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

From the proposed pivot tables, it is possible to take 3 interesting insights:

* There was a much bigger proportion of initiatives in the arts sector where films, video and Theatres represent almost 65% of all initiatives. This indicates that this type of sector may look more for crowdfunding to source their projects which may add as well some competition for projects within this area.
* There were specific sub-categories, with more than 100 initiatives and 100% of success rate such as Rock, Documentaries and Hardware, a great outcome which reveals a great capacity in attracting a good number of backers. This contrasts with 0% success for the sub-categories of Animation, Food Trucks and Videogames despite a similar number of initiatives overall. This can highlight how some areas are currently more attractive to the public than others.
* Success rate depended of the Goal Target within certain brackets. Between $0 to $15k the probability of success seems to decrease once the goal target is higher. Between $15k and $40k the probability of success doesn’t seem to depend much of the goal target with success rate relying more in other features (e.g. category, quality). From $40k the risk of failing to fund the projects increases the higher is the goal target revelling some project may be too ambitious to get funding via crowdfunding.

1. What are some limitations of this dataset?

There are the following limitations of this dataset:

* Who are the backers? The lack of information about the backers which makes it very hard to put in context on why some initiatives were successful and others not. Information such as age group, nationality and profession of the backers would add considerably insight of future initiatives.
* The same can be said about having details of each project, specific features that could be key for success would help analysis and future funding.
* How the different projects were promoted? Were they placed in the same context or some projects had more audience than others?

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

* Looking who attracted more backers overall could be misleading, by comparison one could assume the Art sector were the big winners overall. However, comparing by dollars amount ($) the technology sector was definitely attracted an higher value, in fact this sector account for almost 50% of the value among the successful projects. Refer to “Pivot Data add 1”. Clearly there was a tendency of getting more backers of smaller contributions to the Arts sector and a smaller number of backers for technology but with higher contributions.
* Re-grouping countries by similar characteristics would reveal interesting tendencies. Comparing countries in the following groups:
  + English Speaking – Big Size (US, GB, AU, CA)
  + English Speaking – Small Size (IE, NZ)
  + Non-English Speaking – Big Size (ES,FR,DE,IT)
  + Non-English Speaking – Small Size (NO, CH, NL, SE, DK)

Interesting tendencies would come into surface, for example non-english speaking countries had much more projects in areas of technology and very few in films, theatre or music. This reveals that these areas have less exposure for non-english speaking countries and language shows up as barrier blocking some tentative of funding. This happens especially in countries of minor expression such as the case of Denmark.

* Any graph involving the percentage funded would be an interesting analysis having the potential to highlight initiatives which were real winners.

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

When counting backers for successful and failed projects the Median describes much better what happened in both cases. There are some outliers in both groups that would impact heavily in the overall average. For example, among the success there were 2 projects with more than 20.000 backers which is more than the double of next project with more backers.

The same can be said for the fail projects where there were projects with more 1000 backers which still failed while a significant portion had no backers at all.

Therefore, the median is close to the more frequent values and describes better what happened overall in the group.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There are clearly more variability among the successful campaigns simply because the minimum number of backers is 1 and the maximum is very high.

There were considerable failed projects with 0 backers or below 10 to start with, also the highest values even if occasionally were surprisingly high they are far behind compared to the other successful group. Therefore failed projects would naturally vary through a more limited bracket, many project sitting close to the value 0 and rare for very high values as these were typically successful projects.