**Homework 01-Excel: Kickstarter data analysis report**

Dataset description: Kickstarter is an online platform focusing on crowdfunding of creative projects. It encompasses everything from theater, film and music to journalism and food industry. The dataset that we explore in this homework is a sample of 4000 records for various projects submitted on the Kickstarter website between the years 2009 and 2017. In this report we want to uncover any interesting patterns in the data, comment on any limitations of the dataset and suggest other visuals which may provide additional insight.

Question 1: What are three conclusions we can make about Kickstarter campaigns given the provided data?

1. Volume: The total number of submitted campaigns rose consistently through the year 2015, reaching the maximum of 1226, but dropped in 2016 and 2017. It is important to keep in mind, however, that the records for 2017 are incomplete (include only records for the first three months of the year).
2. Composition and demographics: The three dominant parent categories by total number of launched campaigns are Theater (#1), Music (#2) and Film & Video (#3). Vast majority (74%) were launched in the United States, 15% were lunched in Great Britain and only 3% in Canada, which takes the third place. The total contribution from all other countries is 8%. Among the three dominating parent categories, the prevalent number of campaigns belong to the sub-category Plays. In all three countries with the highest number of submissions (US, CA, GB) Plays are the most prevalent sub-category.
3. Success rate: More than half of all Kickstarter campaigns included in the dataset were successful. Overall trend is that the rate of success peaks in May, while the number of failed campaigns is higher during the summer/early fall (June through October). Percent of successful outcomes and the goal amount are inversely correlated, i.e. success rate is highest for the lowest goals (less than $1000) and the lowest for the highest goals (>$50000). There is also a local maximum at 40-45K, but taking into account a very small sample (only 55 projects were in this goal in this range), it may not be statistically significant.

Question 2: What are some of the limitations of this dataset?

1. The main drawback of this dataset is an imbalanced composition of records by categories, sub-categories and countries. It is heavily skewed toward the projects submitted in the US or GB and towards few select categories. In same cases the sample is so small that any comparisons should be made with caution, since they would unlikely be statistically significant.
2. Incomplete set of records for the year 2017, which contains only records of projects submitted in January, February and March. The “Outcome vs. Month” plot from the bonus assignment takes these data into account, which makes it look (incorrectly) as though more projects are submitted at the beginning of each year than later in the year.
3. Currency varies depending on the country; some categories are not very descriptive and it is not clear what exactly they represent.

Question 3: What are some other possible tables/graphs that we could create?

1. Histogram of average donations (number of backers vs. pledged amount).
2. Average pledged amount vs. parent categories and subcategories (add a filter by country).
3. Hex plot for goal amount vs. number of backers, with intensity representing the average pledged amount.
4. Percentage of the goal met vs. average donations, normalized by the goal amount.
5. Timedelta vs. goal amount (how long did it take to reach the goal), for successful and failed projects.