**REPORT**

**ON**

**<<SPOTIFY DATA ANALYSIS>>**

**B.E.(IT) 4 -Sem**

BY

**SIDDHARTH(160122737190)**

**MANIDEEP(160122737178)**

**SRIVATSA(160122737183)**

UNDER THE GUIDANCE OF

**<<DR.N. SUDHAKAR YADAV>>**

<<Associate Professor>>,

IT Department, CBIT.



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)**

**(Affiliated to Osmania University; Accredited by NBA and NAAC, ISO 9001:2015 Certified Institution), GANDIPET, HYDERABAD – 500 075**

**Website:** [**www.cbit.ac.in**](http://www.cbit.ac.in)

**2023-2024**

**1. Introduction**

In this report, we present an analysis of Spotify data obtained from internet sources. Spotify is a popular music streaming platform that provides a vast collection of songs and playlists. The objective of this analysis is to gain insights into user preferences, popular genres, and trends within the music industry based on Spotify's extensive dataset.

**2. Data Collection**

The data used for this analysis was sourced from the internet, specifically from platforms that curate and provide Spotify-related datasets. These datasets are typically collected through web scraping techniques or API calls to gather information such as song metadata, user playlists, and listening habits. The data collected includes but is not limited to:

* Song titles and artists
* Album information
* Duration of songs
* Popularity scores
* User-generated playlists
* Audio features (e.g., danceability, tempo, energy)

**3. Installation and Setup**

To conduct this analysis, the following tools and technologies were utilized:

* **Python:** Programming language used for data manipulation and analysis.
* **Jupyter Notebook:** Interactive development environment for running Python code.
* **Pandas:** Library for data manipulation and analysis.
* **Matplotlib and Seaborn:** Libraries for data visualization.

The necessary libraries were installed using Python's package manager, pip, and the Spotify API was accessed using Spotipy by setting up appropriate authentication credentials.

**4. Data Analysis**

The analysis comprised the following key steps:

* **Data Cleaning:** The raw data was cleaned by removing duplicates, handling missing values, and ensuring data consistency.
* **Exploratory Data Analysis (EDA):** This involved examining the data's statistical properties, distribution of features, and identifying patterns or trends. EDA included:
  + Distribution of song durations
  + Popular artists and albums
  + Genre distributions
  + Correlation between different audio features (e.g., danceability vs. energy)
* **Visualization:** Various visualizations such as bar plots, histograms, and scatter plots were created to better understand the data and highlight insights.
* **User Preferences:** Analysis of user-generated playlists to identify popular genres and artists based on user listening habits.
* **Trend Analysis:** Examining how music preferences and genres have evolved over time using available historical data.

**5. Insights and Findings**

Based on the analysis, several insights were derived:

* **Popular Genres:** The dataset revealed that genres like Pop, Rock, and Hip-hop/Rap are among the most popular.
* **Artist Preferences:** Certain artists emerged as highly popular based on the frequency of appearances in playlists and song plays.
* **Audio Feature Correlations:** Strong correlations were observed between certain audio features, which could indicate underlying patterns in music composition.

**6. Conclusion**

In conclusion, the analysis of Spotify data obtained from internet sources provided valuable insights into user preferences and industry trends. The use of Python and associated libraries facilitated efficient data processing, analysis, and visualization. This report demonstrates the potential of data-driven approaches in understanding music consumption patterns and shaping recommendations within streaming platforms like Spotify.