**Kickstarter Campaign Results and Analysis**

After analyzing the 4000 Kickstarter past projects, it can be inferred the following conclusions:

1. The most popular applications at Kickstarter are those related to theater, followed by music and technology, which have 1393, 700 and 600 projects, respectively (Fig 1.). For these top 3 categories, it can be observed that for theater and music are more the projects that have success rather than those that are failed, or even canceled; and between both, in terms of proportion, music has the higher rate. It is important to highlight the behavior of technology applications, which has a similar proportion between successful, failed or canceled projects. Additionally, it is worth mentioning that there are not successful projects for journalism, all 24 projects seen on the chart were canceled.

**Fig 1. Category vs State Chart**

1. Following the below trend (Fig 2.), it can be evidenced that the most desired sub-category is “plays”, which has coherence based on the above theater category tendency. It can be also observed that sub-categories such as rock, classical music and metal, are representing only successful projects; and this can justify the above music proportion tendency as well.

**Fig 2. Sub-Category vs State**

1. Now, analyzing the State behavior versus Months (Fig 3.), it can be said that May is the month which has more successful projects, and December is the one with less success. Following the trend, there is an interesting continuous drop within successful applications from May to September. This can be considered as good criteria for organizations that want to apply on any initiative.

**Fig 3. State vs Month**

**Dataset Limitations**

It would be interesting to know the reason of why a project is canceled or failed. On the source dataset its clear that both states show less pledged than the initiative’s goal, but this do not help to understand the difference between both states. If someone desire to know how to avoid canceled or failed projects, it would not be easy to analyze that.

Although the above consideration, I think that more than limitations, this dataset has useful information to leverage and to generate more charts in order to generate a deeper analysis. In the next section, I will give some recommendations about what other charts can be generated in order to have a better understanding.

**Recommendations (other tables/graphs)**

1. Pie Charts are a good idea to analyze proportions. Generating this kind of graphics, we can analyze the portion of each state or categories (Fig 4.). For any kind of results analysis, percentages are a good magnitude to understand and compare the behavior of the data.

**Fig 4. Successful projects per category**

1. It is a good idea to normalize key variables to analyze data per project for proper comparison; for instance, to analyze categories based on backer’s funding as follow:

* Average donation per campaign category:

From Fig 5. It can be seen the aggregate average donation that each backer does per category.

**Fig 5. AVG Donation-Category**

And analyzing the results, it can be mentioned of course, that theater is the category where backers do more donations. From the above the results we already know that this category is the one that presents more applications.

However, if Fig 6. is analyzed, it can be inferred that after normalizing average donation per count of projects (for every state), it can be evidenced that technology projects are those that have the higher average donation as follow:

**Fig 6. AVG Donation-Count Project**

**Bonus**

**Question 1**

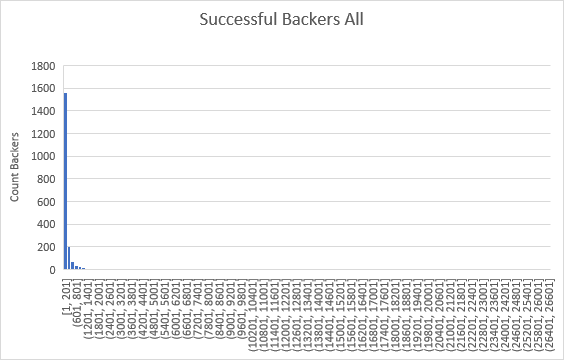
Use your data to determine whether the mean or the median summarizes the data more meaningfully.

**Answer**

In this point Statistics Central Measures were calculated taking in count the whole data, and excluding outliers from the dataset:

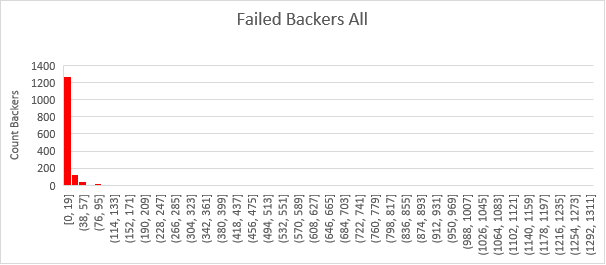
Based on this, we can say that:

For Successful applications and including outliers, we have a Mean of 194.43 backers and a Median of 62 backers. In Fig 7. we can see that there are a lot of data that is out of the average, which makes quite difficult to determine which central measure represents the data; however, if we play with the below chart, we can notice that the median is the one that better represents the data frequency.



**Fig. 7** **Successful-Count Backers Frequency**

This behavior is similar to Failed applications (Fig 8.), which Mean, and Median were 17.71 and 4 respectively:



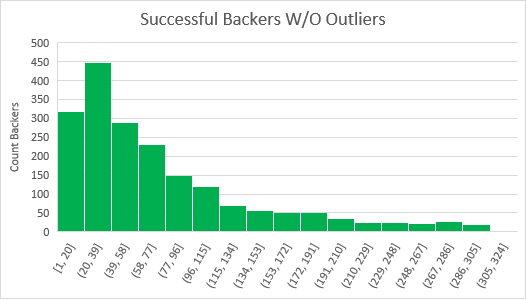
**Fig 8. Failed-Count Backers Frequency**

Having the above limitations, I decided to exclude the un-representative data, making a quartile analysis, and filtering the data based on the minimum and maximum previously calculated. From this exercise, I obtained the following results (see the highlighted numbers):

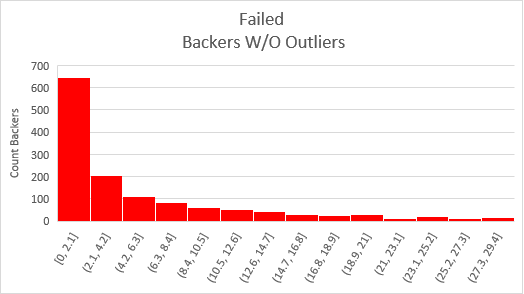
|  |  |  |  |
| --- | --- | --- | --- |
| Successful | | Failed | |
| Median (ALL) | 62 | Median (ALL) | 4 |
| Median (W/O Outliers) | 52 | Median (W/O Outliers) | 3 |
| Mean (ALL) | 194.43 | Mean (ALL) | 17.71 |
| Mean (W/O Outliers) | 75.41 | Mean (W/O Outliers) | 5.41 |
| Minimum (ALL) | 1 | Minimum (ALL) | 0 |
| Minimum (W/O Outliers) | 1 | Minimum (W/O Outliers) | 0 |
| Maximum (ALL) | 26457 | Maximum (ALL) | 1293 |
| Variance (ALL) | 712840.99 | Variance (ALL) | 3773.22 |
| Variance (W/O Outliers) | 4475.87 | Variance (W/O Outliers) | 44.31 |
| ST. DEV (ALL) | 844.30 | ST. DEV (ALL) | 61.43 |
| ST. DEV (W/O Outliers) | 66.90 | ST. DEV (W/O Outliers) | 6.66 |
| MODE (ALL) | 27 | MODE (ALL) | 27 |
| MODE W/O Outliers | 27 | MODE W/O Outliers | 27 |
| Q1 | 29 | Q1 | 1 |
| Q2 | 62 | Q2 | 4 |
| Q3 | 141 | Q3 | 12 |
| IQR | 112 | IQR | 11 |
| MIN | -139 | MIN | -15.5 |
| MAX | 309 | MAX | 28.5 |

**Table 1. Statistics Results**

Doing the above analysis, I found more sense and now it can be determined that Median again is the best measure to represent successful and failed data (see Fig 9. And Fig 10.)



**Fig 9. Successful Backers W/O Outliers**



**Fig 10. Failed Bakers W/O Outliers**

**Question 2**

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

**Answer 2**

There is more variability with successful than unsuccessful campaigns, and this make sense, due that there are more backer counts on successful, and therefore the range between the min and max is also higher than the unsuccessful data.