# Data Exploration

## Data: Kickstarter projects

Within this dataset there is more 300,000 Kickstarter projects ranging from many categories. I found this extremely stimulating due to three reasons. Just the size of the data, with such a large date set I can use different logarithms to find patterns. Secondly, Kickstarter projects are based on the visions of entrepreneurs and revolutionary. New ideas, new products always fascinated me. Thirdly, who does not want to know the underlining treads of projects and their success rates.

## Questions

What main categories are the most successful ones? Does the campaign goal affect its success? Predict if a project will be successful before it is released (Further analysis for this answer)?

## JMP Analyst & Prep:

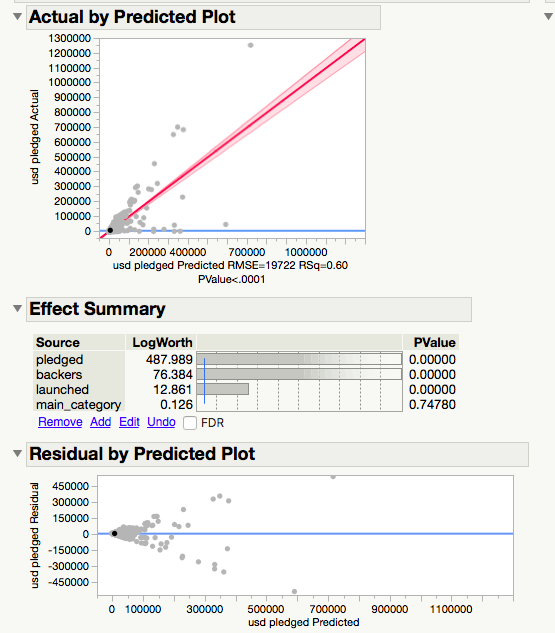
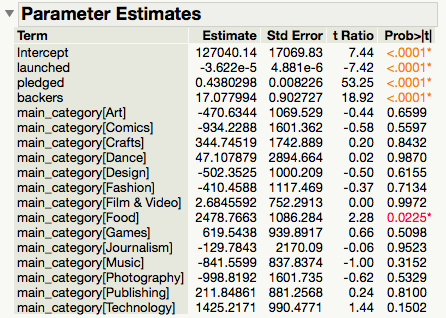
Data Preparations:

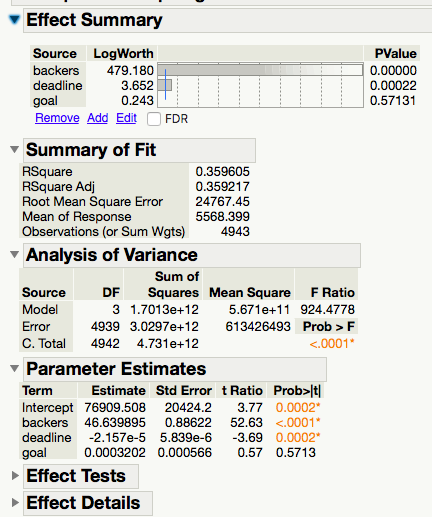
1. The first thing I did was take away steps that were not vital to my analysis. The columns that created noise and did not add insight into the above questions.
2. I took a sublet of 5,000 rows from the 300,000 total rows within the dataset. This was to give me easy to analyze the data. This also became my training data set.
3. Any columns that had many nulls were also deleted.

Analyst:

1. I tried different clustering methods to get a sense of treads in an unsupervised way. I picked to take an unsupervised approach at first to see any patterns or treads that I did not think of.
2. I also moved the data around, many all sorts of different graphs to get a sense of the data what variables where what.
3. The logarithm I picked analysis was multiply step linear regression analysis. Due to several factors: Numeric (continues columns) Values, I wanted to see the strength of connection between Goal amount and pledged amount. As well as multiply other variables.

## Visulation





Results :

1. I could find which category was the most successful one. Which was: Dance
2. One important factor is the number of backers. R^2 = .42394
3. Questions answered : If a campaign is started and backed by many within the dance category it is likely to be successful

