**Written Report**

Based on the Parent category the most successful projects were *Theatre* with 187 and *Film* in 2nd place with 102.

Across the board *Theatre* has the highest number of failed projects at 132 and canceled at 23.

*Film and Video* is 2nd place with the highest number of successful projects at 102, 60 that failed 11 canceled.

The categories that were the most successful were ones that are in the creative arts arena (*Film and Video, Music and Theatre*). One thing they all have in common is that they have the highest number of failed and canceled projects.

*Journalism* was the least successful with 4 but had no cancellations or failures. It’s also the only category with no canceled or failed projects.

Looking at the 7 countries individually shows some interesting results.

* **Australia** – 43 projects – 24 successful with 16 failed and 2 canceled. *55%* success rate.

*Theatre and Film being the most successful categories.*

* **Canada** – 44 projects – 22 successful with 19 failed and 2 canceled. *50%* success rate

*Theatre and Technology being the most successful categories.*

* **China** – 23 projects – 12 successful with 6 failed and 4 canceled. *52%* success rate.

*Theatre and Film being the most successful categories.*

* **Denmark** – 31 projects – 17 successful with 12 failed and 1 canceled*. 54%* success rate

*Theatre and Film being the most successful categories.*

* **Great Britian** – 48 projects – 28 successful with 18 failed and 1 canceled. *58%* success rate.

*Theatre, Music, Film and Video being the most successful categories.*

* **Italy** – 46 projects – 26 successful with 19 failed and 3 canceled. *54%* success rate.

*Theatre and Technology being the most successful categories.*

* **United States of America** – 763 projects – 436 successful with 274 failed and 44 canceled. *57%* success rate.

*Theatre, Music, Film and Video being the most successful categories.*

Analyzing sheet 1 I can see that all of the countries in the report have an average success rate of 54%. So that’s more than half of the projects. Calculating the variance of the average percentage of success rate shows that a variance that has zero indicates that all of the data values are identical. Which indicates that the data points are close to each other as you can see.

Sheet 2 shows that the *Plays* was the most successful Sub-Category with the highest number 187, with *Web* and *Rock* behind with 36 and 49. Least was *Metal* and *Television.*

With sheet 3 you can see that throughout the years of 2010- 2020, the number of successful projects are closer in numbers;

*January* has 49. *February* has 44. *March* has 49.

The number of failed projects reflect a similar pattern;

*January* has 36. *February* has 28. *March* has 33.

The same for canceled;

*January* has 6. *February* has 7. *March* has 4.

Working out the variance and standard deviation for *successful*, *failed* and *canceled* projects, shows us a high level of accuracy based off the low numbers. Meaning that on average the numbers barely differ from each other.

Based off what I have analyzed in this challenge there is some additional criteria that can be included to improve the accuracy of this dataset. For example, if you look at United States of America which is huge, there are also 50 states within that. Figures could be based off each state, which would allow you to have a closer look at which states had more successful, failed and canceled projects. You could also compile information based on region. You could also group all of the states that are in the East Coast, West Coast, South, Mid-West.

An additional table that can be created (I already did one) is one that works out the average percentage of success per country. Once that’s been calculated, work out the variance and standard deviation. That will provide some insight into whether outliers are present in the dataset.

**Statistical Analysis**

After calculating the mean and median number of backers of both *successful* and *failed* projects, it appears that the mean would not be the best for summarizing the data. Based on the statistics of both categories a very high number of variances of backers\_count are present. High variability means that the values are less consistent, so its harder to make predictions. In this case it would not make sense to use the mean since the distribution of the data values isn’t symmetrical and there are clear outliers. The median is the better option since there are clear outliers.