# **Project 2 Description: Spotify Playlist**

**Predicting Spotify Playlist Followers**

## **Challenge Description:**

Spotify is a popular music streaming platform with a vast collection of playlists curated by users and the platform itself. Understanding the factors that influence playlist popularity can provide valuable insights for playlist creators and the platform alike.

## **Objective:**

**Choose one of the objectives:**

1. Playlist Popularity Prediction: Can we develop a model to predict the popularity of playlists based on features such as the number of tracks, number of followers, duration, and number of edits? This could help identify key factors that contribute to playlist success and guide playlist curation strategies.
2. Playlist Content Analysis: Can we analyze the content of playlists to identify popular tracks, albums, and artists across different genres or time periods? This could provide insights into emerging trends, user preferences, and opportunities for content promotion or partnership.
3. User Engagement Prediction: Can we develop a model to predict user engagement metrics, such as the number of followers or edits, based on playlist features and user behavior? This could assist in targeting marketing efforts, identifying high-value users, and optimizing playlist promotion strategies.
4. User Segmentation based on Playlist Preferences: Can we segment users based on their playlist preferences and behavior, such as frequent collaborators, genre preferences, or playlist creation frequency? This could enable targeted marketing campaigns, personalized recommendations, and tailored user experiences to different user segments.

## **Dataset:**

Business data analysts will be provided with a dataset containing information about Spotify playlists, including attributes such as:

* Playlist name
* Number of followers
* Total tracks in the playlist
* Number of albums in the playlist
* Duration of the first track
* Other relevant metadata about the tracks, albums, and artists included in the playlists.

## 

## **Tasks:**

**Data Exploration:**

* Explore the provided dataset to understand its structure, features, and distributions.
* Identify any potential outliers or inconsistencies in the data.
* Visualize the distribution of playlist followers and explore relationship between features and the target variable.

**Model Development:**

* Implement multiple machine learning models to predict the number of followers for Spotify playlists.
* Experiment with different features and combinations of features to train the models.
* Consider using linear regression, decision trees, k-nearest neighbors, support vector machines, and neural networks for modeling.
* Evaluate the performance of each model using appropriate metrics such as mean squared error (MSE).

**Model Evaluation and Comparison:**

* Compare the performance of different models based on their MSE values.
* Identify the strengths and weaknesses of each model in predicting playlist followers.
* Discuss insights gained from the modeling process and the significance of various features in predicting playlist popularity.

**Deliverables:**

Business data analysts are expected to submit the following deliverables:

* **Jupyter notebook or Python script**
  + detailing the data preprocessing, model development, and evaluation steps.
* **Presentation Slides(5 mins presentation & 5 mins Q&A)** 
  + Visualizations and insights derived from the data exploration phase.
  + Summary of model performance and comparison results.
  + Recommendations for playlist creators or Spotify administrators based on the findings.

## **Judging Criteria**

**Innovation in Algorithm Design (30%):**

* Creativity: The degree of originality in leveraging the dataset and designing algorithms for predicting Spotify playlist followership.
* Complexity: The sophistication and technical depth of the algorithms, showcasing innovative machine learning techniques and data processing strategies.
* Scalability: The algorithm's ability to maintain performance and accuracy when applied to large-scale datasets, considering the vast size of Spotify's data.

**Accuracy and Performance (25%):**

* Prediction Accuracy: The precision and reliability of the predictive models in estimating the number of followers for Spotify playlists.
* Efficiency: The efficiency of the algorithms in processing data and delivering accurate predictions, particularly in scenarios requiring real-time application.

**Business Impact Potential (20%):**

* User Engagement: The potential of the predictive models to enhance user engagement metrics on the Spotify platform, such as interaction rates and user satisfaction.
* Revenue Opportunities: The direct or indirect potential for the models to increase revenue through improved ad targeting, premium subscription upsells, or enhanced user retention strategies.

**Market Viability (15%):**

* Applicability: The feasibility of integrating the proposed solutions into existing music streaming platforms, considering technological constraints and market dynamics.
* User Experience Enhancement: The impact of the models on enhancing the overall user experience, including ease of playlist discovery and personalization.

**Presentation and Documentation (10%):**

* Clarity**:** The clarity and conciseness of the presentation, methodologies, and results, ensuring that the approach is easy to understand and follow.
* Comprehensiveness: The depth and completeness of the documentation, covering algorithm design, data analysis, evaluation metrics, and potential business applications.
* Reproducibility: The availability of code or detailed methodologies enabling the replication and verification of results, enhancing the transparency and trustworthiness of the findings.

## **Submission Format**

**Technical Report + Result Summary**

Format: PDF

Length: 1-5 pages

Content: Overview of approach and key metrics. & Detailed description of project components.

**Code Repository**

Hosting: GitHub or GitLab

Content: Organized codebase with dependencies and licensing information.

**Presentation Slides**

Format: PDF or PowerPoint

Length: 10-15 slides

Content: Concise presentation of project overview, methodology, results, and conclusions.