

KICKSTARTER PROJECT 2009-I Q2017 ANALYSIS

R. MIH

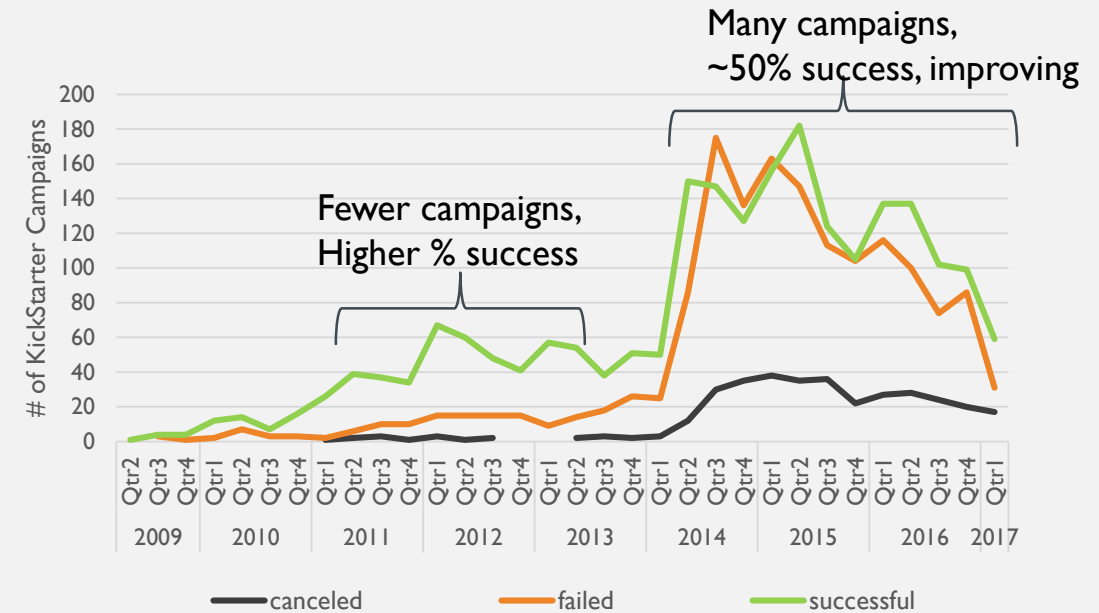
UCB DATA SCIENCE WINTER 2018

BACKGROUND

OVER TWO BILLION DOLLARS HAVE BEEN RAISED USING THE MASSIVELY SUCCESSFUL CROWDFUNDING SERVICE, KICKSTARTER, BUT NOT EVERY PROJECT HAS FOUND SUCCESS. OF THE OVER 300,000 PROJECTS LAUNCHED ON KICKSTARTER, ONLY A THIRD HAVE MADE IT THROUGH THE FUNDING PROCESS WITH A POSITIVE OUTCOME.

SINCE GETTING FUNDED ON KICKSTARTER REQUIRES MEETING OR EXCEEDING THE PROJECT'S INITIAL GOAL, MANY ORGANIZATIONS SPEND MONTHS LOOKING THROUGH PAST PROJECTS IN AN ATTEMPT TO DISCOVER SOME TRICK TO FINDING SUCCESS. FOR THIS WEEK'S HOMEWORK, YOU WILL ORGANIZE AND ANALYZE A DATABASE OF FOUR THOUSAND PAST PROJECTS IN ORDER TO UNCOVER ANY HIDDEN TRENDS.

Evolution in # of Kickstarter Campaigns

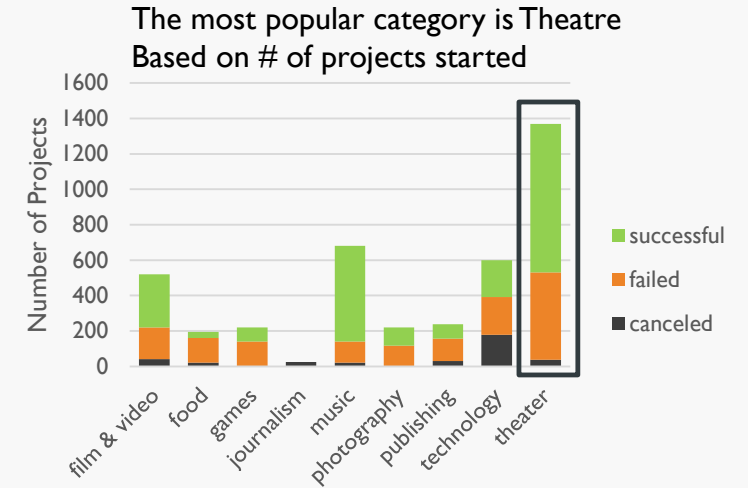
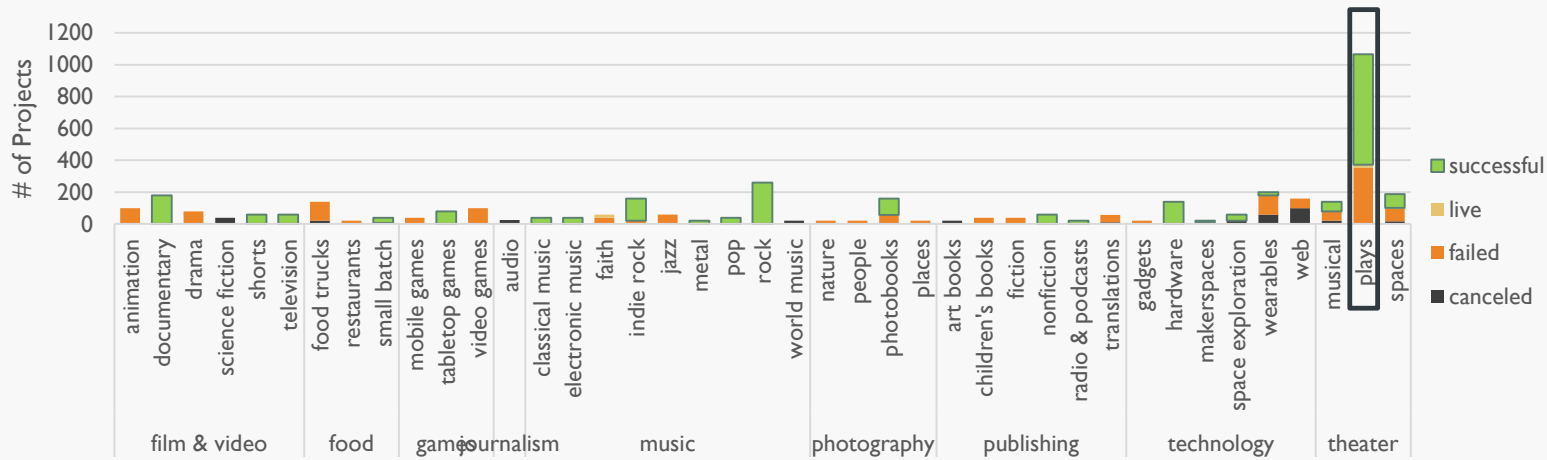


CONTENTS

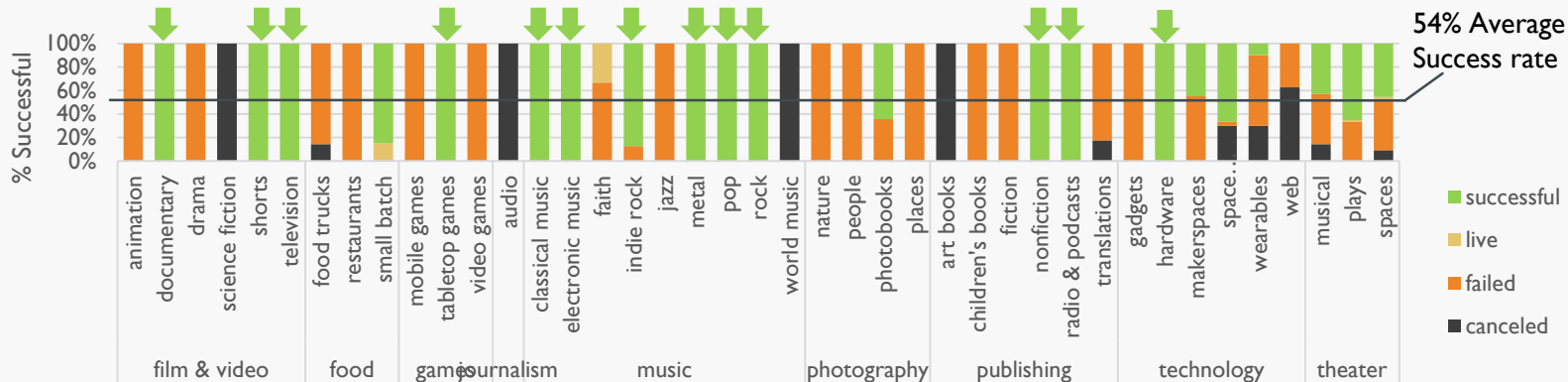
- Conclusions
- Limitations of the dataset
- Next Steps in Analysis
- Bonus Analyses

KICKSTARTER PROJECT 2009-IQ2017 ANALYSIS CONCLUSIONS

- The most popular KickStarter Category for campaigns is: Theatre (33.7%, 1369 submissions)
 - Theatre has over 2X the number of submissions compared to the next most popular category (Music, 16.7%, 680)
- The most popular Kickstarter Sub-Category for campaigns is: Plays (1066 submissions)



- There are a number of “100%” successful sub-categories of KickStarter Campaigns, beyond the “Plays” category
 - However there are few submissions in these categories – Maybe an opportunity?



Recommendation / Next steps

- Reach out to kickstarter communities and encourage more submissions from “100%” successful sub-categories
- Execute Additional Analyses Recommended to better understand why some sub-categories are more successful than others
- Create BKM's to distribute to Kickstarter communities

SOME LIMITATIONS OF THE DATASET

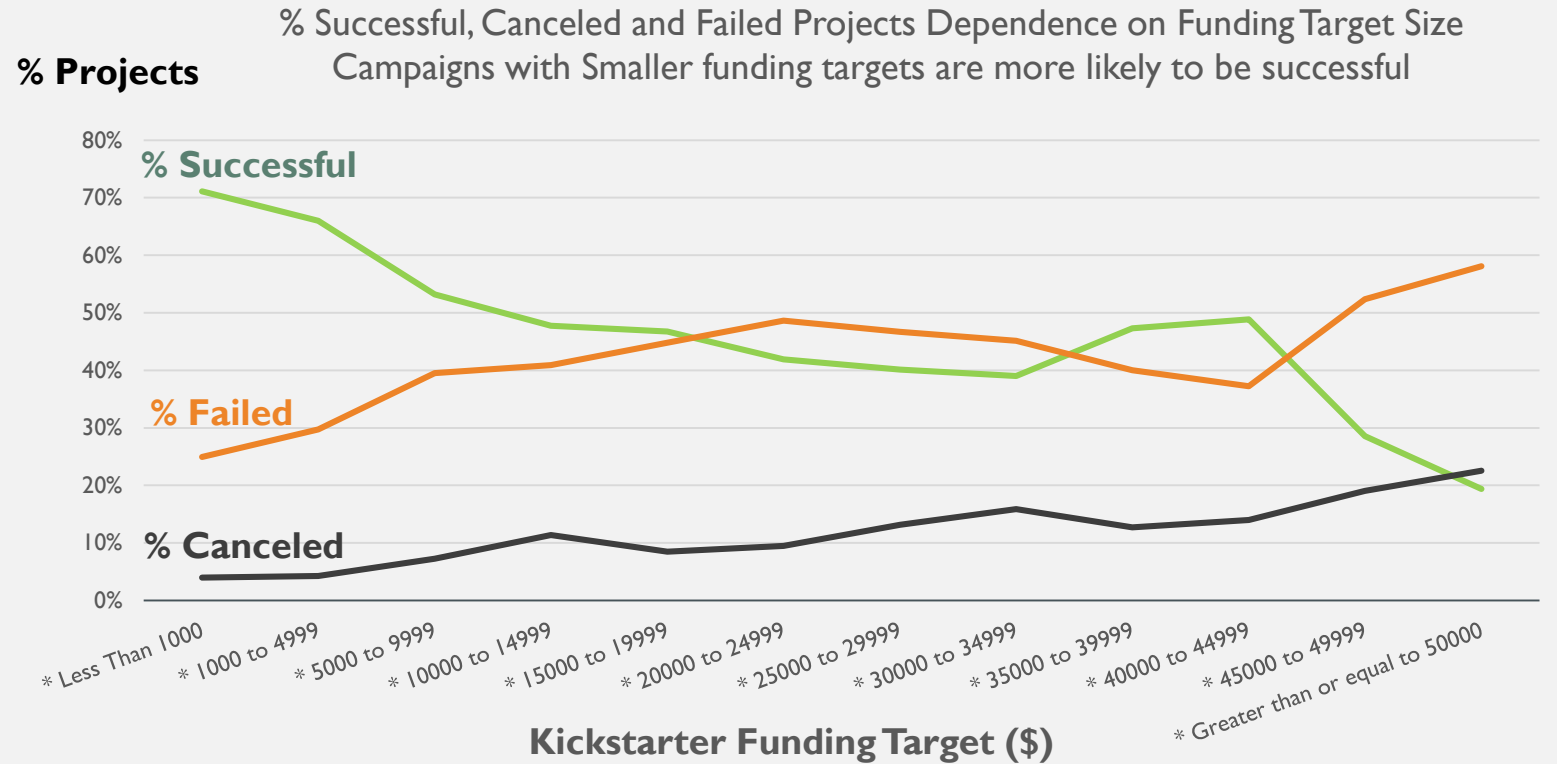
- What are some of the limitations of this dataset?
 - The biggest limitation of this dataset is that we were given ~4000 launched projects in the data to analyse, but it says in the background that there have been 300,000 kickstarter projects launched
 - So that's like ~1.2% of the data. I don't really know how good of a proxy this data is, to the actual data out there.
- The data misses some areas which could help answer the question of ways to improve kickstarter success rates. For example, is there any of this data available:
 - More data on the incentives / rewards for backers (i.e. would likely be the product, or some other award) which may influence the backer
 - Number of types of rewards
 - Expected sales price of the product / reward
 - Price proposed for backers for product / reward
 - More information on the backers – age / gender / regional demographics (for US)
 - More on how the Kickstarter Campaigns were carried out
 - How many emails were sent
 - How many media posts (social and conventional)
 - # of Endorsements
 - Type of Endorsements (famous people, institutions, etc)

ADDITIONAL ANALYSES RECOMMENDED

- What are some other possible tables/graphs that we could create?
 - Fund Targets by Category / Sub-Category (Successful vs others)
 - Number of backers by category and sub-category
 - Average donation by category and sub-category
 - Re-aggregate the data by region (Asia, North America, Europe, etc) and re-do the analysis to see if more trends show up

BONUS ANALYSIS

- * Create a new sheet with 8 columns: `Goal`, `Number Successful`, `Number Failed`, `Number Canceled`, `Total Projects`, `Percentage Successful`, `Percentage Failed`, and `Percentage Canceled`
- * In the `goal` column, create twelve rows with the following headers...
- * Less Than 1000
- * 1000 to 4999
- * 5000 to 9999
- * 10000 to 14999
- * 15000 to 19999
- * 20000 to 24999
- * 25000 to 29999
- * 30000 to 34999
- * 35000 to 39999
- * 40000 to 44999
- * 45000 to 49999
- * Greater than or equal to 50000
- * Using the `COUNTIFS()` formula, count how many successful, failed, and canceled projects were created with goals within those ranges listed above. Populate the `Number Successful`, `Number Failed`, and `Number Canceled` columns with this data.
- * Add up each of the values in the `Number Successful`, `Number Failed`, and `Number Canceled` columns to populate the `Total Projects` column. Then, using a mathematic formulae, find the percentage of projects which were successful, failed, or were canceled per goal range.
- * Create a line chart which graphs the relationship between a goal's amount and its chances at success, failure, or cancellation





BACKUP MATERIALS

BACKGROUND AND INSTRUCTIONS



INSTRUCTIONS

- Using the Excel table provided, you will be modifying and analyzing the data of four thousand past Kickstarter projects as you attempt to uncover some of the market trends.
- * Use conditional formatting to fill each cell in the `state` column with a different color, depending on whether the associated campaign was "successful," "failed," "cancelled," or is currently "live"
- * Create a new column at column O called `percent funded` that uses a formula to uncover how much money a campaign made towards reaching its initial goal.
- * Use conditional formatting to fill each cell in the `percent funded` column using a three-color scale. The scale should start at 0 and be a dark shade of red, transitioning to green at 100, and then moving towards blue at 200.
- * Create a new column at column P called `average donation` that uses a formula to uncover how much each backer for the project paid on average.
- * Create two new columns, one called `category` at Q and another called `sub-category` at R, which use formulas to split the `Category and Sub-Category` column into two parts.
- * Create a new sheet with a pivot table that will analyze your initial worksheet to count how many campaigns were "successful," "failed," "cancelled," or are currently "live" per **category**.
- * Create a stacked column pivot chart that can be filtered by `country` based on the table you have created.
- * Create a new sheet with a pivot table that will analyze your initial sheet to count how many campaigns were "successful," "failed," "cancelled," or are currently "live" per **sub-category**.
- * Create a stacked column pivot chart that can be filtered by `country` and `parent-category` based on the table you have created.
- * The dates stored within the `deadline` and `launched_at` columns are using unix timestamps. Fortunately for us, [there is a formula](http://spreadsheetpage.com/index.php/tip/converting_unix_timestamps/) out there that can be used to convert these timestamps into a normal date.
- * Create a new column named `Date Created Conversion` that will use [this formula](http://spreadsheetpage.com/index.php/tip/converting_unix_timestamps/) to convert the data contained within `launched_at` into Excel's Date format
- * Create a new column named `Date Ended Conversion` that will use [this formula](http://spreadsheetpage.com/index.php/tip/converting_unix_timestamps/) to convert the data contained within `deadline` into Excel's Date format
- * Create a new sheet with a pivot table with a column of `state`, rows of `Date Created Conversion`, values based on the count of `state`, and filters based on `parent category` and `Years`.
- * Now create a pivot chart line graph that visualizes this new table.
- * Create a report in Microsoft Word (Powerpoint OK) and answer the following questions...
- 1. What are three conclusions we can make about Kickstarter campaigns given the provided data?
- 2. What are some of the limitations of this dataset?
- 3. What are some other possible tables/graphs that we could create?