How to run the code

1. git clone https://github.com/Elliott-Hendrickson/CPTS440FinalProjectWSU.git
2. Move to the folder “Data Converting Process”
3. Run “normalization.py”

This is the process of how the code works:
The solution involved a series of data preprocessing steps implemented through Python’s data manipulation and machine learning libraries. The process was as follows:

1. Data Preparation:
    a. Data was read using pandas.
    b. Irrelevant columns, such as 'id', were removed as they did not represent features of the data.
    c. Non-numeric data were identified and isolated for transformation.

2. Data Normalization:
    a. Numeric data normalization was accomplished using MinMaxScaler, ensuring all numeric values were scaled uniformly between 0 and 1.
    b. Non-numeric data were converted to numeric formats using OneHotEncoder, allowing for incorporation into the analysis.

3. Pipeline Integration:
    a. A pipeline was constructed using ColumnTransformer to streamline the conversion process from raw data to an analysis-ready format.

4. Correlation Analysis:
    a. The normalized data was subjected to a correlation analysis using pandas’ .corr() function to calculate and visualize the relationships between features.