Crowd funding Data Analysis

1. Initially I created a graph based on the Parent category and subcategory to determine the count of outcome for Successful, failed, live and cancelled. Among which sub-category **Plays** under the Parent category **theatre** received more successful outcomes compared to other categories.
2. The line graph shows a positive increase in the successful outcome during the month of may, June, July and drops in the month of august and it continues till December. In a average successful outcome is high during the month of July.
3. Based on the goal analysis 100% success was when the Goal was “15000 to 24999” and ”30000 to 34999” and the graph showed a decline of 40% at “10000 to 14999”.

Limitation

1. To determine the future trend of successful outcome we require data for 2021 to 2023. Whereas the available data is from 2010 to 2020.
2. There are more outliers in the “backers\_count” for both successful and unsuccessful data. The outliers might affect the analysis.

Statistical Analysis

1. Calculated the mean , median , variance and standard deviation for both successful and Unsuccessful campaigns based on the backers count.
2. Median is the centre of the dataset whereas mean is the overall average of the dataset. There are many outliers in this dataset. Mean better summarizes the data as it is the centre point and all other data is above or below this number. Mean gets affected by calculating the average along with the outlier.
3. There is more variability in the Successful campaigns compared to Unsuccessful campaigns as the success rate is more than failure rate. Also, there are many outliers in both the successful and unsuccessful outcomes which can be best described by the Box chart.