



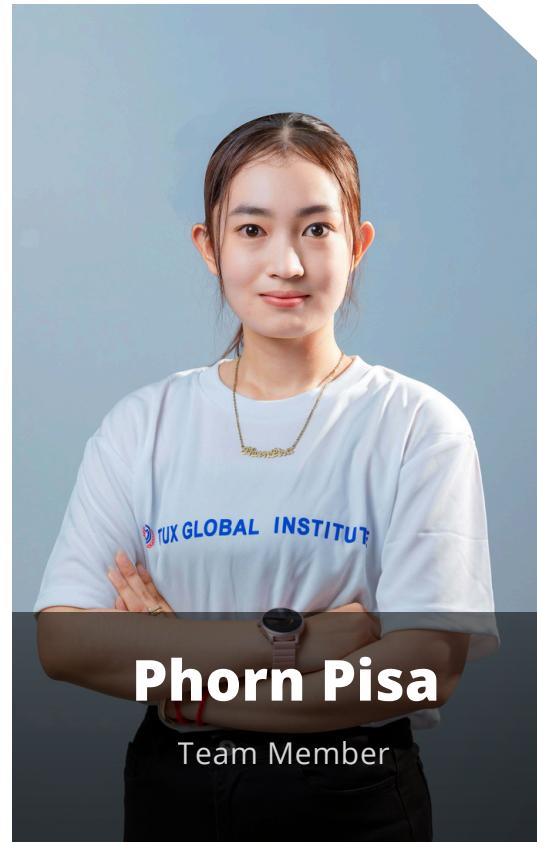
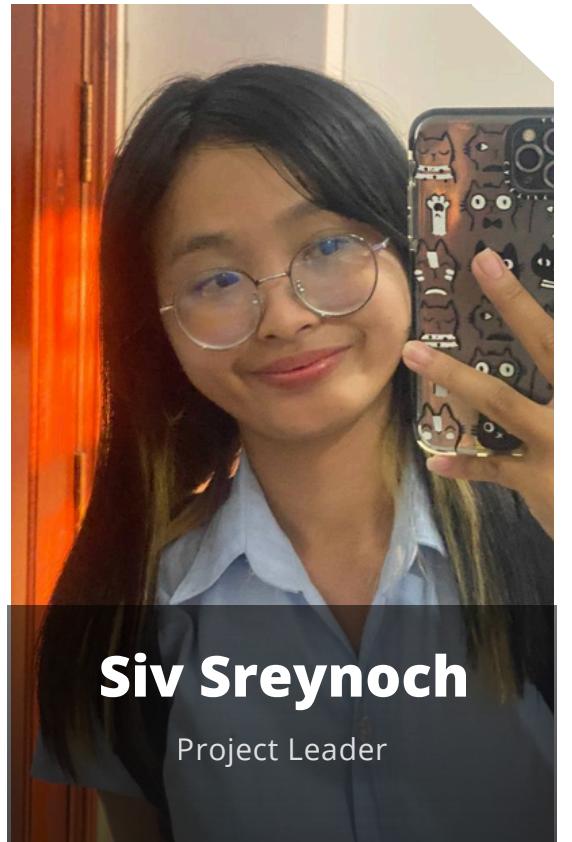
TUX GLOBAL
INSTITUTE

STARTUP SUCCESS ANALYSIS

COURSE
DATA SCIENCES

LECTURER
SOKUNSATYA SANGVAT





CONTENTS :

- 1. INTRODUCTION**
- 2. OBJECTIVE**
- 3. PROBLEM STATEMENT**
- 4. SOLUTION**
- 5. DATA COLLECTION**
- 6. EXPLORATORY DATA ANALYSIS (EDA)**
- 7. MACHINE LEARNING MODEL**
- 8. CONCLUSIONS**

1. INTRODUCTION



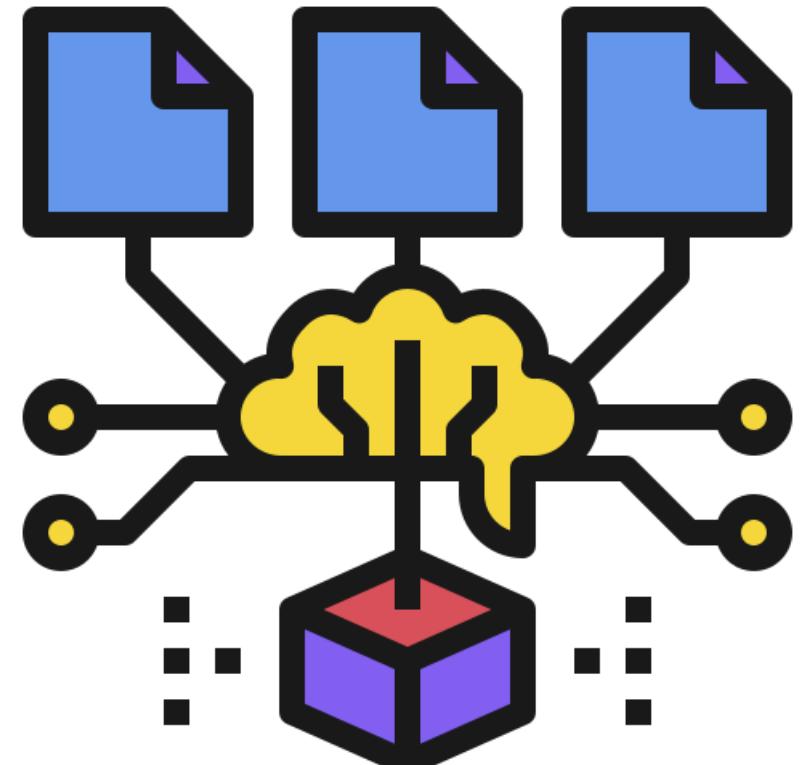
Startup Success Analysis meaning that the systematic study of a lot of factors and conditions that contribute to the success or failure of startup companies. This analysis and prediction of machine learning algorithm involve collecting, processing, and examining data related to startups to identify patterns, trends, and key determinants of success.

2. OBJECTIVE

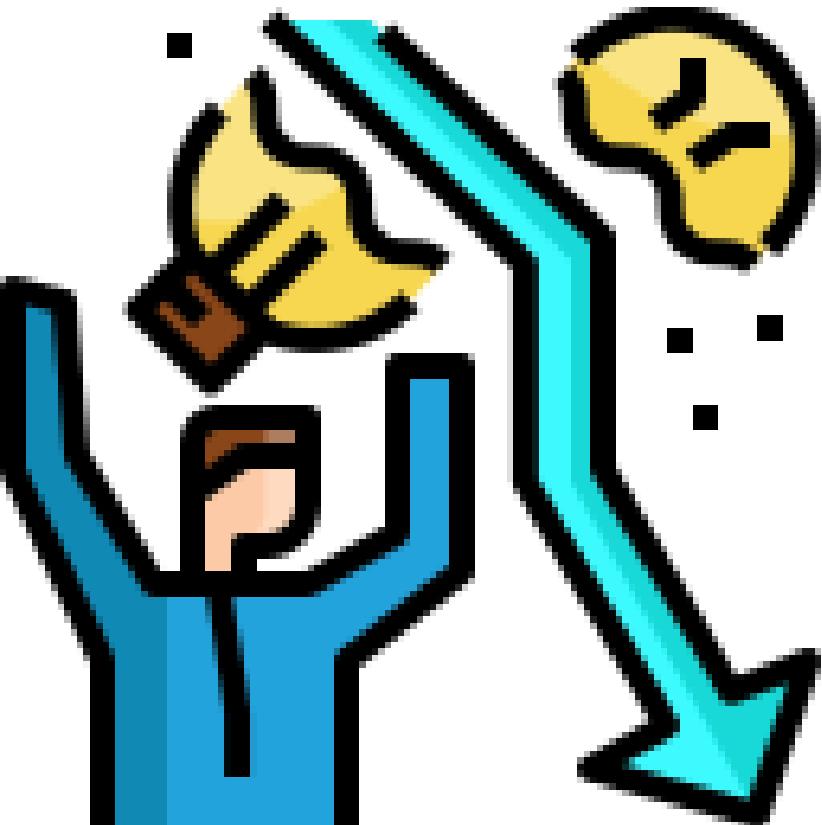
The objective of the startup success analysis is to find factors that impact failure or success for the startup company and build machine learning to predict those factors.



3. PROBLEM STATEMENT



Lack of Predictive
Models for startup
company



High Failure Rate
for new startup



Hard for decision-
making in strategic
planning and other

4. SOLUTION



Build the start-up success EDA and prediction algorithm to predict failure or success.

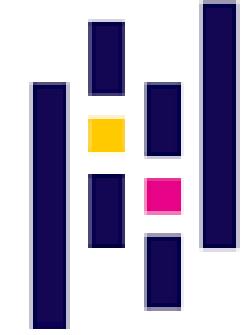
5. DATA COLLECTION

Origin of the Datasets:

- Research Title: EDA Startup Success Analysis
- Resource: kaggle.com
- Publish: 2019

The Kaggle logo is displayed in a large, bold, light blue sans-serif font. The letters are slightly rounded and have a slight shadow effect, giving them a three-dimensional appearance.

DATA CLEANING



pandas

- Define data types
- Checking for number of missing values of each columns
- Drop unnecessary columns
- Fill missing value with NaN
- Correcting same meaning value (e.g. yes, YES -> Yes)



seaborn

DROP UNNECESSARY COLUMNS

Columns that contain more than 30% missing values are needed to removed.

We manage to adjust our datasets from:

- 472 rows × 116 columns

To:

- 472 rows × 105 columns

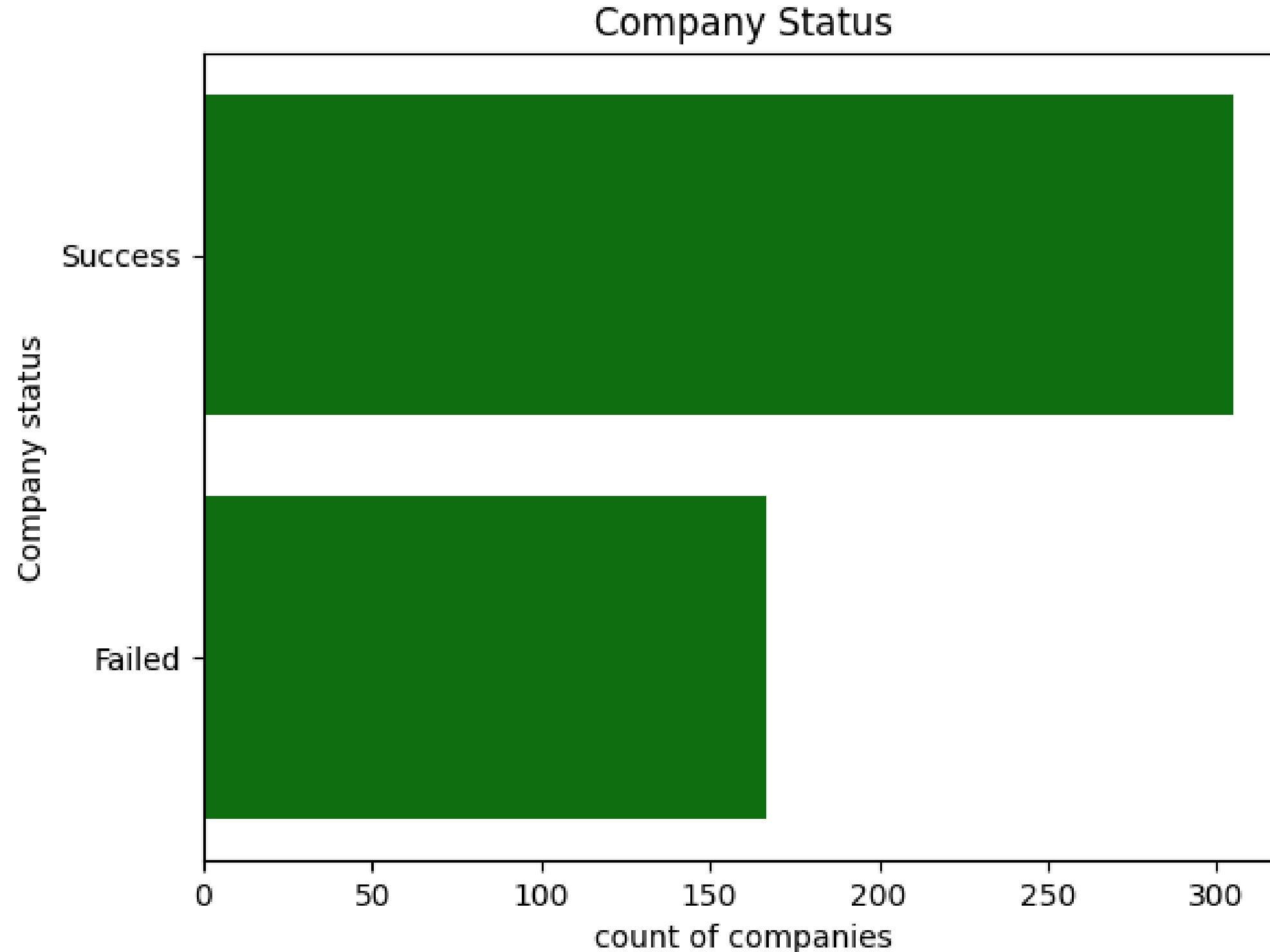
Total of 11 columns removed.

matplotlib

6. EXPLORATORY DATA ANALYSIS (EDA)

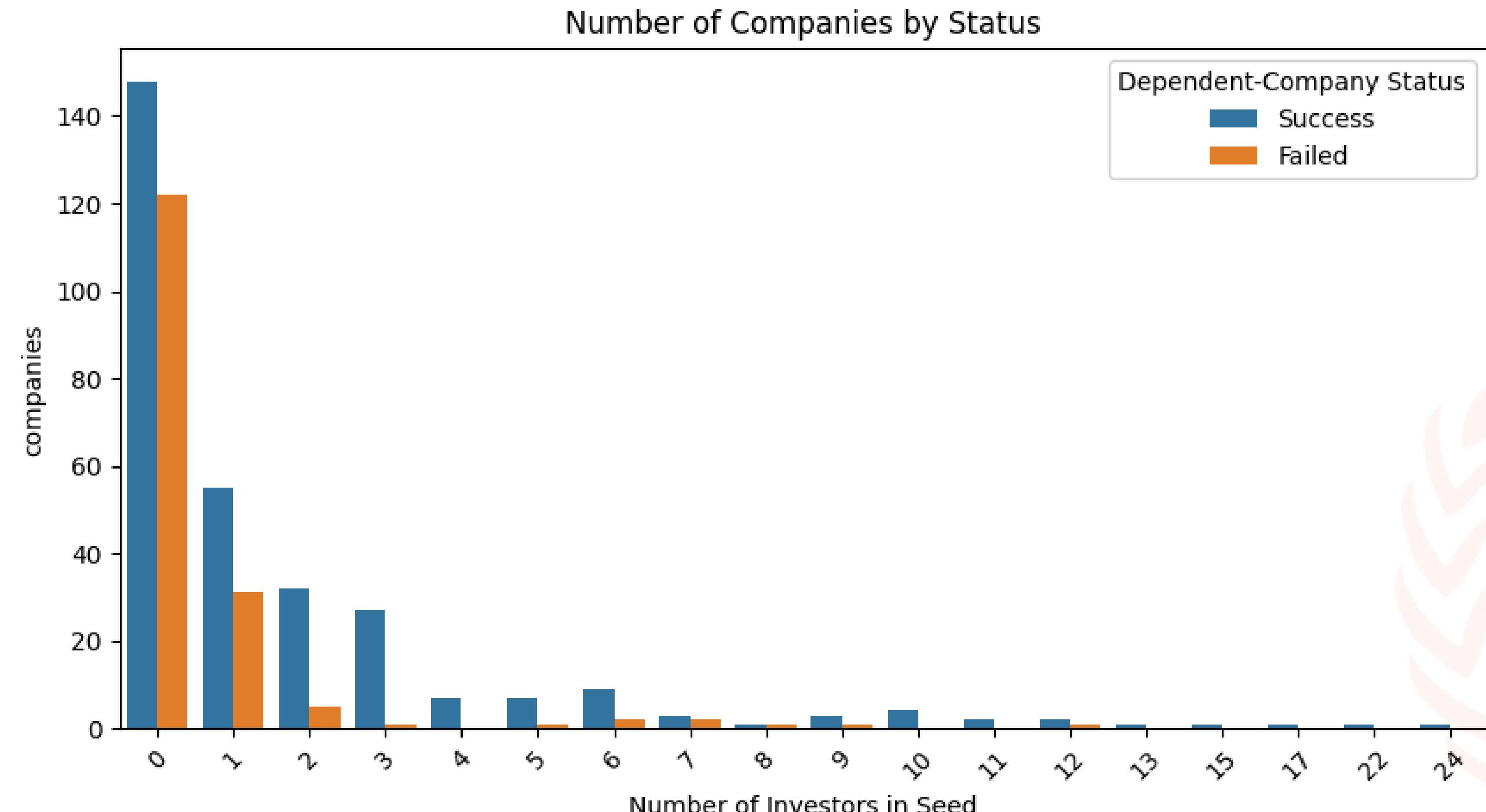
Company status

- Status and Amount of company



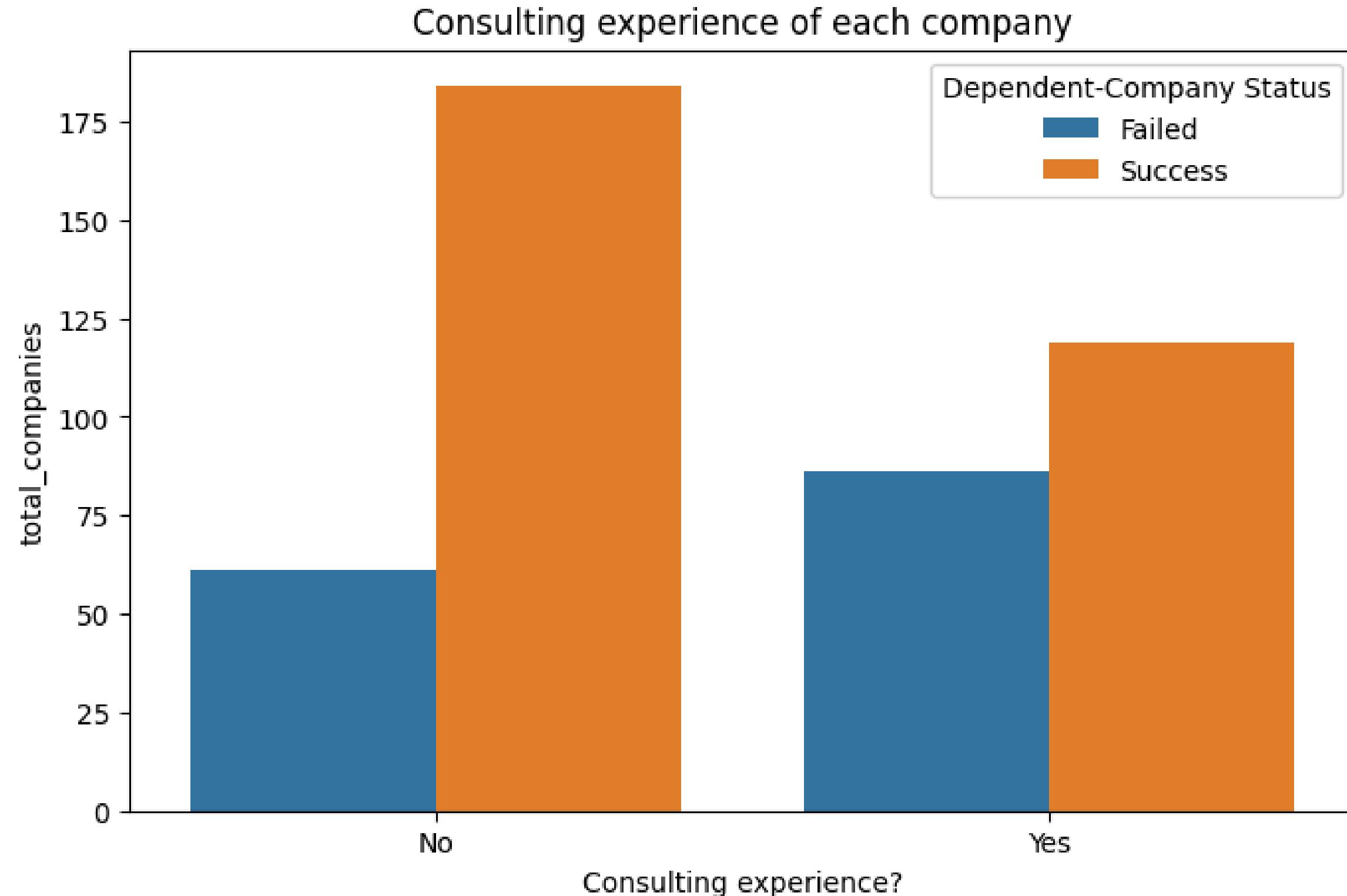
Company status

- Amount of investor in seed of each companies



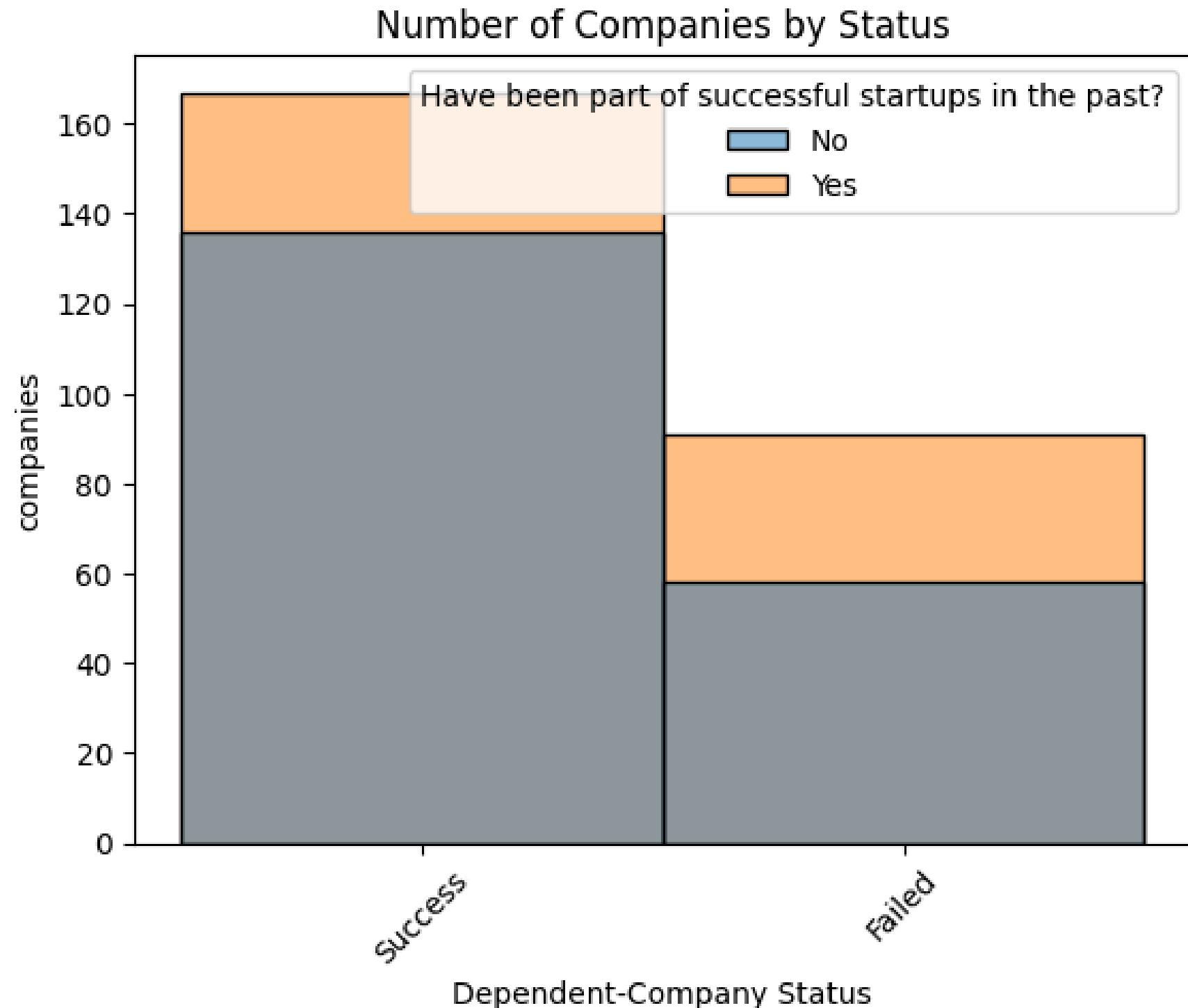
Company status

- Consulting experience of each companies



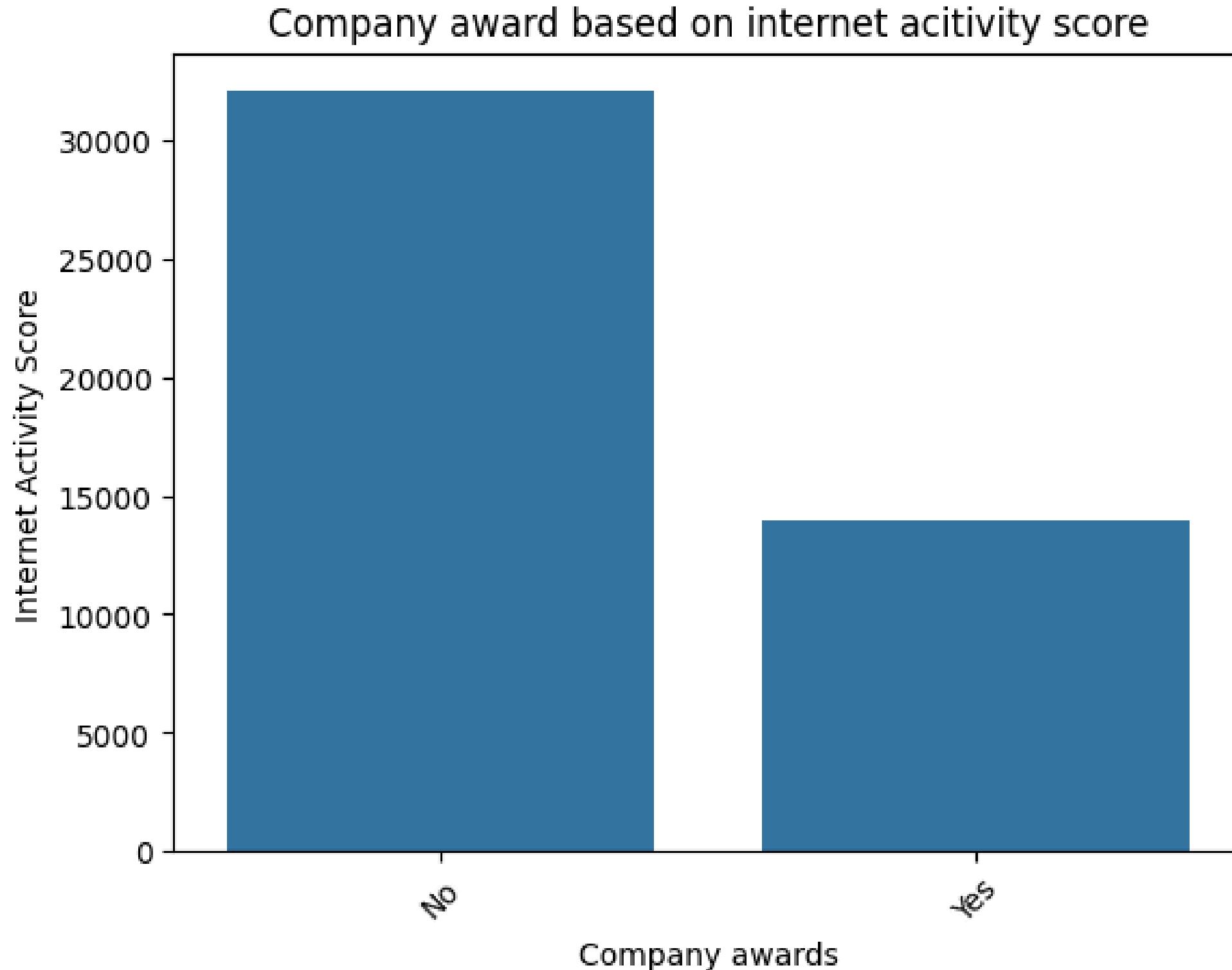
Company status

- start up that used to successful in the past



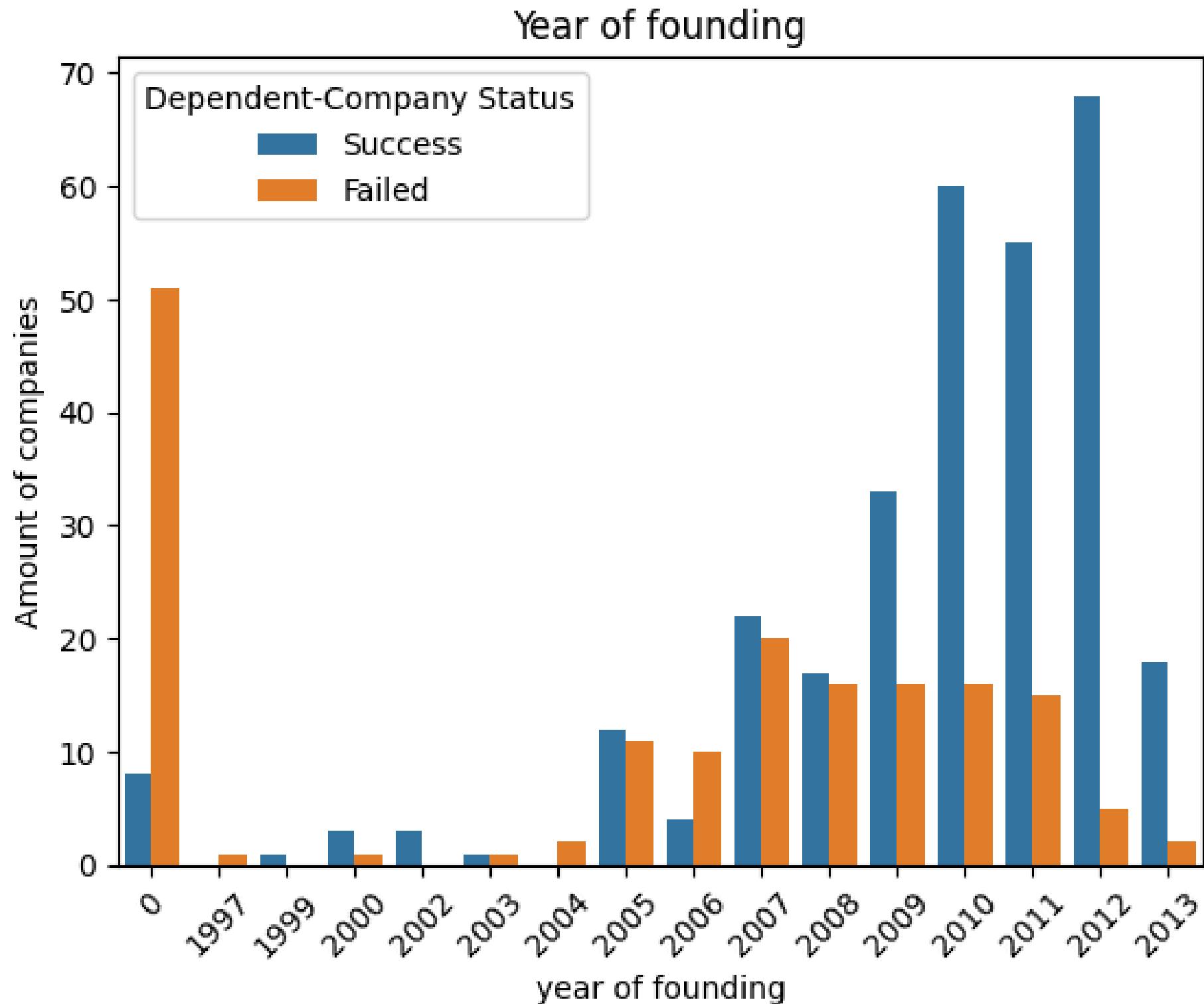
Company status

- Company award base on internet activity score



Information with year

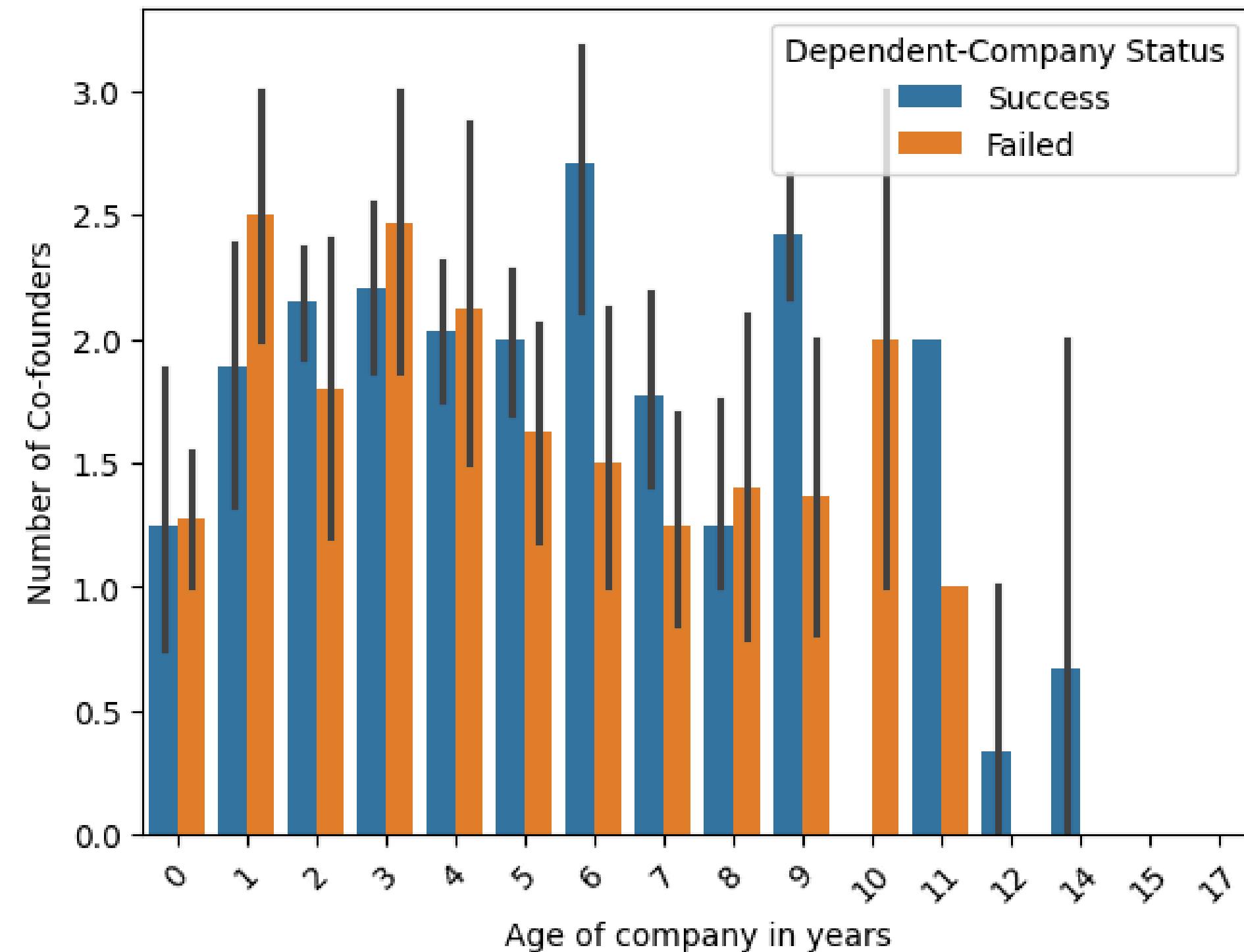
- Year of founding



Information with year

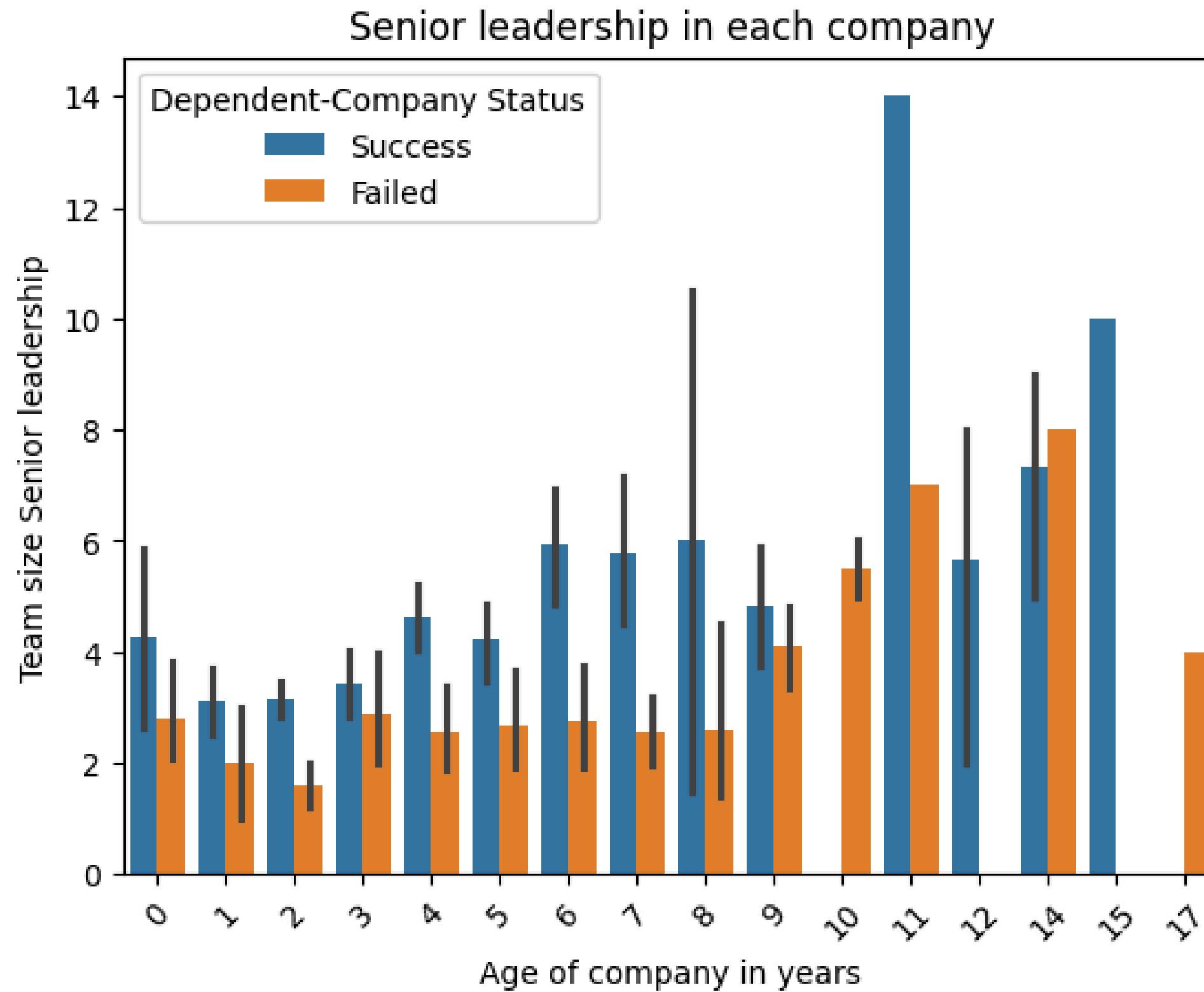
- Co-founder

Co-founder vs the age of the company year



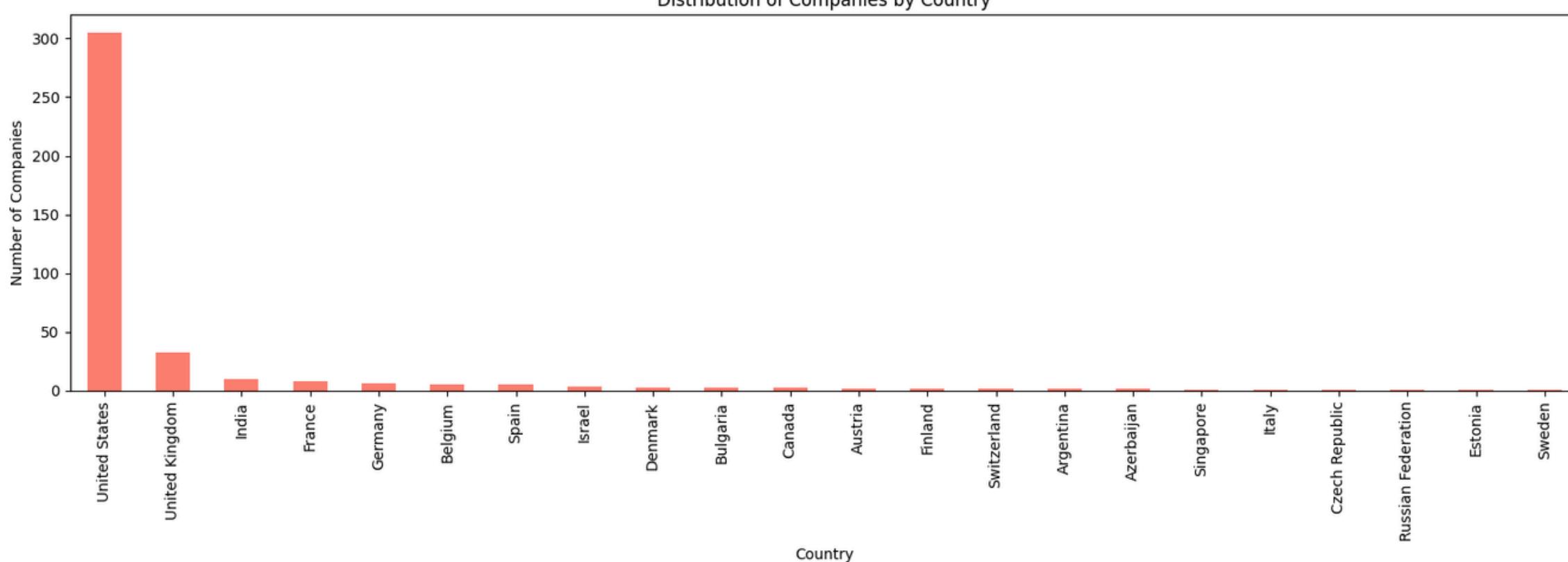
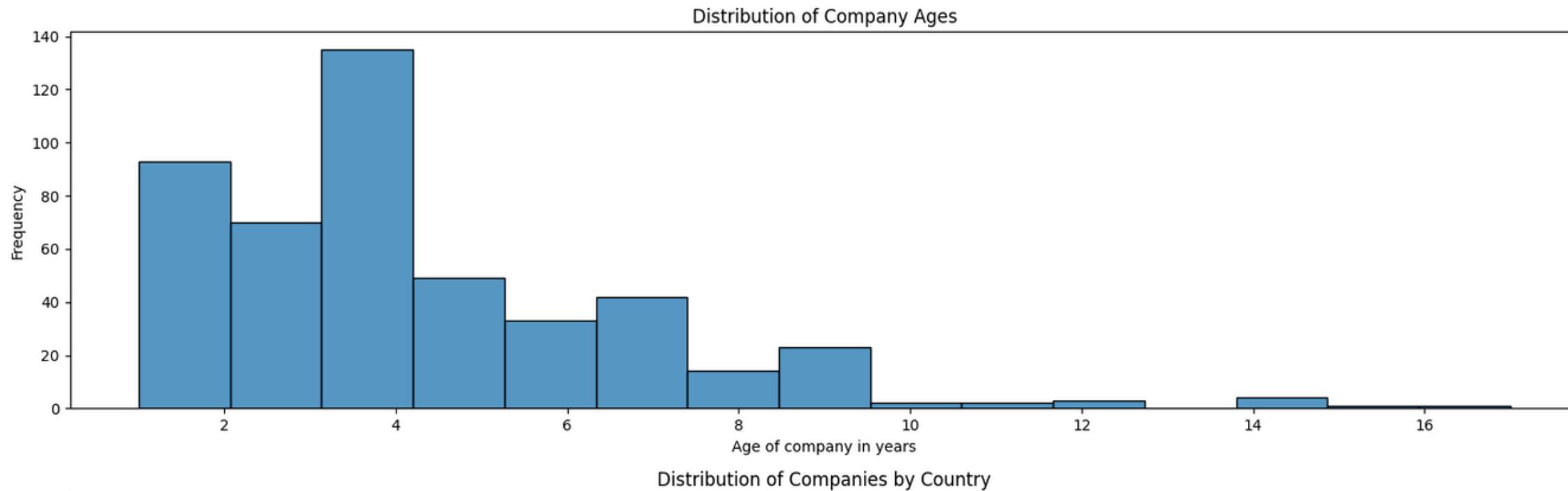
Information with year

- Senior leadership

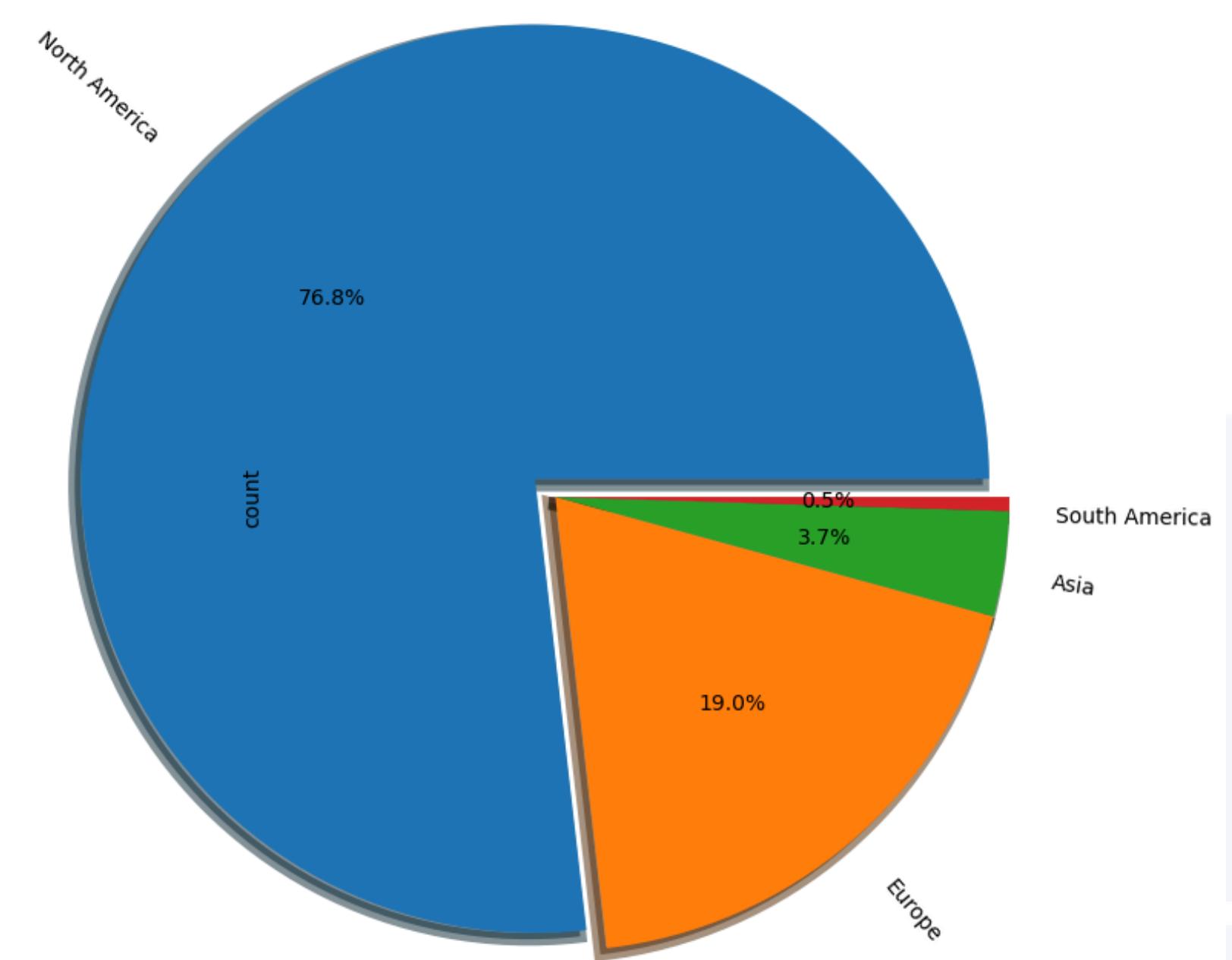
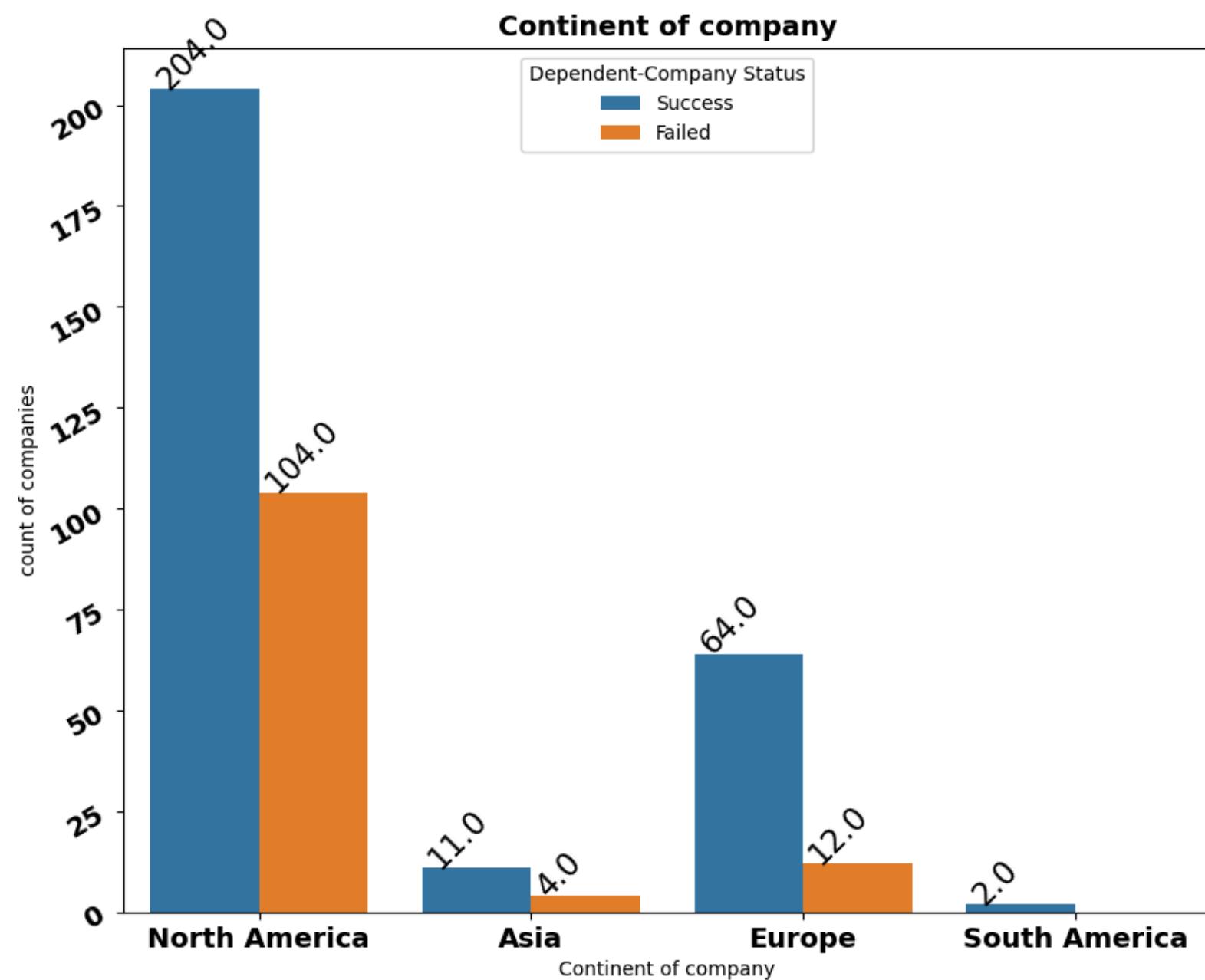


Information with year

- Age of company in years Country of company



Continent of Company



Industry of company

- Industry of companies that succeeded

1. Top Industries:

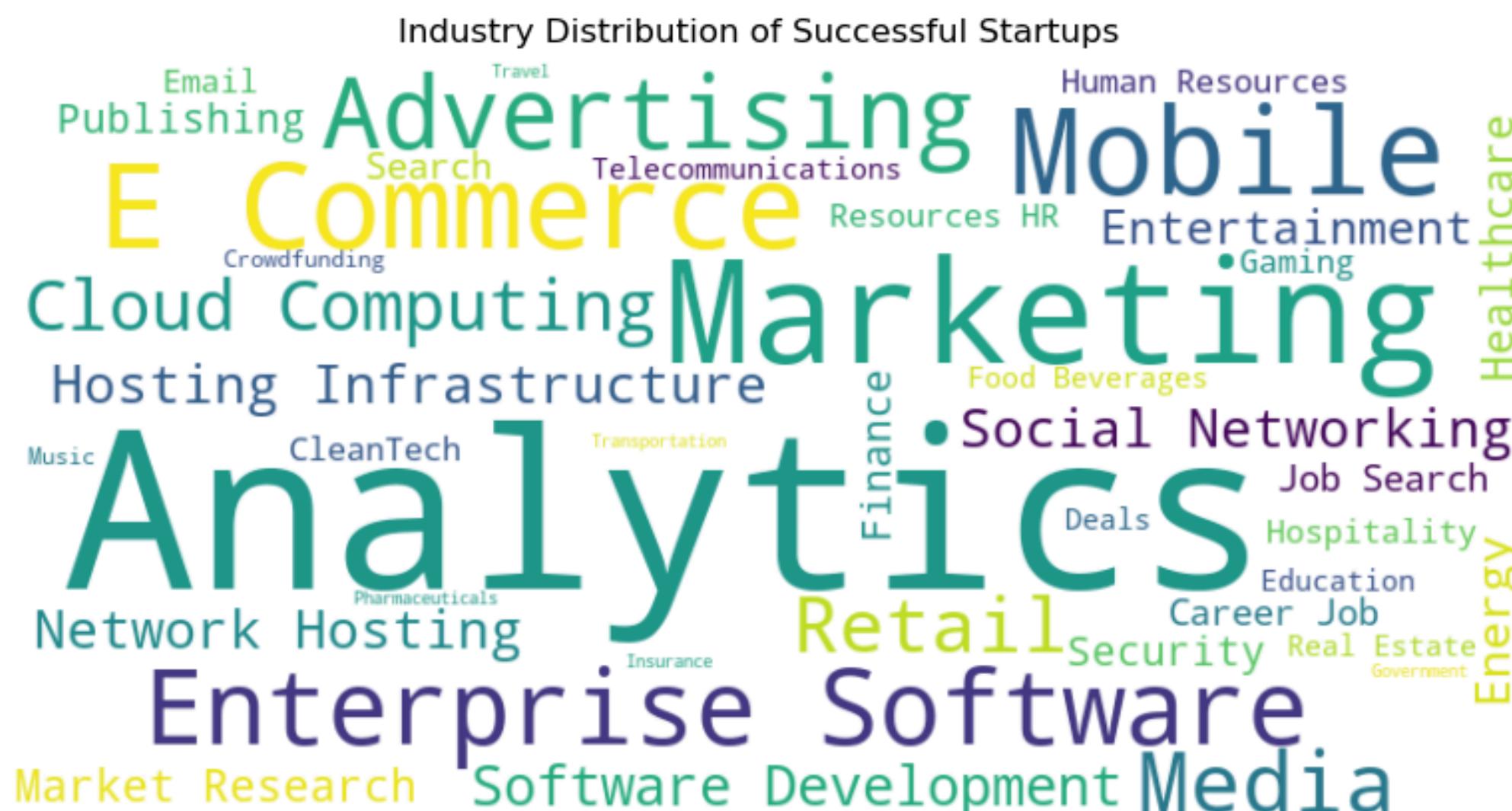
- Analytics
 - Enterprise Software
 - Marketing
 - Mobile
 - E-commerce

2. Significant Industries:

- Advertising
 - Cloud Computing
 - Social Networking
 - Media
 - Retail

3. Other Notable Industries:

- Healthcare
 - Network Hosting
 - Software Development
 - Security
 - Market Research
 - Energy



Industry of company

- Industry of companies that Failed

- ## 1. Top Industries:

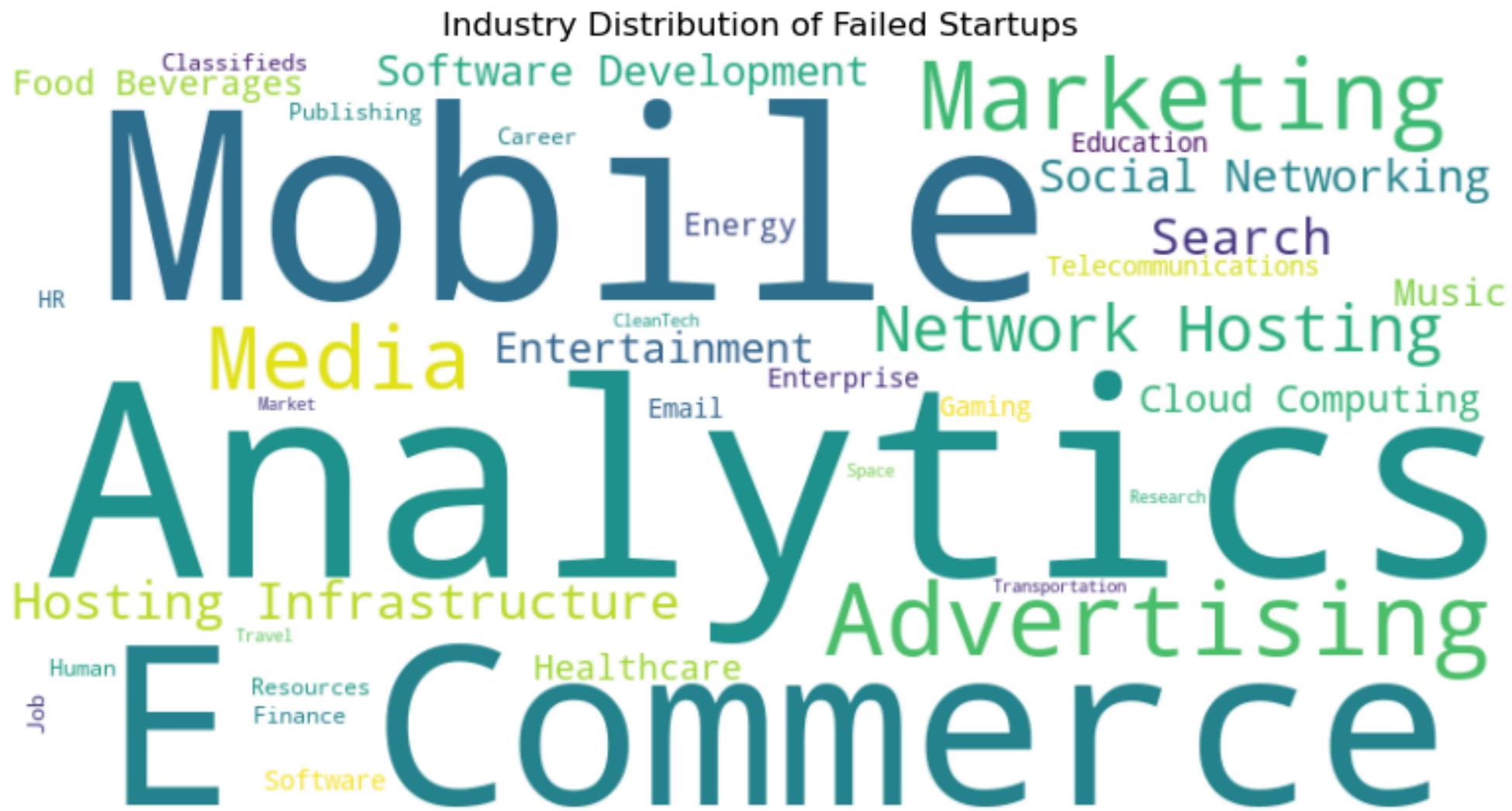
- Mobile
 - Analytics
 - E-commerce
 - Marketing

- ## 2. Significant Industries:

- Advertising
 - Network Hosting
 - Media
 - Cloud Computing

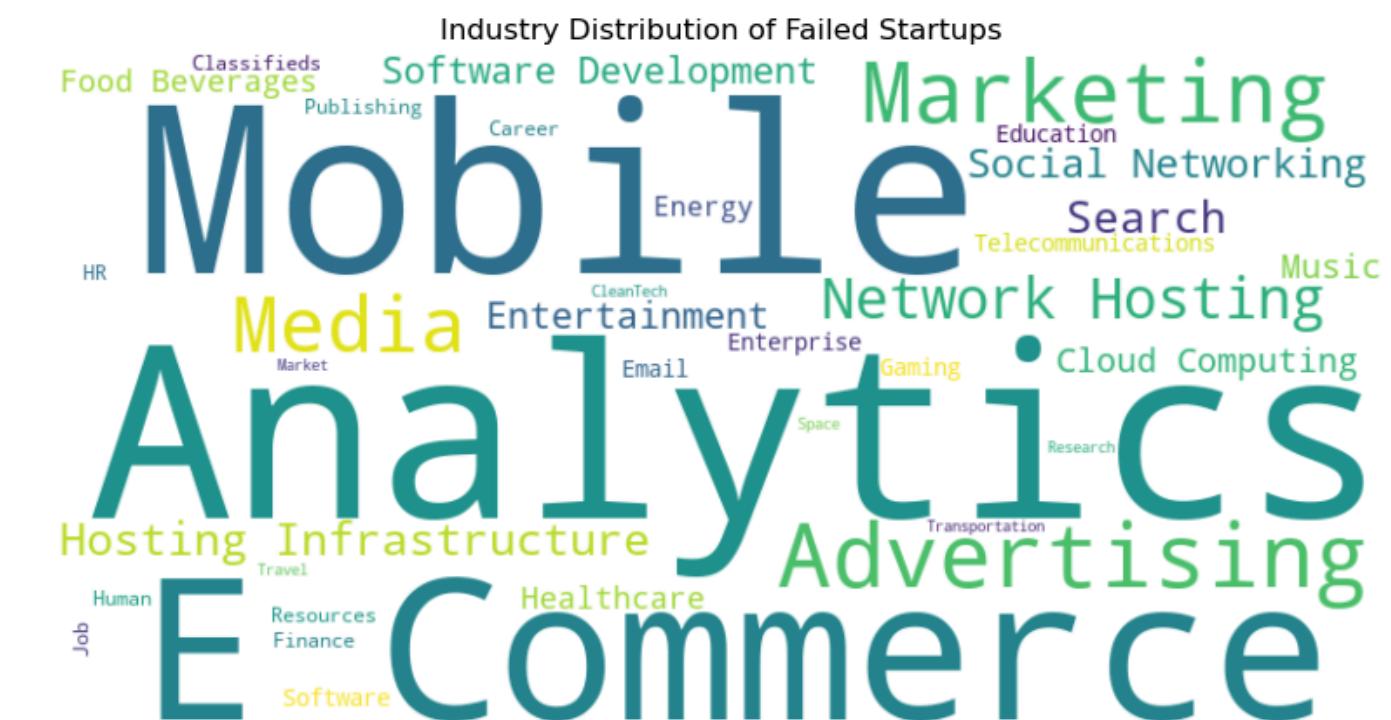
- ### 3. Other Notable Industries:

- Software Development
 - Social Networking
 - Hosting Infrastructure
 - Entertainment



COMPARATION

- **Common Sectors** : Both succeeded and failed startups frequently appear in Analytics, Mobile, Marketing, and E-commerce, indicating high activity and competition in these sectors.
 - **Successful Startups** : Industries like Enterprise Software, Retail, and Network Hosting show strong success rates.
 - **Failed Startups**: Industries like Software Development, Social Networking, and Hosting Infrastructure show higher failure rates.



7. MACHINE LEARNING MODEL

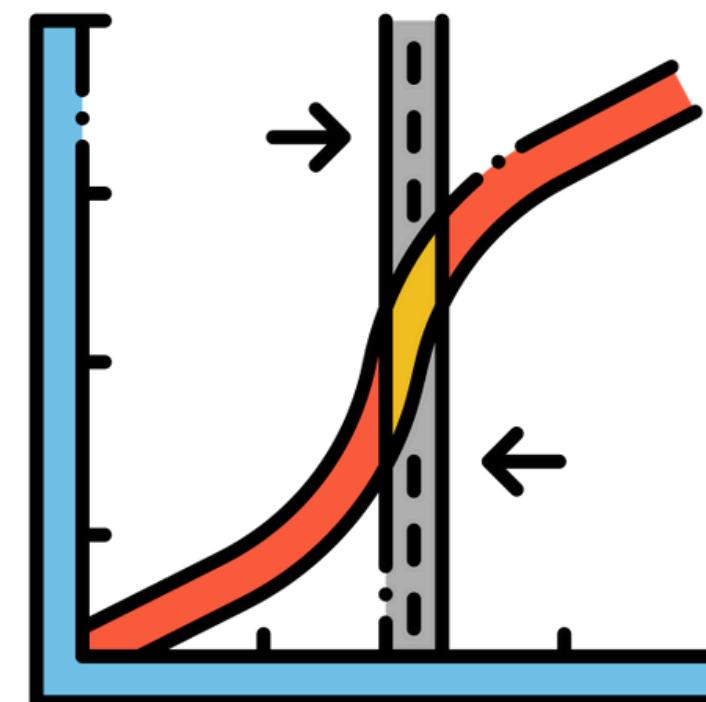
We want to predict and classify startups into categories (success vs. failure)

- ***Logistic regression*** : Logistic regression is a statistical method used for binary classification problems, making it highly suitable for predicting whether a startup will succeed or fail.

```
Best parameters found: {'C': 10, 'penalty': 'l1', 'solver': 'liblinear'}
```

```
Best cross-validation score: 0.88
```

```
Accuracy: 0.8210526315789474
```



7. MACHINE LEARNING MODEL

Logistic regression:

	precision	recall	f1-score	support
Failed	0.70	0.89	0.78	35
Success	0.92	0.78	0.85	60
accuracy			0.82	95
macro avg	0.81	0.83	0.82	95
weighted avg	0.84	0.82	0.82	95

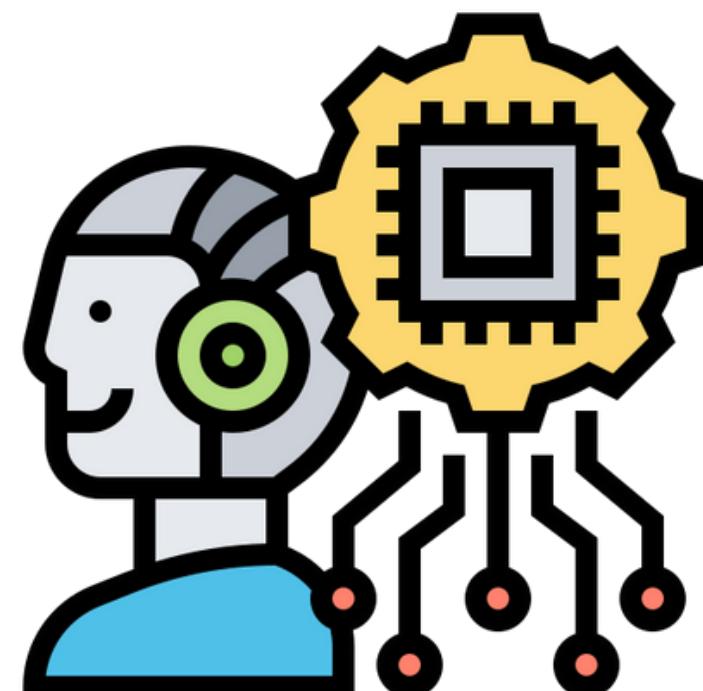
7. MACHINE LEARNING MODEL

- Ensemble Learning: Bagging Classifier and DecisionTreeClassifier

Mean Cross-Validation Score: 0.8767287234042553

Bagging Classifier Classification Report:

	precision	recall	f1-score	support
Failed	0.89	0.89	0.89	37
Success	0.93	0.93	0.93	58
accuracy			0.92	95
macro avg	0.91	0.91	0.91	95
weighted avg	0.92	0.92	0.92	95



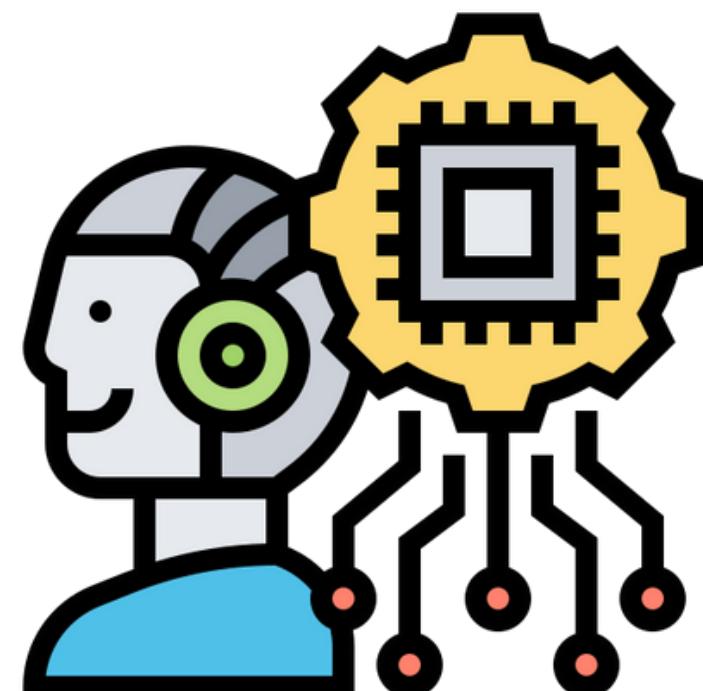
7. MACHINE LEARNING MODEL

- Ensemble Learning: GradientBoostingClassifier and DecisionTreeClassifier

Mean Cross-Validation Score: 0.8979609929078014

--
Gradient Boosting Classifier Classification Report:

	precision	recall	f1-score	support
Failed	0.97	0.92	0.94	37
Success	0.95	0.98	0.97	58
accuracy			0.96	95
macro avg	0.96	0.95	0.96	95
weighted avg	0.96	0.96	0.96	95



3. CONCLUSIONS

- Objective
- Problem statement
- Solution
- Data collection
- Exploratory Data Analysis (EDA)
- Machine learning model



THANK YOU !!!

