**Option 1**:

Problem: Job satisfaction has become a primary issue for companies, especially in relation to matters of attrition and production. Using this data set, I would like to look at which factors contribute to increasing and decreasing job satisfaction and to what extent.

Predictions: By doing an analysis using the data set “Classify and Predict Satisfaction”, we should be able to predict which factors are correlated to attrition and job dissatisfaction the most. We should be able to answer questions to whether higher job satisfaction is correlated to lower rates of attrition, or whether marital status, salary, distance to work, etc. correlates to job satisfaction and how much. Furthermore, this data set has a lot of information on the age, gender, education, and job level of each employee. We can use this to determine the average job satisfaction according to a specific age range. It can be used to answer a simple question such as: are employees who are 50+ more satisfied with their job? This data can also be used to answer more complex questions such as where does job dissatisfaction lie? We can view this by assessing age groups, gender, marital status, and commute time. Are females more dissatisfied with their work relative to males? If yes, does job level have anything to do with it and to what extent?

This data would help clients who are either executives of companies or employers at a managerial level where they would either like to lower their attrition rate or assess how satisfied their employees are currently to prevent potential attrition. Based on a thorough analysis using this data set, these clients can decide what about the factors in their place of employment needs to change accordingly. The goal of this analysis is to use the input from this dataset to predict job satisfaction.

**Option 2**:

One of the data sets that I assessed in particular was “Crime in Chicago”. The data shows you which area of the city the crime occurred in, what kind of crime it was, whether it was domestic, whether there was an arrest, the geographical coordinates, the year, the beat and further information. This data would assist police departments, district attorneys and judges make changes in their day to day decisions. Perhaps, police officers can change their route and give their focus to certain areas or attorneys and judges understand the incidence of specific kinds of crimes in their jurisdiction.

Furthermore, because there is such a large amount of data on the type of crime and the amount of arrests that arose from it, we can determine whether an individual is likely to be arrested in the future for that type of crime. We can also determine in which region crimes are taking place and how often. We can predict how likely a crime is to occur in that location again and even perhaps what kind. There is further information on whether a crime is domestic. This data shines a light on whether domestic crimes lead to higher or fewer arrests and how often they occur.

**Option 3**: One final dataset that I attempted to analyze was “Mental Health in Tech Survey.” Mental health datasets were of those that were more difficult to find. Before finding this data set, I tried looking at datasets for readmissions for psychiatric patients in hospitals. This dataset did not show any variables which could used to predict or determine anything valuable. “Mental Health in Tech Survey” is a better dataset where it looks at several factors such as age, gender, country, whether there is a family history, whether treatment was done, whether they were part of a wellness program, and how easy was taking a medical leave for them. This dataset can be used to predict which variables correlate with an individual having a mental health issue. However, it does not assess whether the individual considers work to be a factor in their condition. It does not assess if there was any traumatic event that had occurred for the individual in the recent past. It also does not assess whether the individual has dependents and what kind of benefits they have. While the dataset does have several variables, it is not specific enough to be able to make accurate predictions.

I have picked option 1 as my final project idea because the dataset thoroughly assesses multiple factors in relation to job satisfaction that can help predict not only satisfaction but also attrition. Several correlations can be made between different factors and on multiple levels. Furthermore, the predictions from this dataset can be valuable to potential clients who would like to understand what to change about their job environment in order to have more satisfied employees and perhaps higher levels of productivity.