Outcomes Based on Launch Date

Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

1. There seems to be a seasonality. Most successful projects were launched during the second quarter of the year while most of the projects that failed or were cancelled were launched during the third quarter of the year.
2. Some sectors were more successful than others. Music had many successful projects but others such as the wearables sub-category of technology didn’t do had most of its projects either cancelled or failed.
3. Geographic location (country) doesn’t seem to be a determinant factor for the success of a Kickstarter campaign.

What are some limitations of this dataset?

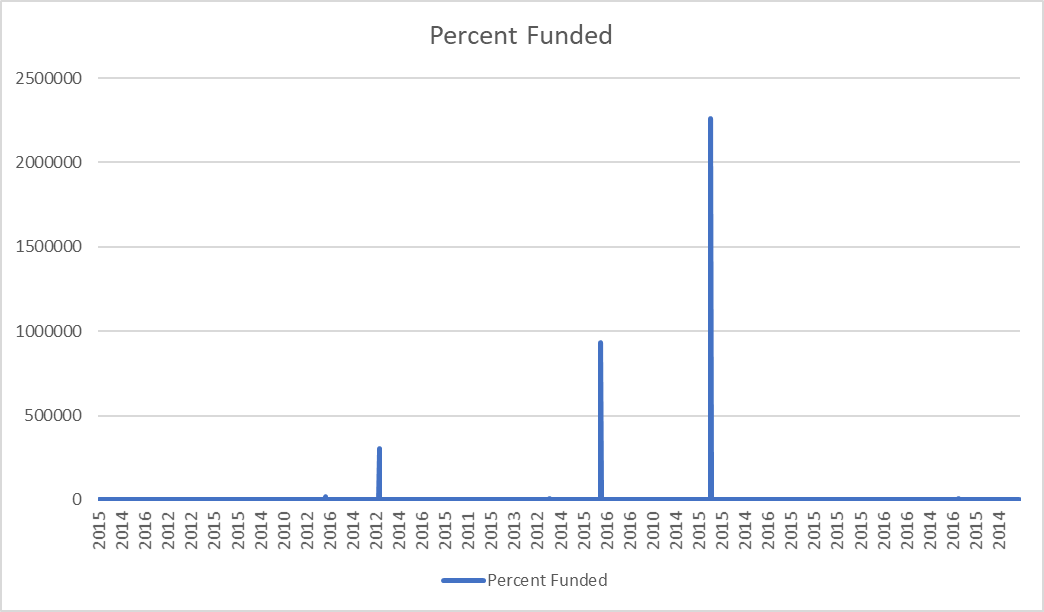
1. It is not clear what is the definition of successful. We can see cancelled projects with a percent funded above 100%. They reached their goal, but they are not successful because they were cancelled. Definition of successful may include a project that was pledged less money than the goal it was seeking.
2. The data has outliers (see next section) may skew results. In particular, when goals were close to 0 and pledged were high.
3. It has no information on returns on investment so we cannot compare how successful a project was compared with another.
4. This data is historic and we don’t know if structural changes happened during this time. If so, it would need to be addressed before calculating any regressions.

What are some other possible tables and/or graphs that we could create?

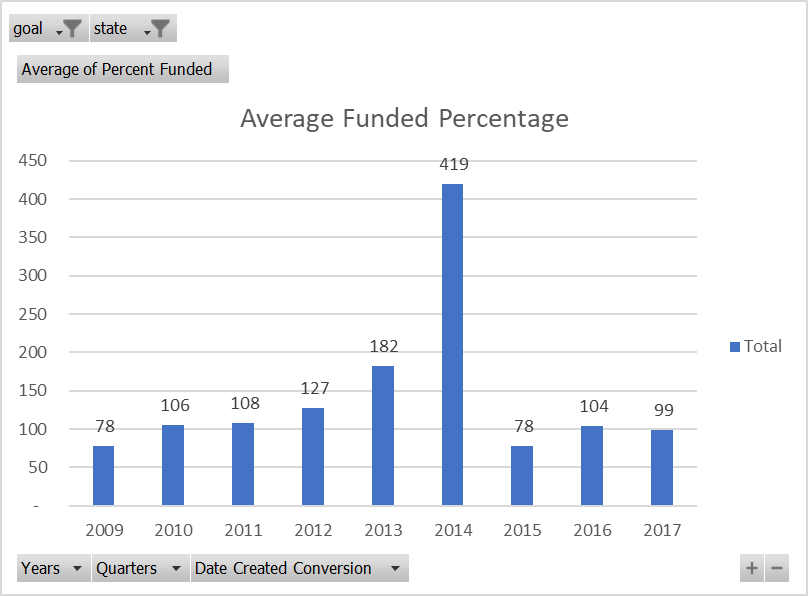
1. A table that shows the Status and Staff Pick to see if the staff picks were a good indicator of success:



1. A chart that shows percent funded to identify potential outliers that may skew results:



This graph shows that some projects received many times more the goal. Most of this projects are related with technology. However, when digging into this data you can find some issues like a projects (e.g. 3840) with a goal of $1 but that received $22,603. This type of outlier skews the results and should be either removed from the dataset or properly standardized/filtered. Once removed a pivot table can be constructed to show in which years the average funded percentage was higher:



1. Calculate a regression to find if some of the variables such as “spotlight” (independent variable) along with other variables is statistically significant in explaining “Funded Percentage” (dependent variable).