

Midterm Project – Data Analysis on Kickstarter

Tingrui Huang

Introduction

In this project, I'm going to use the "Kickstarter Project" dataset to analyze the success of a project (state) and the amount of raised funds (USD_pledged_real) based on its category, goal of funding, number of backers, start date, length(time span), country of the company. Furthermore, if possible, I would make predictions for projects based on the above variables.

Dataset and Source

The dataset is sourced from Kaggle. The dataset contains 379,000 observations with 15 variables. The dataset was updated in Jan 2018 and it is the latest version. The 15 variables include: ID, name, category, main_category, currency, deadline, goal, launched, pledged, state, backers, country, USD_pledged, USD_pledged_real, USD_goal_real.

Analysis Procedure and Methods

The analysis procedure will contain four parts: Background research, EDA, modeling, discussion and conclusion. In terms of model selection, currently I would prefer linear and logistic multilevel models, since there could be random effects on the category and country. But if you have any recommendations on the model selection, please let me know! (I will do more research as well to see if there are other proper models I could use.)

How will this project be related to my career goal

The data analysis part of this project may seem to be straight forward, but the fun part is that I could learn a lot of things by doing this project. For example, I would have chance to explore the most popular projects, topics and people in almost every industry. My career goal is to be a management consultant at companies like the Big Four and MBB. Besides solid analytical and communicational skills, another critical skill a consultant needs to have is to be knowledgeable. By doing this project, I would have the chance to know more about what is happening and what will be happening in many industries around the world. I could find some thing I have never heard or seen before. By analyzing the data, I could figure out what market would be more attractive to people and what market is under huge evolution. It will be super cool!

Possible Improvement

If possible, I would go beyond the dataset and look for more available source on the internet to support my findings. For example, if I found some uncommon outcomes, such as a startup in a niche market raised tons of money, I would probably look for news, google search rate and other relevant information to figure out the reasonable explanation.