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

The following three conclusions can be made:

* In the US and in the timeframe the data is in, theatre was a popular category for Kickstarter projects, but it was most likely to succeed in the music category (success rate of 77%). Journalism is also a hard category to follow through, as all 23 projects were canceled.
* Within the US music category, Classical, Electronic, Metal, Pop, and Rock music were all success with a 100% success rate. Faith and Jazz were not as successful (Jazz has a 0% success rate and Faith has only failed and live projects).
* Projects with goals of $15,000 or less are more likely to meet their target. Projects with more than $40,000 are less likely to meet their target.

1. **What are some limitations of this dataset?**

Limitations include:

* Geographical influence. The data is limited to the 21 countries represented. Other countries may have a different environment and the same conclusions may not be applicable. Most of the data is also collected from the US, thus, the data is swayed towards the US market and environment.
* The consistency and accuracy of assigning a category and sub-category. The category and sub-category are inputted by a user and since each user’s perspective may differ, this may cause some inconsistency.
* Data includes various currency and were not converted into a single currency. For comparison, it is important to evaluate each project under the same metric (currency value).

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

Other graphs can include:

* The count for successful projects in each geographical location (group by continents to simplify data). This would help to understand the trends and interests coming from each country/geographical area.
* Relationship between project success and backers count. Though all backers contribute different amounts, by analyzing the success rate (by count) versus the number of backers, it would help to evaluate if there are any relationships between these two variables.
* Relationships between project success and staff pick. This would help to understand if staff picks have an influence in the project’s success.

**Bonus Assignment:**

In both the successful and failed projects, the mean represents the data more meaningfully. There is high variability in both data sets (variance of 713,167 for successful and 3,776 for failed) and the median compared to the minimum and maximum value is very low. From box and whisker plots, there are also many outliers, thus, the mean can capture these outlier values better than the median.

The variance and standard deviation were calculated under the assumption that the data provided is only a sample set of the entire population (i.e. not all successful and unsuccessful projects were included).

From the statistical analysis, the variability in successful projects is much greater than unsuccessful projects (variance of 712,841 for successful versus 3,773 for failed). The difference in variance does make sense. This is because a project needs money to meet its goal and be labeled successful. To receive that money, the project can have few backers who provide large donations or many backers who provide lower amounts. This is reflected in the high variability of the dataset. For unsuccessful projects, they either have few backers who provide large donations or fewer backers who provide lower amounts thus, a lower variability.