Capstone #3 Proposal

Problem:
What factors and variables will affect the attrition and performance of an employee?

The cost of training an employee for them to be situated in the company has a financial impact because the time it takes for the new hire takes at least 12 months to be fully productive. When an existing employee decides to leave the company. There are many factors that affect an employee's attrition and performance: family, kids, work life balance, job satisfaction, complexity of the work, relationship status, etc.

Developing a classifier that more strongly weighs relevant factors using Logistic Regression, Decision Trees, or Random Forest could help solve this problem. This is an invaluable method because, if successful, it will be extremely important to improve employee satisfaction with the company, improve attrition and performance. The scope of this project is to find ways to best improve the employee satisfaction and retain attrition. Other companies can benefit from similarly specified factors especially when it comes to improving the company's environment to retain employees.

Data:
Having the IBM data as a reference point will allow us to use a representation of the workforce at a large scale company. Given IBMs growth in products and individuals, IBM would serve as a good representation as a baseline for other companies whether to that size and magnitude or hopes to become of that size. Since the data has many variables as below:
- Work life balance
- Relationship satisfaction
- Performance rating
- Job satisfaction
- Environment satisfaction
- Education level
- Manager
- Recent promotions

Data link: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

Deliverable/Approach:
1. Define the factors and variables that best attributes for employee attrition and performance level
2. Finding the minimum threshold that would allow for an employee to stay at a company
3. Train the data set and compare against the test data set to determine whether the attributes and the minimum threshold would determine if the employee is performing well and if the employee would likely to stay with the company (ie. attrition rate)

I will have code that is (hopefully) relatively generalizable to different queries, a slide deck to present my findings, and a good starting point for finding the factors that best define the employees attrition