**Analysis - Challenge 1 (Kickstarter)**

This project regarding Module 1: kickstart is developed on an Excel software base related to the Bootcamp Data Analytics from the University of Toronto. The goal consists of analyzing the data with more than 4,000 crowdfunding projects. All material regarding the findings and data analysis related to this project will be available on Gitbut for public view.

**Purpose**

For this project, students from Data Analytics need to apply all concepts that we learned in class, including the main advanced Excel formulas such as If, Vlookup, Hlookup, Conditional formating, Countif and PivotTable. Also, statistical modelling functions such as mean, Median, Deviation and Interquartile Range, in order to figure out highlights, trends and findings that support the business decision and stakeholders' needs.

**Analysis and Challenges**

**Analysis of Outcomes Based on Launch Date**

The data set related to Outcomes Based on Launch Date has nine categories and 41 subcategories between 2010 and 2017 and divided into successful, failed, cancelled and live. After analysis using the pivot table on Excel, we have some highlights. All count outcomes related to the Category Theater Based on Launch Date have a concentration between 2014 and 2016, representing 94% with 166 of the total. May was the month with the highest count of the successful with 111, considering seven years of data between 2010 and 2017. When we observe the smallest number of successful Outcomes, they all refer to the period from 2010 to 2013; the same happens in the months between November to January.

When we do a broader analysis, analyzing all categories, regarding the total of outcomes, 78% relate to Theather, Music, Technology and Film & video. These same categories are the most successful with 86%. About subcategories, the most representative inside the most successful Categories shared prior plays, rock, documentary, hardware, indie rock and spaces, with 79% of the total.

**Analysis of Outcomes Based on Goals**

The analysis concentrates on Outcomes Based on Goals for the subcategory plays. The first finding is that there are no cancelled projects, which means that all outcomes analyzed are from successful and failed on goals, representing 1047 outcomes, with 66% related to the successful outcomes and 34% of the failed outcomes. The highest concentration related to successful projects is in the range up to 4999, with 73% of the successful projects. The failed Outcomes Based on Goals are in the range up to 9999, with 267 projects representing 30% of the total failed projects. Although the highest percentage failed in the range between 45000 to 49999, we consider it an outlier because there is just one project in this range analyzed.

**Challenges and Difficulties Encountered**

One of the challenges found during the data analysis is using filters for specific results or variables such as Theater and plays. This challenge increases the risk of compromising the data accuracy in case of not filtered correctly during the analysis process. Analyzing all data and sharing the findings and highlights should give a general view of the results found and could be helpful in comparison. Another suggestion for future analyses is to prepare two analyzes. The first one has the general data with findings. The second one uses specific filters; these processes should support both analyses for better business decision-making.

**Results**

* Two conclusions relate to the Outcomes-based on Launch Date

The Theatre category's first conclusion is that it is the most successful outcome based on the launch date. It means that the stakeholder, when suggesting analyzing this specif category, was suitable. The number corroborates with the results finding the Theather is the most successful category, especially during the year range from 2014 to 2016 and the month range between November and January.

The second conclusion is that when we analyze other categories such as Theather, Music, Technology and Film & video, we find that all these are successful categories, representing 86% of all successful outcomes categories.

* Conclusions about the Outcomes-based on Goals

It is vital to analyze and observe the outliers on the results; for example, when we analyze the percentage failed, we have a result with 100% on the outcomes range between 45000 to 49999; however, it represents just one project. In order to present the result more representative concerning the total, we consider adding the highest results in the column number failed and total projects, considering the same line and the same range.

The result shows that the first three ranges up to 9999 represent 30% of the failed projects with 267 projects.

* Limitations of this dataset

This assignment is related to the Bootcamp in Data Analytics course with very fast-paced week by week. This project must follow what we learned in class following the application of some formulas related to Excel. The main limitation found during the development project was the deadline and some specific features used on Excel such as Vlookup, Hlookup, Conditional formating, Countif and PivotTable. For future analyzes, we should consider expanding this analysis using new formulas and advanced functions on Excel such as Forecasting and Prediction and the advanced Excel function STOCKHISTORY.

* Other possible tables and graphs that we could create

In order to have an overview of another's categories related to successful outcomes, I would like to suggest creating a table using as a filter all categories, by year and just for successful outcomes, and sorting by most significant to smallest. Relate to graphs; It will be very visual using the Area Chart because it shows the participation in successful outcomes of each category by area visually. For example, the analysis performed should help business decisions considering other categories to make a decision or better evaluation for future campaigns.

As a suggestion, the second table and graph create a table using filter all subcategories, by all successful outcomes as a horizontal axis, and years as the vertical axis and sort by most significant to smallest. The graph for this table is Stacked Bar Chat; to have a more precise information graph. We suggest changing the axis options, bounds Minimum zero, Maximum 1000, and relate to axis options units central 50 and Minor 10. In this graph, we have a better view than most subcategories such as plays, rock, documentary, and hardware, which are most representative between 2014 and 2016.