Excel Challenge – MLUC

Question 1.

Three possible conclusions from the data:

The most popular online crowdfunding campaigns relate to theatre (344 campaigns), music (175 campaigns) and film and video (178 campaigns).

In particular, the theatre category had the most successful campaigns with 187 (54% successful). However, proportionally journalism campaigns were the most successful overall as a percentage (100%) with only 4 campaigns. The worst performing categories for crowdfunding were games (44% successful) and food (48% successful).

Campaigns run between April and July have the most successful completions, whereas campaigns run between October and December tend to be experience the least successful completions.

Question 2.

There are many limitations of the dataset, it features campaigns run from 2010-2020, however very few examples exist from 2020 (2) which may not highlight current trends. Furthermore, the dataset reflects a heavy bias towards USA produced campaigns which made up approximately 75% of all campaigns, and may not therefore be reflective of the global situation. The Australian data in particular shows significant differences in the distribution of types of campaign vs the other regions (albeit with a relatively small sample size) and may represent outlier data. The dataset does not state the origin of the data, and it is unclear whether the averages are representative of both websites or skewed for a particular one. Given the source is unclear, it is unclear if any of the sources reflect a bias in the type of campaigns available.

Question 3.

Other tables and graphs of interest may include:

Staff Picks vs successful campaign

Spotlight vs successful campaign

Goal vs successful campaign

Investment/Goal (converted to USD) vs successful campaign

Category of campaign over time

Removal of Australian data and relook at the data.

Bonus Statistical Analysis

Question 1.

The data is skewed, the median with IQR would summarise the data. Means and standard deviation would be more appropriate in a normal distribution.

Question 2.

Variance measures the distance from the mean; in the context of the data set – there is clearly a larger range for successful campaigns compared to failed campaigns. Therefore, the greater variance is expected. However, the data set is skewed and extremes of values increase the variance significantly. It is relatively uninterpretable in this context beyond there is a wide range of values.