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

Theater, music, technology & film & video are the top 4 categories with 78% of total campaigns, being Theater, music and film & video the categories with the highest rates of successful projects, while categories like technology, photography, games, publish and food, have more failed than successful projects, all projects related to journalism had been canceled.

Theater subcategories have highest successful rates but also with important failed rates. Music subcategories have high successful rates, with exception of jazz and world music. While third most important category, technology, all campaigns have been successful campaigns in subcategories like hardware, and tabletop games. Subcategories from film & video like documentary, television, and shorts have also only successful rates.

There is no seasonality related with the state of a campaign, whoever it seems that months May-July have the highest rates of successful campaigns along the years.

1. What are some limitations of this dataset?

We don’t have enough information to know exactly which was the cause that made a campaign successful or not, which external variables as economy, trends, etc. could affect the campaign. On the other hand, we don’t know what the criteria from people was in order to support or don’t a campaign.

Is not easy to visualize percentage funded and successful projects depending on the investing

Even if we have t amount of donation and average donation, this data is in different currencies, we must homologate it.

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

Group information to know which percentage of coverage had each category or subcatecory to know which kind of categories/projects rise more funds.

Graph which categories, subcategories have become more relevant in recent years, or which ones have loose relevance

Have a graph where we can see how many projects are above, exceeding or below its objective, divided by category, to know which categories are more prone.

We can do a calculation to know the total number backers and days that the campaign was running and if this is correlated with the amount raised.

**Bonus Statistical**

The median is more meaningful to understand our data, as we have scattered data, we can say that successful campaigns must have an average of 194 backers to be successful vs the half of the campaigns were successful with only 62 backers.

Regarding variability, both states successful and unsuccessful campaigns have a wide range of variation, it does make senses to have such high numbers in variance and standard deviation, we must clean data and work without outliers.