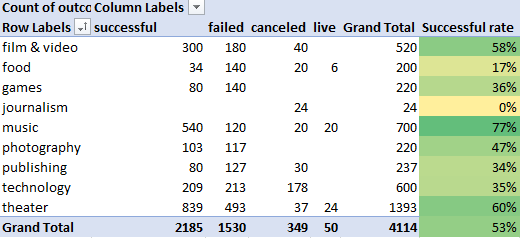
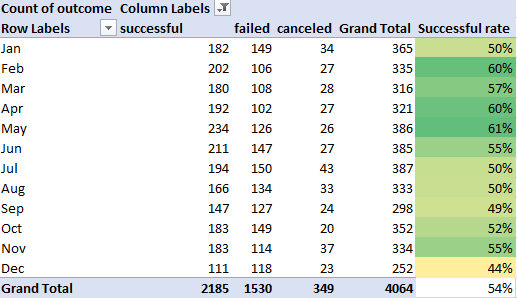
**Assignment Questions:**

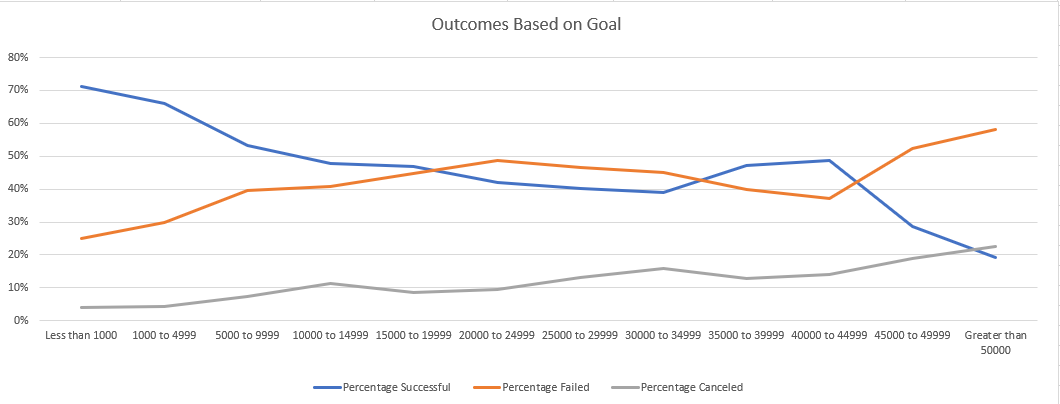
1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**
   1. From the stacked column chart by category, we can see category “Theater” has the most projects. However, category “music” has the highest successful rate.



* 1. Project starts between Feb and May has higher successful rate with May has the highest number, and project starts in December has the lowest successful rate.



* 1. When the goal amount is smaller than 20k, the successful rate is higher than the failure rate. Between 20k to 45k, the successful rate and failure rate are very close in percentage. When the goal is higher than 45k, the failure rate is higher than successful rate. See the graph below.



1. **What are some limitations of this dataset?**
   1. This is a sample data with 4000 projects out of 300, 000. The assignment description didn’t mention how the sample is selected, so there is possibility the sample is not a good representation of the whole population.
   2. On the currency column, we can tell there are different currencies, but in the goal column, there is only a number, so cannot tell the amount is all in dollars or in the local currency format, that will impact the conclusion drew from the analysis.
   3. Goal amount seems widely spread in this dataset, it is positive skewed. Therefore, break the goal amount by different ranges, and then analyzing the data within the range, we might have more meaningful and relevant conclusions.
2. **What are some other possible tables and/or graphs that we could create?**
   1. Look at the length of the Campaign (days between start and end of the project date) and see if that has any impact on the successful rate of the projects.
   2. Average donations per category to identify which category has the most average donations.
   3. Line graph to show the trend year over year per category
   4. Line graph to show the trend year over year per country

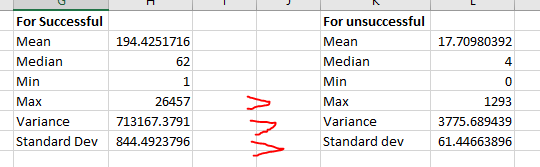
**Bonus Questions:**

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

In this dataset, median is more meaningful. For successful campaigns, the range of data is from 1 to 26457, and the mean is 194.43 which is a lot higher than the median of 62, which indicate data is skewed to the right. The mean will be higher than the median due to the high amount of backers for successful campaigns.

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

By comparing the range, variance, and standard dev, we can tell that successful campaigns have more variability since range is bigger, variance and standard dev are all at a higher number.



It makes sense because successful campaigns will tend to have more backers, and the most successful ones probably will have way more backers than average unsuccessful campaigns, which will make the range is much bigger than the unsuccessful campaigns. The very successful campaigns will skew the data to the right, so the successful campaigns has more variability.