**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: (In alphabetical order)

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 of the Project:**

**Identify Dataset for Analysis**:

- Gather raw data from different sources ( CSV files, web, etc.). Ensure that the data is relevant to the problem..

**Data Preprocessing:**

- **Data preprocessing** is a key step in data analysis and machine learning that involves cleaning, transforming, and organizing raw data to make it suitable for building models. It includes handling missing values, scaling features, encoding categorical variables, and splitting the data into training and test sets to ensure the model performs well.

**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.

**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.

**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.

**Project Submission:**:

- Prepare and Submit artifacts for the project - (Reports, Videos etc)

Name of Your Selected Dataset: **Most Streamed 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**: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023
- **Size**: The dataset contains **106 KB** of data.
- **Number of Features**: The dataset has **24 columns** (features).
- **Number of Records**: The dataset consists of **approximately 950+ 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.

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.)<br><br>Participate in the discussion/meetings<br><br>Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis<br><br>Created GitHub Accounts | Dataset Brainstorming (University Grade, Spotify Music, Few more from Kaggle like Net Fix Visualization etc.)<br><br>Participate in the discussion/ meetings<br><br>Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis<br><br>Accessed GitHub | Dataset Brainstorming (University Grade, Spotify Music, Few more from Kaggle like Net Fix Visualization etc.)<br><br>Participate in the discussion/ meetings<br><br>Discussion on Objectives for the project like EDA, Linear Regression, Correlation Analysis<br><br>Accessed GitHub |

| | | |
|---|---|---|
| Discussion on how to do python coding (to use classes, or not etc.)<br><br>Identify the immediate deliverables (P1, P2,P2) and track them.<br><br>Perform Analysis and Tabulate the results<br><br>Prepare Report, Video and Submission | Discussion on how to do python coding (to use classes, or not etc.)<br><br>Perform Analysis and Tabulate the results<br><br>Prepare Report, Video and Submission | Discussion on how to do python coding (to use classes, or not etc.)<br><br>Perform Analysis and Tabulate the results<br><br>Prepare Report, Video and Submission |

Comments/ Roadblocks: Initial plan is for 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.