# 19ZO02-Social and Economic Network Analysis

Project Report

**Topic: Genre and Popularity Analysis of Songs**

**Team Members:**

19Z226 M.MANOJKUMAR

19Z228 NANDHA KISHORE. V

19Z234 RAHUL RAJ. D

19Z236 RASWANTH. E. A

19Z242 SANJAI S

# BACHELOR OF ENGINEERING

**Branch: COMPUTER SCIENCE AND ENGINEERING**

Of Anna University



PSG College of Technology Coimbatore – 641004

# Problem Statement:

Objective is to analyze genre, artists and other attributes based on popularity and predict the popularity using other attributes in dataset and genre using Artists. For this we make use of dataset from Kaggle for representing the song name, artists, genre, popularity and other attributes related to songs as a graph with

* + - Artists and Genre serve as Nodes
    - Songs serves as Edges.
    - The Edges are weighed according to the popularity of genre and artists of songs.

# Dataset Description:

* The dataset is comprised of song names, genre, artists and popularity of the songs.
* This dataset was extracted from the Kaggle database.
* Dataset Attributes:
  + - index - Row ID
    - Title – song name
    - Artists
    - Genre – Type of song
    - Popularity
    - Danceability, valence, loudness, liveness, energy, acoustic, length,
* Dataset: <https://www.kaggle.com/datasets/iamsumat/spotify-top-2000s-mega-dataset>

1. Tools **used:**

* **Python:** We have used the Python Language for the coding part because of its User-friendly Data Structures.
* **Google Collab**: Google Collab is particularly well suited to machine learning, data analysis, and education since it enables anyone to develop and run arbitrary Python code through the internet. Python code may be written and run through a browser using Google Collab.
* **Packages used:** matplotlib, sklearn, pandas, cborn, networkx.

**4.Challenges Faced:**

* There was some error in the code, so we were unable to visualize the graph Initially.
* Even though our code was debugged and ran, the expected output in terms of degree and centrality were all 0.
* Since we were new to Network X, it was tiring to understand and visualize the graphs.

# 5.Contribution:

|  |  |
| --- | --- |
| **Name (Roll number)** | **Contribution** |
| M.Manojkumar(19Z226) | Graph Visualization, Analysis and coding |
| Nandha Kishore V (19Z228) | Collecting dataset and Preprocessing |
| Rahul Raj D (19Z226) | Documentation and Statistical analysis |
| Raswanth E A(19Z236) | Graph Visualization |
| Sanjai S (19Z242) | Dataset Analysis |

**Annexure – I:**

<https://github.com/MManojkumar16/SENA-PROJECT>

**Annexure- II:**

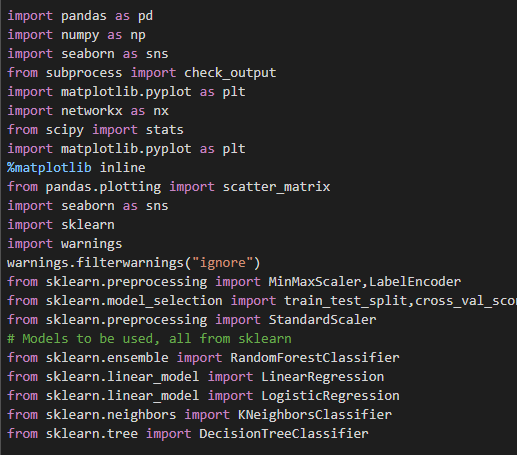
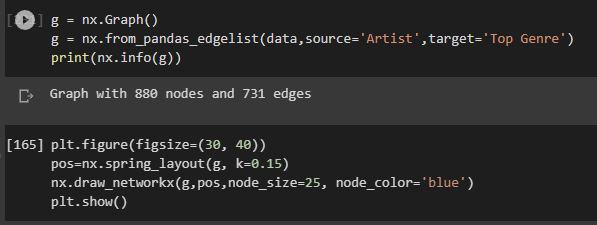




Figure 1: Loading the Libraries and describing the dataset information



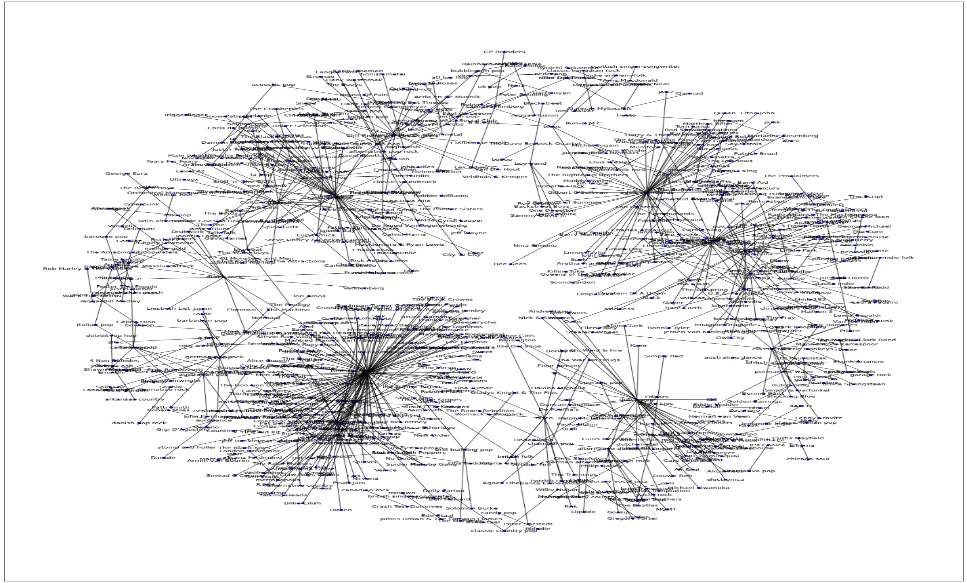


Figure 2: Shows how Artists and Genre are related by using Network Graph.

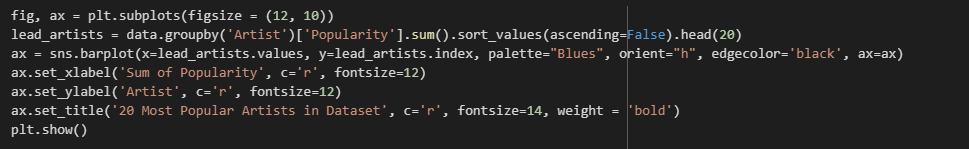
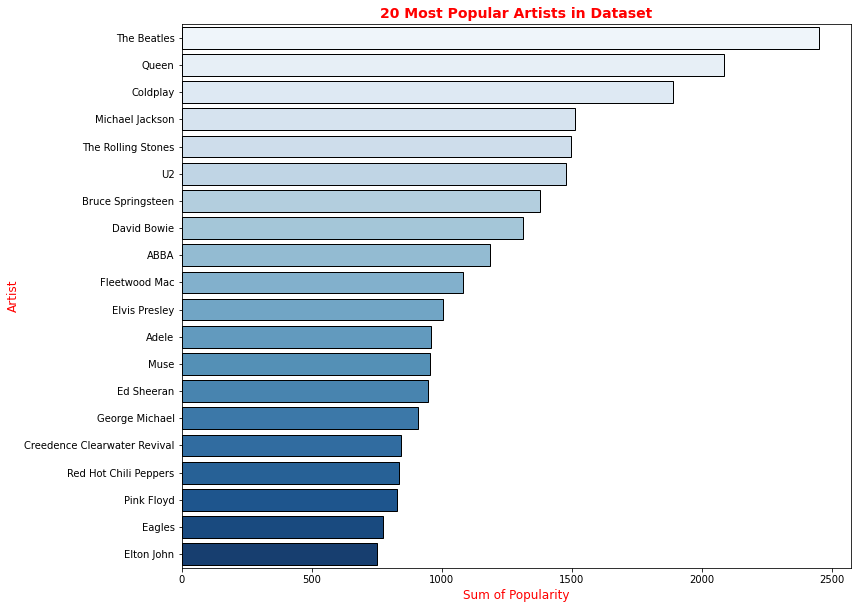
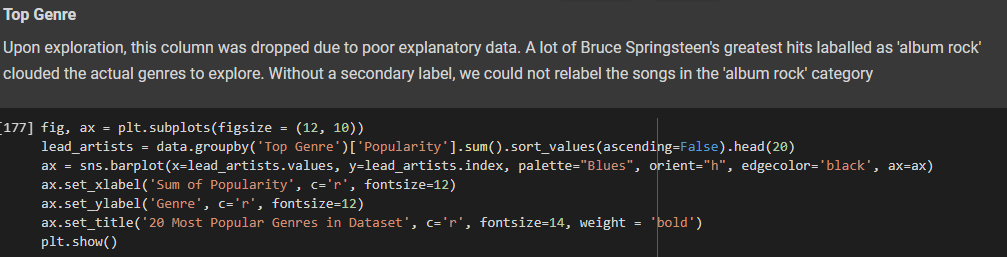
 

Figure 3: Shows 20 Most Popular Artists in the dataset.



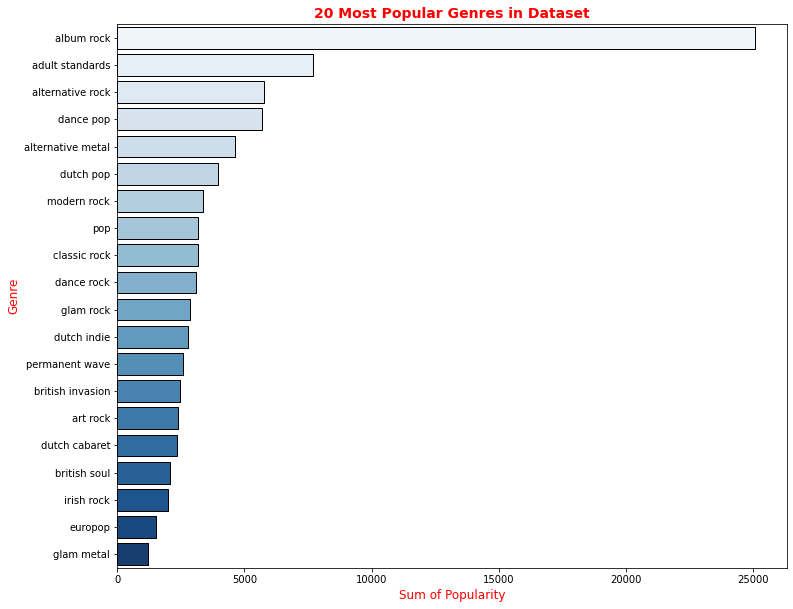


Figure 4: Shows 20 Most Popular Genres in the Dataset.

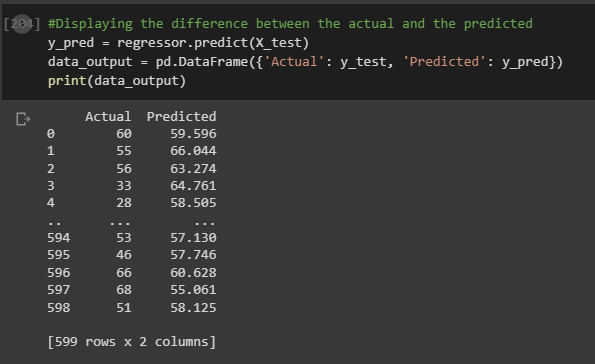
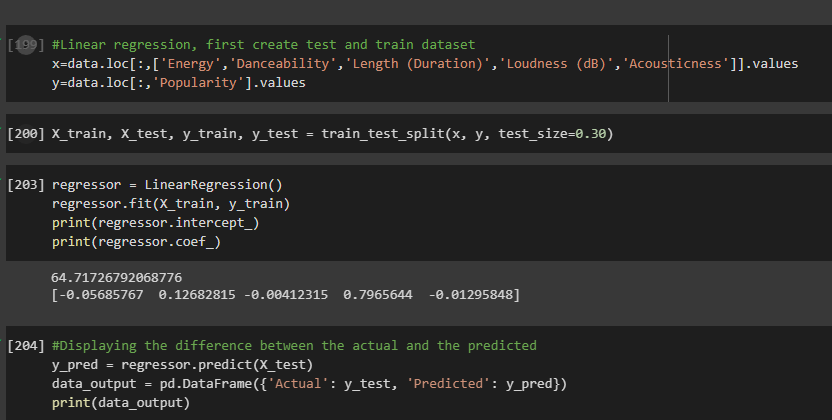
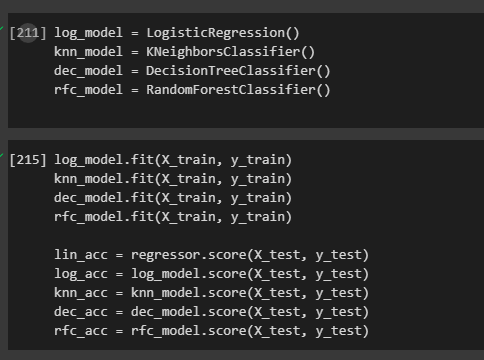
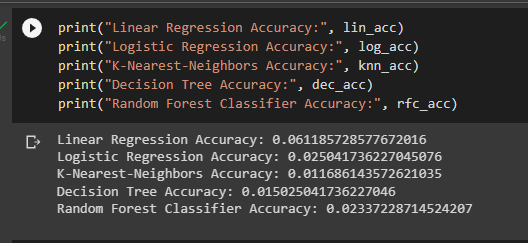
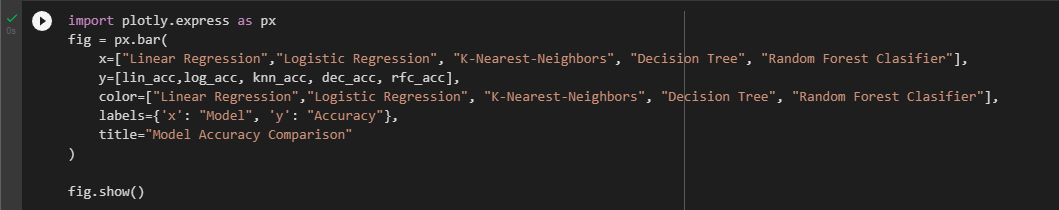
****

Figure 5: Shows the Prediction of Popularity using Linear Regression Model.

** **

****

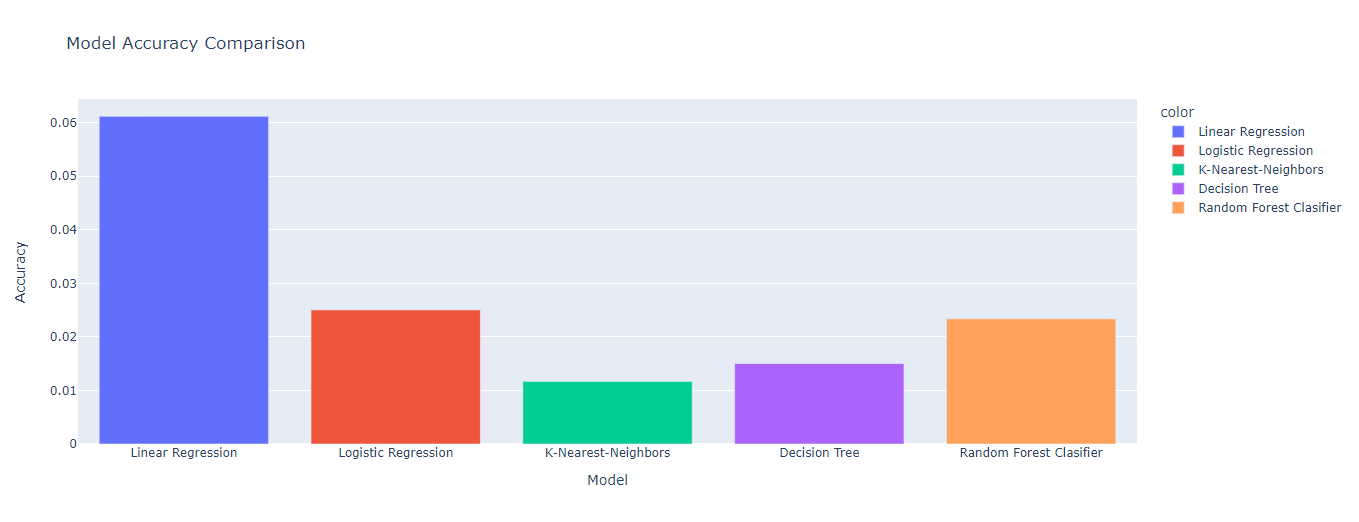
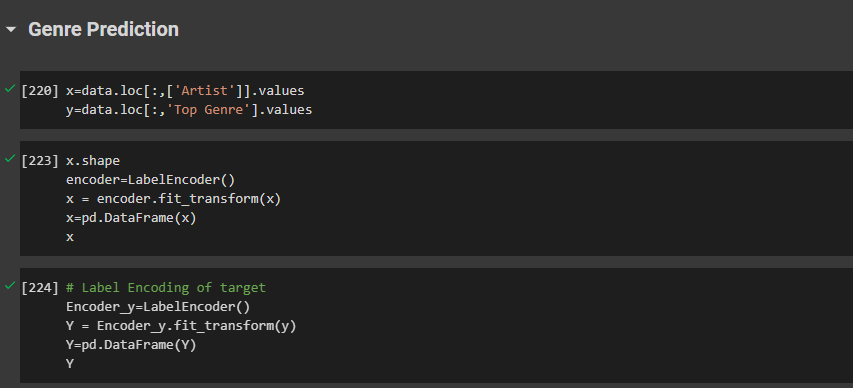
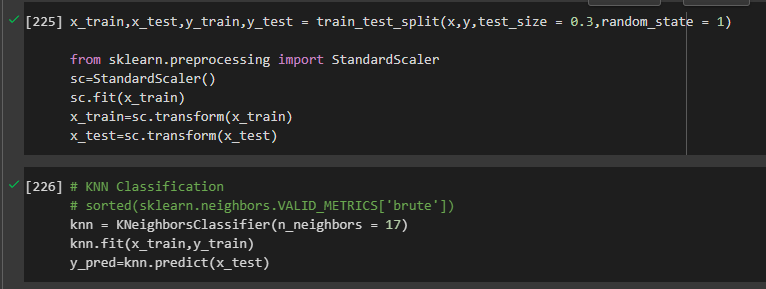


Figure 6: Shows the Predictive accuracy of popularity using all other modules and displaying the same.

**** ****

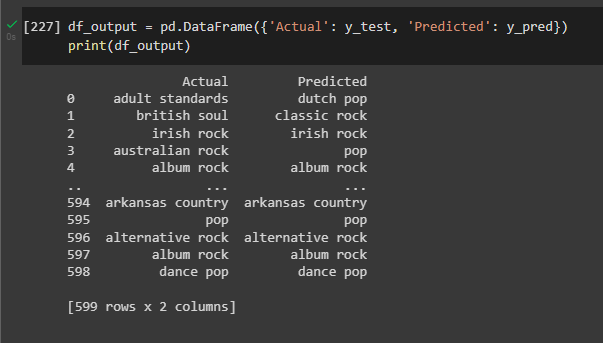
****

Figure 7: Shows the Genre Prediction based on Artists using KNN Classification Model.

# References:

1. Kaggle Dataset — <https://www.kaggle.com/datasets/iamsumat/spotify-top-2000s-mega-dataset>
2. NetworkX —https://networkx.org/documentation/stable/reference/algorithms/index.html
3. Pandas- <https://pandas.pydata.org/>
4. <https://medium.com/web-mining-is688-spring-2021/network-of-genre-and-artists-in-spotify-98c896569dee>
5. <https://arxiv.org/abs/1706.01953>
6. <https://ieeexplore.ieee.org/document/9675998>
7. <https://dl.acm.org/doi/10.1145/3474085.3475495>