# Excel Homework: Kickstart My Chart

Aydin Jalilov

Report

Before starting any data related analysis it is important to conduct validation of the data received. The validation of the data set of 4114 Kickstarter projects have been conducted by checking the blank cells in each available column. This was done by using COUNTBLANK function for each column range from 2 to 4114(e.g. COUNTBLANK(B2:B4114)). The idea behind this validation method is to identify projects with missing properties and remove them from the data set to prevent any erroneous result in the final product. The duplicate data in “name” column were also checked with the help of conditional formatting. Even though there are project with duplicate names those are ones that were not successful at first attempt and were initiated again with modification to the goal. This is useful information and therefore was kept in the data set.

The assignment requires the addition of “average donation” column to the existing excel sheet. For the average donation it is important that the currency is considered. The data provides several types of currency and the average donation should address the currency difference in some way. A dynamic currency rate table was uploaded from <https://www.x-rates.com/> and placed in “Currency\_exchange” sheet. The table updates itself automatically every time the excel file is opened also the table can be updated manually by the user by selecting the table and clicking Refresh on the Data tab. Then by using several If statements and Vlookup functions each currency was converted to USD.

A screenshot of a cell phone

Description automatically generated

The assignment seeks answer to the following questions:

* Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
* What are some limitations of this dataset?
* What are some other possible tables and/or graphs that we could create?

Below are the answers to these questions:

1. Given the provided data, the following conclusion were made:
   1. Out of 4114 data in this data set 3038 represent US based kick starter projects. Therefore, the end result of any analysis largely shaped by the US data. The bonus exercise revealed some interesting facts about the project popularity based on the goal amount. It looks like projects with goal amount in range 0 – 5000 USD are more successful than others. It is notable that the projects with goal amount in rage 5000-50000 USD have approximately the same success rate. It also may be because the projects with higher goal are backed by more people and the amount of the pledged money by an individual does not differ much. But It cannon be seen from this analysis. Analysis on average money pledged by individuals per project can shed some light on this hypothesis.

A close up of a map

Description automatically generated

* 1. It is remarkable that the projects under sub categories “art books”, “audio”, “science fiction”, “faith”, “drama”, “jazz”, “video games”, “web” and “animation” had zero(0) success – all failed or canceled, while projects in sub-categories “rock”, “electronic music”, “documentary”, “hardware”, “tabletop games”, “shorts” and “television” were 100% successful.

A picture containing sky

Description automatically generated

* 1. The sub-category “hardware” raised on average the most money and has 7 projects in top 10 highest pledged projects in this data set. Categories “hardware”, “tabletop games” and “documentary” have the most number of backers. More information below in 3.

1. Limitations are as follows:
   1. The lack of information on how this data set were gathered is a potential issue. Data retrieval from a bigger data set must be based on a stochastic process. Stochastic process is an ongoing process where the next state might depend on both the previous states and some random element. In other words the subset should be randomly chosen from a larger set of data without altering the proportions of the elements. Following link (<https://thehustle.co/archive/02102019d/>) concludes that the 37% of the total Kickstarter projects were successful while based on our data set the percentage of the success is approximately 53%. Based on the given link, projects under category “Theater” share 3% of the total number of projects while in our data set it is 34%. In order to exhibit the same properties as the larger population of the same data it is important that randomized selection of the projects keeps the proportions of the number of projects in each category and subcategory.
   2. Because the data set contain values (goal, pledged) in different currencies, it is important to convert them into USD.
2. The following additional analysis were conducted:
   1. Added Duration of Solicitation Column to the end of the table. The column represents the days the project ran on kickstarter to collect money they need.
   2. Analyzed the average of pledged money for sub-categories. The analysis shoed that the projects under sub-category “hardware” raised on average 130.535 USD on successful projects. When sorted from largest to smallest the amount pledged in USD top 3 projects appear to raise over a million USD per project (2344134, 1279105 and 1052110 USD respectively). Worth noting that there are 7 projects from “hardware” sub-category in top 10 largest pledges. The analysis is contained in sheet named “extra\_average\_pledged” and “hardware”.

A picture containing screenshot

Description automatically generated

* 1. Analysis on what sub-category raised more interest of the audience is shown in sheet named “extra\_backers\_subcat”. The analysis shows that “hardware” sub-category has the most backers by far followed by the “documentary” and “tabletop games”.

A screenshot of a cell phone

Description automatically generated