**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

* Just by filtering and sorting the data we can see that Plays and Music related campaigns get more success.
* At the same time Plays also see a big failure rate, therefore the Genres probably matter but that data field is not provided.
* Theater, Music and Films/videos have the highest number of backers, but the rate of success is almost 50-50. This seems to be consistent across countries.
* Journalism though don’t get many backers, but the success rate is 100%.
* Plays get the highest number of backers.
* Denmark sees the highest success rates in comparison with other in term of any sort of campaign being run.
* Year 2020 saw the least number of campaigns, and none got the success probably due to COVID.
* Campaigns that do not target over $5000 see the most success, however the success rate is almost equally good up to targets of $10000, whereas the success rate takes a nosedive as the target Goes beyond $15000.

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

* Some data fields like the name, Blurb, Unix code date, category & sub-category were qualitative shortfalls.
  + Name is not indicative of what is the first and last name among the hyphen separated names.
  + Similarly, the Blurb and category data has data collection and entry issues where more information is fused within one field.
  + Unix code dates add an extra step in preparing the data for analysis.
* Data set seems to have a very spread-out range, and many outliers can be seen by just sorting the data, this implies that if this data set is not the complete set, then the analysis might result as skewed as this dataset.
* We cannot get context of the spotlight data fields as there is no data dictionary provided. Similarly, Though Plays have high numbers of success as well as failures, further context for type of plays can help us better understand what types of plays see more success or failure.

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

* **Box plot –** To see the spread visually and the positive or negative skewness.
* **Pie charts –** To visualize the share of each category and its share so that we can better judge which campaigns have better chances of success.

**Whether the mean or the median better summarizes the data?**

* The mean is significantly higher than the median, in the boxplot we can see that there are many outliers that are pulling the mean up.
* The large standard deviation and variance also indicate a widespread in the number of backers.
* The interquartile range shows that the middle 50% of the data is spread out over 1,152 backers.

**Given this, the median might be a better measure of central tendency for this dataset, as it is less affected by the extreme values. The mean might give a skewed perception of the typical number of backers due to those outliers.**

**Determine if there is more variability with successful or unsuccessful campaigns?**

* Mean and Median: Both the mean and median are higher for successful campaigns, indicating that successful campaigns attract more backers.
* Minimum and Maximum: The minimum number of backers for failed campaigns is 0, while successful campaigns have at least 16 backers. The maximum number of backers is higher for successful campaigns, showing that they can attract a significantly larger number of backers.
* Variance and Standard Deviation: Both variance and standard deviation are higher for successful campaigns, suggesting a wider spread in the number of backers. This indicates that successful campaigns can vary greatly in their number of backers.
* Interquartile Range (IQR): The IQR is higher for successful campaigns, which means the middle 50% of successful campaigns have a wider range of backers compared to failed campaigns.**Successful campaigns tend to have more backers, and there's a greater variability in the number of backers they attract.**