Executive Summary

Of Data Analysis of the Kickstarter Project

For our data science project, we were given the task to analyze the data of Kickstarter projects and provide information to future potential users about the likelihood of success in their own Kickstarter project. The data comprised of more than 300,000 past of current Kickstarter project users with information such as name, main category, country, category, goal, pledge, state of the project and etc.

We started the project by reprocessing the data in order to achieve the information that we would be using to analyze the data. We used category, main category, the deadline, the launched date, goal and state of the project. We decided not to use pledge, backers, USD_pledge, USD_pledged real and USD_goal real because these attributes are not available for the potential users at the start of the project. We also didn't choose Currency as an attribute for data analysis because the projection for currency and country would almost be similar. The data from the 'category' field was not used and analyzed in our project as we found there were more than a hundred of categories which would be too much of classification to analyze the data. Instead, we just used the 'main_category' as it contains all the categories within it. Altogether, there were 23 countries, 17 main categories, 159 categories, and 5 states. The five states were Successful, Failed, Canceled, Suspended, Live and Undefined. Since the project with 'Live' and 'Undefined' states are yet to provide a definitive outcome, we disregarded data with these states.

We calculated the success percentage of the project data and found it to be 35.98%. Similarly, the percentage of failed data were to be 53.11%, with canceled projects and suspended projects having a percentage of 10.42% and 0.5% respectively. However, while analyzing the success rate in each attribute, we grouped 'canceled' and 'suspended' projects with failed projects.

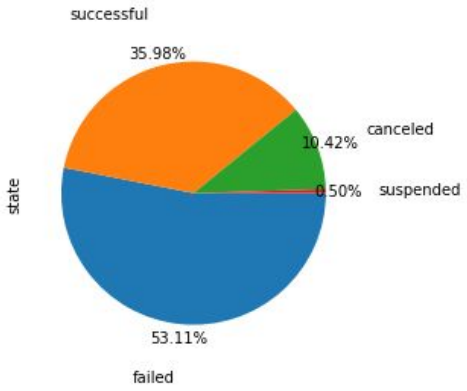
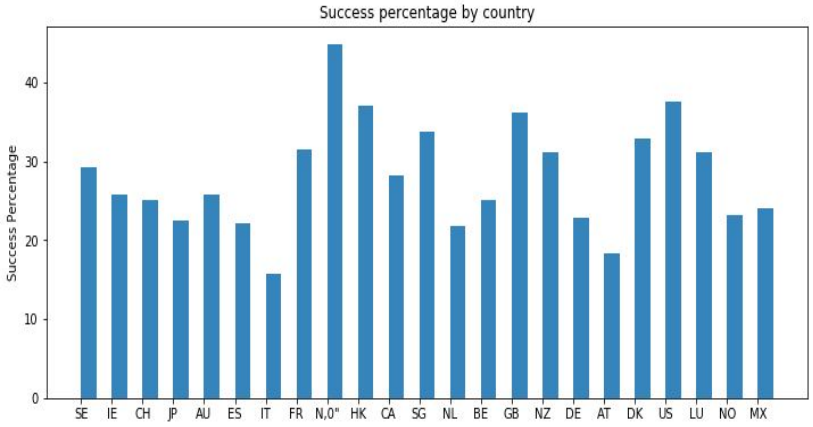
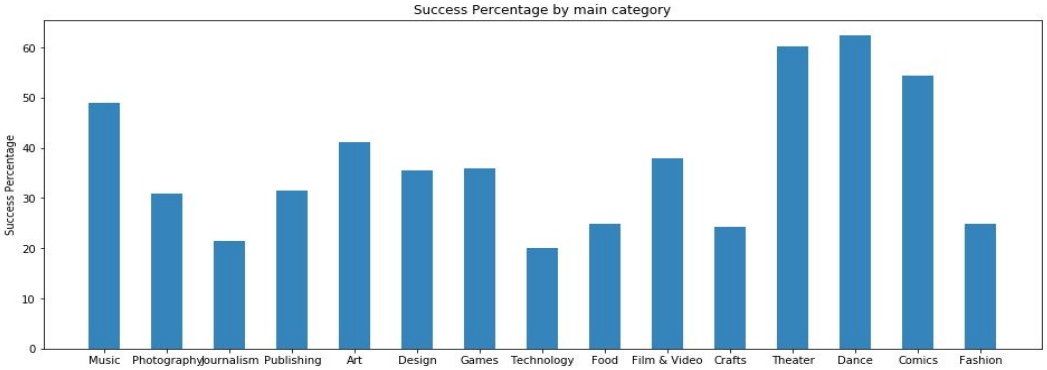


Fig: Pie chart of states' percentage

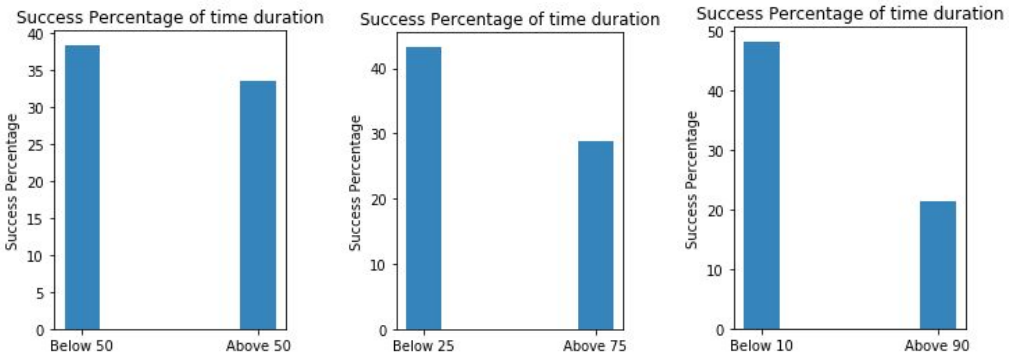
We decided to independently analyze each attribute and determine the likelihood of success for these attributes. Out of the 23 countries, the country with the highest number of Kickstarter project was ‘US’ with 290887 projects excluding the live and undefined projects. The country with the lowest project was ‘JP’ with just 31 out of 372300 projects. However, in terms of the success percentage, the highest country was found to be N,O’’ with a success percentage of 44.87%. US was in second place with a success rate of 37.57%. In terms of the success rate, the lowest country was IT with 15.67% whereas AT is the second lowest with a success rate of 18.38%. JP meanwhile was the fifth lowest country with a success percentage of 22.57%.



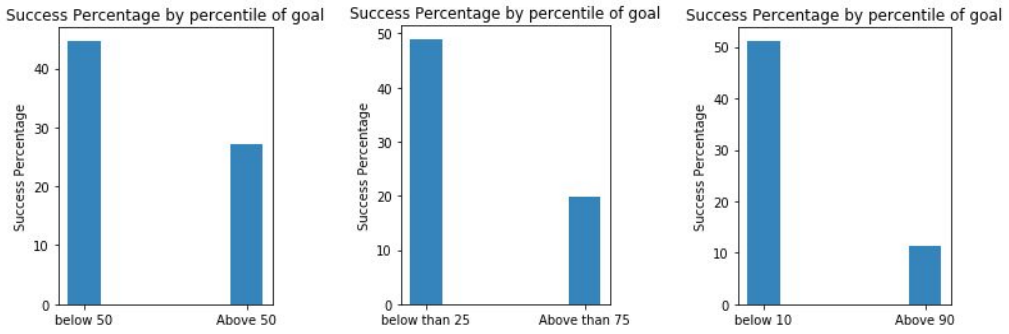
Through a similar approach, we also determined the success percentages for each main categories and categories. Out of the 17 categories, the category with the most number of projects was Film and Video with 62399 projects while the second highest number belonged to Music with 49403 out of 372300 projects. The category with the lowest number of projects belonged to Dance with 3749 out of 372300 projects. However, in terms of the success rate, Dance had the highest rate with 62.32% with Theater coming in second with 60. 14%. The category with the lowest success rate belonged to Technology with 19.98%. Music, meanwhile had the fourth most successive rate with 48.98% whereas Film and Video had a success rate of 37.85%.



As for time launched and deadline, we differentiate these times to get the time duration set by the users to their respective project. We then used percentiles to gather the success rates of low time duration and high time duration. As a reference for comparison, we first used 25th and 75th percentile. For duration having less than 25th percentile, the success rate was 43.33% whereas data with a duration above the 75th percentile, the success rate was 28.76%. As we changed the percentile from 25th to below 10th percentile, the success rate increased to 48.13% whereas the success rate decreased to 21.29% when changed from 75th to above 90th percentile. At the median which was around 30 days, the success rate was 38.46% below percentile and 33.50% above it.

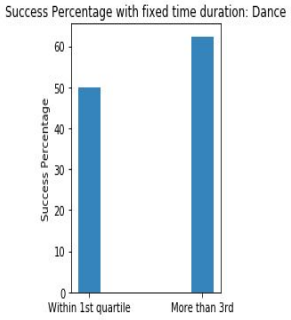


Likewise, we also used a similar percentile approach for goals. For projects with goals less than the 25th percentile, the success rate was 48.99 %. In contrast, projects with goals more than the 75th percentile had a success rate of 19.91%. When the 25th percentile was changed to less than 10th percentile, the success rate increased to 51.29 %. Similarly, when the 75th percentile was changed to 90th percentile, the success rate decreased to 11.25 %. At the median (goal=5500), the success rate was 44.71 % below the percentile and 27.09% above the percentile.



Furthermore, we also checked the impact of the goal and time duration on attributes on main categories and countries. For that, we took the data with Dance category and kept the goal

duration within 25th percentile. The success rate of data with time duration having less than 25th percentile was 70.24% and 64.04% for more than 75th. Similarly, when the time duration was fixed (less than 25th), the success rate for data less than 25th percentile was 50% and more than 25th percentile was 62.37%. We also did the same format for country US. We got 57% and 37% for goal less than 25th and 75th respectively on fixed time duration and 56% and 43% for time duration less than 25th and above 75th respectively for fixed goal interval.



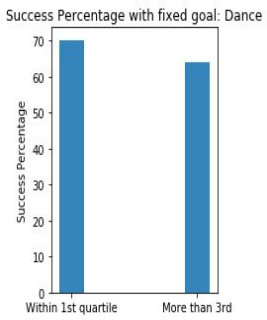
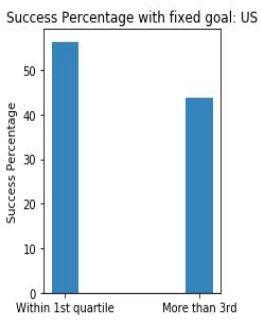


Fig: Category: Dance



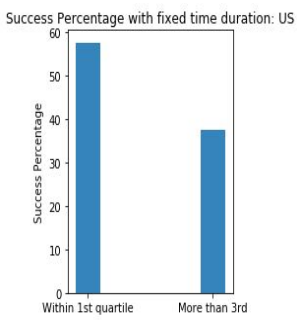


Fig: Country: US

Conclusion:

After analyzing the data, we see that countries such as N,O", US, SG, GB have high success rates whereas countries such as IT, NL, AT have low success rates. Especially for US, despite being a country with the most number of projects, US still had a relatively high success rate over most countries. Similarly, we also see that main category such as Dance, Theater, Comics and Music had relatively high success rates whereas Technology, Journalism, Fashion and Craft had a relatively low success rate.

In terms of goals, we found that the success rate increased as we chose data with fewer goals and success rate decreased as we chose data with greater goals. So, establishing fewer goals has a greater success rate than having large goals. Similarly, establishing less duration for project has a greater success rate than having a large project duration. We also found that for a particular category and country, time duration and goals had some impact on the success rates but there were no drastic changes especially in the case of time duration. From the case of US and Dance, we also saw that the success rate deviated more when there were changes in goals rather than in time duration.

Contribution:

The project was a group project done by Alien Shrestha and Rishabh Shrestha. Most of the ideas were brainstormed through cooperation. However, we divided the task in terms of coding and summary. Alien Shrestha did the coding for determining the success percentage of category, and country and the pie chart whereas Rishabh Shrestha did all the coding for goal and time duration attributes as well as for the plot of all attributes. Alien Shrestha was responsible for outlining and writing the executive summary.