Module 1 Challenge Answers:

Three Conclusions we can draw about crowd funding Campaigns?

* Based on the 1000 projects listed, ~56% of projects are successfully funded, ~36% failed and less than 1% are cancelled. Meaning that there is a relatively good chance if the project appeals to Kickstarter backers that a project will be funded.
* Certain industries currently utilize this mode of funding more than others ie. the Arts (theatre, music, film & video). Perhaps an area of possible growth is available in other industries after additional analysis.
* Summer months have a slightly higher success rate.

What are the limitations of this Dataset?

* Cancelled projects do not state reasoning as to why they were cancelled.
* Lack of intel on cancelled or unsuccessful projects and if they were later restarted etc…
* A total of only 1000 projects is listed. Kickstarter has over 230,000 successful projects and just as many or more that were unsuccessful.
* Actual fulfillment of projects after the project is completed is not tracked. Ie. Does the item actually get delivered or fulfilled to backers?
* Repeat backers are not listed.
* Information of $ that was not collected or revoked last minute is not listed.
* Kickstarter cut of money raised is not listed.
* Other categories of projects that do not necessarily fit a mould of Games, Music, Film etc. are not listed in this Dataset.

What are some other possible tables and or/graphs that we could create, and what additional value would they provide?

* Effect of Currency used on results, Effect of Country running the campaign on results. I.e. Are American based projects more successful than European projects. This would help to determine which currency is better to start a project with.
* Effect of Staff Pick or Staff Spotlight on success and over success of a project. This kind of scatter graph would show if its worth spending the money to have extra coverage and advertising of the project? Is the extra spend worth the extra money to get visibility?
* Something like the effect of Time Started vs Time Ended on overall success of the project and length of projects and times of year they were started i.e. before Christmas or after the January Credit card crunch etc. (This could be a bar graph showing success based on size of ask) This would help with determining when optimal times to start projects for highest rate of successful funding would be.
* The size of the goal required vs Result of funding of the Kickstarter project. (Scatter relationship between size and success?) Seeing this would help gauge where to start a project for a better chance of funding

Does the Mean or Median Summarize the Data better?

In a data set like this, the Median is a better representation of where reality lays in terms of results. The result of the Median at 201 for Successful Projects and 114 for Failed Projects lies closer to the distribution of all the numbers. A good chart to use to illustrate this distribution would be a Histogram which would show a better view of where majority of data falls. See Excel for Histogram for reference. This makes sense as larger successful and unsuccessful projects tend to skew the Mean away from the actual average size of projects on Kickstarter.