Rachel Downs, Muhammad Ghauri
MIS 373
Maytal Saar-Tsechansky

Predicting Kickstarter Campaign Success

Project Objectives:

Kickstarter is an online crowdfunding platform aimed at helping people get their ideas funded while building a community of fans to support their ideas. While Kickstarter publishes many advice and best-practices articles on their blog, over half of campaigns still fail.

Why does this matter?

Unlike their competitor, Indiegogo, Kickstarter campaign projects follow an "all or nothing" funding model. This means that if a Kickstarter campaign fails, both the project creators are disappointed, as well as the people who did contribute because the project will not be completed in any capacity.

This project's objectives are to:

1. Understand the marketplace of Kickstarter including timing of campaigns posted, types of projects, location of campaigns, description of campaigns and more
2. Provide insight into attributes that set campaigns up for a higher rate of success to inform campaign creation in the future
3. Build a predictive model that allows Kickstarter to identify high-failure-risk campaigns before they fail and provide supplemental advice and material to the creators
4. Identify a business opportunity for Kickstarter by finding campaigns just below the predicted threshold of success and helping them get to their goal, helping both parties earn revenue

Why does data mining makes this problem approachable?

With over 4,700+ projects live right now, it would be impossible for Kickstarter employees to monitor each campaign individually to support their success. Applying data mining to this problem allows Kickstarter to scale their support and outreach to their creator community with minimal, targeted work.

Methodology:

This dataset contains data on 20,632 Kickstarter campaigns on the site as of February 1st 2017. Important attributes are described below:

- Project: a finite work with a clear goal that you'd like to bring to life (aka campaign)
- Funding goal: amount of money that a creator needs to complete their project
- Name: name of project on Kickstarter
- Blurb: the short description displayed under the name of your project and on the browse page
- Pledged and backers: amount of money that a project has raised and people that have supported it at the point of the API pull
- State: successful, failed, cancelled, live or suspended

- SuccessfulBool: whether the project was successful or not
- Deadline, state changed, created at, launched at: deadline given for successful funding, state changed when campaign went to success or failure, time the project was created at, time the project was launched at
- Other attributes in this dataset: country, currency, category

To attain a deeper understanding of our data and to have more attributes to explore, we also created the following features out of the data for our analysis:

- Name and blurb (description) length including and excluding “stop words” - name_len_clean, blurb_len_clean
- Day of week and hour of the day for creation, launch and deadline date - deadline_weekday, created_at_weekday, lauched_at_weekday, deadline_hour, created_at_hour, launched_at_hour
- Days between creation and launch, and days between launch and deadline - create_to_launch, launch_to_deadline

Findings:

Understanding the marketplace:

Through our exploratory analysis, we developed some key understandings of Kickstarter’s marketplace that can inform marketing efforts towards campaign creators.

Kickstarter is predominantly domestic.

Although the crowdfunding platform boasts its global reach, 68.5% of Kickstarter’s campaign creators are United States-based with the next most popular countries being Great Britain, Canada and Australia.

Campaign timing is more systematic than you would think.

The timeline of a Kickstarter campaign includes creation of the campaign where creators set up the funding page and marketing material, launching the campaign where creators publish the campaign to the world for others to start backing, and the campaign deadline where funding stops. Creation and launching is more popular in the beginning of the week (Appendix A). In fact, 61% of campaigns are launched Monday-Wednesday. In contrast, campaign deadlines are more commonly placed at the end of the week with over half of deadlines falling Wednesday-Friday, and 18% of deadlines on Fridays alone. The timing of campaign creation, launch time, and deadline all coincide with the average work-hours, suggesting that people create Kickstarter campaigns during the workday (Appendix B). One notable exception here is, as you can see, there is a spike in campaign deadlines around 11pm.

Kickstarter campaigns have a strong technology and entertainment focus.

If you’re familiar with Kickstarter, you’re not likely not surprised that the top four categories by count on Kickstarter are Web, Hardware, Software and Gadgets. However, among the top 10 categories are also a few entertainment-oriented project like Plays and Festivals. We found that entertainment-focused projects did make up 16.4% of all projects with a category listed.

Providing insight:
After a general understanding of the Kickstarter "campaign creators", we analyzed specific aspects of a campaign to determine what attributes lead to higher successful funding rates.

Campaigns with longer descriptions and names are more likely to be successful.
After removing stop words, successful campaign names are on average approximately 1 word longer than failed campaign names (p-value <.01) (Appendix D). In fact, campaigns with names less than or equal to 3 words are 2.1 times more likely to fail than those longer than 3 words. Successful campaign descriptions are also significantly longer than failed campaign descriptions (p-value <.01) (Appendix E).

Successful campaigns invest more time in creating the campaign.
The median number of days spent between creation and launch for successful campaigns is 19, as compared to the median of 12 days spent for failed campaigns (Appendix F). Furthermore, taking longer than 1 week to create your campaign makes your campaign 1.83 times more likely to succeed.

Setting the right goal is as important as you'd think.
The median goal for successful campaigns from this data is $5,750 while the median goal for failed campaigns is nearly $17,000. We also found that 28% of failed campaigns had a goal of over $50,000. When comparing successful and failed campaigns, we found that the goal matters - especially when the goal is very high. In fact, campaigns with goals under $50k are 2.4 times more likely to successfully get funded.

Tuesdays and Sundays are the best days to launch a campaign.
Only about 15% of campaigns are launched on a Tuesday or Sunday. However, our data shows that campaigns launched on one of these days are 1.22 times more likely to be successfully funded (Appendix G).

Staff picks have significant impact on campaign success.
Kickstarter's "staff picks" are given high-value front page real estate as "*Projects We Love*". Since these projects are given such high visibility, it's no surprise that staff pick projects are 9.6 times more likely to be successful than those that aren't. While it's understandable that not all projects can be staff picks, we will touch on how Kickstarter can leverage the power of staff picks to improve its platform's success rate in the implications section.

Building a predictive model:
After the creation of additional features mentioned in the Methodology section, the dataset had 60 columns. To create our predictive model, first we had to prune certain attributes out of the 60. The two categories of attributes that were removed from the dataset used when generating a predictive model were:

1) Attributes that WEKA (Waikato Environment for Knowledge Analysis) was unable to import

2) Attributes that had a significantly high correlation with campaign success, characteristic of attributes that require a campaign to *already* have been successful

Leaving the second category of attributes in the dataset resulted in nearly 100% classification accuracy on both the training and test data regardless of which of five modeling techniques were used. To determine which attributes fell into this category, we used both a correlation matrix between all attributes in the data and SuccessfulBool as well as WEKA's InfoGain attribute evaluator in conjunction with its ranker. Intuitively, we wanted to make sure we were building our predictive model using attributes Kickstarter would be able to act on proactively at the launch of the campaign, rather than retroactively at the end. A few of the high correlation attributes were obvious; these included:

- State - This was just a classifier attribute that said whether a campaign had already succeeded, failed, or was cancelled
- Spotlight - A campaign can only be spotlighted after it has already succeeded
- Backers Count - A successful campaign has several times more backers as failed or cancelled campaigns
- USD Pledged - A successful campaign has several times more pledged funds as failed or cancelled campaigns
- Pledged - This attribute was almost identical to USD Pledged and thus had a similar effect on classification accuracy

Some of the less obvious contributors to SuccessfulBool that we excluded were:

- Staff Pick - This is an indicator of whether a campaign was designated by Kickstarter team members as a "favorite" while it was active
- Launch to State Change Days - How long a campaign was active between when it was launched and when it either failed, succeeded, or was cancelled

After unrealistically accurate attributes and attributes that would be unknown at the start of a campaign were removed from the dataset, the remaining data was split into a training and test set of 67% and 33% of the dataset, respectively. Five modeling techniques were used and compared to determine which yielded the highest test accuracy. These included:

- Bagging (30 base models using the J48 classifier)
- Naive Bayes
- K Nearest Neighbors (5 neighbors)
- Random Forest (6 attributes, 60 models)
- J48 Classification Trees

Model Name	*Classification Accuracy*	*Precision (Weighted)*	*Recall (Weighted)*
Bagging	67.8249 %	0.662	0.678
Naive Bayes	69.715 %	0.773	0.697
K Nearest Neighbors	65.804 %	0.619	0.658
Random Forest	65.8186 %	0.598	0.658
J48	71.9541 %	0.708	0.720

J48 classification trees had yielded the most accurate model when used on the provided test set. However, upon further inspection, it was found that the tree produced had nearly 800 leaves and the size was almost 1100. To ensure that the tree was not overfitting, the Confidence Factor in the J48 model was made extremely small (.005) so as to ensure a well pruned tree. The result was a tree that had a length of 39 and 31 leaf nodes.

The final tree model had a classification accuracy of 71.7214 %, precision of 0.706, and recall of 0.717 making it only slightly less accurate than the larger tree and still more accurate than the other four modeling techniques (Appendix H).

Business Proposition:

Using this classification tree model as the classifier, we output the predictions WEKA generated when given our final training and test sets. There were around 7000 predictions made that were cleaned, imported into Excel, and distinguished under the columns Example Number, Actual Class, Predicted Class, and Probability. We replaced the default probability statistic (say, x) with 1 - x if the predicted class was 0 so as to be able to rank predictions by the highest likelihood of being a successful campaign.

WEKA would convert the prediction from a 1 (successful campaign prediction) to a 0 (failed campaign prediction) after a threshold of 0.516. Though of the 2338 successful campaigns in the dataset, almost 800 were correctly identified in the first 1200 predictions, there was a total of almost 1600 successful campaigns outside of the predicted success region. In the next 300 predictions alone there were 166 actually successful campaigns left.

If we take the median goal value of a successful campaign ($5750) and multiply it by the 134 failed campaigns Kickstarter in that high potential segment and multiply that by the 10% commission Kickstarter earns for each successful campaign, the firm could have earned up to $77,050 just by extending help to the projects at the cusp of the current prediction threshold.

Thus, we recommend that Kickstarter recoup some of this lost revenue simply by extending its "Staff Pick" tag to the campaigns slightly below the current prediction threshold. As shown by Providing Insight portion of the analysis, staff picks are nearly 10 times as likely to succeed as campaigns without the flair. And since this tag costs nothing to Kickstarter and there is no monetary penalty for applying this tag to a failed campaign, Kickstarter only has upside by implementing this idea.

Of course this is with the consideration that the tag should be extended in moderation so as to avoid having it lose the value and recognition that comes with being a staff selected campaign. Kickstarter should not just make every campaign a staff pick, but adding a few more campaigns to the *Projects We Love* section that meet the signs of a good campaign but are just barely predicted to fail would earn Kickstarter funds at virtually no cost.

<u>Implications:</u>

At the end of the day, Kickstarter's goal should be to have as many successfully-funded campaigns as possible. This is especially important because if a campaign doesn't reach its funding goal, it receives none of the money and the project is very likely to cease to exist. Another key component of Kickstarter campaigns that makes early prediction that much more useful is that because of the "crowdfunding" nature of Kickstarter, popularity cues and virality early in the campaign are essential to acquire the "social proof" needed to keep a steady flow of backers as the campaign progresses.

Through this analysis, Kickstarter can now do the following:

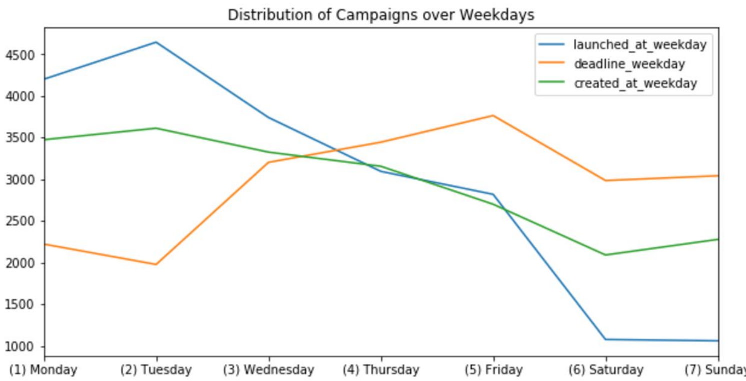
- Understand their target market and usage trends to make general marketing more effective - i.e. concentrating marketing dollars around the work schedule
- Create generalized "tips and tricks" as marketing assets based on insights around what campaign attributes tend to lead to more successful campaigns
- Be proactive about marketing to high-failure-risk campaigns including featuring high potential campaigns, identified through our predictive model, as "staff picks" to increase success rate

The most powerful implication of this research is how Kickstarter can increase its leverage of their "staff picks" to increase their platform's success rate. As detailed above, Kickstarter can leverage this "lead scoring" to target campaigns that are on the precipice between success and failure and earn potentially $77,050 from just this timer period analyzed by extending to only 150 more campaigns. The potential revenue is even greater if Kickstarter chose to extend Staff Picks to even more than those 150.

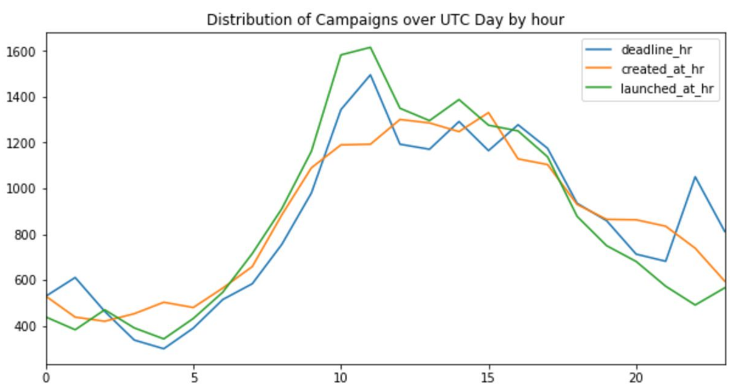
Our data shows that failed Kickstarter campaigns resulted in $31,707,603 of opportunity cost *just from the time period we analyzed.* With this research, Kickstarter can better leverage their platform's tools and produce targeting marketing to increase the success rates and resulting funding from the world's most powerful crowdfunding tool.

Appendix:

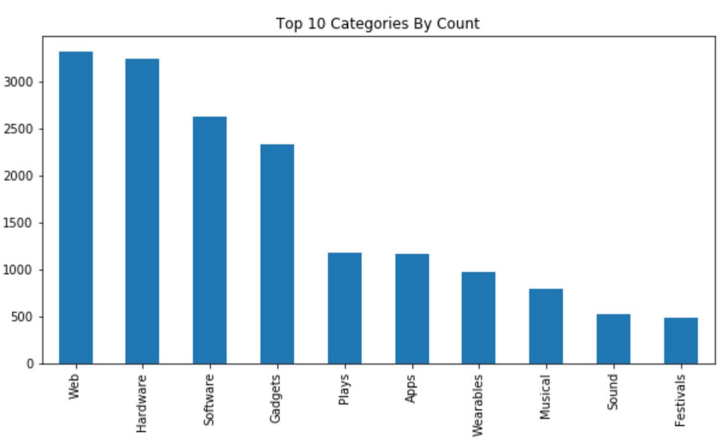
Appendix A:



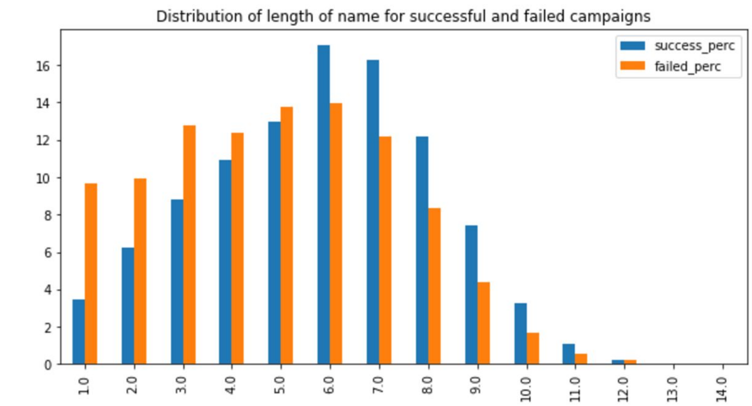
Appendix B:



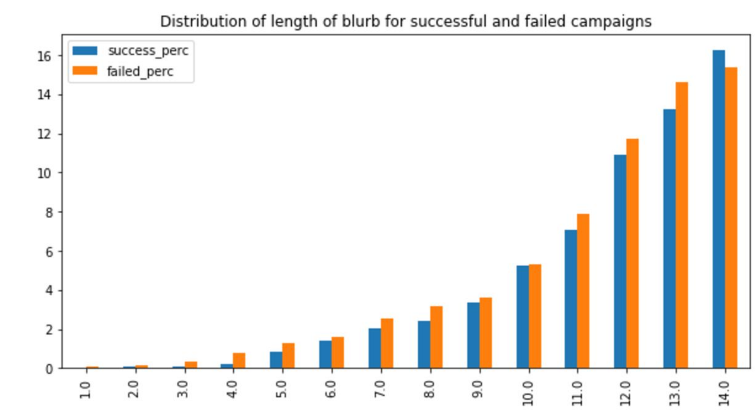
Appendix C:



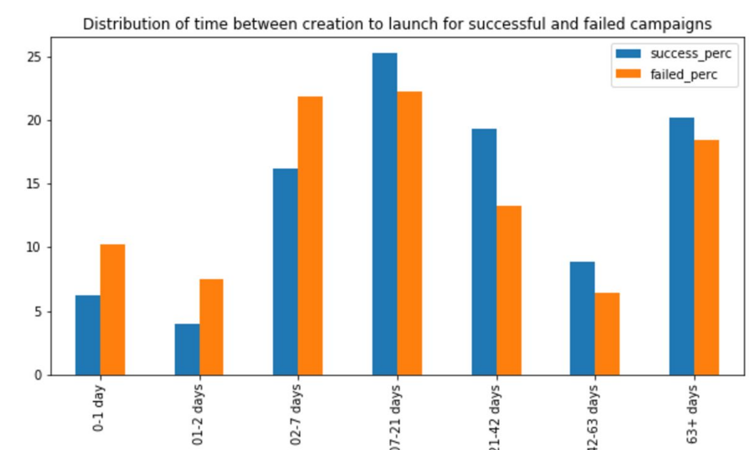
Appendix D:



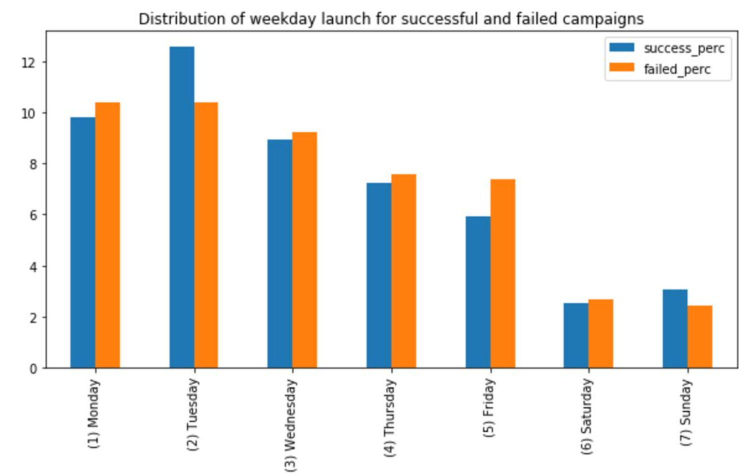
Appendix E:



Appendix F:



Appendix G:



Appendix H:

```
disable_communication = FALSE
|   goal <= 3333.33
|   |   create_to_launch_days <= 2: 0 (668.0/208.0)
|   |   create_to_launch_days > 2
|   |   |   category = Academic: 1 (0.0)
|   |   |   category = Places: 1 (0.0)
|   |   |   category = Blues: 1 (0.0)
|   |   |   category = Restaurants: 1 (0.0)
|   |   |   category = Webseries: 1 (0.0)
|   |   |   category = Thrillers: 1 (0.0)
|   |   |   category = Shorts: 1 (0.0)
|   |   |   category = Web: 0 (59.01/7.68)
|   |   |   category = Apps
|   |   |   |   deadline_yr <= 2016
|   |   |   |   |   static_usd_rate <= 1.64098: 1 (138.08/42.74)
|   |   |   |   |   static_usd_rate > 1.64098: 0 (11.94/1.31)
|   |   |   |   deadline_yr > 2016: 0 (11.94/1.42)
|   |   |   category = Gadgets
|   |   |   |   blurb_len_clean <= 6: 0 (8.57/0.34)
|   |   |   |   blurb_len_clean > 6: 1 (167.2/76.68)
|   |   |   category = Hardware: 1 (377.89/133.74)
|   |   |   category = Festivals: 1 (0.0)
|   |   |   category = Plays: 1 (0.0)
|   |   |   category = Musical: 1 (0.0)
|   |   |   category = Flight
|   |   |   |   launched_at_yr <= 2013: 1 (3.66)
|   |   |   |   launched_at_yr > 2013: 0 (42.79/15.38)
|   |   |   category = Spaces: 1 (0.0)
|   |   |   category = Immersive: 1 (0.0)
|   |   |   category = Experimental: 1 (0.0)
|   |   |   category = Comedy: 1 (0.0)
|   |   |   category = Wearables: 1 (75.33/35.53)
|   |   |   category = Sound: 1 (64.03/28.7)
|   |   |   category = Software: 0 (429.37/162.85)
|   |   |   category = Robots: 1 (125.55/44.22)
|   |   |   category = Makerspaces: 1 (37.66/14.76)
|   goal > 3333.33: 0 (11315.0/2670.0)
disable_communication = TRUE: 0 (218.0)
```