**Module 1 Excel Challenge:**

**Crowdfunding data analysis**

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

Film &Video, Music and theatre make up 70% of crowdfunding projects.

The lowest number of projects are launched after Northern hemisphere Autumn. (Although looking at AU, the project numbers fall in the AU autumn.)

People are very interested in theatre (assuming that the data is accurate)

The best success rates are achieved in June and July.

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

Further information is required on how well the sample represents the population. How large is the population? While it is recommended to have a maximum of 1000 for sample size with so many variables such as the one in this data set it is questionable if this is adequate.

With all countries be compared, some countries had a very small number of projects.

Theatre, with no subcategories, has the highest number of projects and is almost double the next highest category. We would need to confirm the accuracy of this data.

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

Additional data analysis that could be performed include:

* How long were the projects crowed funded for?
  + Longer duration allows more people to back a project. Data analysis indicates that projects with 30 to 50 day duration has a marginally higher success rate than those up to 30, however those above 50 days had a much lower success rate. : Note could be the result of deadline date being extended for losing project, hoping that they get over the threshold with more time
* Percentage success/failure vs category/subcategory
* Percentage of success based on launched month, category and funding duration
* Number of backers per category/subcategory
  + This helps identify the market size that could fund a project. Where are there more opportunities?
* Data analysis based on country?

**Bonus:**

**Use your data to determine whether the mean or the median summarises the data more meaningfully.**

As this is not a normal distribution, Median is a better representation of the number of backers that a project can expect.

Due to the large outliers, the mean is significantly higher than the Median.

**\*Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

Both campaigns had very high variance, predominantly due to not being a normal distribution, and outliers.