**Customer data wrangling Project Proposal**

1. What is the problem you want to solve

Given a dataset, containing customer demographics, score and other data, identify which variables affect most the higher customer score.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

The client is a marketing company, that wants to understand the image of their best customers, based on their demographics, assigned scores, skills, job title, and industry.

1. What data are you going to use for this? How will you acquire this data?

The client has provided the data

1. In brief, outline your approach to solving this problem (knowing that this might change later).

* Data Cleanup
  + Missing values - exploring each variable for missing values, and identify the best strategy to replacement/enrichment.
  + Duplicate values – eliminate duplicate values
  + Outliers – investigate the outliers
  + Comma separated values – break down the values in one cell into multiple cells and populate an appropriate indicator values (1 or x) for the appropriate values
* Conduct Dimensionality Reduction
* Conduct Discriminant Analysis to identify the influencing variables.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

* Code
* Input data
* Output data