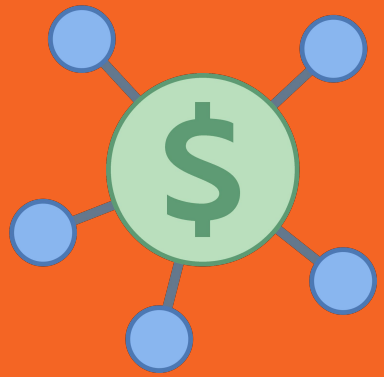
Kickstarter Project Proposal



Team members:
Monica Boomgaarden, Rodrigo Garcia, Nicole Lincoln

Introduction

Kickstarter first launched on April 28th 2009

Kickstarter's purpose is to bring creative projects to life through community backing

Since its inception, over 19 million backers and $5.3 Billion dollars pledged

What data can be mined and what interesting questions can be asked?

Objective

We will be looking at a dataset of Kickstarter campaigns through 2019. We will be answering the question, “which factors increase the likelihood of success of a Kickstarter campaign?”.

Prior work?

There have been some analyses already done on kickstarter projects:

Daniel Kupka “Mining Kickstarter Data: Analysis and Prediction

https://medium.com/@daniel.kupka/we-analyzed-331-000-kickstarter-projects-here-s-what-we-learned-about-crowdfunding-success-63b341b025ac

Prior work?

Justin S Gage “ Kickstarter Analysis and Prediction”

https://www.kaggle.com/gagejustins/kickstarter-analysis-and-prediction

Data Source

We will be using the "Kickstarter Projects" dataset. This data is supplied by Mickaël Mouillé, who used a twitter bot to gather the information for the dataset. The dataset can be found at:

https://www.kaggle.com/kemical/kickstarter-projects.

Attribute Overview

- Id
- Name
- Category
- Currency
- Deadline (Date)
- Goal (Amount)
- Launched (Date)
- Pledged (Amount)
- State (Success/Failed/Cancelled)
- Backers
- Country
- USD pledged

Some Interesting Questions

- Do certain Kickstarter categories have a higher likelihood of success than others?
- Do backers in different countries have any categorical preferences?
- Is there a relationship between kickstarter amounts and success? At what value do these amounts begin increase the likelihood of failed kickstarter campaign?
- Does the time of year affect the outcome of a kickstarter campaign?

Dataset Integrity & Proposed Work

For cleaning and preprocessing, we will need to carefully review the dataset for missing data, however, after a cursory review, the data is relatively complete, and not a lot of cleaning is necessary

Dataset Integrity & Proposed Work

Using some of the metadata provided from the database creator, we will likely be able to determine if there are any attributes that are not necessary for our data analysis.

We will likely need to create some new attributes in order to accurately compare campaigns. We will also need to use binning for some of the monetary and date values.

Tools to be used

We will be using a combination of libraries available for Anaconda such as matplotlib, pandas, numpy, seaborn etc. Visualizations will also be generated using the aforementioned libraries with additional support from visualization software such as Tableau where necessary.

Evaluation Criteria

We can evaluate our results by determining whether there are any attributes from our dataset that show a positive correlation with the "state" attribute, which determines whether a campaign was successful or failed.