Individual self assessment

This is an individual assessment of the final project for the University of Miami Data Analytics and Visualization Bootcamp. The project explored a dataset to analyze and predict salaries for Data Science and Business Analytics jobs that Bootcamp graduates could become candidates for. The project consisted of data cleaning, machine learning and data visualization form students Nancy Arguello and Erick Cabrera Baez.

**Self-Assessment**

The project utilized a dataset obtained from Kaggle.com and it there was a significant data cleaning effort by Nancy using the Pandas library in Python, she also defined an ERD to layout a database that ended up not being necessary to connect since the analysis was done out of a single dataset contained in a csv file. But the tables and keys were defined.

Erick completed the machine learning models also via Python by using the SKLearn library in a Jupyter Notebook; and the data visualizations in Tableau using the data from the dataset. There were three different versions of the dataset explored and analyzed for the machine learning models, one filtering only the jobs labeled in the Data Science area, another with the jobs labeled in the Business Analytics data and another one including all salary area labels.

The greatest challenge with the project was definitely the machine learning models reaching an acceptable score to determine the correlation between the independent variables and the dependent variable. After multiple data cleaning and filtering efforts and many considerations about the independent variables to be used, we ended up with a score as high as 0.49.

**Team Assessment**

We maintained constant communication analyzing the necessary steps to have the technical requirements of the assignment covered. We distributed the work as equally as possible considering it was only a 2 member group. We were good at pivoting from one option or tool to another when there were challenging parts of the task. Big thanks to class instructor Sam Browne who was a big support with the technical aspects and challenging parts of the project.

**Summary of Project**

With the dataset provided by analyzing via linear regression and random forest models after many data cleaning we came to the conclusion that there is not a lot of influence from experience and years at a company in the base salary perceived by Data Science and Business Analytics roles. Considering the low scores obtained by the models and sustained by the analytics in the visualizations, the longer you stay in a company, the fewer expectations to perceive a significant salary jump.