**Kickstarter Campaign Analysis**

Submitted by: Adi Agarwal

**Initial analysis**

Success ratio is defined as the number of successful campaigns over total number of campaigns, sum of successful and failed. It dropped by 38% in 2014, which is correlated with the increase in total campaigns in the same year as more countries from EMEA and ANZ regions started to participate in crowdfunding. More projects failed in initial years for countries in these regions.

There are few features correlated to campaign success. More significant features are **‘duration’** of the campaign and **‘goal’** (or money requested for the campaign). Success is when goal is achieved within predefined duration of the campaign. The ‘duration’ spread is smaller for successful campaigns when compared with failed campaigns. This is true for mean and median values as well, across all demographic regions and **‘categories’** (product category feature). Similar results for **‘goal’** when compared between successful and failed campaigns.

**Statistical analysis**

Median **‘goal’** amount is 48% lower for successful campaigns when compared to failed campaigns. There are some major outliers where campaigns have goal for as high as $100 MM, hence chose median instead of mean. ‘Duration’ of campaigns is not skewed similar to ‘goal’, because campaigns normally have to select deadline date of 1 to 3 months from the launched date. Mean ‘duration’ is lower by 8.5% for successful campaigns.

There was no significant distribution type for both of these features. Visual inspection via histogram shows close to exponential distribution for successful campaigns. However, it is more appropriate to assume there is no identifiable distribution and use non-parametric tests. Spearman test proves **statistical significant correlation** of downward sloping 0.095 between campaign success and duration of campaign. And, downward sloping 0.22 when checked against goal amount.

I split the data into successful and failed campaigns, to compare if there is a difference in duration and goal distribution. Since data is skewed as mentioned earlier, I used Mann Whitney as a non-parametric test to compare median values.

**Hypothesis testing**

**Hₒ**: The median goal amount is same for successful and failed campaigns.

**Hₐ**: The difference in median goal amount is statistically significant between successful and failed campaigns.

P-value is zero and we reject the null hypothesis. Thereby concluding there is a **statistically significant difference between median goal amount for successful and failed campaigns.** P-value is zero when the same test is run for duration of the campaign.

**Interesting finding**

‘Pledged’ amount is the money campaign actually raised. This is one of the features in data along with ‘goal’, which is the request amount of the campaign. If pledged is more than goal within given deadline, campaign is a success. I looked at the pledged as a percent of goal. For all failed campaigns this has to be less than 100%. Let’s call this new feature ‘pledge\_per\_goal’ or PPG. PPG value is exponentially distributed. Only 3% of the failed campaigns have raised more than half of the goal amount. Does this mean, campaigns overestimate their abilities to raise funds? Also, what is the probability of success p, if a campaign is able to raise x% of goal amount? I will further explore the data for milestone submission.