Kickstarter Statistical Modeling

1.1 Is there difference between the amount requested between different categories?

The variable amount requested (usd_goal_real) is normally distributed, we can use a t-test to compare different variables. However, there are 9 main categories, so we at minimum need to run 36 tests (9 * (9-1) / 2). As you run that many tests, the probability of false positive increases; thus, we need to correct this. Thus, we would need to first use anova to determine if there is any difference between different categories. Then, we will perform post hoc analysis to determine if there are any differences between different groups.

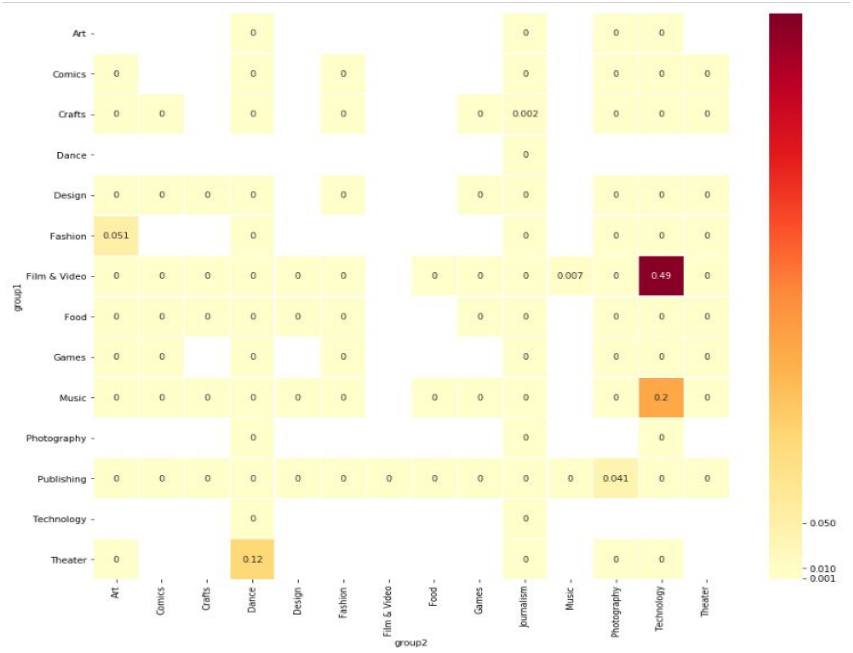
Diagram 1: General Anova Test, Post - Hoc Analysis

Results: General ANOVA ($p < 0.5$) indicates that there is a difference between the groups. Post - hoc analysis shows that the mean amount requested is significantly different across groups. However, we see that Music isn't significantly different from Theater, Photography, or Dance. Furthermore, Design is significantly different from all other groups.

1.2 Is there a difference between the amount pledged between different categories?

The variable amount pledged (usd_pledged_real) isn't normally distributed. Thus, I used permutation tests to get a sense of whether there is a difference between different groups when it comes to the amount pledged.

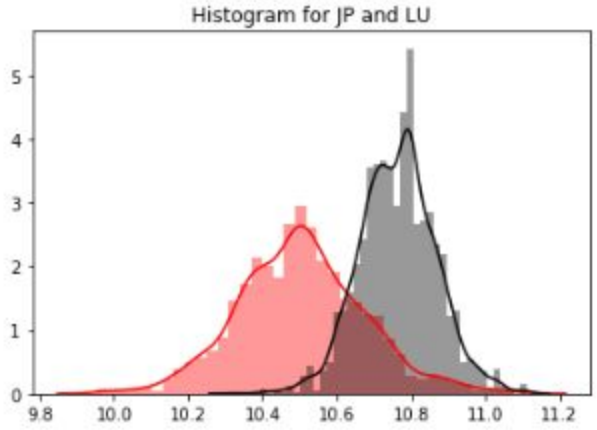
Diagram 2: Permutation Tests Results



Results: The permutation tests indicate that there are few groups where the difference between the means isn't significant. For Example, Technology isn't significantly different from Film & Video and Music when it comes to the amount pledged. Dance and Theater aren't also significantly different.

1.3 Is there a difference between the amount pledged for successful projects for LU and JP?

However, there are too few successful projects in JP and LU! In fact, JP has only 10 successful projects while LU has 20! The sample is too small to even do a permutation! Thus, Bayensian statistics will be utilised to a) generate 3000 trials and then evaluate the difference between the observed mean and the simulated mean.



Result: After comparing the observed mean and simulated mean, the p-value was 0.51 indicating that there isn't any difference between the amount pledged between JP and LU