# Problem #1: Kickstarter

Kickstarter is the world's largest funding platform for creative projects. Since the site launch on April 28, 2009, 13 million people have backed a project, $3.3 billion has been pledged, and 369,069 projects have been posted. Meanwhile, only 131,814 projects (36%) have successfully met their target goal. There’s a lot that could be learned by reviewing historical data surrounding Kickstarter launches!

Which machine learning model is most effective for classifying whether a project will meet its goal? Across the different models, which features have the most importance in the classification of the final status? How could someone use this information to launch a Kickstarter campaign that has a high likelihood of succeeding? Using natural language processing, what key words can be used to describe a project that results in high success rates? Equally important, what key words result in high failure rates and therefore a person should avoid using in their Kickstarter launches? For the Kickstarter projects that reach their goal, what machine learning model is most effective for predicting the length of time (in days) that the project will take to meet its goal. How does launching a project on a weekday compare to launching a project on a weekend in regards to the length of time to meet a goal. How does concluding a project on a weekday compare to concluding a project on a weekend in regards to the length of time needed to meet a goal?

**Pitch:** 63% of Kickstarter projects fail to meet the target goal. These projects aren’t like horseshoes and hand-grenades; there is no reward for a launch that gets “close” to a target fundraiser goal amount. This means that new launches need to have a carefully selected goal along with the right mix of variables to make launching successful. What is that right mix?

**Pitch (part b) …** You’re in need of an investment or else your prototype won’t make it to market. You need to secure an additional investment as soon as possible. What can you do to make your project quickly reach its’ goal!

There are a few data sets that could be used to assess the problem statements above. One is a [pull of 108,000 observations of Kickstarter launches](https://www.kaggle.com/codename007/funding-successful-projects). The data set includes the project description, goal, launch date and end date for each observation. The data is very clean with few missing values and a response variable for each record. The description and key words variables provide some interesting possibilities for natural language processing.

Another avenue for retrieving [Kickstarter data sets comes from webrobots.com](https://webrobots.io/kickstarter-datasets/), which is managed by a team dedicated to scraping the web for useful information. The data sets is available in either JSON or CSV for every month since April 2014. The data sets provide much more detail including location, links to photos, whether it was a staffed picked project, and more. It would be preferable to use this data set as it provides a lot more information. One problem is that the .csv data set is split across 10+ files while the JSON data is just on single file. The files would either need to be merged prior to building a model or learning how to import a JSON file.

# Problem #2 – Speed Dating

Speed dating is a formalized matchmaking process of dating system whose purpose is to encourage people to meet many new people. The ideas is simple, individuals rotate through a number of quick dates with hopes of developing a quick bond and spawning a lasting relationship. For some, finding the right partner is not an easy task. Many might be wondering how they can stand out in a speed dating setting. There are so many variables in a typical speed dating setting, some that are in your control and some that are not.

What machine learning model provides the most accurate classifications of whether two participants will want to meet again? What’s the most important variable in predicting the likelihood that a couple will want to meet again? What position in the speed dating rotation provides the greatest opportunity for a second date? What’s more important, appearance or intelligence when it comes to partner compatibility? What are women looking for in men and what are men looking for in women?

**Pitch**: You plan on attending an upcoming speed dating event, but you haven’t found success during your previous visits. What can you do to give yourself the best chance at finding love and going on a second date?

On Kaggle.com is a [data set posted capturing speed dating events from 2002-2004](https://www.kaggle.com/annavictoria/speed-dating-experiment) from experimenters at Columbia University. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end of their four minutes, participants were asked if they would like to see their date again. The data set provides 8,300+ observations and over 150 features. However, there are some challenges with the data set. One major challenge is that the data comes from surveys spanning three years. Over that period, the questions went through slight variations and the speed dating events had “special” events that might have skewed results for that event. Another challenge is dealing with the many missing values. Working with the nuances in the data set will be challenging and require careful design upfront.