A logo for college computing

Description automatically generated

**Predicting Start-up Success**

|  |  |
| --- | --- |
| *Student Full Name:* | Karen Montserrat Andraca Guzman |
| *Student Number:* | 2024345 |
| *Module Title:* | Strategic Thinking |
| *Assessment Title:* | CA 2 Capstone Report |
| *Assessment Due Date: Data of Submission* | 15-12-2024 |
|  |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Abstract

The goal of this capstone project is to create a prediction model that will evaluate the probability that a start-up company will become a unicorn, which is defined as having a valuation of more than $1 billion. We will examine a number of factors that affect start-up success, such as industry sector, funding history, and financial performance, using publicly accessible data from sources including Crunchbase and Kaggle.

After identifying important trends and behaviours through exploratory data analysis (EDA), pertinent features will be chosen for modelling. To develop a strong prediction model, we will use machine learning techniques learning in class like logistic regression and random forests.

The outcomes, which highlight significant features that support a start-ups potential for quick development and success, will offer useful information to investors and entrepreneurs. In addition to improving knowledge about unicorn scene, this study intends to provide practical suggestions for encouraging investment and creativity in up-and-coming businesses.

Contents

[Introduction 1](#_Toc184921914)

[Chapter 1 2](#_Toc184921915)

[What is a start-up - Chapter 1.1 2](#_Toc184921916)

[What is a “Unicorn Company” Chapter 1.1.1. 2](#_Toc184921917)

[Chapter 2 3](#_Toc184921918)

[General Objective 3](#_Toc184921919)

[Specific Objective 3](#_Toc184921920)

[Methodology 3](#_Toc184921921)

[Work Plan 3](#_Toc184921922)

[Expected Outcomes 4](#_Toc184921923)

[Requires Resources 4](#_Toc184921924)

[CA – 2 4](#_Toc184921925)

[Data Preparation 5](#_Toc184921926)

[Findings 6](#_Toc184921927)

[Conclusion 7](#_Toc184921928)

[References 8](#_Toc184921929)

# Introduction

The last ten years have seen a dramatic change in the start-up scene, with an increase in the number of businesses becoming unicorns—those valued at more than $1 billion. These unicorns have become significant players in the world economy, fostering innovation, generating employment, and boosting Economy. Strong investment, talent access, and encouraging government policies all contribute to each country successful startup ecosystem. But even with the enthusiasm and promise, most entrepreneurs fail to reach this level of success.

Investors, founders, and politicians must all have an excellent knowledge of the characteristics that set successful enterprises apart from those that fail. The goal of this capstone project is to create a prediction model that identifies the crucial qualities and factors influencing a start-ups chances of becoming a unicorn. Through the examination of past data from many sources, we aim to find patterns that help direct strategic planning and decision-making.

# Chapter 1

## What is a start-up - Chapter 1.1

I want to begin by examining the idea of a start-up and how, although it is a very basic description, it also fits into the modern world and its business practices. [[1]](#footnote-1)Eric Ries defines a startup as an organization dedicated to creating something new under conditions of extreme uncertainty.

To put it another way, a start-up is a business that is still in its beginning and aims to provide an innovative service or product that will lead to success and the positioning required to make money. However, the founders are inexperienced and frequently lack the resources needed to establish a traditional company, requiring apply with institutions or investors, known as "angels" in this industry, which suggests a significant challenge given the risks and uncertainty of the business.

### What is a “Unicorn Company” Chapter 1.1.1.

We begin by defining a unicorn company, which is essentially a start-up that has been established in the market and has a valuation of at least $1,000 million dollars without being public companies (listed on the stock exchange). As excellent examples, we have the following at the top of the global rankings:A screenshot of a phone

Description automatically generated

Source: CB Insights.

The Chinese media and entertainment company ByteDance, which owns the social media platform TikTok, is the world’s most valuable unicorn company at $225 billion.

Meanwhile, the largest unicorn company in the United States is SpaceX, led and founded by Elon Musk, valued at $150 billion. SpaceX is the only company on the list in the industrial category.

OpenAI, founded in 2015, is the world’s highest-valued artificial intelligence pure-play company. Best known for its groundbreaking program ChatGPT, it is worth $80 billion.[[2]](#footnote-2)

# Chapter 2

## General Objective

Develop a model that predicts the likelihood of a startup becoming a unicorn. By examining data and patterns of unicorn companies, we want to forecast how a budding business can become a unicorn.

## Specific Objective

Analysing the business sectors in the and the ranking of unicorn companies around the world to identify which one is most likely to produce a favourable environment for new start-ups and prosperous expansion is our next goal.

And lastly, apply the models we learnt in class to evaluate and interpret the data and confirm whether it makes sense for an investor to place an investment on an start-up.

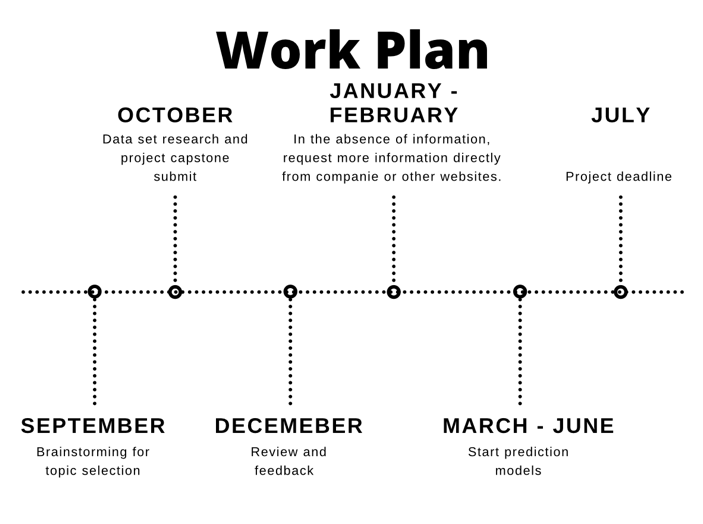
## Methodology

As part of the planned methodology, data will be gathered from a variety of websites, including Crunchbase, Kaggle, and a few others to supplement if I am unable to gather all the information I need. Once I have the information I need, I have chosen to test the following models: To assess my versions' success and failure in light of the binary classification, I have selected the logistic regression model in order to increase prediction accuracy and view successful and unsuccessful examples while accounting for independent variables. In order to assess the significance of each variable in the prediction and to have enough data for the prediction, I have chosen to employ the random forest model, which will examine subsets of features to add diversity and lower the chance of overfitting.

## Work Plan

To determine whether a startup has the potential to become a unicorn company, it will be required to collect important data, such as the profitability of products and services, and the sources of revenue. In addition, it is meant to analyse the market's size, trends, and direct and indirect competitors. We could go over the financial area's projections, expenses, income growth rates, and investor engagement and investment levels.

Here there’s a time line of the work plan.



## Expected Outcomes

According to the findings we anticipate, investors and entrepreneurs will have a better idea of what steps to take to make their start-up businesses stand out and, in the best case scenario, become unicorn companies. Given that some of the required data is sensitive information specific to each company and that there are likely to be fewer examples in the open websites, the capacity and volume of information that will be gathered will determine the best and most accurate panoramas.

## Requires Resources

If we use the "Unicorn Companies Overview" database, which is available for free and in csv format on the Kaggle website, as a source of information, we can look at information like each company's valuation, industry they are in, city of origin, year of foundation, amount of money raised, and investors involved. This will allow us to determine the most likely scenario for a start-up to succeed in becoming a unicorn by looking at typical examples of the biggest unicorns in the world.

The project is dedicated to being transparent in its approach, clear information will always be provided on the origin of the data and its analysis, essential ethical characteristics are taken into consideration that are important to guarantee the integrity and responsibility of the use of the data when carrying out this project to predict the success of a start-up, and the school and teachers are exempt from any improper usage of data, pointing out that this project is merely educational and for use only by the subject. This project will continue maintain its ethical standards and be receptive to criticism and a review of its applied methodologies.

# CA – 2

The reasons behind a startup's failure  
To begin this second section of my capstone project on predicting a start-up's success, I will first discuss the characteristics of a start-up company that fails in its attempt to succeed. This is because in the first section, I discussed how I hope to evaluate a start-up's success, what features need to be evaluated in my data set, and generally what action plan I had estimated to do.

According to an article in FastCompany, "Why Most Venture Backed Companies Fail," 75 percent of venture-backed startups fail. This statistic is based on a Harvard Business School study . In a study by Statistic Brain, Startup Business Failure Rate by Industry, the failure rate of all U.S. companies after five years was over 50 percent, and over 70 percent after 10 years.

This study also asked company leadership the reason for business failure, giving a list of four main reasons for failure with sub-categories below those.

* Lack of focus
* Lack of motivation, commitment and passion
* Too much pride, resulting in an unwillingness to see or listen
* Taking advice from the wrong people
* Lacking good mentorship
* Lack of general and domain-specific business knowledge: finance, operations, and marketing
* Raising too much money too soon
* All of these focus on the decision-making of the entrepreneur and general business knowledge.

We can begin by stating that we conducted a thorough search for the data set that includes as much information as possible in order to obtain a better analysis. We found a fairly robust data set, with 37 columns that I can highlight as the most important information about the company, investors, number of employees, country of origin, and size of all employees, among other more pertinent data for the analysis. However, this data set contains too much information that will not be useful, and when we perform a phyton cleanup, we intend to reduce the data set with the necessary information without exceeding with meaningless entries or out of the scope of the capstone project

Once these null values have been identified, I should think about how much they can affect me. If the percentage of null values in my data is less than 15%, I would leave the information unclean because it is not worth the effort when the information does not really affect my data. However, if the percentage is higher than 15%, I must replace these null values with specific values, such as 0, the mean, or the mode.

## Data Preparation

In phyton, the first thing you'll notice is that we have The dataset is divided into 472 rows and 37 columns, with data types including object—strings or characters that need to be valued and float64 —numerical values—found inside the columns.

We start by deploying the data set's head, where we can see the first entries, headers, and features. Then, we can use the describe and info functions to learn more about the type of data we have. I can use the missing values function to identify which columns contain unique values or null information that are contaminating my data set. Once these null values have been identified, I should think about how much they can affect me. If the percentage of null values in my data is less than 15%, I would leave the information unclean because it is not worth the effort when the information does not really affect my data. However, if the percentage is higher than 15%, I must replace these null values with specific values, such as 0, the mean, or the mode.

It is crucial to concentrate on a good data set cleaning because, in particular, this data set contains too much information, which we intend to analyse in order to reach the best evaluation of the project. During the data set cleaning process, we plan to change column names, look for non-existent values, and decide whether they should be replaced, modified, or simply deleted. If the columns have very long values, we will analyse whether to replace or delete the information.

In order to determine the percentage of accuracy we can call prediction and recall, the machine learning model that will be used to predict next semester is anticipated to be determined. Of these, logistic regression and decision trees have appeared to be the most appealing and could be applicable to my data set.   
We can do more extensive experiments on the Jupyther notebook the following semester and apply the concepts and codes covered in class to get the intended results. After that, I'll present a timeline of the anticipated procedures to be followed.

# 

## Findings

The percentage of success that the companies in our data set exhibit is one of the first findings we discovered. Over the course of the upcoming semester, we will conduct a thorough analysis of all the elements that a start-up company must follow in order to succeed, using these companies as examples.

A graph with blue squares

Description automatically generated

## Conclusion

In contrast to capstone number 1, which had six stages, this new delivery focusses on a more thorough analysis and organisation of my new data set. We examined the types of information we have and how we plan to clean it up so that we can work with it more thoughtfully and efficiently. We also planned to develop the necessary models to make more accurate predictions the following semester.

# References

Ries, E. (2011). The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. United Kingdom: Crown.

Visual Capitalist. (2021, March 18). Ranked: The top 10 most valuable unicorn companies. <https://www.visualcapitalist.com/cp/ranked-the-top-10-most-valuable-unicorn-companies/>

Harvard Business School. (2024)*: [*Why Most Venture Backed Companies Fail*]*. Harvard Business School. Available at: <https://www.hbs.edu/news/Pages/item.aspx?num=214> [Accessed 10 Dec. 2024].

1. Ries, E. (2011). The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. United Kingdom: Crown. [↑](#footnote-ref-1)
2. Visual Capitalist. (2021, March 18). Ranked: The top 10 most valuable unicorn companies. <https://www.visualcapitalist.com/cp/ranked-the-top-10-most-valuable-unicorn-companies/> [↑](#footnote-ref-2)