Analyzing Kickstarter Campaign Performance

Background

Kickstarter (http://www.kickstarter.com/) is a popular crowdfunding site that was launched in 2009. Kickstarter was originally conceived to help artists raise funds to finance creative projects (e.g., films, works of art, performances). Many of the projects on Kickstarter still relate to the arts, but others are projects to raise money for new products or for new service businesses.

When an entrepreneur or artist starts a Kickstarter campaign, he or she sets the fundraising goal, the funding period (defined by a start date and an end date), and the awards set for different levels of funding pledges. For example, someone who is running a Kickstarter campaign to fund the production of a documentary might offer "a thank you in the credits" to anyone who pledges $20 or more.

Important detail: if a campaign does not reach its fundraising goal, no pledged money can be collected.

In this exercise, you will be examining the success of campaigns based on year and category. Success can be thought of two ways—first, as meeting the fundraising goal and second, as the amount of money raised if the goal is met. To do the analysis, you have a data set in a spreadsheet. The data set has approximately four years of Kickstarter projects. Details about the data set:

- Each row in the data set is a project (or campaign). In each row, there are many details, including "goal" (the fundraising goal) and "pledged" (the total amount backers committed to give if the fundraising goal is met).
- Projects are also organized into a "category" (e.g., Narrative Film) and a "category_parent" (e.g., Film & Video).

You will work with this data set to answer questions about the historical performance of campaigns. Please consult the "Selected answers" for questions 1-5 on the last page of this document to confirm that you are on track as you work. ***There are many ways to get the right answers****. Challenge yourself to find efficient and robust ways.*

Questions

1) Create a Pivot Table to answer the following questions. *Remember that you can change the operation on each field to Count, Sum, or Average through the* ***Field Settings***.
    a) How many Photography campaigns were there, including ones that did not meet their goal?

    b) What was the total amount collected across all campaigns?

c) Within the **category_parent** of Crafts, which **category** (Candles, Crochet, etc.) had the highest AVERAGE amount collected per campaign?

d) Across all **category_parents**, which **category** had the MOST campaigns?

e) Which **category_parent** had the highest AVERAGE number of backers per campaign?

f) Create a graph that shows the AVERAGE amount collected per campaign for each **category_parent**.

2) How many projects are in the data set in total?

3) How many of them reached their funding goal (i.e., had pledges that reached or exceeded the goal)? What percentage of all the projects reached their funding goal?

4) There are four years of data: 2012, 2013, 2014, and 2015. How many projects in each year? What percentage of projects in the data set reached their funding goal each year? (Use the *ending date* to determine the year of the project.)

Year (of End Date)	Number of Projects Ending in that Year	% of Projects that Reached Funding Goal
2012		59.27%
2013		62.09%
2014	26,247	
2015	32,058	

5) Is the average of the four percentages in the table above equal to the percentage from question 3) (the percentage of all projects that reached their funding goal)? Why or why not?

6) Using all the years in the data set, create a carefully designed graph that answers this question: which of the fifteen category_parent labels had the highest percentage of projects reach their

funding goal? ***Challenge***: how can you efficiently show the number of projects in the labels on the graph, like the answers show? [Hint! & for string concatenation.]

7) Open-ended CHALLENGE question. Consider working with a buddy on this question. **Did projects that had higher goals tend to collect more money?** (Recall that a project only collects money if they meet or exceed the goal.)

   There are many ways to answer this question with this data set. Try to create a beautifully designed graph that **best** answers the question and conveys depth in the answer.

   Consider these approaches:

   a) Look at a scatterplot of goal level vs. money collected. Correlation? Regression line? Are those helpful visualizations for answering the question?
   b) Another approach is to break down the projects into buckets based on goal level. For example, a table like this:

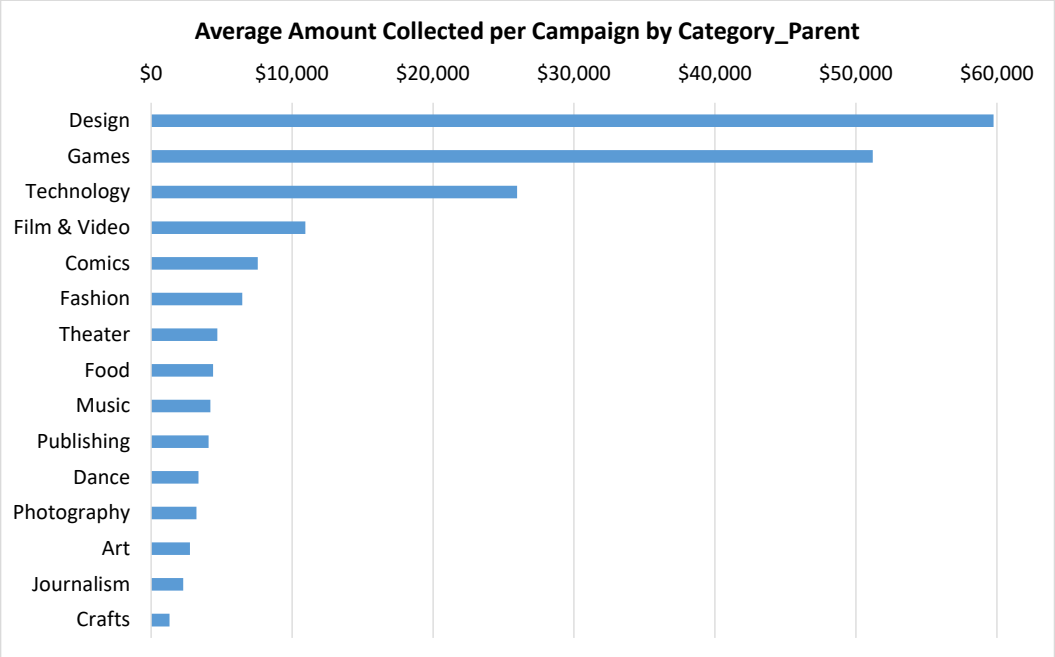
Goal Level	Number of Projects	% Reached Funding Goal	Average Funds Collected*
$0 -$10,000	62,929	53.21%	$3,671.37
$10,000.01 - $20,000			
$20,000.01 - $30,000			$21,682.75
Etc.			

   * Projects that meet their goal collect the pledged amount. Projects that do not meet their goal collect no funds. Average Funds Collected: the average amount of money collected per project, *including* projects that collect nothing because they did not meet their goal.

8) Open-ended CHALLENGE questions. When there are more projects asking for funds, do they each tend to collect less money on average? When there are more projects asking for funds, is the total amount of money raised higher?

Selected Answers

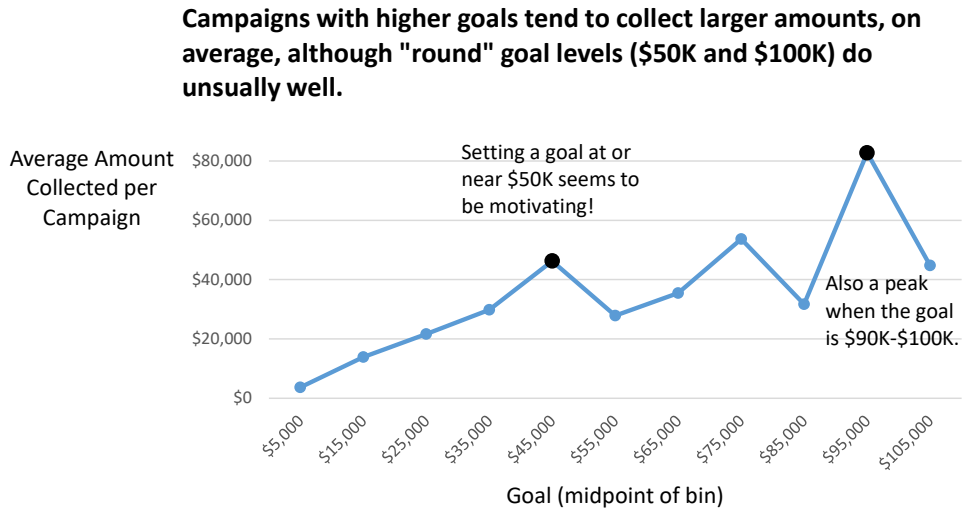
1.
    a. 1,894. (To find this, add any field to your Pivot Table using the Count operation.)
    b. $1,119,815,248. (Look at Grand Total for Sum of Amount Collected.)
    c. With the category_parent of Crafts, the category Letterpress collected the most on average ($8,093).
    d. The category with the most campaigns was Fiction. Fiction had 2,739 campaigns. The Fiction category belongs to the Publishing category_parent.
    e. The Games category_parent had the highest average number of backers (836.4).
    f. Average amount collected for each category parent.



2. 90,122 projects
3. 43,420 projects reached their funding goal; 48.18%
4. See hints in the table.
5. No, because the average of the four percentages does not take into account the number of projects for each year. For instance, if one year has more projects than another, then its % should affect the overall average more.

6. Your graph should show that the top *category_parent* is Design, with a success rate of 72.50%.

7. One possible solution:



Note, though, that the *correlation* between the Goal level and the Amount Collected is only 0.0112, which suggests no relationship. Is that low correlation a conflict to the message in the graph above?

8. For this question, **one approach** is to break the projects into time windows (e.g., annually or quarterly or even monthly) and to look at the relationship between number of projects in each time window and average amount of money collected in each time window.