Excel-Challenge

Stephen Bennett

**Question:   
Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**  
  
1) The sub-category ‘plays’ are extremely popular with these crowdfunding campaigns. Taking up nearly 34% of the data set with the 2nd place following up at 8.5%. That’s still 4 times the amount of the 2nd most. In the sense of ad marketing – I would be very interested in this fact.

2)August had both the lowest success rate (48%) and the highest failure rate (41%). Compared to the average of 57% success and 36% failure rate. Using the data provided, I would not recommend starting campaigns in August. (See additional tests sheet tab)

3) Spotlighting had only a little effect on the chances of success or failure. 3% higher success and 3% lower failure rates when given a spotlight. I would like a much larger data set to confirm my observations.

**Question:  
What are some limitations of this dataset?**

It could be risky to assume that all campaigns had equivalent goals. What if the failed campaigns had more difficult goals set for them? This could make the data look more favorable for the success campaigns.

I think it would also be beneficial to look at goal amounts vs country. Counties that have higher GDPs could be more likely to hit the goals versus those with lower dollar values or lower GDP’s.

Not all categories / sub-categories may have the same online presence as another. This could lead towards higher failure rates for non-techy or older client bases that would not interact with a Kickstarter campaign.

**Question:  
What are some other possible tables and/or graphs that we could generate, and what additional value would they provide?**

While working on the summary statistics table portion of the assignment I did this exact situation without realizing. I wanted to be 100% sure how the data variance was – either skewed or symmetrical. I added both a box plot and a normal distribution bell curve graph. They each had the same results: The data was HEAVILY skewed to the right.

This helped me answer the primary question of if mean or median was more useful with confidence. (It was median due to the skew).