SOUTHERN METHODIST UNIVERSITY

Data Science bootcamp

Homework assignment -1

Microsoft Excel data Analysis

YaYa Dagnoko

Background

Over $2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Getting funded on Kickstarter requires meeting or exceeding the project's initial goal, so many organizations spend months looking through past projects in an attempt to discover some trick for finding success (1). For this assignment, we have a raw data set of several campaigns and that includes successful, failed, cancelled and also on-going ones. We’ll use Microsoft Excel as data analytics tool in order to discover the different trends.

Given the provided data, what are the three conclusions we can draw about the Kickstarter campaigns

The main conclusions about the kick starter dataset are:

1 – There’s a direct correlation between the amount of the pledge, the number of backers and the outcome of the campaigns

This trend is apparent when you take a look at the number of backers by amount. In a successful campaign, 71% of bakers preferred to pledge an amount less than 1000 USD as opposed to 58% of backers in a fail campaign decided to pledge an amount greater than $50K. A first conclusion is that most backers are more comfortable with lower amounts of pledges as there’s always a risk of project failure and unforeseen events that could cause them to lose their money.

2- The backers were much more confident investing in film/video and Music campaigns. There’s almost a 2 to 1 ratio between successful and failed campaign for entertainment projects we can see that there’s 300 successful projects VS 180 failed ones in the film/video category and 540 successful VS 120 failed campaign for music. The main reason is that investors are aware of the film and video industry and no specific training or additional knowledge is required to understand and appreciate the visual arts.

3- Most campaign happened during the summer season May, June and July with respectively 234, 211 and 194 successful campaigns. Not sure what would be the reason for such trend

4- The respective standard deviations for the successful and failed campaigns are 844.30 and 61.43 this shows that the distribution is skewed for the successful campaigns as the standard deviation is large. This is probably due to the high number of backers for lower amounts less than $1000

The standard deviation for failed campaigns is smaller in comparison, therefore, are much closer to the mean. The distribution would be a normal on a graph

II- what are some of the limitations of the dataset

The main limitation for the dataset is that it doesn’t include the age of the backers. This information would be easy obtained using sign in data. It would be helpful to discover the trend of the demographics and would help uncover the investment trends by age groups as well as the categories.

III- What are some other potential tables and/ graphs that could be created

A scatter graph would be a good fit to uncover the relation between the amount of the pledge and the number of backers on both successful and failed campaigns.