**AAI-500 – Team Project Status Update Form**

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: 08

Team Leader/Representative: \_\_\_\_\_\_Gangadhar Singh Shiva\_\_\_\_\_\_\_\_\_

Full Names of Team Members:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Akshobhya Rao, BV \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Gangadhar Singh Shiva \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Himanshu Kumar \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Title of Your Project: Exploring the Impact of Musical Features on Track Popularity: A Spotify Data Analysis"

Short Description of Your Project and Objectives:

The Spotify dataset provides rich information about various musical tracks, including their popularity, characteristics, and attributes. This data contains features such as **track duration**, **danceability**, **energy**, and **loudness**, which offer valuable insights into how musical elements influence listener behavior and track popularity.  
Objectives:  
**Exploratory Data Analysis (EDA)**:

* + The first goal is to explore the dataset through visualizations and summary statistics to understand the distribution of features like **popularity**, **energy**, and **danceability**. This will help uncover initial patterns and trends in the data.

1. **Linear Regression**:
   * A linear regression model will be built to predict the **popularity** of tracks based on their musical features, such as **tempo**, **energy**, and **danceability**. This will help us understand which features are the strongest predictors of a track's success.
2. **Correlation Analysis**:
   * Correlation analysis will be performed to identify relationships between numerical features, such as **energy** and **danceability**. This will help determine how various features are related to one another and whether they have an impact on a track's popularity.

Name of Your Selected Dataset: **top-spotify-songs-2023**

Description of Your Selected Dataset (data source, number of variables, size of dataset, etc.):

The **Most Streamed Spotify Songs 2023** dataset from Kaggle contains comprehensive information about the most popular tracks on Spotify. This dataset includes **24 features** and covers various aspects of songs, such as their popularity, artist, album, and musical characteristics like energy, tempo, and danceability.

**Dataset Details:**

* **Source**: [Kaggle - Most Streamed Spotify Songs 2024](https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-songs-2024)
* **Size**: The dataset contains **106.27 kB** of data.
* **Number of Features**: The dataset has **24 columns** (features).
* **Number of Records**: The dataset consists of **approximately 500+ records** (songs), depending on the latest updates.

**Key Features:**

* **Track Name**: The title of the song.
* **Artist**: The name of the artist who performed the song.
* **Album**: The album the song is part of.
* **Streams**: The number of times the song has been streamed on Spotify.
* **Danceability**: A measure of how suitable a track is for dancing, ranging from 0 to 1.
* **Energy**: A measure of intensity and activity in the song.
* **Valence**: Describes the mood of the song, with higher values representing positive, happy tracks.
* **Tempo**: The speed of the song, measured in beats per minute (BPM).
* **Duration**: The length of the track in milliseconds.

**Goals of the Analysis:**

1. **Exploratory Data Analysis (EDA)**:
   * We will explore key trends and patterns in the dataset using visualizations and summary statistics. This will help uncover general insights into song characteristics, such as which songs have the highest streams and what makes them popular.
2. **Linear Regression**:
   * A regression model will be used to predict the **number of streams** based on features like **danceability**, **energy**, and **tempo**. This will help understand which musical features contribute most to a track's popularity.
3. **Correlation Analysis**:
   * We will analyze the relationships between numerical features like **energy** and **streams** to see how they are related. This analysis will help us identify which attributes of songs are connected.

Are you using and practicing GitHub as a code hosting platform for version control and collaboration? If yes, provide the link here:

<https://github.com/gshiva1975/AAI-500-Project-Group-8>

How many times have your members met in the last two weeks? 5 Times

List the specific contributions that each team member is providing for the Final Team Project in the table below.

* **NOTE:** ALL students on the team should contribute equally to the Final Team Project.

|  |  |  |
| --- | --- | --- |
| Team Member 1 (Gangadhar Singh Shiva) | Team Member 2 (Akshobhya Rao, BV) | Team Member 3 (if applicable) (Himanshu Kumar) |
| Dataset Brainstorming (University Grade, Spotify Music, Few more from Kaggle like Net Fix Visualization etc.)  Participation in the meetings  Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis  Created GitHub Accounts  Discussion on how to do python coding (to use classes, or not etc.)  Identify the immediate deliverables (P1, P2,P2) and track them.  Perform Analysis and Tabulate the results | Dataset Brainstorming (University Grade, Spotify Music, Few more from Kaggle like Net Fix Visualization etc.)  Participation in the meetings  Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis  Accessed GitHub  Discussion on how to do python coding (to use classes, or not etc.)  Perform Analysis and Tabulate the results | Dataset Brainstorming (University Grade, Spotify Music, Few more from Kaggle like Net Fix Visualization etc.)  Participation in the meetings  Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis  Accessed GitHub  Discussion on how to do python coding (to use classes, or not etc.)  Perform Analysis and Tabulate the results |

Comments/ Roadblocks: Initiative plan is, all team members to start with EDA analysis like cleaning the data, then merge the data, share the workload equally like performing analysis in parallel, merge the analysis, create reports and complete the project in time,