**Python for Data Analytics Final Project**

My goal for analyzing the data is to find a correlation between a song’s popularity when considering its loudness, danceability, energy, and duration. The dataset I am using for this project comes from a Kaggle site that consists of a wide range of 125 songs of different genres that were taken from a database on Spotify. This database consists of multiple columns of information which include Track Id, Artist, Album Name, Track Name, Popularity, Duration, Explicit, Danceability, Energy, Key, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, Time Signature, Track Genre. But for the sake of my research for this project I am just focusing on the Duration, Energy, Popularity, Loudness, and Danceability. I chose the Spotify dataset because I love music and it is one of my favorite hobbies. According to the findings of the dataset there is a positive correlation between Danceability and Energy. There is also a slight positive correlation between popularity and loudness. Using a histogram, I discovered the range for most amount of popularity occurrences is between 20 and 30. I also found there is a very small positive correlation between popularity and energy. There is a negative correlation between duration and popularity. The Danceability stats compared to the number of occurrences have normal distribution. There is a positive correlation between energy and the number of occurrences. Claremont Colleges had similar conclusions to my research. They found that a songs popularity increases when the energy of a song, loudness, and danceability increases.

The process of my code consisted of first printing the full data frame list. Before researching the statistics, I removed any duplicates of a track id. Following that I printed the Duration of each song column and found its mean, median, and mode. I did the same process for the Duration, Energy, Popularity, Loudness, and Danceability of each of the songs in the data set. In order to visualize my findings, I created a Scatter plot for Danceability vs. Popularity, a scatter plot for Loudness vs. Popularity, a histogram of Popularity vs. the number of occurrences, a scatter plot of Energy vs. Popularity, a scatter plot of Duration vs Popularity, a histogram of Danceability vs number of occurrences, and a histogram of Energy vs number of occurrences. I chose to use both histograms and scatter plots because they were able to accurately display the information given the large sum of data. I do feel my analysis was successful. This is because the research I found online backed up my research findings.

Citations:

*International Music Preferences: An Analysis of the Determinants of Song Popularity on Spotify for the U.S., Norway, Taiwan, Ecuador, and Costa Rica*. (2019). Brendan Joseph Suh. Retrieved September 26, 2023, from https://core.ac.uk/download/pdf/216833222.pdf