IEEE NSUT

TRACKS ANALYSIS OF SPOTIFY

Anam Shoeb	2024UCA1864
Mayank Kumar Jha	2024UIC3545
Gulshan Prasad	2024UIC3573

AIM OF THE PROJECT:

To perform an Analysis with EDA and feature engineering with graphic visualization using matplotlib on tracks.csv, a file containing data about components of tracks of Spotify.

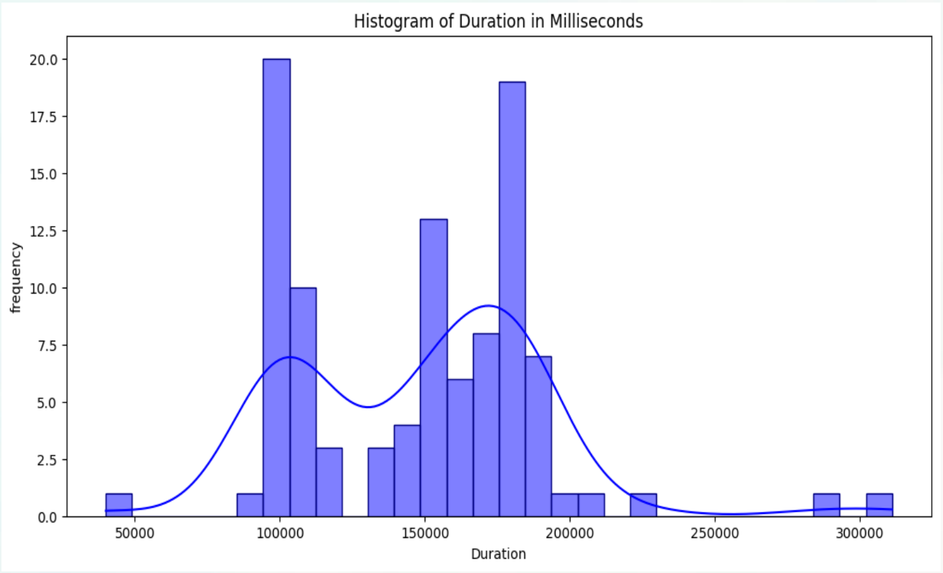
Provided "tracks.csv", a file containing data about components of tracks of Spotify, contains columns:

id, name, popularity, duration_ms, explicit, artists, id_artists, release_date, danceability, energy, key, loudness, mode, speechiness. acousticness, instrumentalness, liveness. Valence, tempo, time_signature

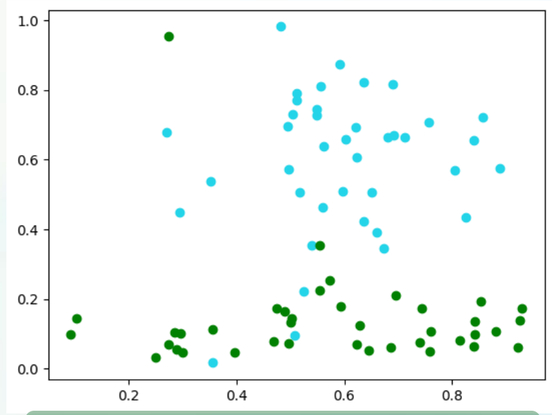
Working on:

- Creating new columns
- Visual representation of comparison between columns,
- Training model
- Numeric Calculations/Operations

RESULTS:



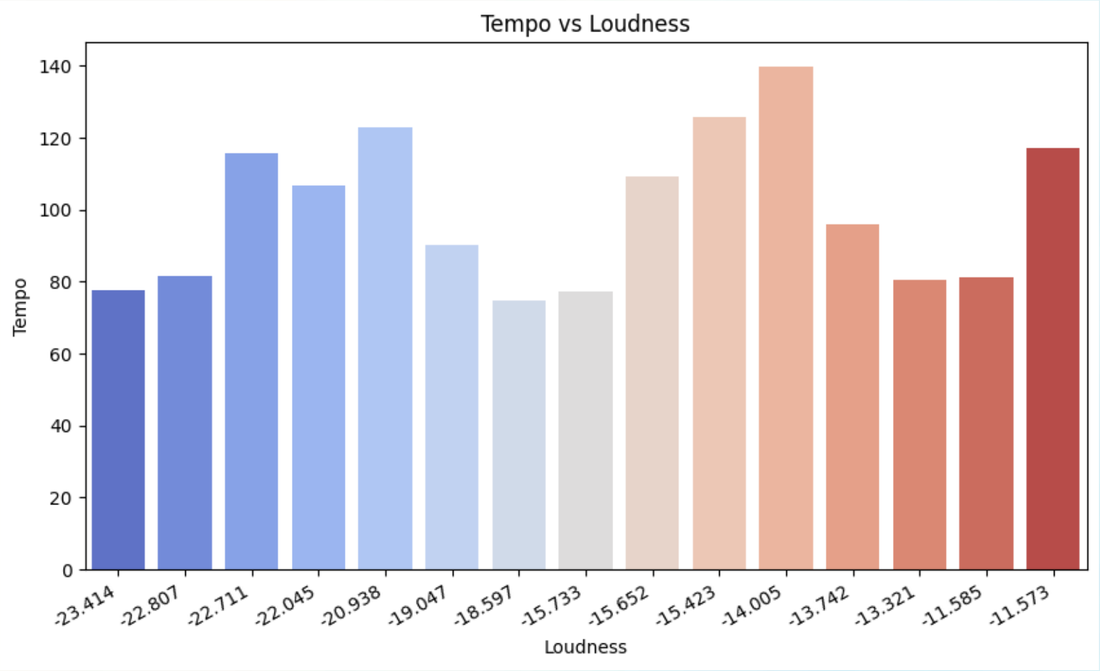
This Histogram represents the distribution of music tracks durations in milliseconds. Tracks around 100,000 ms (1 minute 40 seconds) and 200,000 ms (3minutes 20 seconds) have the highest frequency, indicating listeners prefer songs in the 2-3 minute range. [Frequency is the quantity/count of no. of tracks in the duration]



Danceability vs Energy

Valence vs Liveness

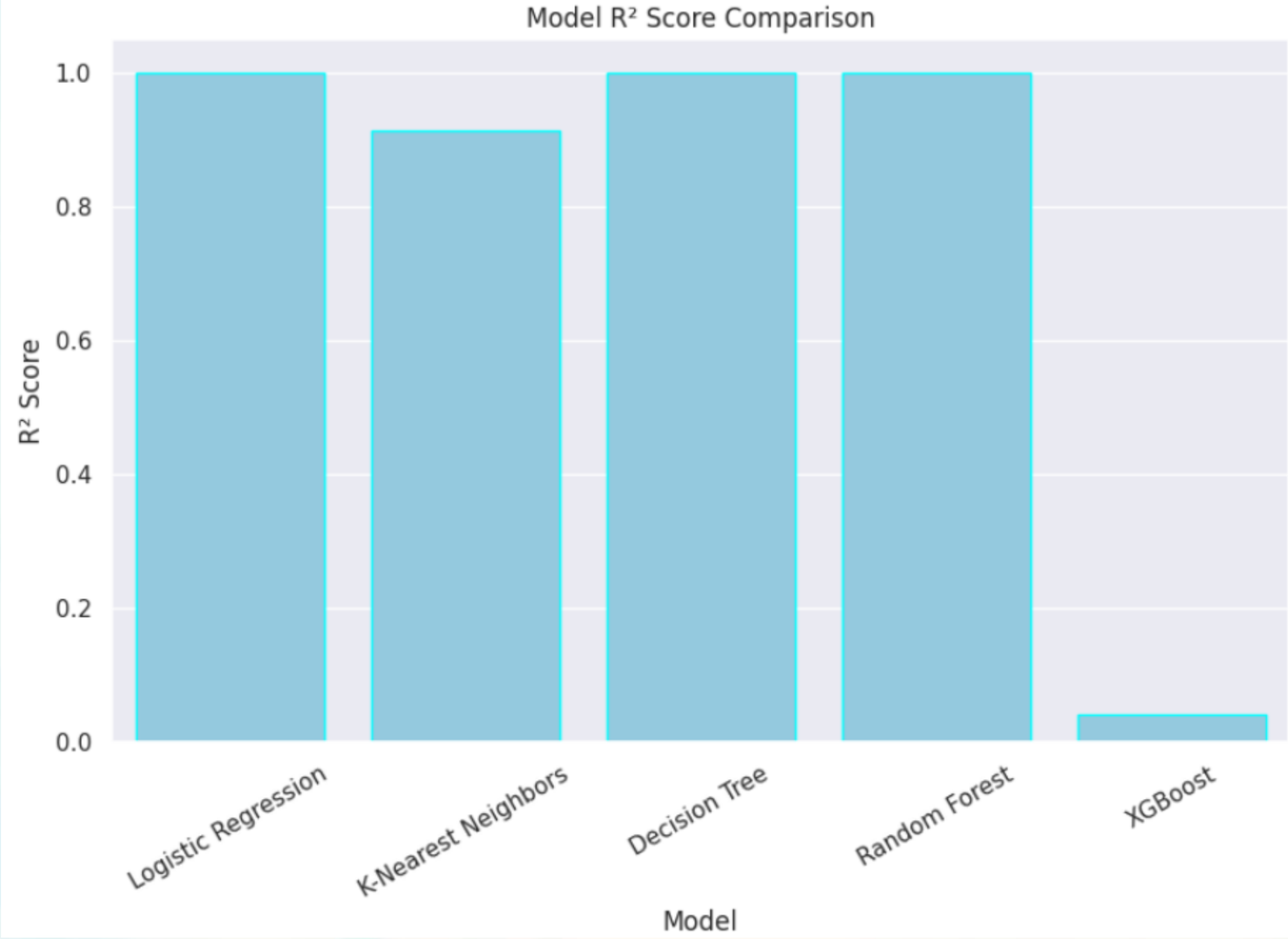
This Bar Graph represents how the pace of music is related with loudness. Loudness around -14Db has >120 pace that concludes people are more likely to listen Phonk, Rock, EDM type tracks. And Loudness around -23Db has <80 pace that concludes people are less likely to listen Lo-Fi and Ambient type tracks.
[Music tracks are quitter than 0Db (reference level) so they're measured in -ve]



RESULTS:

Evaluating and comparing the performance of multiple machine learning models on a given dataset.

And the comparing them visually representing R-squared scores as below:



	MSE	RMSE	R2
Logistic Regression	1.000000	1.000000	1.000000
K-Nearest Neighbors	0.913231	0.913231	0.913231
Decision Tree	1.000000	1.000000	1.000000
Random Forest	1.000000	1.000000	1.000000
XGBoost	0.473253	0.687934	0.040879

The best-performing regression model based on the highest R-squared (R^2) score is **'Logistic Regression'.**

```
Mean Squared Error: 0.4732532060466591
Root Mean Squared Error: 0.687934012857817
Mean Absolute Error: 0.626053027389554
R² Score: 0.040879218307066756
```

IEEE NSUT

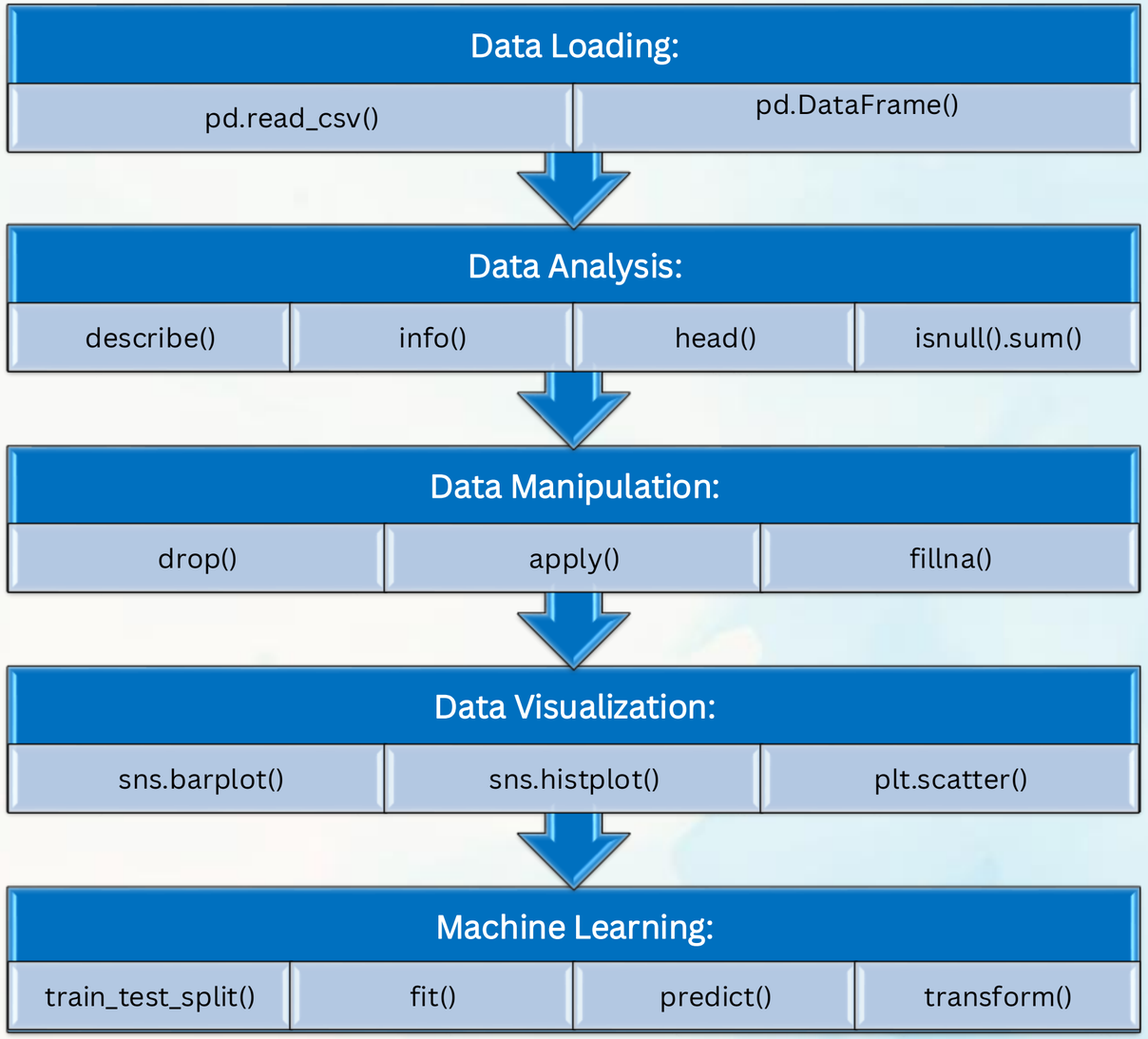
TECH STACK

Language/Computing environment

- Python
- Colab Notebook
- Google Drive

Libraries :

- Pandas
- Matplotlib.pyplot
- Sklearn
  - LabelEncoder
  - OneHotEncoder
  - train_test_split
  - LinerRegression
  - StandardScaler
  - LogisticRegression
  - KNeighborsClassifier
  - DecisionTreeClassifier
  - RandomForestClassifier
  - accuracy_score
- Numpy
- Xgboost
- Seaborn



IEEE NSUT

FUTURE SCOPE

The future scopes of this project involve expanding into more sophisticated machine learning applications, integrating with emerging technologies, and enhancing user experience through personalized and interactive music systems.

- Popularity Prediction: Use machine learning models to predict song popularity based on audio features, helping artists optimize their compositions.
- Audio Features: The dataset includes features extracted from audio signals, such as danceability, energy, loudness, and valence. These features are crucial for understanding the musical content and predicting user preferences
- Mood-Based Playlists: Create playlists for relaxation, workouts, or parties by analyzing energy and valence levels.
- Genre Classification: Classify songs into genres using audio features like tempo, acousticness, and instrumentalness.
- Music Recommendation Systems: Develop personalized playlists based on user preferences for features like danceability, energy, and valence. Also improve algorithms for suggesting songs tailored to specific moods or activities.