# 

**Project Based Learning Report**

on

**Create a Spotify Music Analysis Visualization Using Python Pandas**

Submitted in the partial fulfillment of the requirements

For the Project based learning in (**Essentials of Data Science**)

in

Electronics & Communication Engineering

By

**PRN Name of the Student**

2114110477 Aditi Kumari

2114110481 Summi Kumari

2114110492 Meenakshi Pothireddy

Under the guidance of

Prof. Vikas Kaduskar

Department of Electronics & Communication Engineering

Bharati Vidyapeeth

(Deemed to be University)

College of Engineering,

Pune – 4110043

**Academic Year: 2021-22**

**Bharati Vidyapeeth**

**(Deemed to be University)**

**College of Engineering,**

**Pune – 411043**

**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING**

**CERTIFICATE**

Certified that the Project Based Learning report entitled, “**Create a Spotify Music Analysis Visualization using Python Pandas”** is work done by

**PRN Name of the Student**

2114110477 Aditi Kumari

2114110481 Summi Kumari

2114110492 Meenakshi Pothireddy

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in “**Essentials of Data Science”** of Bachelor of Technology Semester IV, in ECE Div-1.

**Date: 25/04/2023**

**Prof. Vikas Kaduskar Dr. Tanuja S. Dhope**

**Course In-chargee PBL Co-ordinator**

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr. No | Title | Page No. |
| 1. | Problem statement | 1-2 |
| 2. | About Spotify | 3-4 |
| 3. | Software Used | 5-5 |
| 4. | Result | 6-12 |
| 5. | Outcome & Conclusion | 13-13 |

**Problem statement:-**

Create a Spotify Music Analysis Visualization using Python Pandas?

**Solution :-**

What is song recommendation?

What is song recommendation?

Song recommendation refers to the process of suggesting songs to a listener based on their musical preferences, listening history, and other relevant factors. This can be done through various means, such as through music streaming platforms that use algorithms to suggest songs based on a listener's listening history, or through human-curated playlists that are created by music experts or enthusiasts.

Song recommendations are useful for discovering new music that a listener may enjoy, and can help to broaden their musical horizons. They can also be a way for artists to reach new audiences and gain exposure for their music.

**What is role of data science in song recommendation?**

Data science plays a crucial role in song recommendation. Music streaming services use data science techniques to analyze a user's listening behaviour, such as the songs they listen to, the artists they follow, the playlists they create, and the genres they prefer. This data is then used to build a personalized music recommendation system.

Data science algorithms are used to identify patterns and relationships in the data that can be used to make predictions about what songs a user might enjoy. These algorithms use various techniques such as collaborative filtering, content-based filtering, and deep learning to make recommendations.

Collaborative filtering is a popular technique used in music recommendation systems, which involves analyze the listening behaviour of other users who have similar preferences to the current user. This helps to identify songs that the user may like but has not yet discovered.

Content-based filtering involves analyze the characteristics of the music itself, such as the tempo, melody, and lyrics, and recommending similar songs to the user based on those characteristics.

Deep learning techniques are also used in music recommendation systems, which involve training neural networks on large datasets of music to learn patterns and relationships in the data, which can be used to make more accurate recommendations.

In summary, data science plays a crucial role in developing effective song recommendation systems that can provide personalized music recommendations to users, thereby enhancing their overall music streaming experience.

**What are different data science techniques in song recommendation systems with example?**

There are several data science techniques that can be used in song recommendation systems. Here are some examples:

1. Collaborative Filtering: Collaborative filtering involves recommending songs based on the listening behaviour of other users who have similar preferences to the current user. For example, if User A and User B have similar listening habits and both enjoy a certain song, the system may recommend that song to User A.
2. Content-Based Filtering: Content-based filtering involves recommending songs that are similar in characteristics to the songs that a user has previously enjoyed. For example, if a user enjoys songs with a fast tempo and upbeat melody, the system may recommend similar songs that match those characteristics.
3. Matrix Factorization: Matrix factorization is a technique used to identify hidden factors or features in a dataset. In the context of music recommendation systems, matrix factorization can be used to identify the underlying musical features that are important to a user, such as tempo, melody, and genre.
4. Deep Learning: Deep learning involves training neural networks on large datasets to learn patterns and relationships in the data. In music recommendation systems, deep learning can be used to analyze the characteristics of songs and identify patterns in the listening behaviour of users. For example, a neural network could be trained to identify songs that are popular among users who have similar listening habits.
5. Hybrid Approaches: Hybrid approaches combine different data science techniques to provide more accurate recommendations. For example, a hybrid approach could combine collaborative filtering and content-based filtering to provide recommendations based on both user behaviour and song characteristics.

In summary, there are several data science techniques that can be used in song recommendation systems, and the choice of technique depends on the specific requirements of the system and the type of data available.

**Spotify music visuallisation:-**

******

**Spotify** [audio streaming](https://en.wikipedia.org/wiki/Streaming_media) and media services provider founded on 23 April 2006 by [Daniel Ek](https://en.wikipedia.org/wiki/Daniel_Ek) and [Martin Lorentzon](https://en.wikipedia.org/wiki/Martin_Lorentzon). It is one of the largest music streaming service providers, with over 422 million monthly [active users](https://en.wikipedia.org/wiki/Active_users), including 182 million paying subscribers, as of March 2022..

Spotify offers [digital copyright restricted](https://en.wikipedia.org/wiki/Digital_rights_management) recorded music and [podcasts](https://en.wikipedia.org/wiki/Podcast), including more than 82 million songs, from record labels and media companies. Spotify is currently available in 180+ countries, as of October 2021. Users can search for music based on [artist](https://en.wikipedia.org/wiki/Musician), [album](https://en.wikipedia.org/wiki/Album), or [genre](https://en.wikipedia.org/wiki/Music_genre), and can create, edit, and share [playlists.](https://en.wikipedia.org/wiki/Playlist)

Unlike physical or download sales, which pay artists a fixed price per song or album sold, Spotify pays [royalties](https://en.wikipedia.org/wiki/Royalties) based on the number of artist streams as a proportion of total songs streamed. It distributes approximately 70% of its total revenue to rights holders (often [record labels](https://en.wikipedia.org/wiki/Record_label)), who then pay artists based on individual agreements. According to [Ben Sisario](https://en.wikipedia.org/wiki/Ben_Sisario) of [*The New York Times*](https://en.wikipedia.org/wiki/The_New_York_Times)*,* approximately 13,000 out of seven million artists on Spotify generated $50,000 or more in payments in 2020.

**DATASET:-**

We have downloaded dataset about spotify music from github.com site which is

Spotify\_Dataset.csv.

We have performed analysis visulization on google collab.

**Library used:-**

For mathematical computation:-

1. Numpy library **-** numpy is used to perform various mathematical operations on arrays.
2. Pandas Library **-** pandas provides various data structures and operations for manipulating numerical data and time series.

3)Scipy-stats - All of the statistics functions are located in the sub-package scipy.stats and a fairly complete listing of these functions can be obtained using info(stats) function. A list of random variables available can also be obtained from the docstring for the stats sub-package.

**For data visuallisation:-**

1. Matplotlib library from which pyplot module is used for plotting library used for 2D graphics.
2. Seaborn library - seaborn is a library for making statistical graphics in Python.
3. Plotly - Plotly is a Montreal based technical computing company involved in development of data analytics and visualisation tools such as Dash and Chart Studio. It has also developed open source graphing Application Programming Interface (API) libraries for Python.

**Software** used is **Google Collab **

Google is quite aggressive in AI research. Over many years, Google developed AI framework called **TensorFlow** and a development tool called **Colaboratory**. Today TensorFlow is open-sourced and since 2017, Google made Colaboratory free for public use. Colaboratory is now known as Google Colab or simply **Colab**.

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

## **What Colab Offers us?**

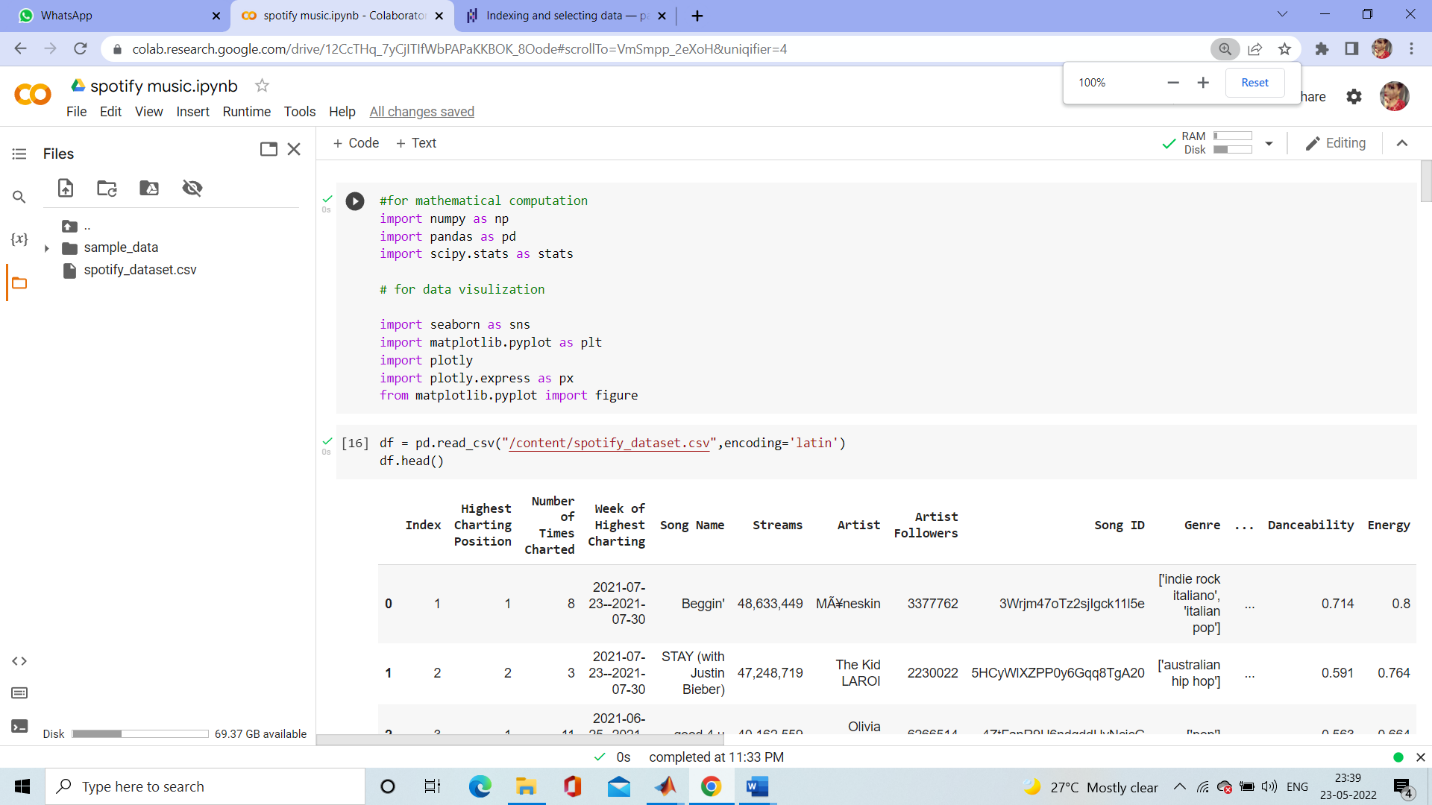
As a programmer, you can perform the following using Google Colab.

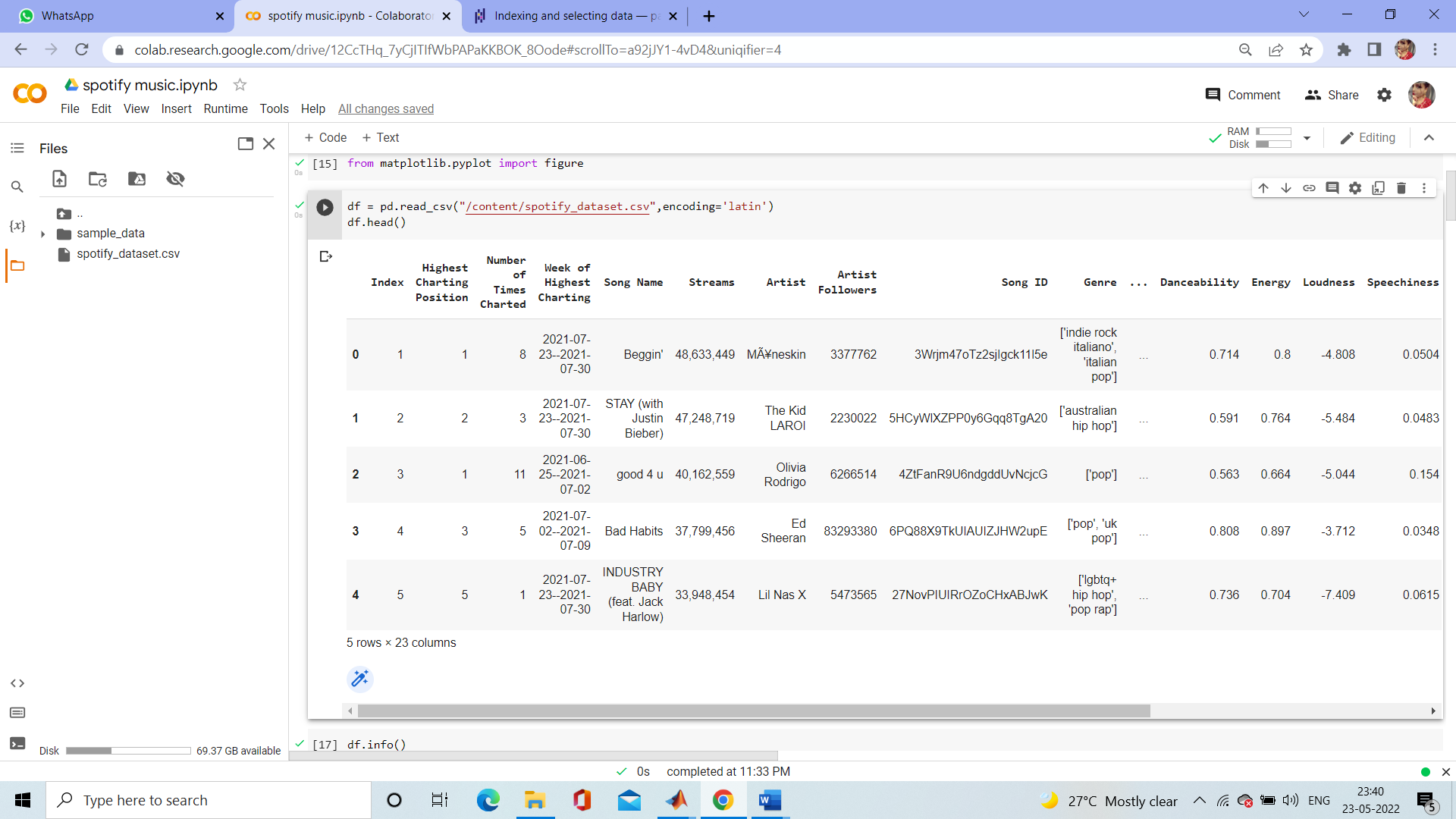
* Write and execute code in Python
* Document your code that supports mathematical equations
* Create/Upload/Share notebooks
* Import/Save notebooks from/to Google Drive
* Import/Publish notebooks from GitHub
* Import external datasets e.g. from Kaggle
* Integrate PyTorch, TensorFlow, Keras, OpenCV
* Free Cloud service with free GPU

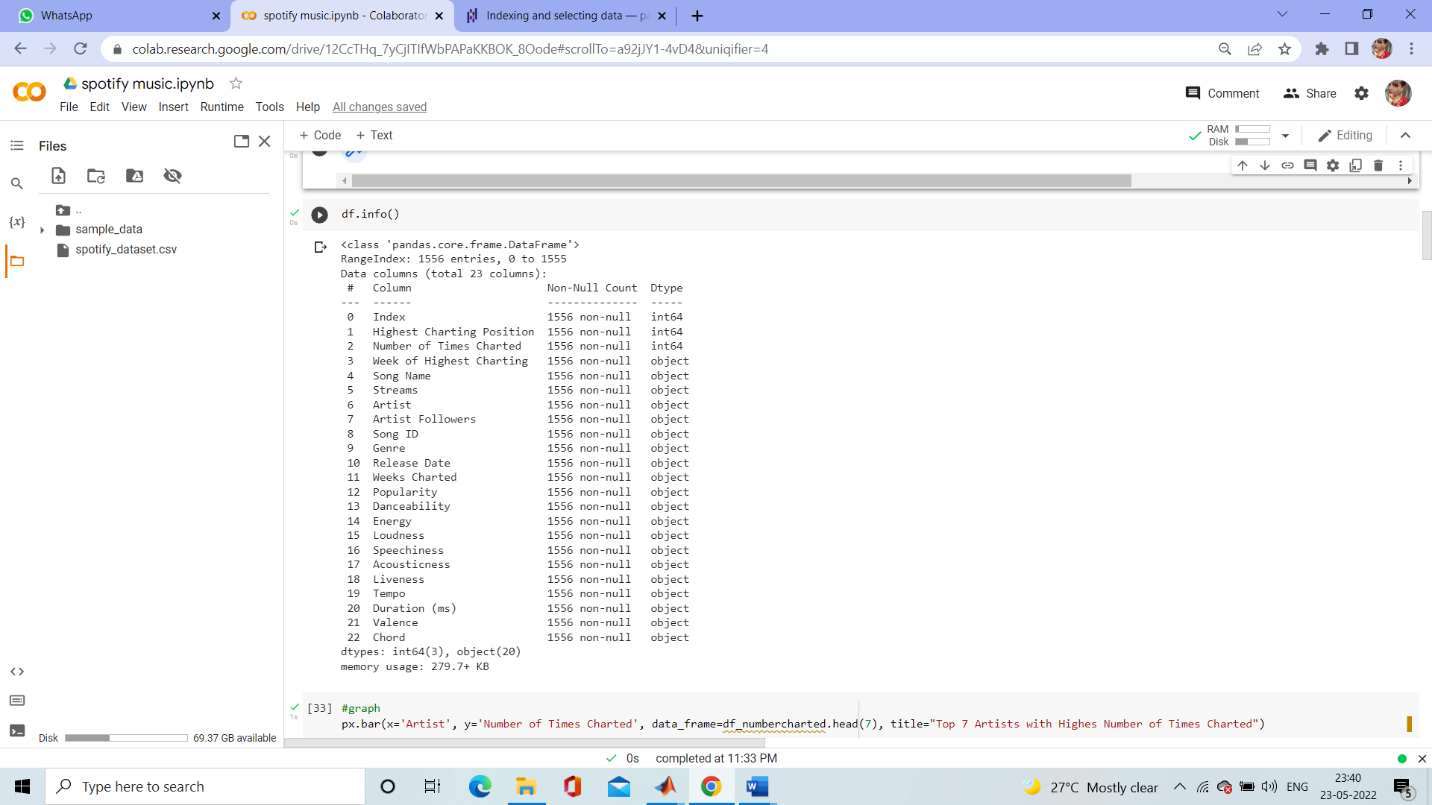
**Result with analysis**

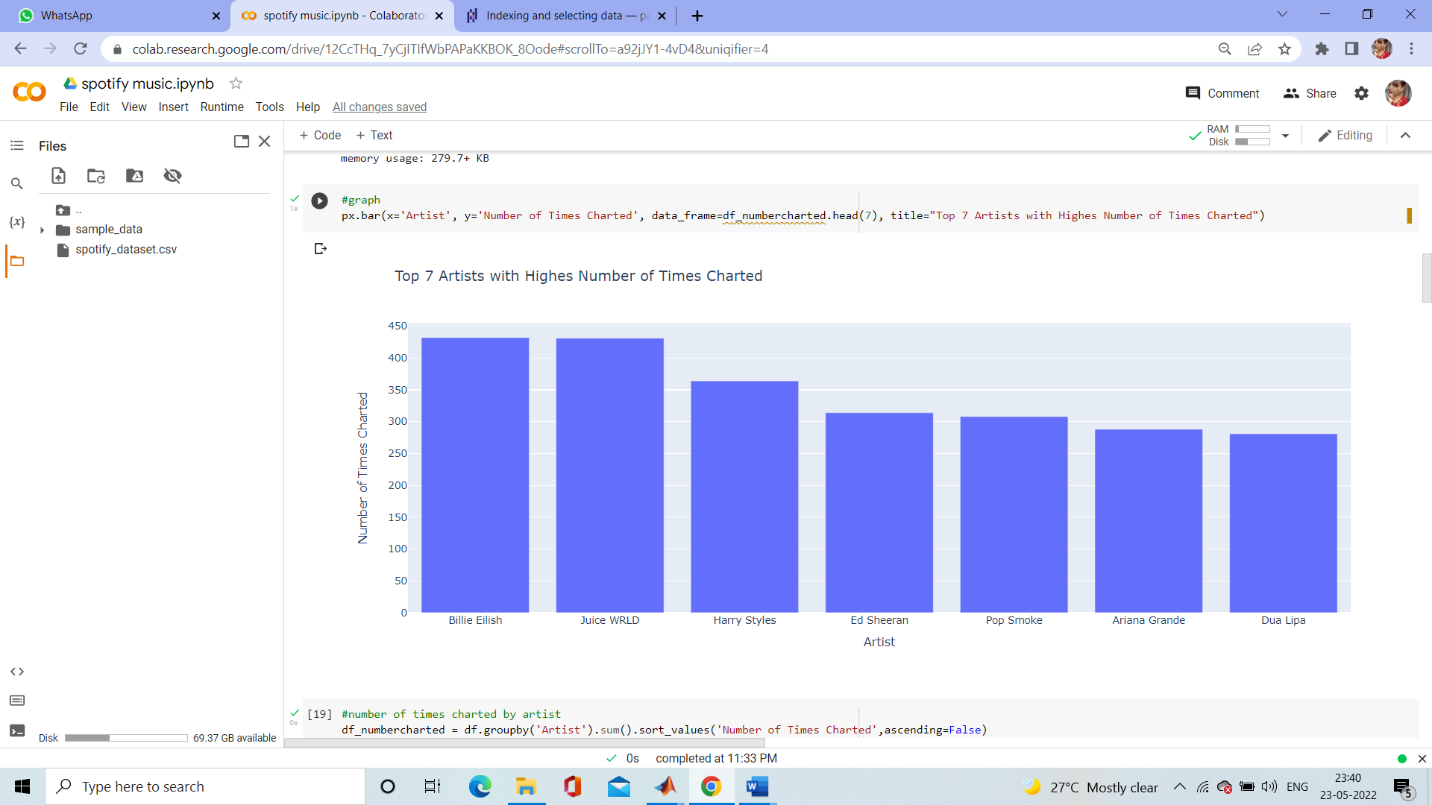
**Analysis of the code: -**

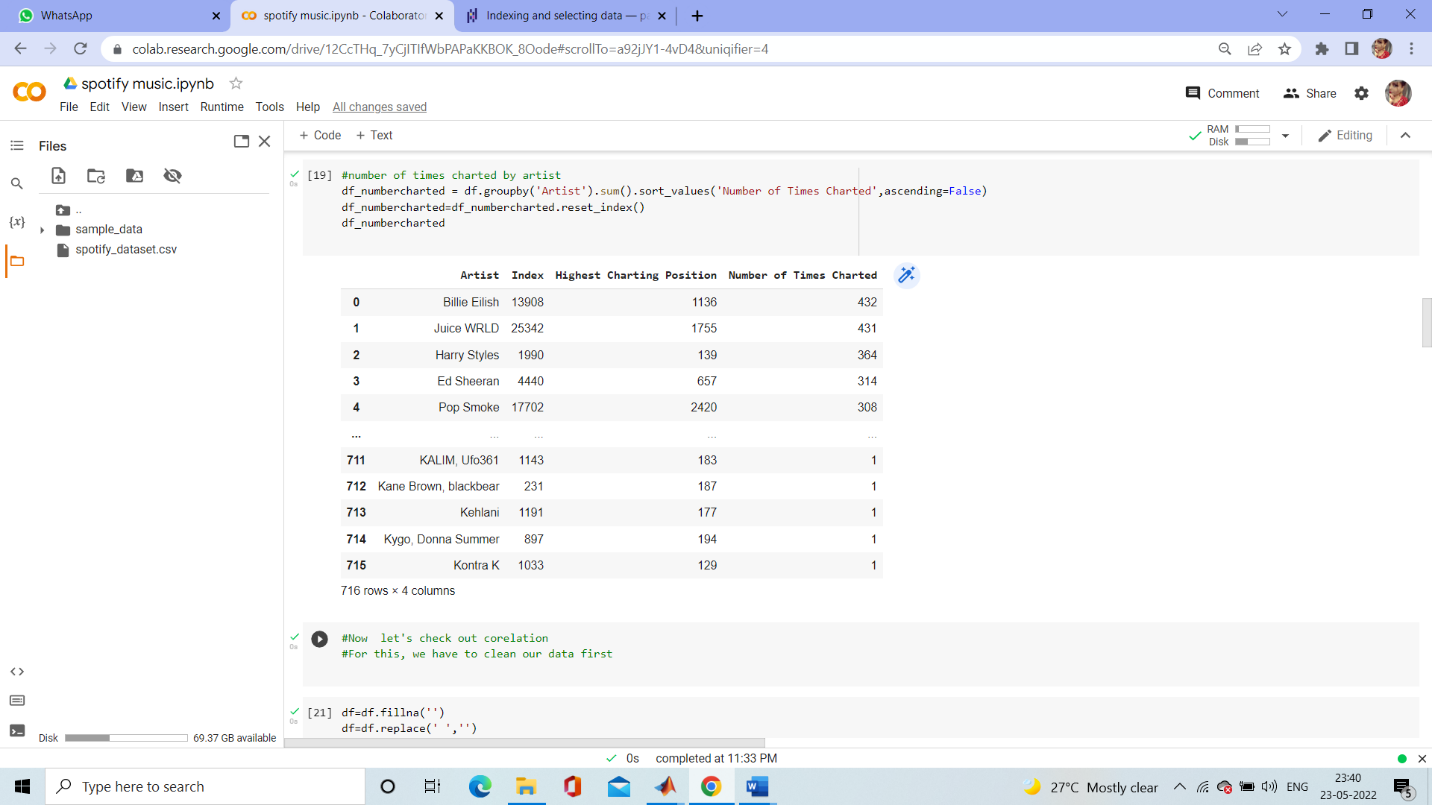
* First, we Import the libraries
* Secondly, Download the dataset and add that to the path to load the dataset. we use panda library and used head() function for displaying first five row of dataset.
* We get more information by using df.info().then for checking null values in dataset we used is null() function.
* Then we find the graph of number of time charted by artist by using px.bar() function.
* Then we create a correlation using heatmap()
* Then we use the library plotly to plot the graph of danceability by use px.line().
* Then we plot graph by using px.bar()
* At last we use pandas library to get information about genre and plot the pie chart.

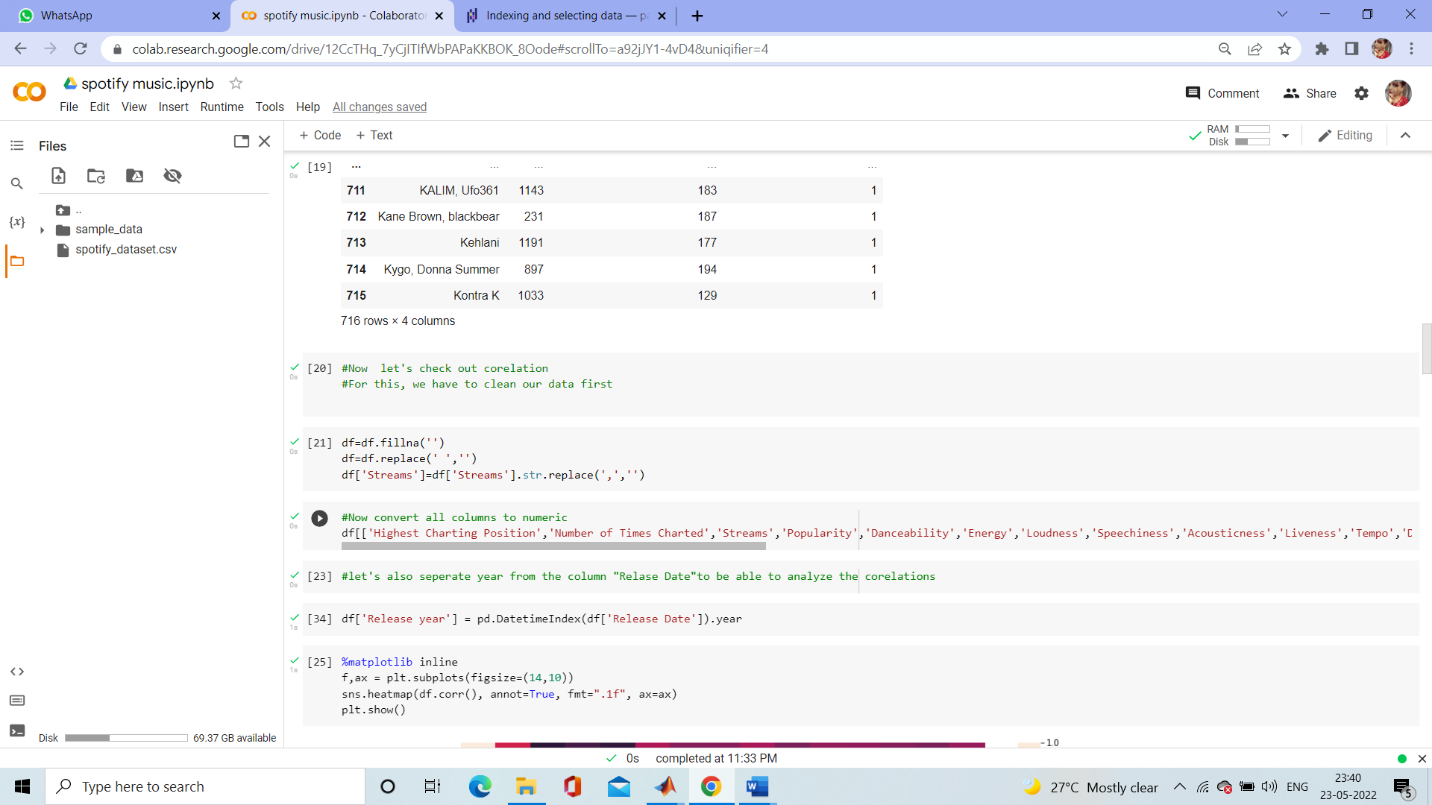


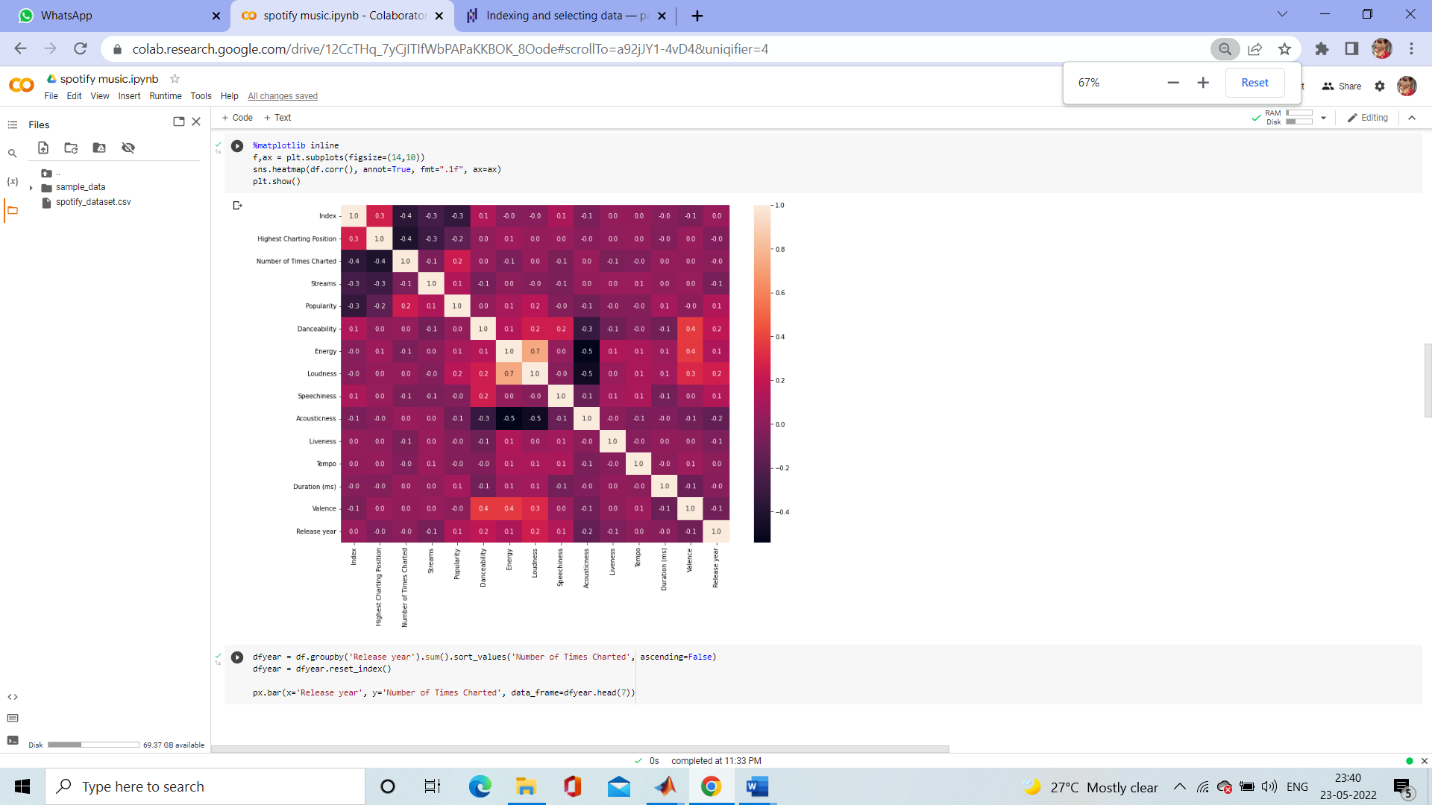


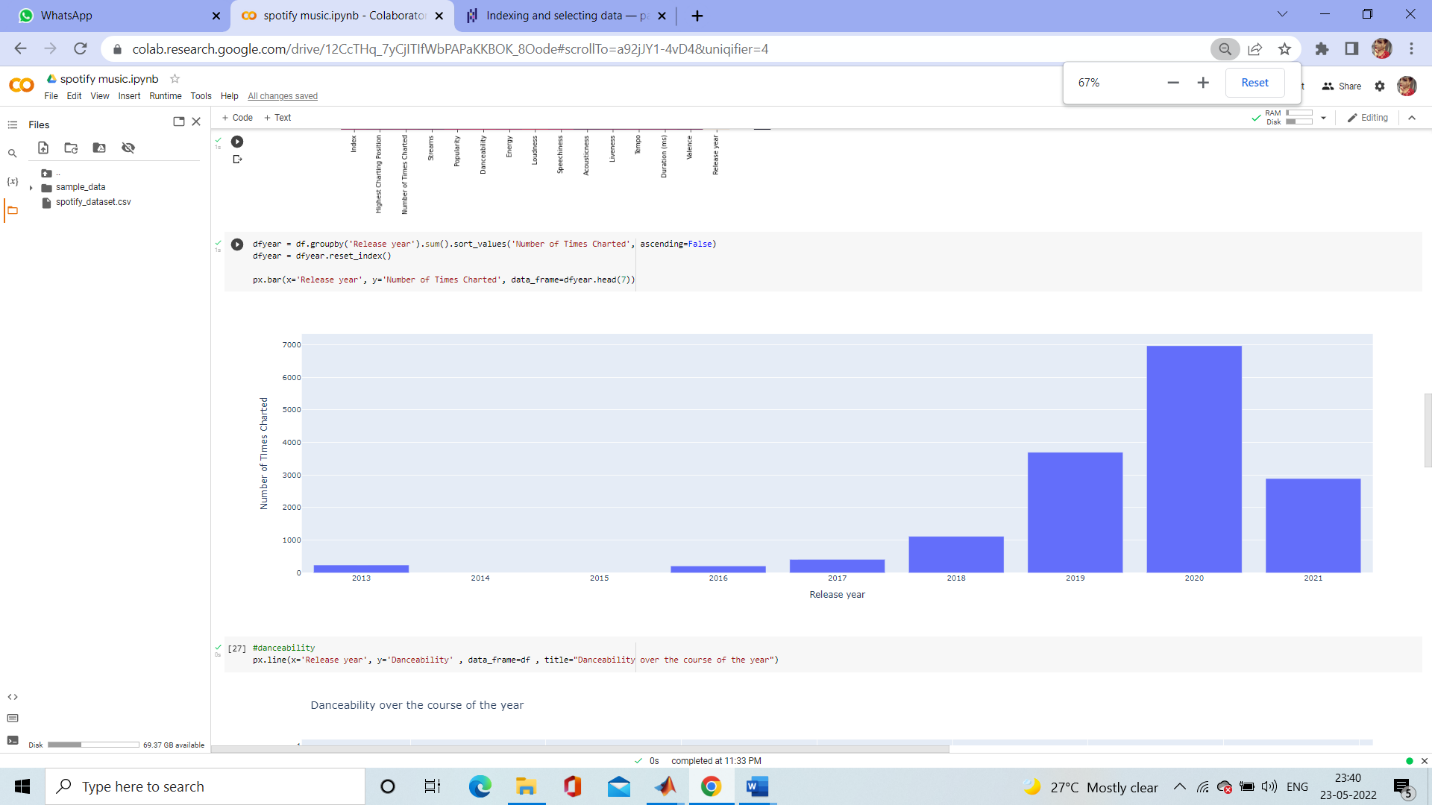


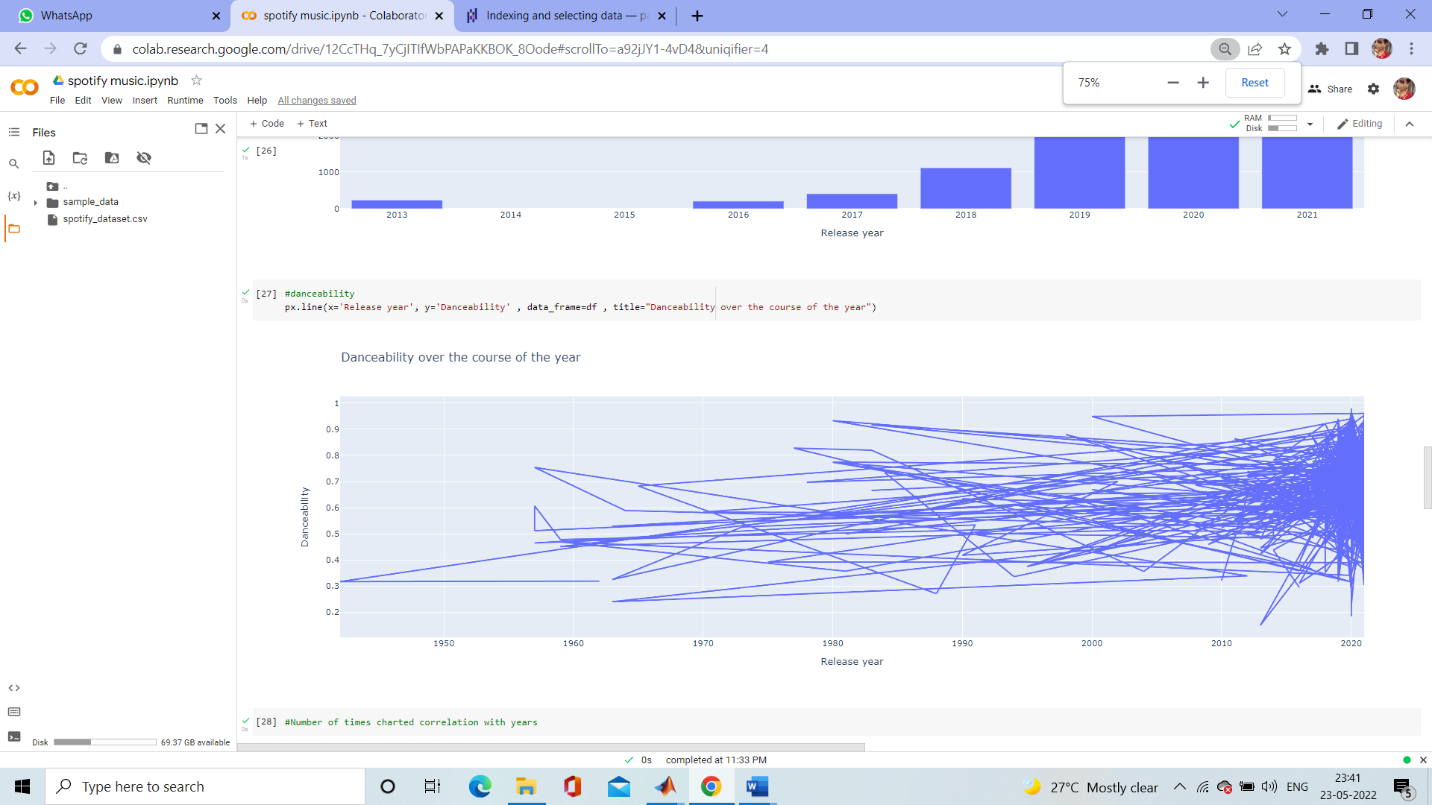


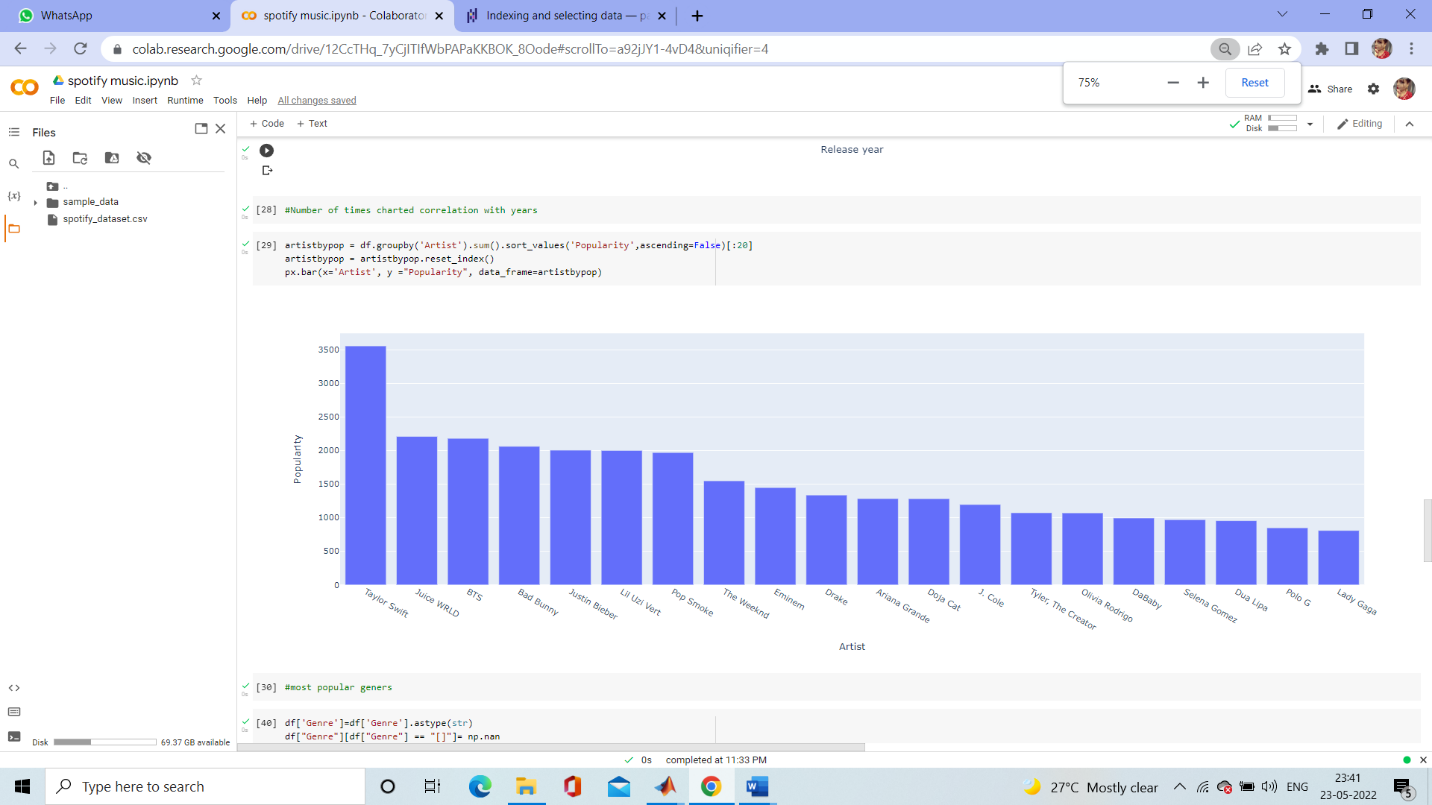
****

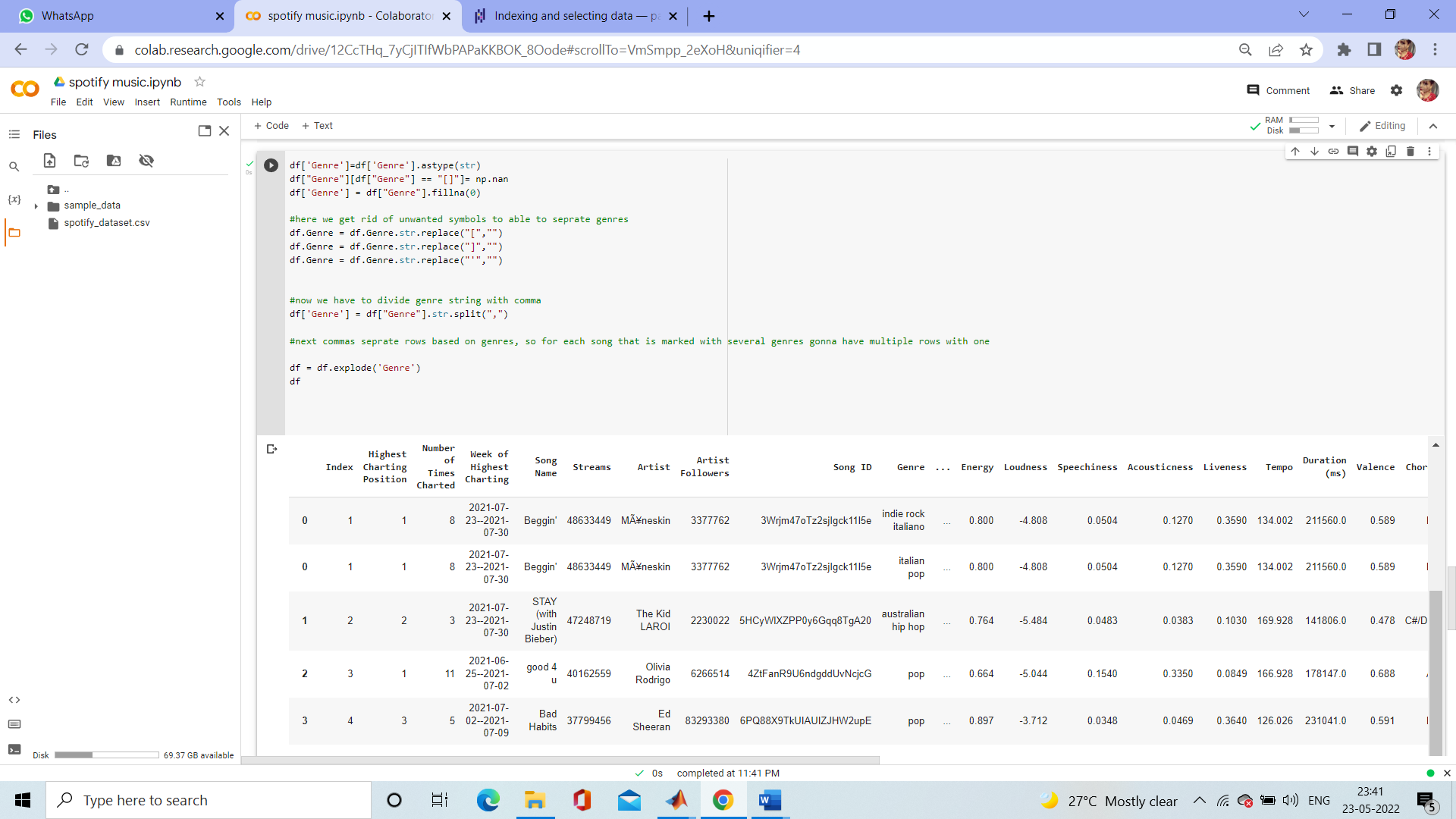
****

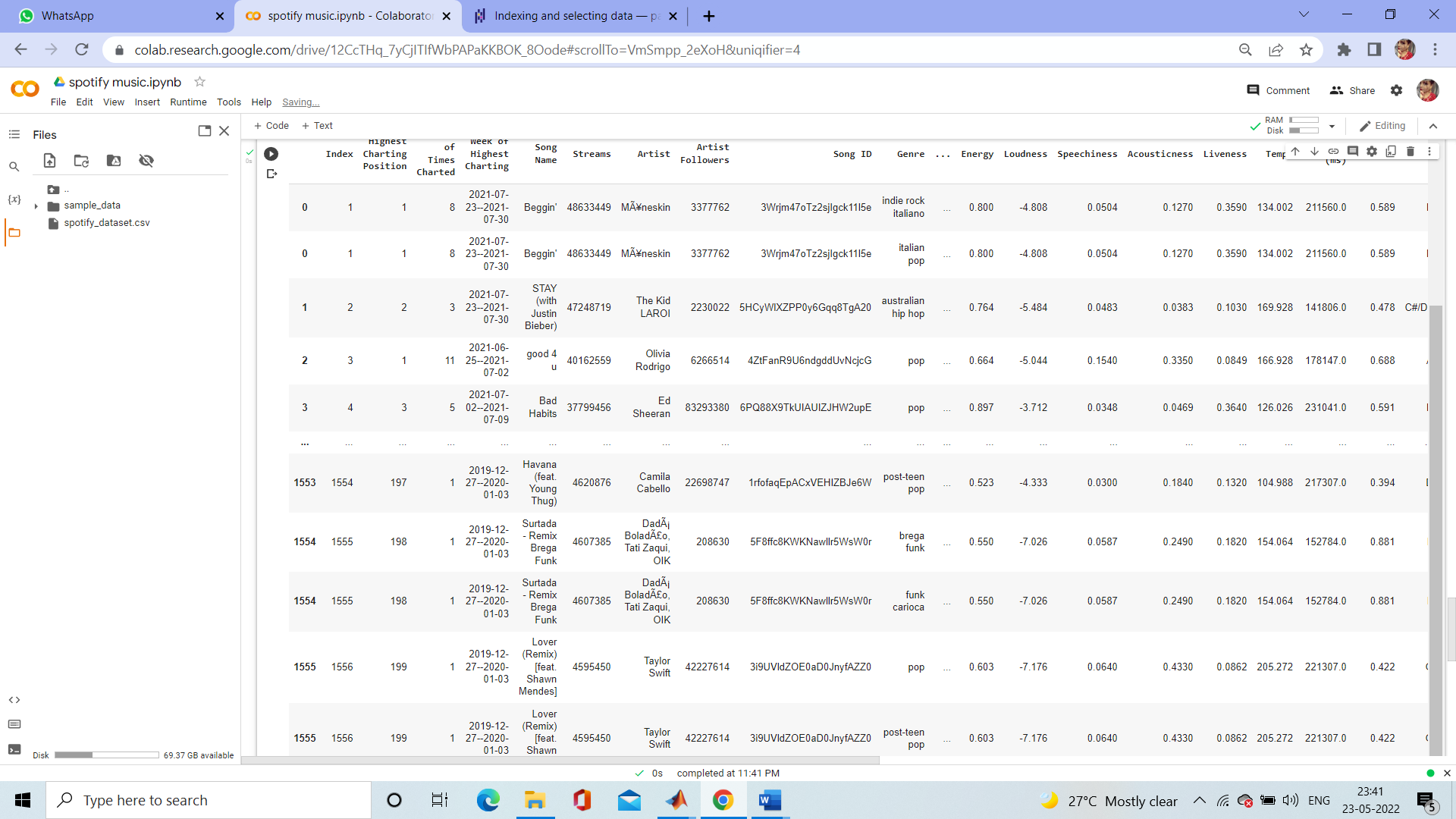
****

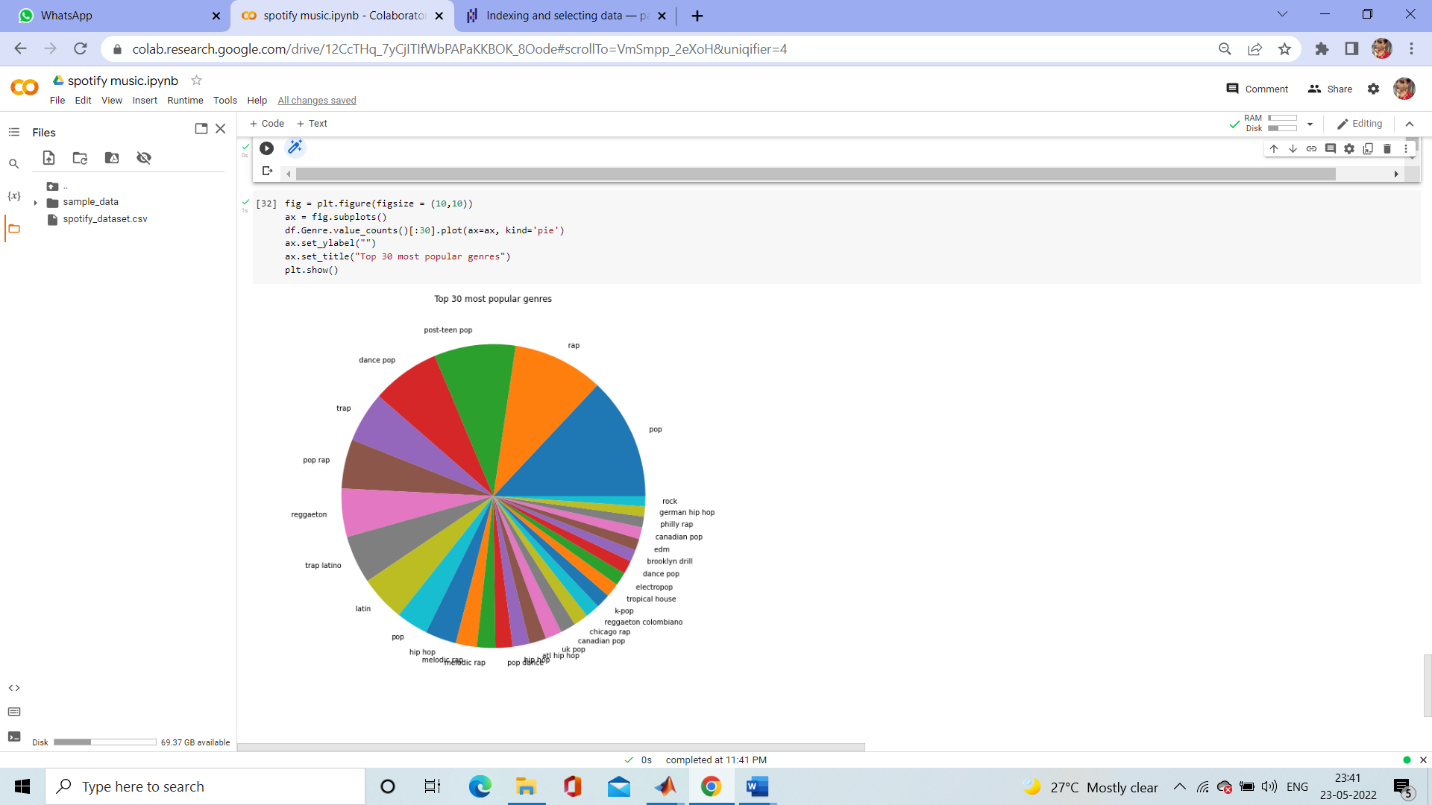
****

****

****

****

****

****

**Outcome:**

From this project, we learnt to describe a flow process for data science problems and classified data science problems into standard typology. We also learnt about correlating results to the solution approach followed and assessing the solution approach.

**C01 :**

**Project Conclusion:**

From this project, we gained the knowledge of software – Google colab. We learnt to analyse the datasets and afterwards, visualizing them. We learnt about various plots .