**Analytic Paradigm**

1. **Decompose the Ask**

Since getting funded on **Kickstarter** requires meeting or exceeding the project's initial goal, many **organizations spend** months looking through **past projects** in an attempt to **discover** some trick to finding **success**. For this week's homework, you will organize and analyze a database of four thousand **past projects** in order to **uncover any hidden trends**.

**Answer**: Using Kickstarter historical data, build models to predict reliable ways to deploy successful projects.

1. **Identify Data Sources**

StarterBook.xlsx spreadsheet.

1. **Define Strategy and Metrics**

Instructions provided by instructor.

1. **Build Data Retrieval Plan**

StarterBook.xlsx spreadsheet.

1. **Retrieve the Data**

StarterBook.xlsx spreadsheet.

1. **Assemble and Clean**

Instructions provided by instructor.

1. **Analyze for Trends**
2. ***What are three conclusions we can make about Kickstarter campaigns given the provided data?***

**PivotTable#1**

Starting the analysis using StarterBook data table, I was able to establish the first filter on the dataset. Considering the fact that the size of this dataset is large, an approach of the top 3 elements was used in order to focus more on where the majority of the data is distributed. Therefore, we can divide the data into the following categories:

* **Categories**
* Theater (34%)
* Music (17%)
* Technology (15%)
* **Countries**
* US (74%)
* GB (15%)
* CA (4%)

Since the main goal of the assignment is to find success, if we group the state of the projects and the top 3 countries, we can build a performance measurement indicator for the projects by state:

* **US**
  + Successful (54%) and Failed (36%)
* **GB**
  + Successful (61%) and Failed (34%)
* **CA**
  + Successful (44%) and Failed (44%)

For the top 3 countries, the percentage of successful projects is higher than the percentage of failed projects leading us in a direction that we’re trying to achieve. Choosing countries where the percentage of failed projects is greater or equal to the number of successful projects can lead us to risk.

Another important indicator that can help us to reduce the risk is using a top 3 failed categories. A failed category is defined as:

total failed projects > total successful projects.

* **Failed Categories**
* Food (70%)
* Games (63%)
* Photography (55%)

**PivotTable#2**

Using the pre-defined filter (PivotTable#1) by limiting the countries to: US, GB, CA, and limiting the categories to: Theater, Music and Technology, we can build the top 3 sub-categories:

* **Sub-categories**
  + Plays (40%)
  + Rock (10%)
  + Spaces (7%)

Note that Plays and Spaces categories are members of Theater category and Rock is a member of Music category. Using this subset of the data can lead us to an understanding of eliminating one of the top 3 categories (Technology).

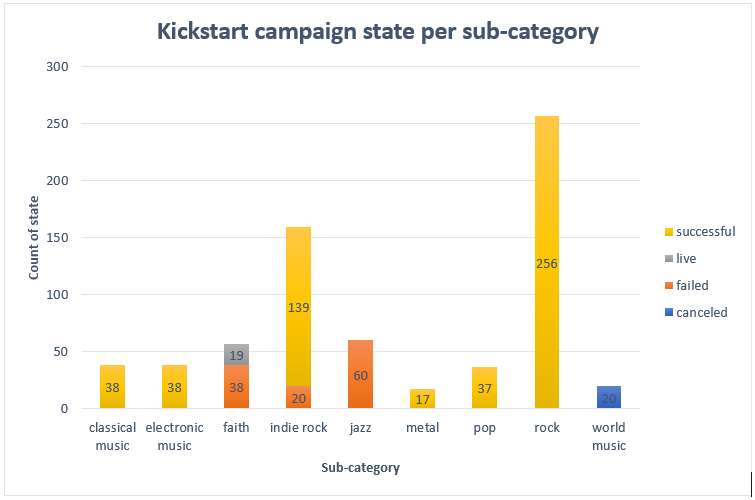
Grouping the state of the projects and the sub-categories in order to build a performance measurement indicator for the projects by state we have:

* **Plays (Theater)**
  + Successful (66%) and Failed (32%)
* **Rock (Music)**
  + Successful (100%) and Failed (0%)
* **Spaces (Theater)**
  + Successful (47%) and Failed (41%)

In the same way that we built the top 3 failed categories, we can create also the top 3 failed sub-categories:

* **Jazz (Music)**
  + Successful (0%) and Failed (100%)
* **Web (Technology)**
  + Successful (0%) and Failed (100%)
* **Faith (Music)**
  + Successful (0%) and Failed (100%)

Based on the filtering applied on the dataset, Music category presents a good indicator for successful projects. It also includes other sub-categories where the perfect score is reached.



**PivotTable#3**

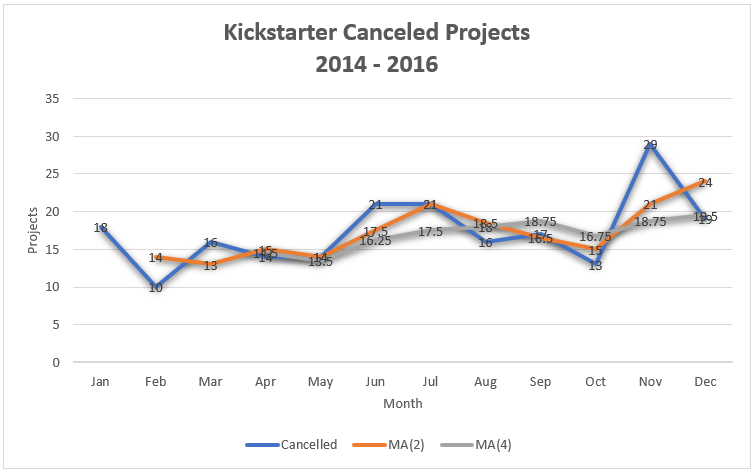
For this last dataset, time-series data is available. Using the pre-defined country and category filters, we can have a better understanding of how the data is distributed across the years. Applying again to the top 3 transactions per year we have the following data distribution per year:

* **2015 (31%)**
* **2016 (24%)**
* **2014 (23%)**

This indicator is important to be used as a guide to see if the product might or might be appropriate for the year in question. It also provides the age of the dataset. For instance: there are several products where the timeframe of popularity is limited.

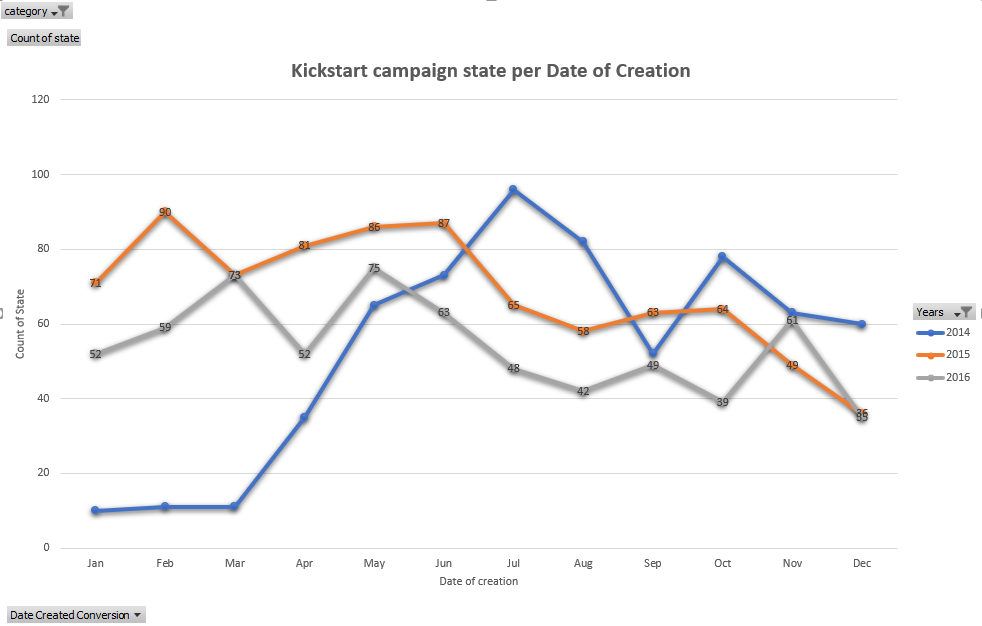
Once the year filter is applied for the dataset, the following behaviors can be observed:

* **Number of cancelled projects tends to stay constant**



Applying a moving average data transformation for cancelled projects series makes the trends clearer to be recognized.

* **Maxima of projects tends to occur during Q1-Q2**

****

The year of 2014 presents a different maxima pattern compared with the other years but a similar pattern for the minima.

* **Failed projects higher than successful projects tends to occur during Q4**

This event can be observed for the following dates:

* + **Dec-2014**
  + **Oct-2015**
  + **Dec-2015**
  + **Dec-2016**

Note that the event can be observed in Dec for different years.

**Bonus#**

Using the bonus dataset, we can observe that:

* Goal presents an inverse relationship with the percentage success. As the goal increase, the percentage of success decreased.
* Goal presents a direct relationship with the percentage failed. As the goal increase, the percentage failed increases.
* Goal presents a direct relationship with the percentage cancelled. As the goal increase, the percentage failed increases.
* Goal presents a direct relationship with the percentage of failed. As the goal increase, the percentage of cancelled increase.

1. **Acknowledge Limitations**
2. ***What are some of the limitations of this dataset?***

* The lack of:
* The rate of change for the pledge amount in respect of time.
* Number of page views of projects is respect of time.
* Backers profile information such as email or location.
* Rewards information
* Pages updates
* Outdated dataset

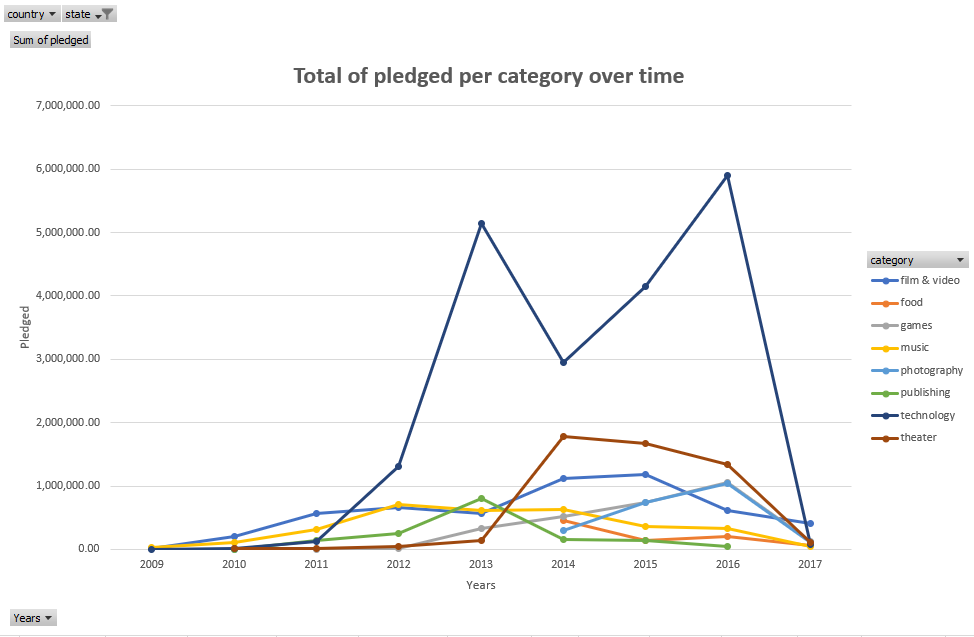
1. **Make the Call or Tell the Story**

Therefore, we can conclude that:

* The majority of Kickstarter projects are happening in the US and it can be divided into 3 main categories: Theater, Music, and Technology. The categories where a higher failure risk can be observed are Food, Games, and, Photography.
* Rock, Classical and Electronic are sub-categories that demonstrate low failure risk. Jazz and Faith are sub-categories with a high failure risk.
* Good project launch dates for Kickstarter are inside Q1-Q2 and risk launch dates are inside of Q4.

1. ***What are some other possible tables/graphs that we could create?***

* **Total of pledged per category over time**

****

* **Funded projects per category when the percent funded is higher than 100%**
* **Average of projects execution**
* **Scatterplot to identify the relation of backers and pledged amount**