Increasing Crowdfunding Efficiency

Design Document

Group 8: Ayoola Ogunsola, Sameer Vinayak, Malvika Tripathi, Vishwak raam raghu



# Introduction

For our final project, we will be exploring the traits of Kickstarter crowdfunding campaigns in order to identify what makes a campaign successful. In this document, we will highlight the different aspects of our project, and give details related to the structure that we plan to use (different files, libraries, classes, etc.). At the conclusion of our project, we hope to have analyzed thousands of past Kickstarter campaigns in order to draw insights about successful - and unsuccessful - crowdfunding campaigns.

# Dataset

The data used for this project was obtained from Web Robots, which scrapes the campaign data from Kickstarter itself. (​https://webrobots.io/kickstarter-datasets/)​ .

For this project, we will be using the October 2016 datasets. Each dataset contains the following:

* name: project title (string)
* pledged: actual amount pledged (float)
* goal: the preset funding goal of the campaign (int)
* blurb: project description (string)
* state: whether a campaign was successful, unsuccessful, cancelled, or is currently live (string)
* country: abbreviated country code (string of length 2)
* currency: currency type of amt.pledged (string of length 3)
* staff\_pick: whether or not the campaign was featured on the website as a staff pick (boolean)
* location: city, state (or country) (string)
* spotlight: whether or not the campaign was a “spotlight” campaign on the website (boolean)
* percentage.funded: percent of the goal that was/has been raised (float)
* backers\_count: the number of backers for each campaign (int)
* average\_backing: the average amount pledged by each backer for a particular campaign
* deadline and launch\_date: the date the campaign started and was set to end. This can be used to calculate the length of the campaign as well.

Within each dataset, there are some key attributes that will be used in our analysis. For each campaign, the amount pledged attribute, when paired with the goal attribute, allows us to determine which campaigns were fully funded, and thus considered a success (and by how much). Additionally, the goal amounts can be used to determine whether lofty goals or more conservative ones tend to be met more often. Each campaign also has the possibility of being named a staff pick or spotlight pick, and understanding the benefit of achieving this recognition could help a campaign tremendously.

Each campaign entry also has a description (blurb). It will be quite useful to observe whether the length of the description has an impact on campaign success. Perhaps shorter descriptions, while less informative, keep a potential backer (donor) more engaged in the idea. ​ ​The dataset also includes the number of backers per campaign, which will allow us to calculate the average donation made for each campaign.

Along these lines, we are also provided data on each campaign’s start and end date. By knowing how long each campaign is scheduled to last (before the campaign actually begins), we can observe how length impacts each campaign. Perhaps shorter end dates drive more money to the campaign, since donors would not have to wait as long to receive their pledge reward. Lastly, we will explore how location impacts a campaign’s success as well. Perhaps associating a campaign with Silicon Valley boosts the funding a little bit.

# Design

## 1. OOP

The classes in our custom library will be designed as follows:

Superclass: Kickstarter Campaign

* Attributes shared by both datasets. Example: amt.pledged, by, currency, location, title, url
* Methods: Getters for data that we will analyze.

Subclass: Live Campaign

* Attributes unique to campaigns that are still active. Example: country, end.time, percentage.funded, state, type
* Methods: Additional getters for data that we will analyze.

Subclass: Most Backed Campaign

* Attributes unique to the campaigns that were the most successful. Example: category, goal, num.backers, num.backers.tier, pledge.tier
* Methods: Additional getters for data that we will analyze.

## 2. File I/O

We will begin by using the ​**CSV** module in Python to read in our two data files. Then, we expect​ to load the data into a ​**Pandas** dataframe in order to explore the data in a more organized​ manner.

## 3. Data Structure

Next, using ​**matplotlib**​, we can visualize some of the key aspects of the data. One potential visualization that comes to mind is grouping the pledge goals into different tiers and making a bar graph or histogram based on how many successful campaigns exist for each tier. This will help to identify any trends, and the same can be done for other categories, like staff picks and campaign length.

As a finished product, we hope that our program will be able to answer some of the following questions, using visualizations to aid us:

* Do longer campaigns tend to fare poorly or are they more successful?
* What is the impact of a campaign being chosen as a staff pick or spotlight pick?
* Should campaigns aim for higher, more lofty goals or more conservative amounts?
* Does location play a factor in a campaign’s success?
* How does the campaign description length impact the success of a campaign?