

# Module 1 Challenge

- Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  - The most successful categories seem to be theater, film and video and music. Also, they seem to be most of the categories in general. Within film and video, the category documentary is more successful. Within music: rock stands out from the rest of its category.
  - The most successful results for food campaigns took place in summer. It may be since they were food trucks. The same happens with music in general maybe due to more concerts of outdoor activities during that period. In order to be funded for a similar project it would be better doing it during those summer months.
  - The better goal range for a project to be successful seems to be around 1000 to 10,000, whereas projects with goals greater 50000 show higher risk of failing, but surprisingly also a lot of them are successful. Those two blocks seem to be more successful in total numbers but also account for most of the data. We could deduce from this that setting realistic goals will result in likeness to succeed.
- What are some limitations of this dataset?
  - The main limitation is that the data is not equally represented. We can find clusters of categories that appear more often which could give us an idea that most of the projects belong to those categories but make difficult to analyze the actual success of other less represented categories that show high percentage of success (like journalism).
  - For some categories we have collected too few data. To give an example for the journalism category, although are projects were successful, only 4 were analyzed. This means that this category is not often used for crowdfunding and maybe should be encouraged further given its success.
  - When comparing goal amount most of the data is distributed between the lower or higher amounts making it difficult to study those ranges in the middle. These middle ranges show a higher successful percentage, but they are less in total numbers to extract more conclusions from

them. It would be interesting to collect more data from this goal ranges.

- What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  - We could have analyze within each parent category what sub-category is more successful.
- Use your data to determine whether the mean or the median better summarizes the data.
  - The mean is larger than the median, so the data is skewed to the right. When data is skewed, the median normally expressed better the results but because both conditions (successful and failed) are similarly skewed both parameters seem to be useful to explain the differences. There are more successful campaigns than failed ones.
- Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  - The Variance and Standard deviation are higher in the successful campaigns. Also, the difference between minimum and maximum value is higher in the successful group. This might be expected as also there are more successful campaigns than failed ones.