**Major Data Analysis**

#### Clustering Users

how user segmentation (clustering) was carried out to comprehend user behaviour patterns.

Gaussian Mixture Model (GMM) clustering:

* For detecting the various behavioural groups of users, we employed the Gaussian Mixture Model (GMM) clustering algorithm.
* Total listening time, average listening time, number of sessions, and active duration were taken as inputs for clustering.
* These features were normalized to have similar scales.
* GMM algorithm was configured to 5 clusters segment users by activity and engagement patterns. - They were assigned to one of the 5 clusters after clustering, which gave us an understanding of their behaviour (e.g., low activity, high activity, frequent churners).

Churn Distribution Across Clusters

Here, observes the distribution of churn in the clusters discovered to see if user behaviour is associated with churn.

Churn Rate by Cluster:

- By examining churn by clusters, we discovered that there is more churn for certain clusters, i.e., that users within these clusters are more prone to cancel subscriptions.

- Analysis assists in determining which segments of users need more focused retention strategies.

- A stacked bar plot graphically depicted the proportion of churned and retained customers in each cluster.

Feature Engineering and Model Preparation Detail data preparation for modeling, feature engineering, and the creation of the final dataset that was used for classification.

Feature Selection:

* After clustering, the model's final dataset contained features such as total listen time, average listen time, sessions, active duration, and user clusters.
* The target variable, churn, was also present in this dataset.

Data Splitting:

* The data was split into training (70%) and testing (30%) sets in a way that the data regarding churn distribution is balanced.

Preprocessing:

* Missing values were filled with column-wise means (for numerical features).

- Standard scaling was applied to the numerical features to prepare them for the machine learning models.