**Individual Project (Spring 2020)**

**Background**: as entrepreneurship is getting prevalent nowadays, many people want to start their own companies and without the needs to listen to their bosses tell them what to do. Especially in the US, entrepreneurship is highly motivated and there are still many new business ideas waiting to be found and put into ground.

**Datasets Introduction**: People are wondering what the keys could be contribute to those founders’ success, who possess high valuation startups. The two datasets are about 573 founders’ background information and 1216 founders’ startups’ valuations. After joining 2 tables and cleaning redundant data, there are 553 distinct observations left.

**Decision Tree:** After first explored the basic statistics about the dataset, I used first decision tree then logistic model to predict which features could influence the startups’ valuation equals to or more than $60,000,000. The tree gave me the first best predictor is whether the founders worked at Google, second is whether they attended Stanford or Berkeley for undergrad, then ever worked is sales position, and followed by ever served as TA/Teacher/Professor/Mentor? and years of employment.

**Logistic Regression Model:** using the same dummy y, it gave me the similar results. Worked at Google?, Stanford or Berkeley?, and Ever served as TA/Teacher/Professor/Mentor? are the 3 relatively significant variables.

**Conclusions**: using the above 3 variables (positive effect), plus Worked in Sales (negative effect) and Years of Employment (important but neutral) total 5 most important features, I derived from them but not limited to the surface meaning and analyze the logic and intuition behind.

**Graphic**: except for initial graphical description about the dataset. After I derived the results from the models, I went back to the original data and according to the features I created average line charts for startup valuations to see if the model results could match what the datapoints directly show.