outlier Handling



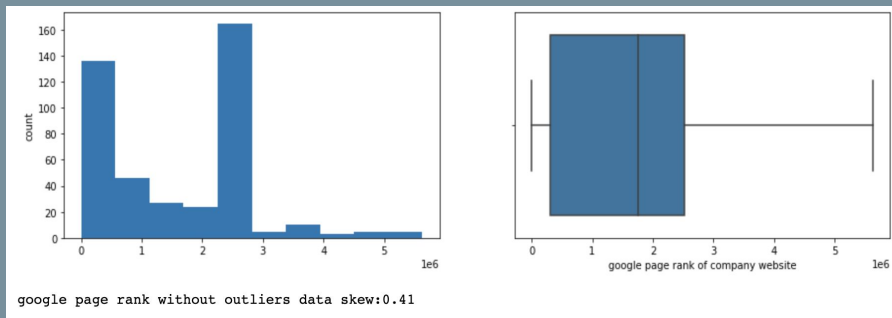
Outlier handling:

$$\text{IQR} = Q3 - Q1$$

$$\text{Lower bound} = Q1 - 1.5 * \text{IQR}$$

$$\text{Upper bound} = Q3 + 1.5 * \text{IQR}$$

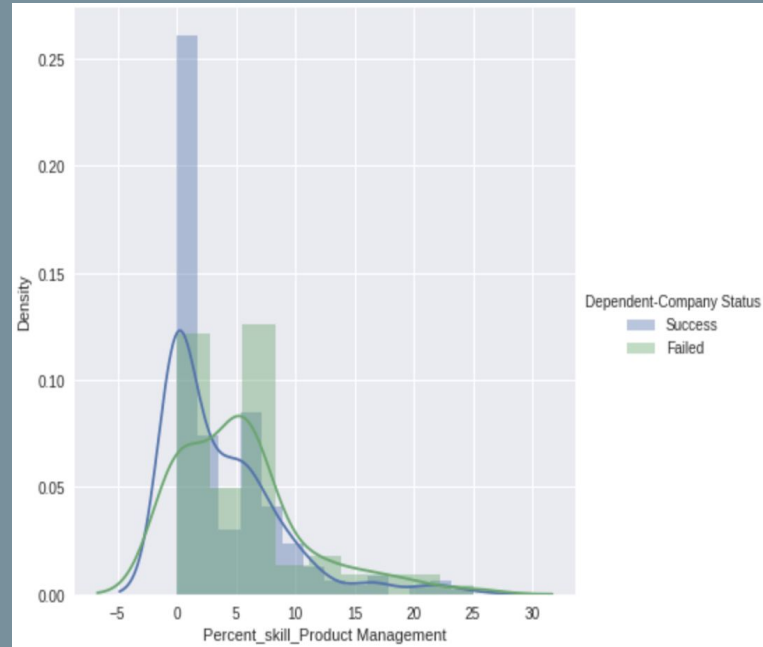
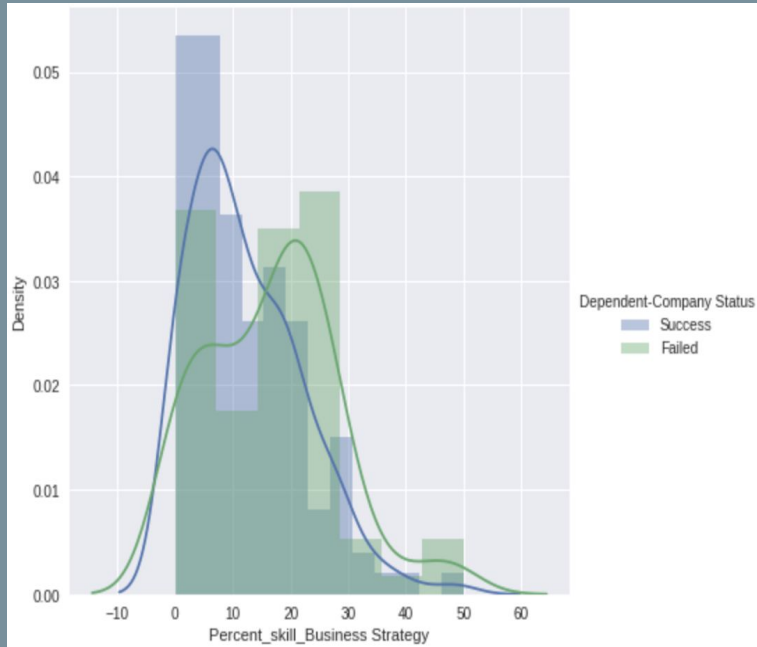
Filter out data beyond the bounds



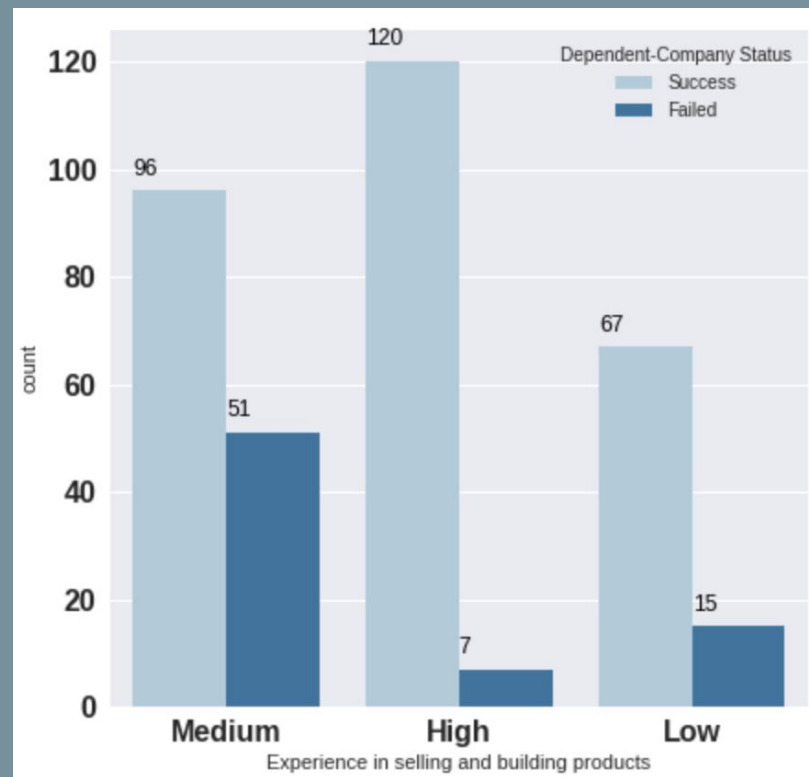
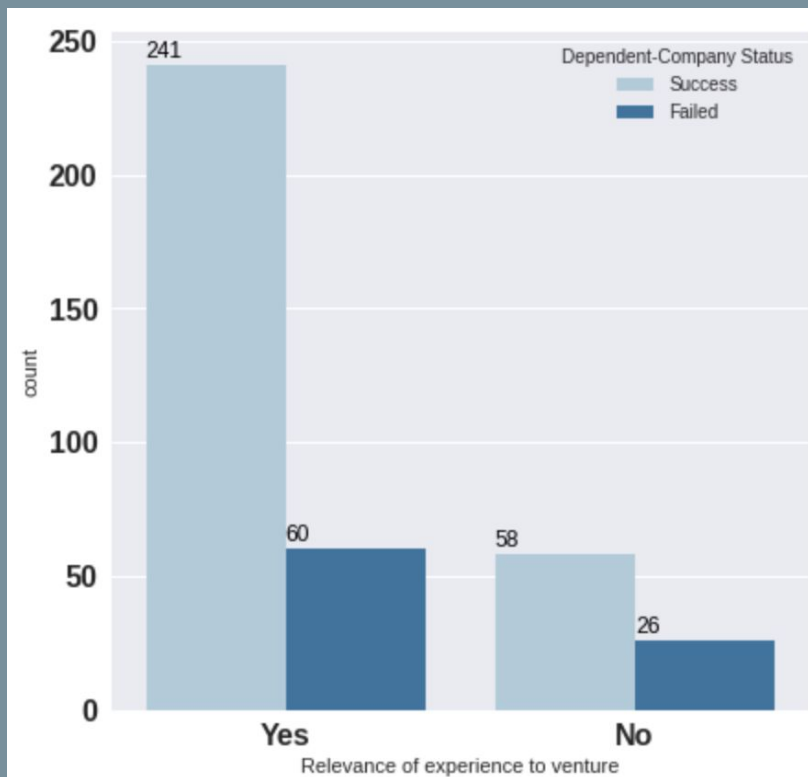
***Note: IQR ~ Interquartile Range**

Data Visualization and Analysis

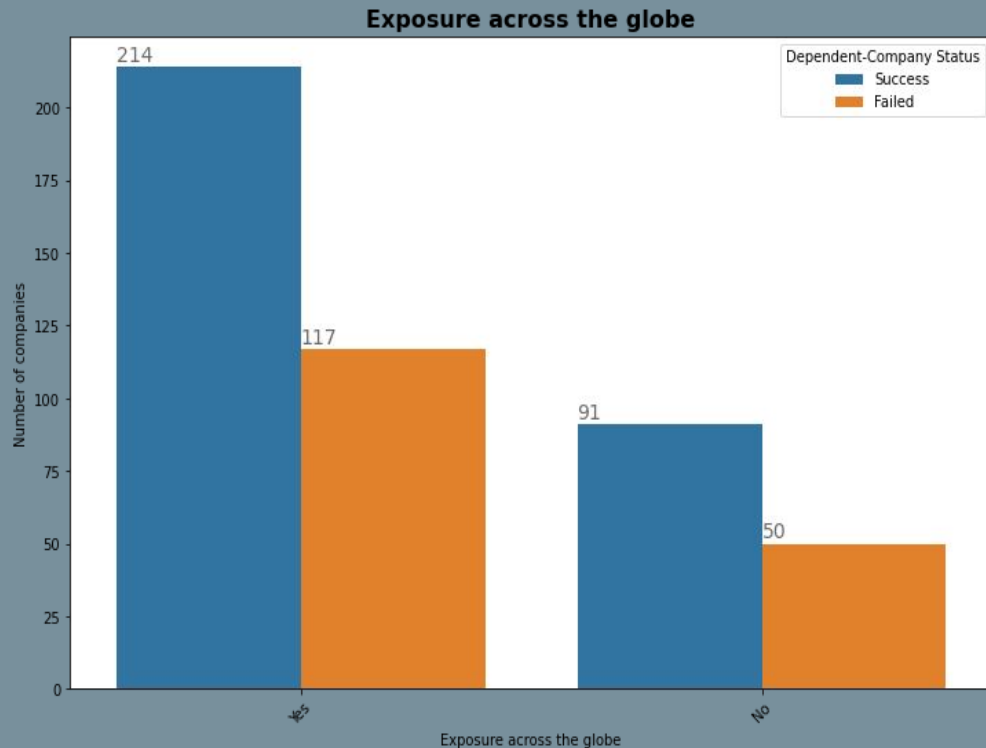
Business Strategy and Product Management



Success Rate vs. Founders Previous Experience



Exposure across Globe vs Success Rate



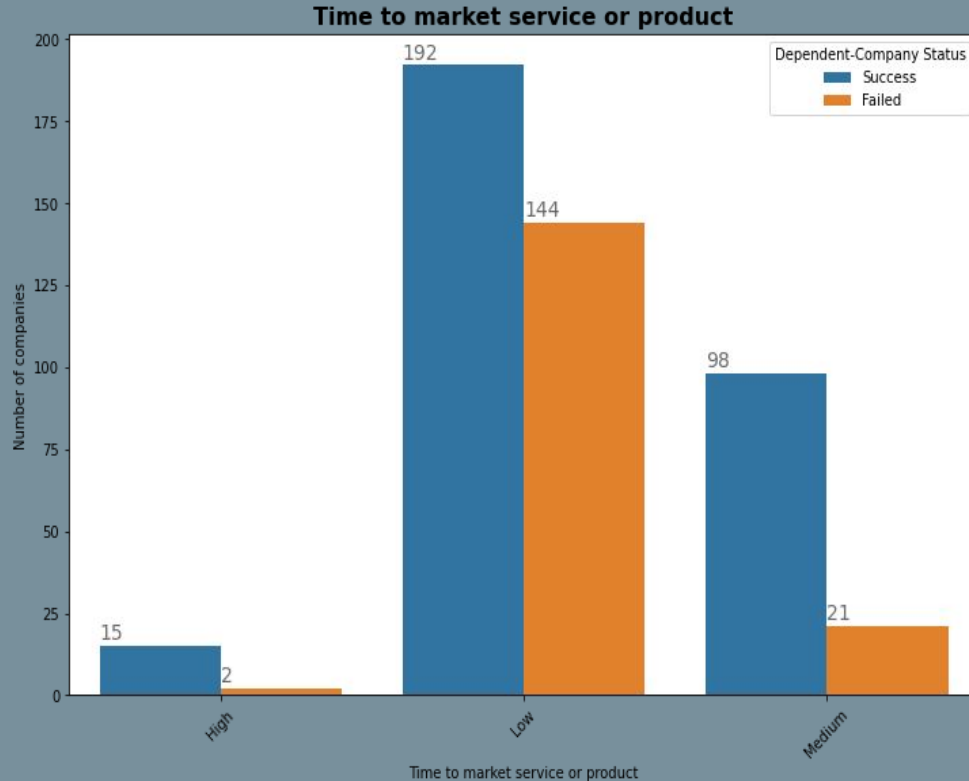
#Successful startups w/ exposure across globe $\approx 2 \times$ successful startups w/o.

#Failed startups w/ exposure across globe $\approx 2 \times$ successful startups w/o.

#Successful startups w/ exposure across globe $\approx 2 \times$ failed startups w/

#Successful startups w/o exposure across globe $\approx 2 \times$ failed startups w/o

Time to Market Service or Product vs Success Rate

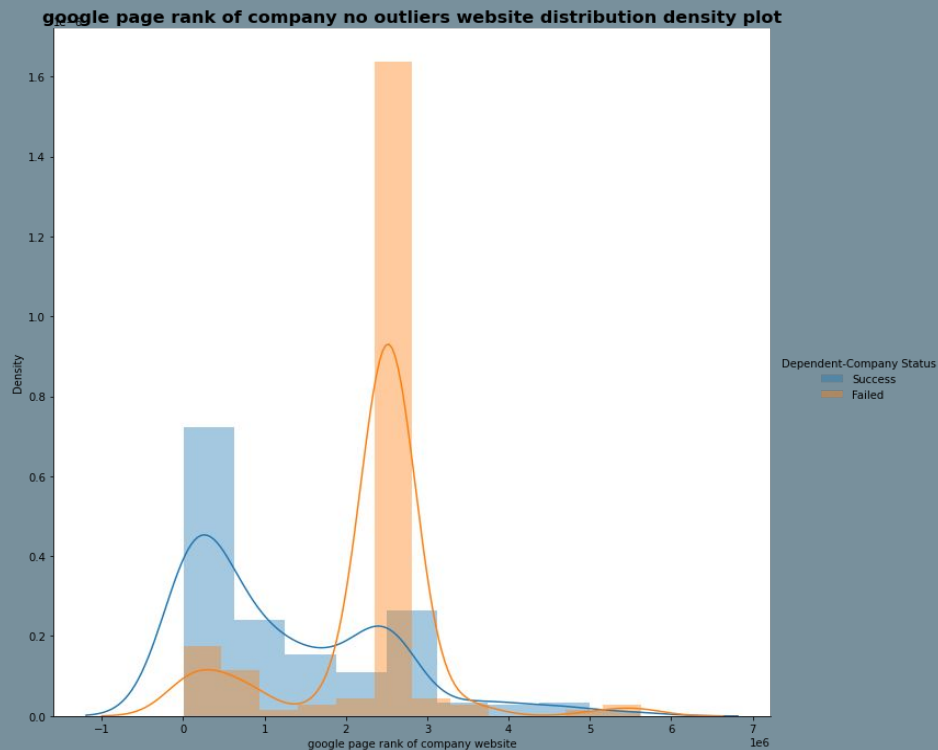


Most successful startups have low medium time (fast)

successful startups w/ low time is higher than failed startups w/

successful startups w/ medium time is significantly higher than failed startups w/

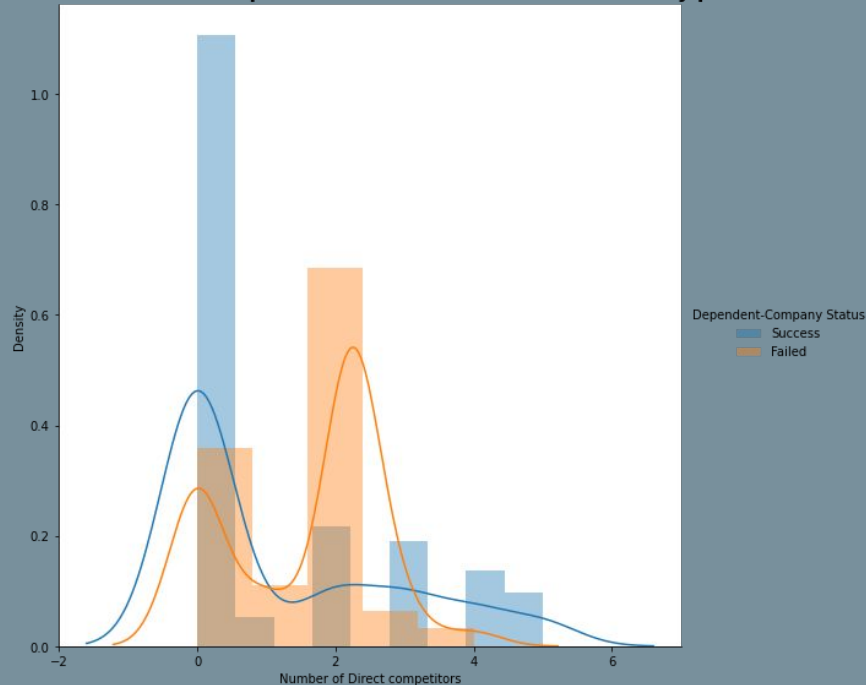
Google Page Rank of Company Website vs Success Rate



Most successful startups have better page rank (nearer to the top rank) than failed startups.

Number of Direct Competitors vs Success Rate

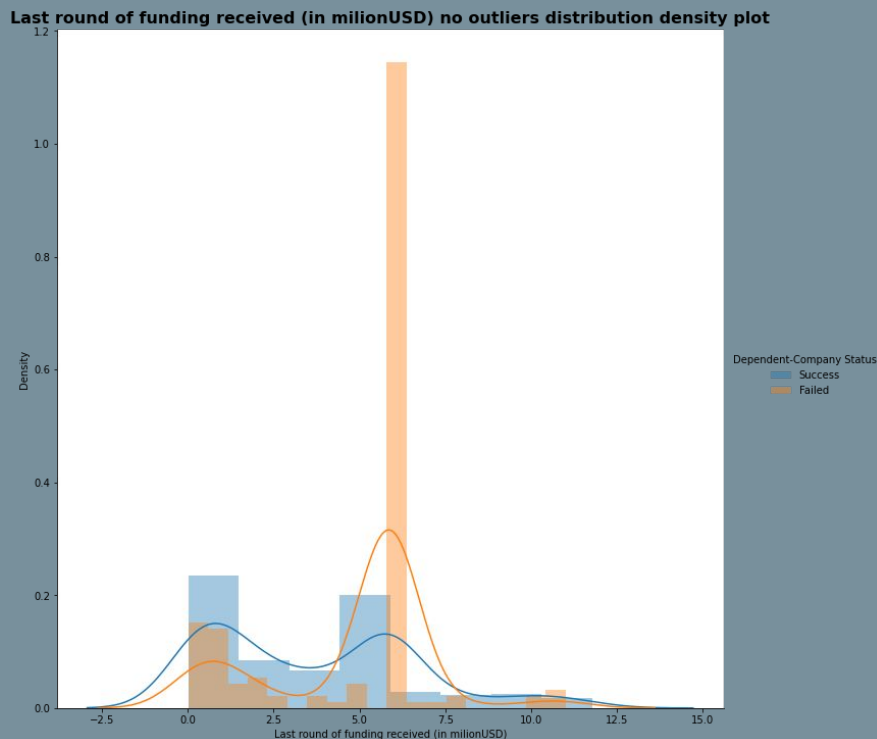
Number of Direct competitors no outliers distribution density plot



Most successful startups have fewer competitors.

Most failed startups have 2 direct competitors.

Last Round of Funding Received vs Success Rate



In general, # successful startups are higher than # failed startups with last round of funding received.

However, significant higher # failed startups w/ last round of funding received about 6 million USD

Multivariate Analysis



Correlations between any 2 of these variables are below 0.50.

“google page rank of company website” and “last round of funding received” have correlation value of -0.11

Summary

1. The data is relevant to United States market.
2. Success rate of a startups tends to decrease if they focus more into business strategy.
3. Startups take a low time to market their service or product. The percentage of successful startups is much higher than failed startups for medium time marketing.
4. More competition decreases the rate of success for a startup.
5. The number of successful startups are higher than the number of failed startups with last round of funding received.

Thank You