BANA-288

Group #4

Group Members: Susan Mikhail, Matthias Ronnau, Bryan Sam, Daniel Yee

April 19, 2021

BANA-288 Predictive Analytics Project Proposal

**Goals**

The goal of our final project for BANA-288, Predictive Analytics, is to predict when an employee will leave a company, as well as to determine which factors are most important in influencing this decision. For instance, do individuals with smaller salaries, all else equal, exhibit a great chance of leaving than their better paid counterparts? We seek to answer this and other questions like it. Understanding these relationships can better help companies retain their employees.

**Forecasting Variable of Interest**

The forecasting variable of interest for our dataset,”Attrition”, measures whether or not someone left the company (binary 0/1). One possible way of achieving our goal would be to use a logistic regression to predict whether a given individual will leave the company or not.

**Observation**

A row in the dataset is for a single individual; no individual has multiple rows. Each record has information containing an individual’s age, how often they travel, the department they are in, their salary, and other information. In total, there are 35 columns: 1 dependent variable (“Attrition”) and 34 independent variables.

**Correlates**

For our models we will initially try to use all relevant variables, as some of the 34 independent variables, such as “EmployeeNumber”, are not relevant to model building. That being said, we hypothesize that “Age”, “JobSatisfaction”, and various economic variables (income, stock options, etc.) are likely to strongly influence our dependent variable. We believe that an employee’s age is important, as younger people may be more willing to leave a company (or may be less engrained within it). Additionally, an individual’s job satisfaction should be a key indicator, as dissatisfied employees should leave at higher rates than satisfied or happy employees. Finally, we expect economic variables to have a strong impact on one’s propensity to leave the company: people making less money may be more willing to jump ship for a higher paying job.

**Data**

We found data on Kaggle that was simulated by IBM data scientists: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>. In total, there are 1,471 rows of data. Luckily, we do not expect there to be much data cleaning necessary.

**Plan**

While we still need to assign certain tasks, each of us requested the other members of the group because we are aware of the quality of work that they do. In the next few weeks we will meet to set deadlines for specific parts of the project (cleaning, EDA, visualizations, etc.) so that we are not doing the entire project the week before it is due. Our tasks will include:

##### Ask the right questions

1. Analyze different subsets of data
2. Explore trends - Summary statistics with visualization
3. Build Models - Training and Validating the data
4. Conclusion and Recommendations