Mark Shinozaki
3/29/23
Cpts 315 – Project Proposal

1. Data mining Task: What is your data mining task? This task could be a series of exploratory questions that you want to investigate or analyze. What is your motivation behind choosing this problem for your project?

   - The data mining task for this project is to preform exploratory data analysis on the New York State Power Ball and Mega Millions winning numbers, and to build a predictive model to predict future winning numbers. The motivation behind choosing this problem for this project is to gain insight into the patterns and trends in the winning numbers that can help in predicting future winning numbers, and to learn and apply data mining techniques to real-world problems. I understand by using lottery data, the predictive model can only be so accurate due to the lottery being a game of chance and the numbers randomly selected

2. Dataset: What is the source of your data?

   - The source of the data is from Data.gov, which is an open data platform provided by the United States government. The dataset includes the winning numbers of Power Ball and Mega Millions lottery games in New York State from 9/26/2020 to 3/27/23 and 9/25/2020 to 3/28/23 respectively.

3. Methodology: How will you solve the data mining task? You should have some idea of the algorithms or software tools you plan to investigate.

   - To solve this data mining task, I plan on using Python and several libraries such as pandas, numpy and potentially seaborn and scikit-learn. The Pandas library will be used to load and preprocess the data, numpy will be used for numerical operations, seaborn could potentially be used for data visualization, statical analysis, and a feature engineering to gain insights into the data and identify patterns and trends. There may be tools added and not added depending on input from the professor or usability of these applications to produce an adequate predictive model.

4. Final product: What will be the outcome of this project? How will you measure the success of your course project? Will this project help you explore or learn something new?

   - The outcome will be a productive model that can predict future winning numbers of Power Ball and Mega Millions lotter games in New York State. The success of the course project will be measured based on the accuracy of the predictive model and the insights gained from the exploratory data analysis. This project will definitely help me explore and learn new data mining techniques and apply them to this real-world problem and possibly other real-world problems.