**Final Project, Part 1: *Proposal***

**Introduction and Problem:**

According to [Yahoo](https://smallbusiness.yahoo.com/advisor/16-surprising-small-business-statistics-infographic-190434232.html), about 543,000 new businesses are started each month. Also, small businesses in the U.S. employ 57 million people ([StartBlox](https://www.startblox.com/blog/8-small-business-statistics-2018" \t "_blank), 2018).  There are many startups in the U.S., and they are big part of the U.S. business overall. Over 99% of U.S. employers are small business owners ([Cision](https://www.prnewswire.com/news-releases/small-business-owners-us-is-facing-a-retirement-crisis-300215234.html" \t "_blank), 2016). These facts lead to the question how startups are growing in the U.S. market.

In order to understand, what variables will lead success of startups, I would like to know how much R&D, administration and marketing expenses other startups the startups have invested and profits they have generated. Within the dataset I found, I will be able to learn about 50 startups’ R&D, administration and marketing expenses as well as their profits by states.

**Objects in this case study:**

To find the variables which lead the success of startups.

1. Include the following elements:
   * Hypothesis/assumptions:
     + High investments = High profits/high success rate
     + High marketing and R&D expenses will lead startups to their success.
   * Goals and success metrics:
     + To find the variables which lead the success of start-ups
   * Risks or limitations:
     + Data-set limitation: other variables which are not includes in the datasets, such as variable expenses, business types, start date, and etc. could be the important predictors to predict the target variable.
     + 50 startups are not enough data to draw any conclusion regarding the correlation between expenses (Marketing, R&D and Admiration) and profits.
2. Identify at least one relevant internal dataset and confirm that you have (or can get) the right access permissions.

* The [data set](https://www.kaggle.com/farhanmd29/50-startups/version/1) is from Kaggle.