A Project Report On

**ANALYSIS ON SPOTIFY DATA**



**Kishan Vaghasiya (92300584098)**

**Academic Year 2023 – 2024**

|  |
| --- |
| Internal Guide |
| **Prof. Rushika Patt** |

MarwadiUniversity

Rajkot-Morbi Road, At & m PO: Gauridad, Rajkot 360 003. Gujarat. India.

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapters** | **Particulars** | **Page No.** |
| **1** | **INTRODUCTION** | 3 |
| **2** | **PROBLEM STATEMENTS** | 4 |
| **3** | **ORGANIZATION OF DATA** | 4 |
| **4** | **DATA VISUALIZATION THROUGH SUMMARY** | 5 |
| **5** | **METHODS/ TECHNIQUES** | 6 |
| **6** | **RESULT OF ANALYSIS USING GRAPHS/CHARTS** | 7 |
| **7** | **CONCLUSIONS** | 17 |
| **8** | **FUTURE ANALYSIS** | 18 |
| **9** | **APPENDIX** | 18 |

**1.INTRODUCTION**

**Data analysis** is defined as a process of cleaning, transforming, and modelling data to discover useful information for decision-making. The purpose of data analysis is to extract useful information from dataset and taking the decision based upon the data analysis.

The Spotify dataset offers a rich source of information about songs, albums, artists, and user interactions within the Spotify platform. It includes attributes such as song features, popularity metrics, artist details, and user behavior data. Analyzing this dataset provides insights into music trends, user preferences, and engagement patterns, facilitating strategic decisions in the music industry and beyond.

Data analysis is the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. In today's data-driven world, data analysis plays a crucial role across various fields including business, science, finance, healthcare, marketing, and more.

A Spotify dataset for data analysis can offer a wealth of insights into music preferences, trends, and user behavior within the Spotify platform. This dataset typically includes various attributes related to songs, albums, artists, and user interactions.

**2.PROBLEM STATEMENTS**

* **Highest-Lowest Analysis:**

1. Determine the highest and lowest values for various audio features

2. Highest and lowest streams of songs

3. Highest and lowest Count of Artist

4. Highest and lowest total song during 2020-2023

* **Summary Analysis:**

7. Total song release during 2022-2023.

8. Total streams during 2022-2013.

**3.ORGANIZATION OF DATA**

This dataset **“Analysis On Spotify Dataset 2010-2024”** represents the data of songs online or offline in many countries over the world.

There are 500 rows and 10 columns which are as follow track name: artist(s)\_name,artist\_count,released\_year,in\_spotify\_playlists,in\_spotify\_charts,streams,in\_apple\_playlists,in\_apple\_charts,mode

|  |  |  |
| --- | --- | --- |
| **Raw Data Counts** | Rows | 500 |
| Columns | 10 |
| **Filtered Data Counts** | Rows | 0 |
| Columns | 0 |
| **Net Data Counts Available** | Rows | 500 |
| Columns | 10 |

This Data describe total no. of 500 songs, which have been release in online or offline in different countries over the world in between years 2010 to 2024.

**4. DATA VISUALIZATION THROUGH SUMMARY**

* **From where i have collected this sales dataset?**

I have collected songs dataset from this website: https://www.kaggle.com/datasets

* **How many rows and columns are there in dataset?**

This sales dataset consists of 500 rows and 10 columns.

**5. METHODS/ TECHNIQUES**

**5.1 Identification of Techniques and Tools Used:**

Python is a versatile and high-level programming language known for its simplicity and readability. Created by Guido van Rossum and first released in 1991, Python has since become one of the most popular programming languages in the world, used extensively in various fields such as web development, data science, artificial intelligence, scientific computing, and more.

One of Python's key strengths is its elegant syntax, which emphasizes readability and encourages clear, concise code. This makes it an excellent choice for both beginners and experienced programmers alike. Python's design philosophy, often referred to as "The Zen of Python," emphasizes simplicity, clarity, and practicality.

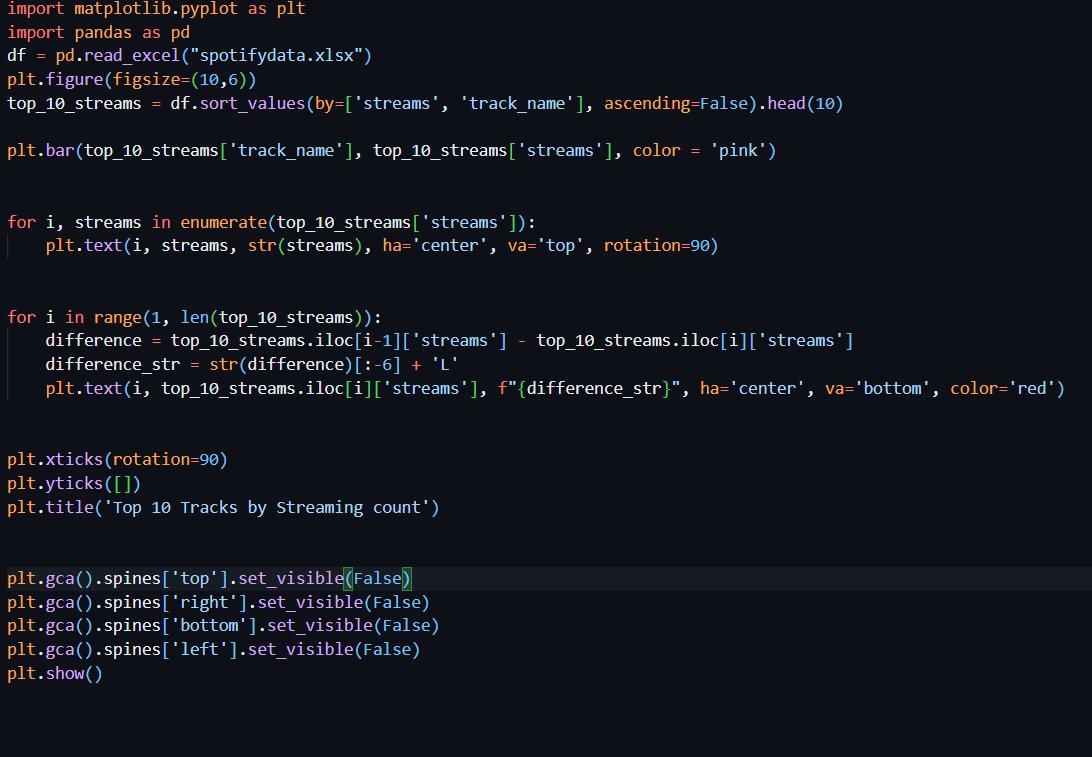
Python is an interpreted language, which means that code can be executed directly without the need for compilation, making the development process faster and more iterative. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming, giving developers flexibility in how they approach problem-solving.

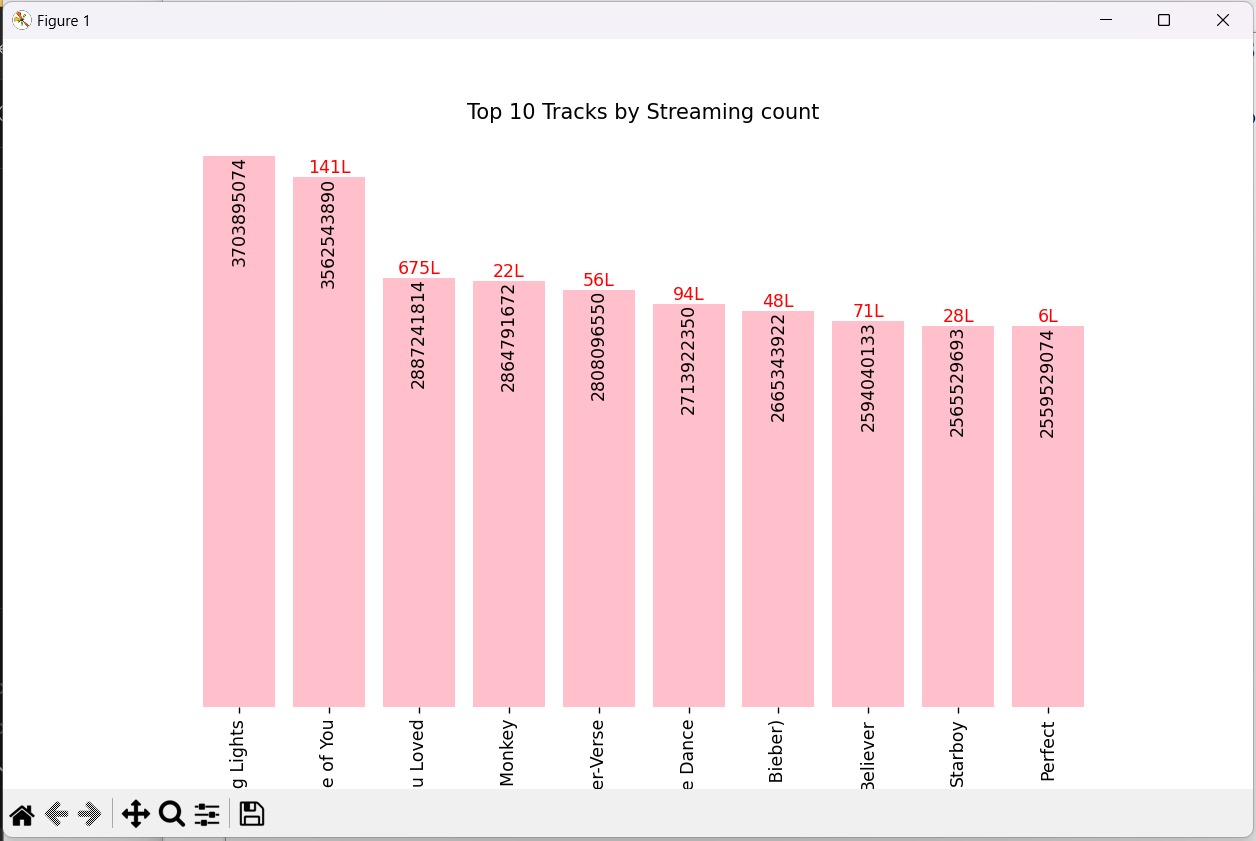
Python's extensive standard library provides a wide range of modules and packages for tasks such as file I/O, networking, database access, and more, allowing developers to accomplish complex tasks with minimal effort. Additionally, Python's vibrant community has created a vast ecosystem of third-party libraries and frameworks, further expanding its capabilities and making it a powerful tool for building applications of all kinds.

Overall, Python's simplicity, versatility, and strong community support have contributed to its widespread adoption and enduring popularity among developers worldwide.

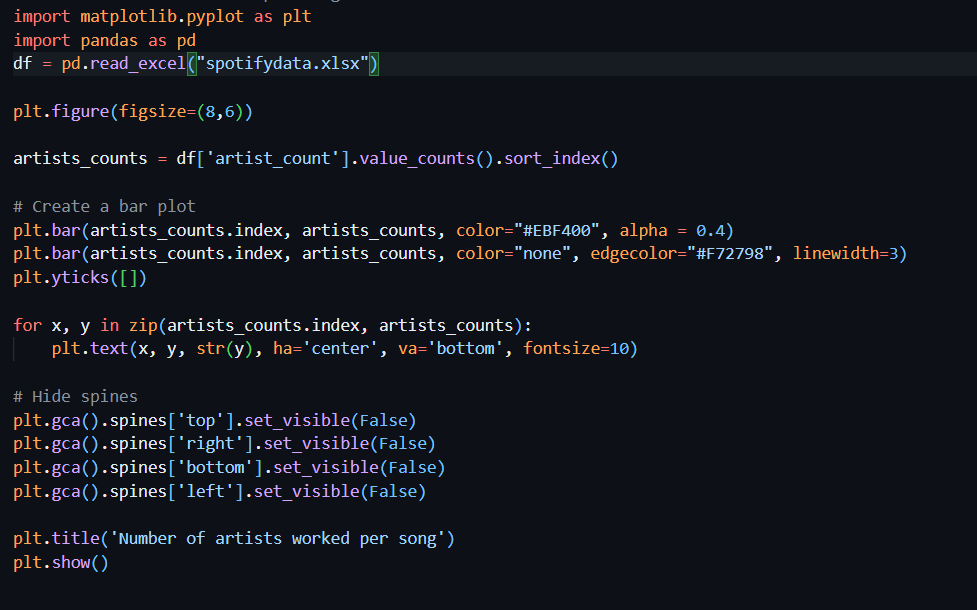
**RESULT OF ANALYSIS USING GRAPHS/CHARTS**

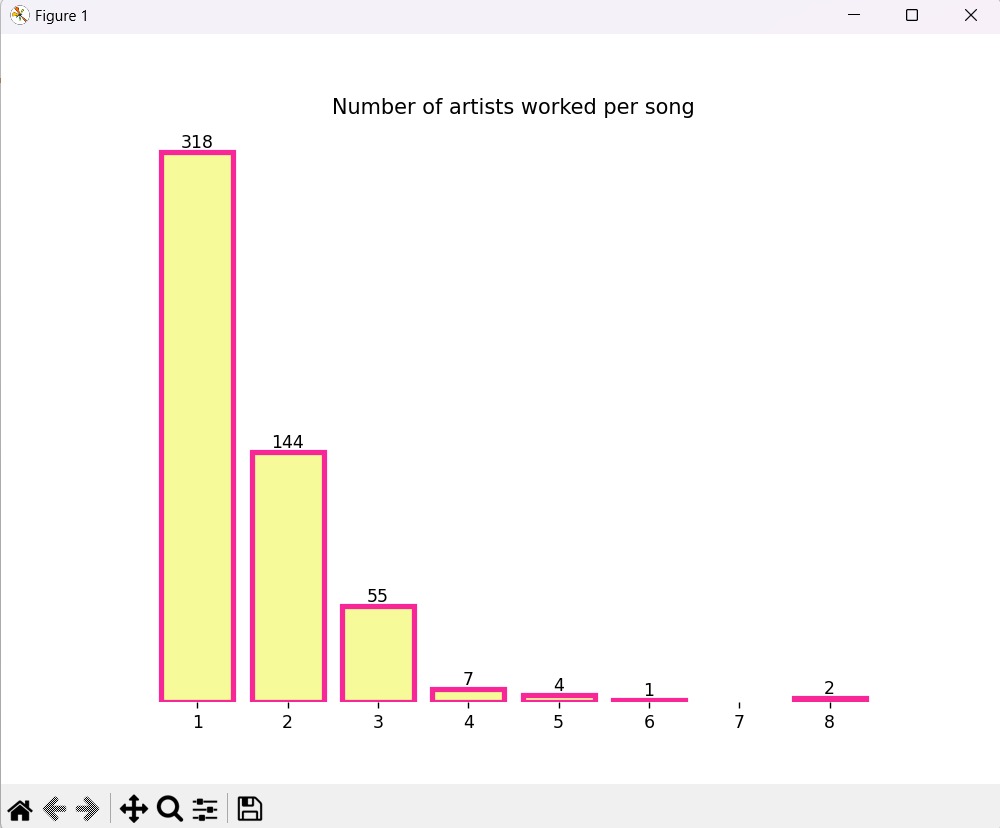
* **Top 10 songs by streaming count**



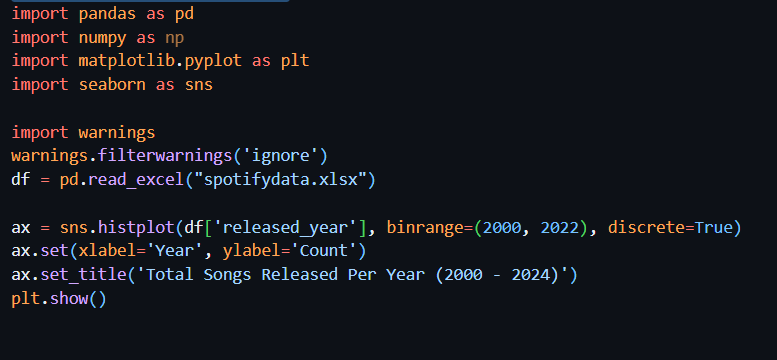


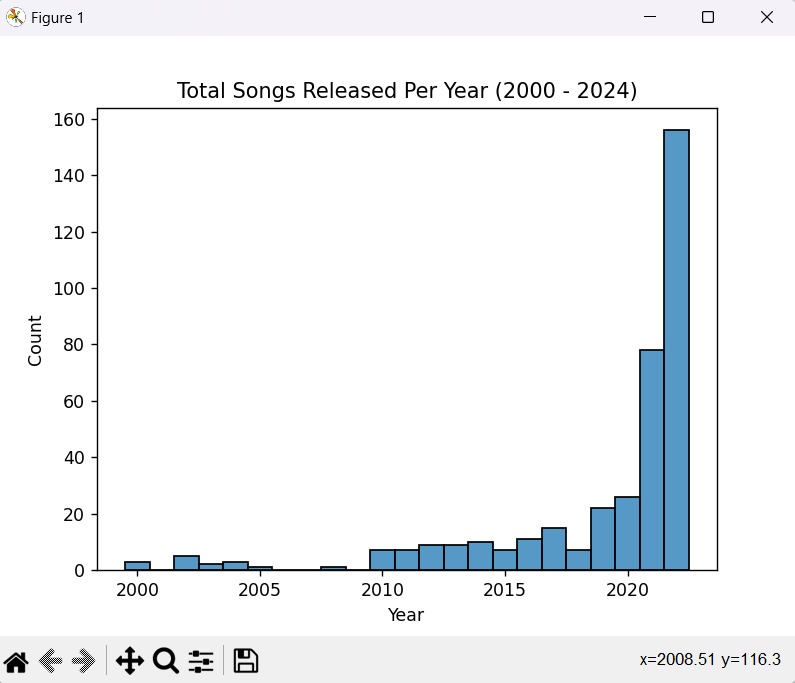
* **Numbers of artist worked per song**

****

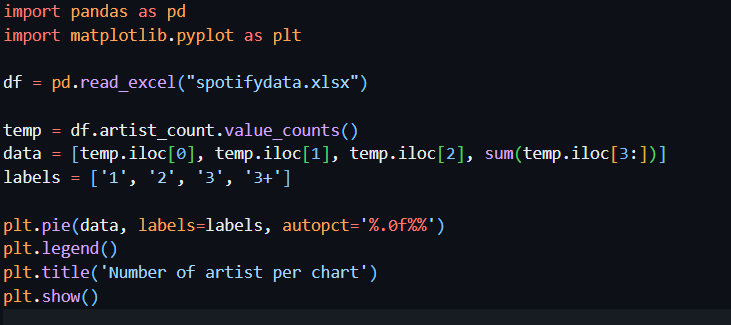


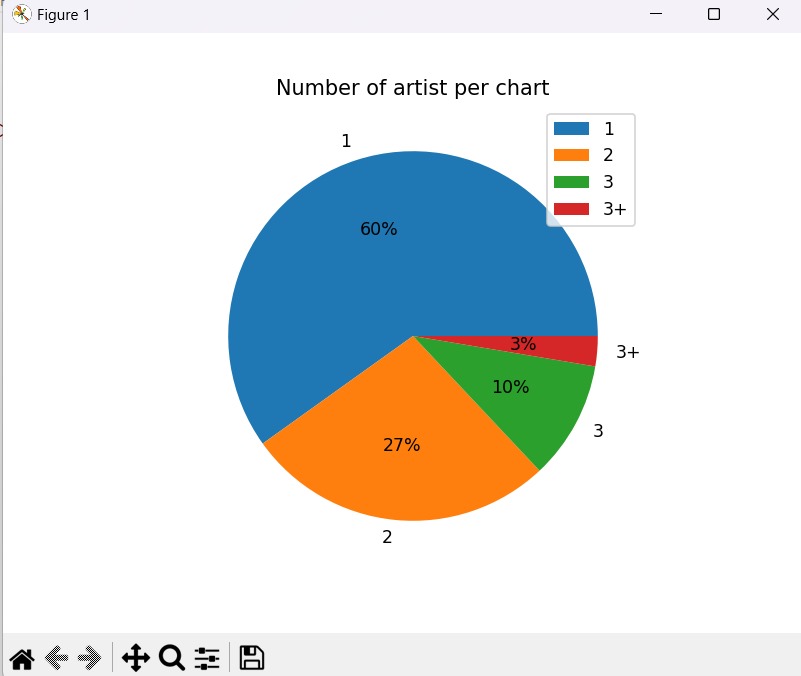
* **Total song released per year**



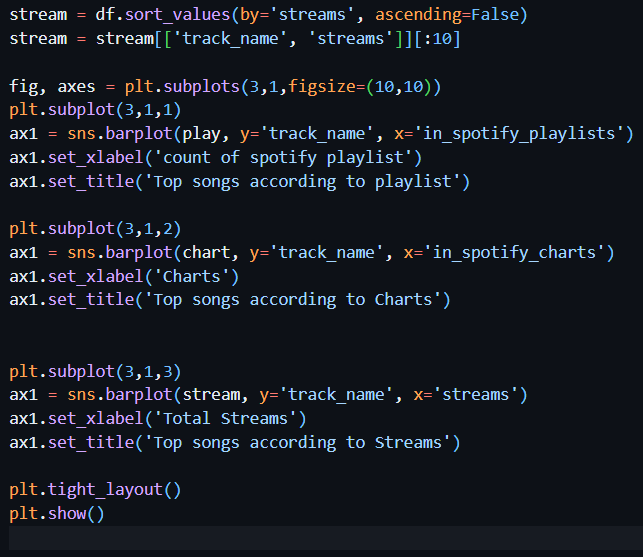


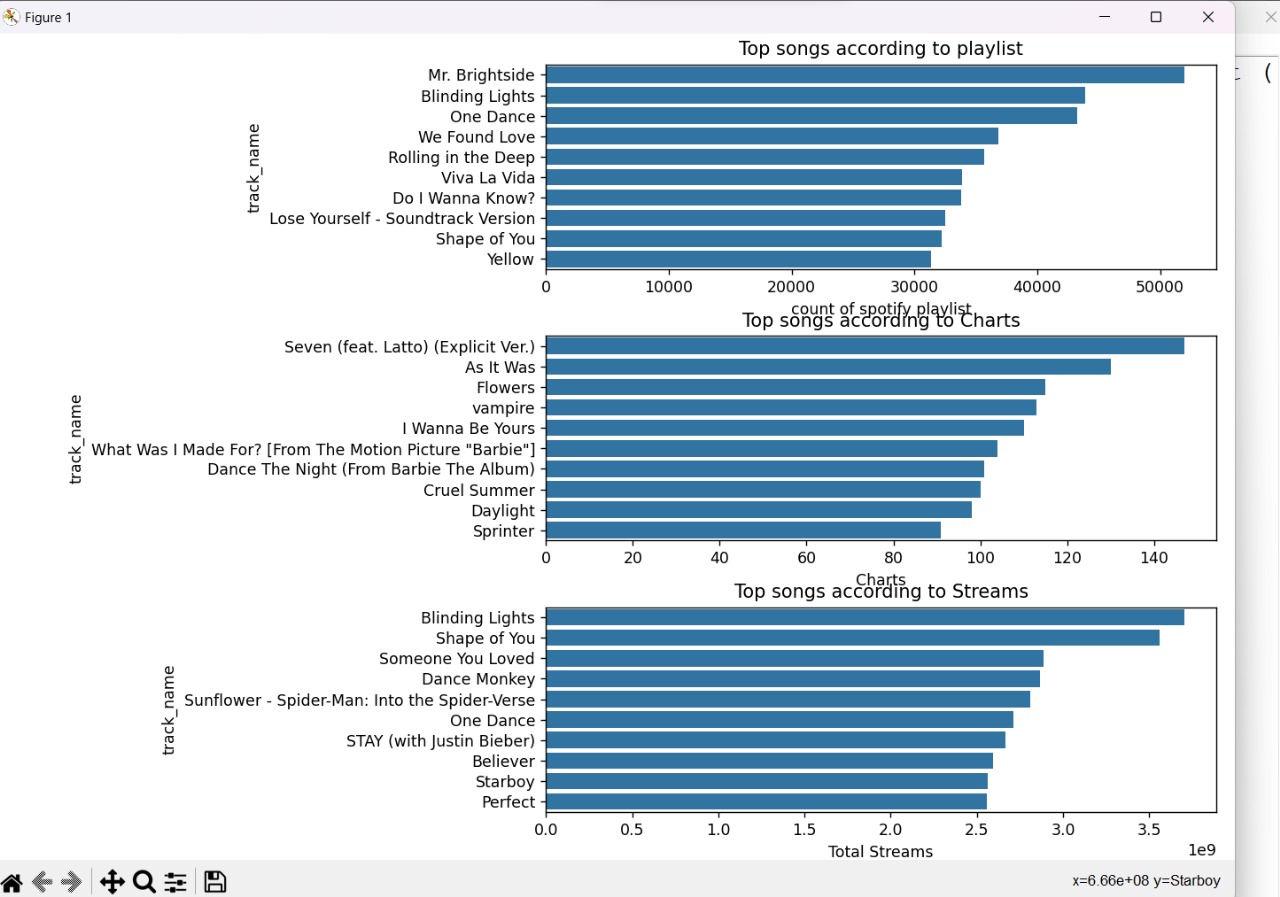
* **Number of artist per chart**

****



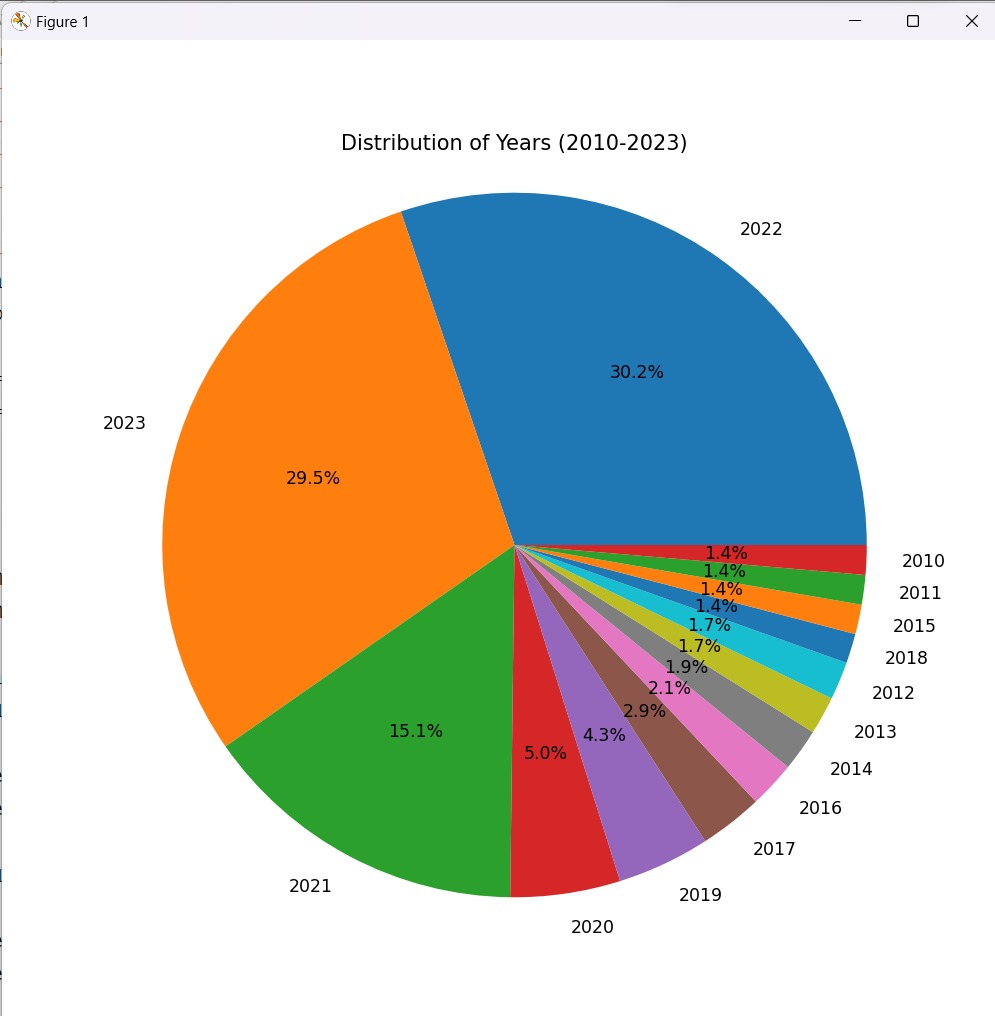
* **Top songs according to Charts and Streams**



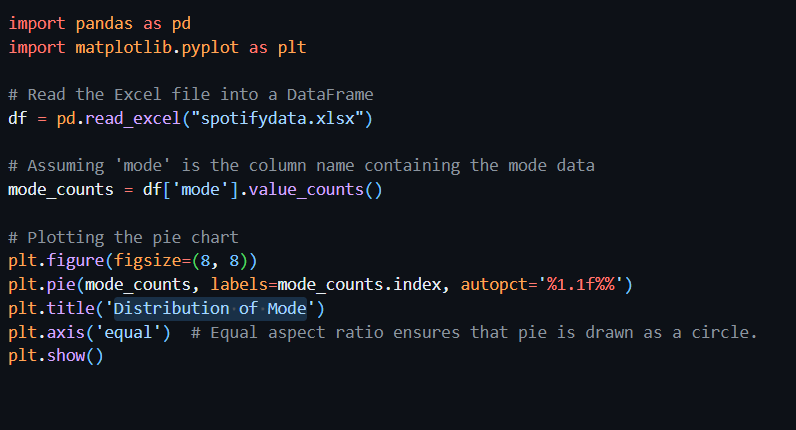


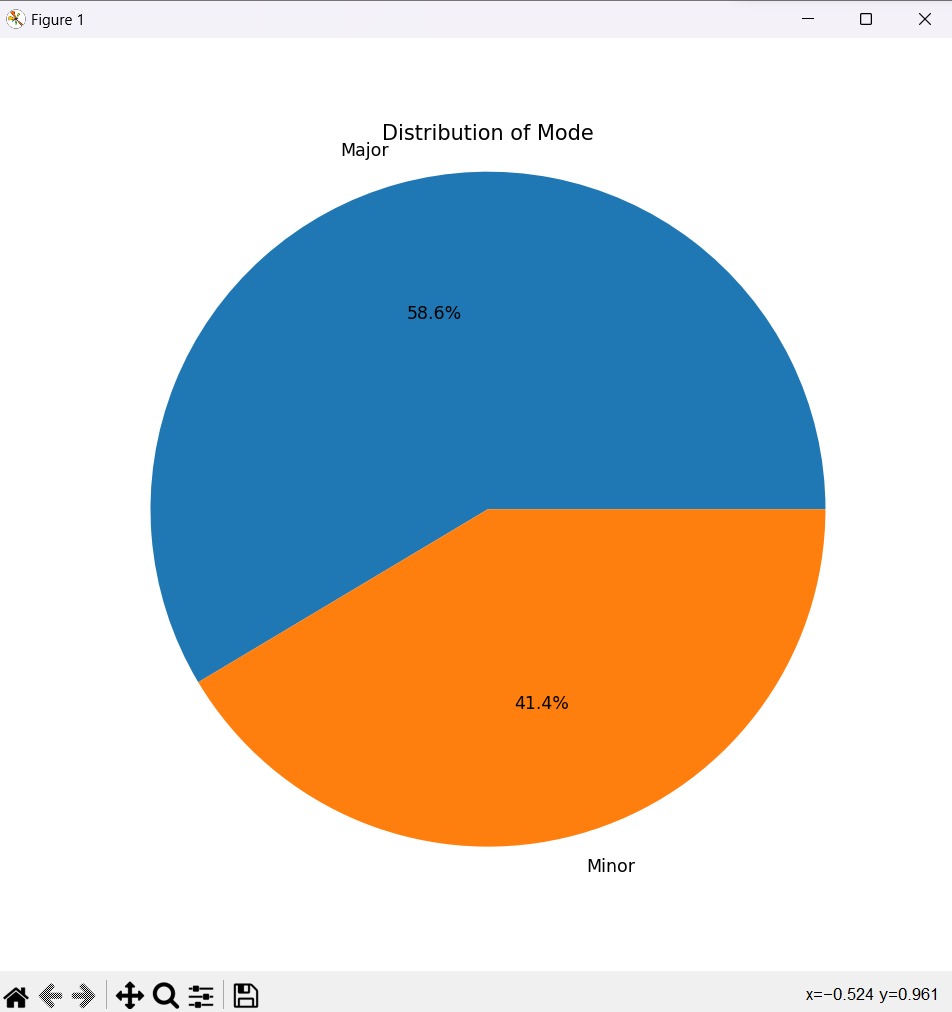
* **Distribution of Years (2010-2023) songs**





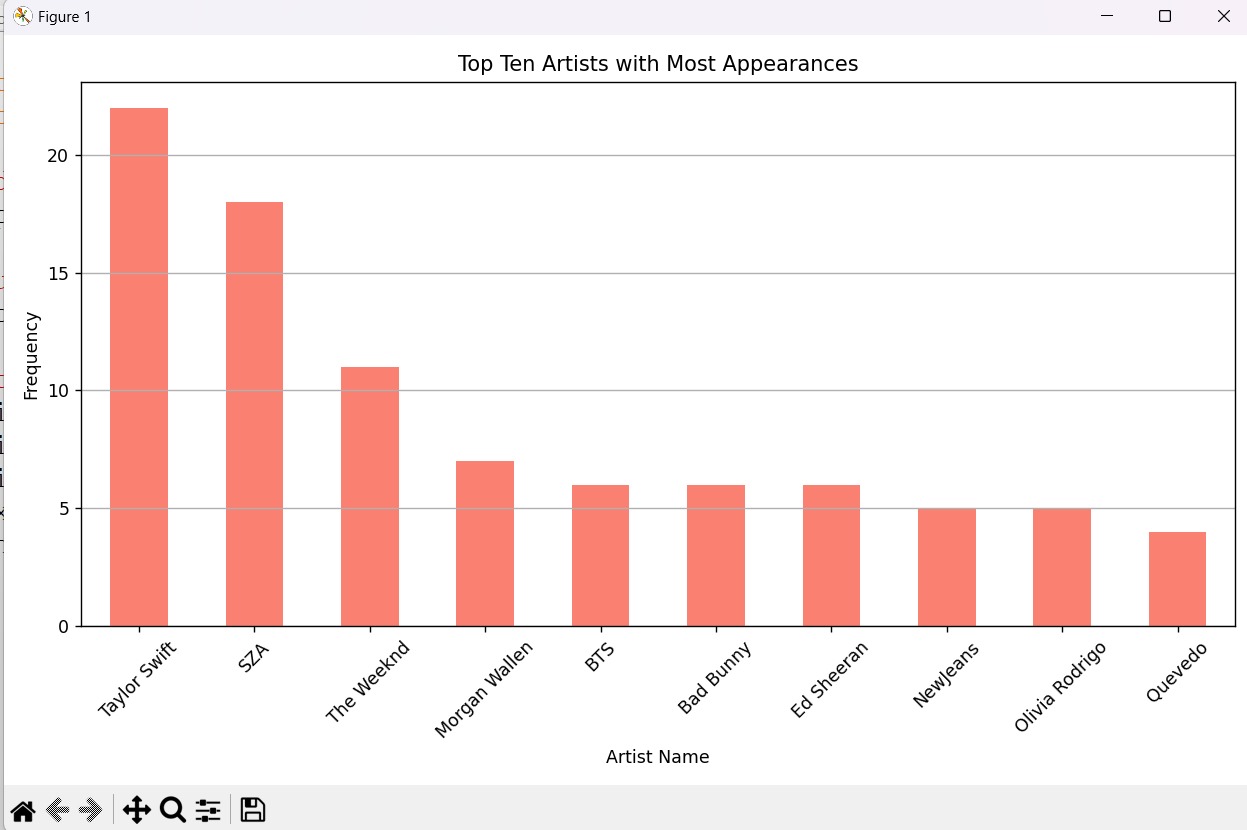
* **Distribution of Mode of the songs**



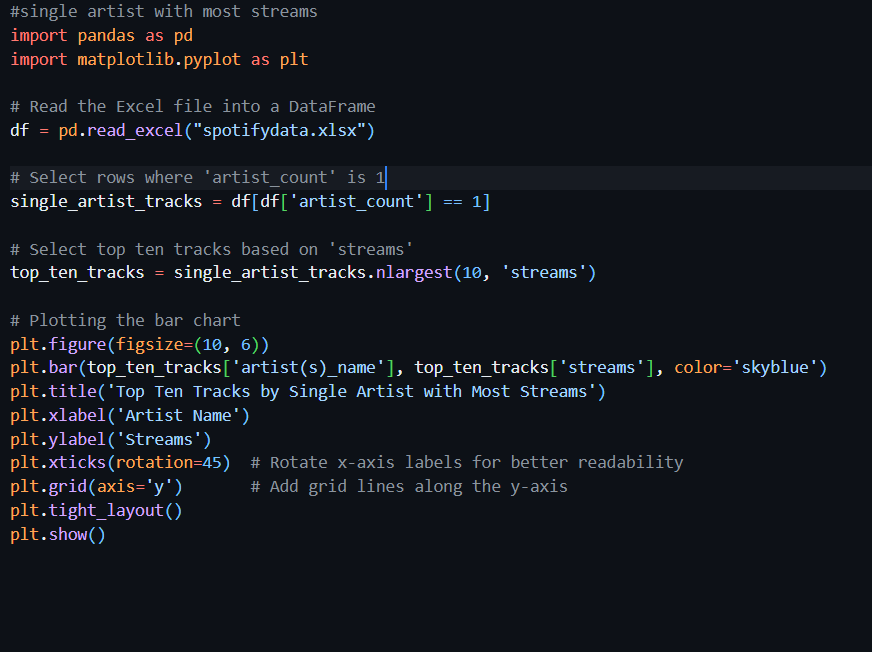


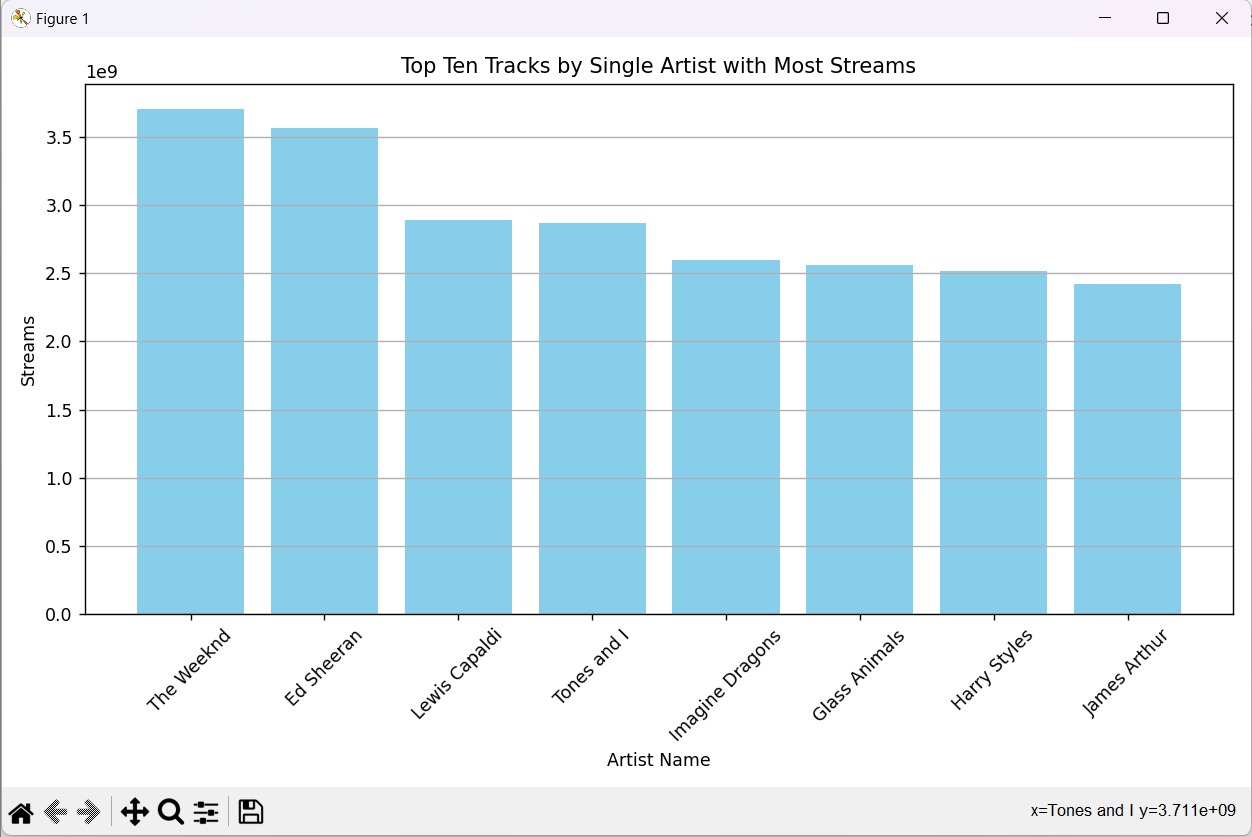
* **Top Ten Artists with Most Appearances**





* **Top Ten Tracks by Single Artist with Most Streams**





**CONCLUSIONS**

Here by, I declared that this project work entitled **Analysis On Spotify data set During 2010-2024** has been completed. I have performed various types of analysis on the dataset using Python. I have performed this data life cycle; collecting, cleansing, organizing, storing, analyzing, governing. I have done various basic, graphical and numerical analysis on the dataset using python functions.

Problem statements are built up and various charts; line chart, pie chart, bar chart, histogram, box plot, scatter chart are prepared for analysis. I have implemented statistical concepts for better understanding and analyzing the dataset; mean, median, standard deviation, variance, range, decision tree, time series, etc.So, I have tried my best for analyzing the dataset in better way and this analysis may be helpful in future also for taking any decision.

**10. FUTURE ANALYSIS**

I have performed various types of preliminary, numerical and graph based analysis through python functions on rows and columns of Spotify dataset. I have done year analysis, song wise analysis, stream wise analysis, artist analysis, playlist wise analysis…

* Total no. of songs which have been received in particular year.
* Total search according to streams, country and year.
* Total artist according to year, streams, country.

Also using advanced and efficient statistical concepts and functions of r, future analysis on this dataset may be possible.

**11. APPENDIX**

**11.1 Acknowledgment Of Sources:**

I have collected Spotify dataset from this website: https://www.kaggle.com/datasets. **This dataset is generated through random logic in python idle. This is not real Spotify data and should not be used for any other purpose other than testing.**

**11.2 Dataset:**

[**ANALYSIS ON SPOTIFY DATA SET DURING 2010-2**](file:///E:\Mansi\Integrated_MCA-8th_Semester\Software%20Project%20-%208\ANALYSIS%20ON%20MART%20DURING%202010-2017.xlsx)**024**