

Descriptive Stats:

Each of the functions provides a specific analysis or summary of the data in the "taylor_swift_spotify.csv" dataset. They utilize the pandas, matplotlib, and seaborn libraries for data manipulation and visualization. The functions do not return any values but rather print the results directly.

- **print_album_names()**
 - **Purpose:** Prints out the unique album names from the "album" column of the dataset.
 - **Returns:** A list of album names
- **print_avg_scores()**
 - **Purpose:** Prints the mean (average) of numeric columns in the dataset (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, time_signature, duration_ms).
 - **Returns:** Prints a chart of the avg of the numeric columns
- **print_std()**
 - **Purpose:** Prints the standard deviation of numeric columns in the dataset (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, time_signature, duration_ms).
 - **Returns:** Printing the SD of the numerical values
- **print_max_min()**
 - **Purpose:** Prints the songs with the maximum and minimum values for each numeric score (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, time_signature, duration_ms).
 - **Returns:** Print of the max and minimum value of all the numerical values
- **hist_pop()**
 - **Purpose:** Generates and displays a histogram for the "popularity" column in the dataset.
 - **Returns:** Histogram for popularity
- **hist_dance()**
 - **Purpose:** Generates and displays a histogram for the "danceability" column in the dataset.
 - **Returns:** Histogram for danceability
- **hist_energy()**
 - **Purpose:** Generates and displays a histogram for the "energy" column in the dataset.
 - **Returns:** Histogram for energy
- **hist_loud()**
 - **Purpose:** Generates and displays a histogram for the "loudness" column in the dataset.
 - **Returns:** Histogram for loudness

- **hist_speech()**
 - **Purpose:** Generates and displays a histogram for the "speechiness" column in the dataset.
 - **Returns:** Histogram for speechiness
- **hist_acoustic()**
 - **Purpose:** Generates and displays a histogram for the "acousticness" column in the dataset.
 - **Returns:** Histogram for acousticness
- **hist_liveness()**
 - **Purpose:** Generates and displays a histogram for the "liveness" column in the dataset.
 - **Returns:** Histogram for liveness
- **hist_valence()**
 - **Purpose:** Generates and displays a histogram for the "valence" column in the dataset.
 - **Returns:** Histogram for valence
- **hist_tempo()**
 - **Purpose:** Generates and displays a histogram for the "tempo" column in the dataset.
 - **Returns:** Histogram for tempo
- **hist_duration()**
 - **Purpose:** Generates and displays a histogram for the "duration" column in the dataset.
 - **Returns:** Histogram for duration
- **main()**
 - **Purpose:** Provides the final print of all of the functions
 - **Returns:** All functions' final results

Predictive Stats:

The functions collectively perform predictive analytics, training and evaluating regression models to predict the popularity of Taylor Swift's songs based on audio features. The logistic regression model predicts whether a song's popularity is above or below the median. The results and diagnostic visualizations provide insights into the models' performance and the importance of different features.

- **lin_regression()**
 - **Purpose:** Performs linear regression to predict the popularity of Taylor Swift's songs based on selected audio features. It also calculates and prints the Mean Squared Error and R^2 score for the model. Additionally, it generates a scatter plot of observed vs predicted popularity.
 - **Returns:** Scatter plot observed vs predicted popularity and the mean squared error and r^2 score
- **log_regression()**
 - **Purpose:** Performs logistic regression to predict whether a song's popularity is above or below the median popularity. It conducts a grid search over

hyperparameters to find the best model and prints the best parameters, cross-validation accuracy, and standard deviation of accuracy. It also stores the best model, parameters, and scaled datasets as global variables.

- **Returns:** prints the best parameters, cross-validation accuracy, and standard deviation of accuracy. It also stores the best model, parameters, and scaled datasets as global variables
- **coeff()**
 - **Purpose:** Retrieves and prints the coefficients and feature importance of the logistic regression model. It displays both the raw coefficients and the sorted feature importance based on absolute coefficient values.
 - **Returns:** Coefficients and feature importance chart
- **log_plot()**
 - **Purpose:** Generates a scatter plot comparing predicted probabilities and actual values from the logistic regression model. This helps visualize the performance of the model in predicting the positive class.
 - **Returns:** A scatter plot comparing predicted probabilities to actual values
- **main()**
 - **Purpose:** Executes the main program by calling the **lin_regression()**, **log_regression()**, **coeff()**, and **log_plot()** functions in sequence.
 - **Returns:** Final values of all functions are printed

Diagnostic Analytics:

These functions collectively perform diagnostic analytics on the Taylor Swift dataset, exploring relationships between audio features and popularity and conducting statistical tests to assess the significance of observed differences.

- **corr()**
 - **Purpose:** Generates a correlation matrix and heatmap for selected columns in the dataset, providing insights into the correlation between various audio features and the popularity of Taylor Swift's songs.
 - **Returns:** Heatmap of different numerical values and the degree of correlation to the popularity of Taylor Swift's song
- **t_test_acoustic()**
 - **Purpose:** Performs a t-test to compare the acousticness of songs with high and low popularity. It categorizes songs into high and low popularity based on the median popularity and then conducts a statistical test to evaluate the significance of the difference in acousticness between the two groups.
 - **Returns:** Prints the T-Statistic, P-Value, and a statement indicating whether the difference in acousticness between high and low popularity songs is statistically significant.
- **scatter()**
 - **Purpose:** Generates scatterplots between each independent variable (audio features) and the popularity of Taylor Swift's songs. This provides a visual

representation of the relationship between each audio feature and the popularity score.

- **Returns:** Scatterplot for each independent variable and popularity
- **main()**
 - **Purpose:** Executes the main program by calling the **corr()**, **t_test_acoustic()**, and **scatter()** functions in sequence.
 - **Returns:** All final results