**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

Name of Students in Alphabetical order with Seat Number /PRN Number

**PRN Name of the Student**

2014111074 Harshita Kumari

2014111078 Aditi Maheshwari

2014111081 Himanshu Mishra

Under the guidance of

Prof. Dnyandeo Lakhare

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**

2014111074 Harshita Kumari

2014111078 Aditi Maheshwari

2014111081 Himanshu Mishra

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in **Digital Communication** of Bachelor of Technology Semester IV, in ECE Div-1.

**Date:**

**Prof. Sonali Pawar Dr. Tanuja S. Dhope**

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

**Dr. Arundhati A.Shinde**

**Professor & Head**

**ELECTRONICS & COMMUNICATION ENGINEERING**

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No | Title | Page No. | Signature |
| 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:-**

What is Data Science? Why learn Data Science?

**Solution :-**

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models. The data used for analysis can come from many different sources and presented in various formats.

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply [machine learning](https://www.datarobot.com/wiki/machine-learning/) [algorithms](https://www.datarobot.com/wiki/algorithm/) to numbers, text, images, video, audio, and more to produce [artificial intelligence (AI)](https://www.datarobot.com/wiki/artificial-intelligence/) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate [insights](https://www.datarobot.com/wiki/insights/) which analysts and business users can translate into tangible business value.

Reasons to learn Data Science are: -

1. Learning about data science provides an opportunity for you to recreate yourself.
2. **We live in a digital world, everything is data-driven.** There is data sciencein**business, accounting, education, science, engineering,  healthcare, technology, energy sector, government,** and so on.
3. **Data science is also a very promising field with lots of high paying job opportunities.**
4. **Basic data science skills are important for personal use.**
5. Great potential to branch out with different options.
6. Become a decision-maker, not every job opportunity will give you the power to make informed business decisions. For a data scientist, that is the core responsibility.
7. Less competitive because it is a highly analytical role, competition is less, but demand is not. With a limited talent pool, there is always a challenge for businesses to hire in these roles.

**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?**

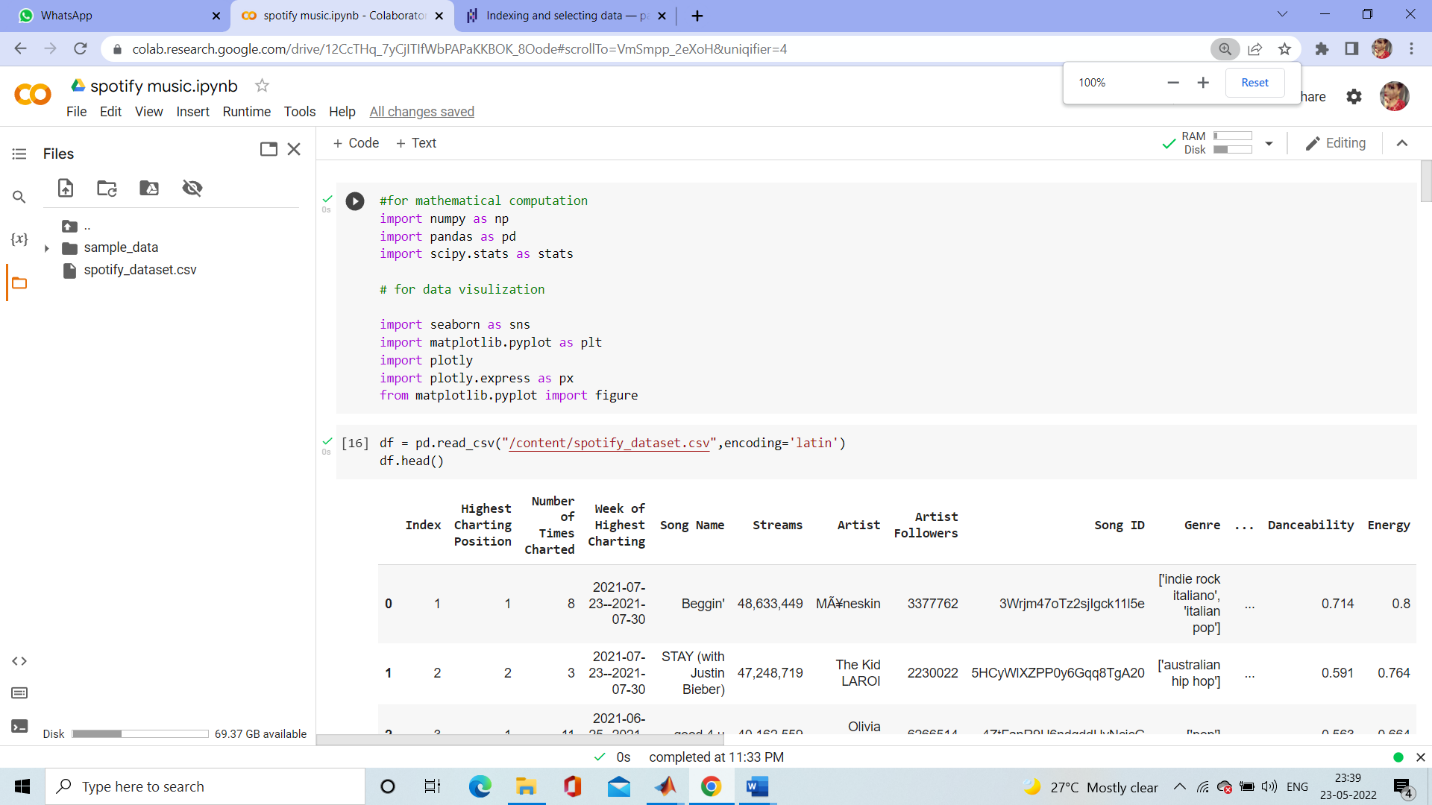
As a programmer, you can perform the following using Google Colab.

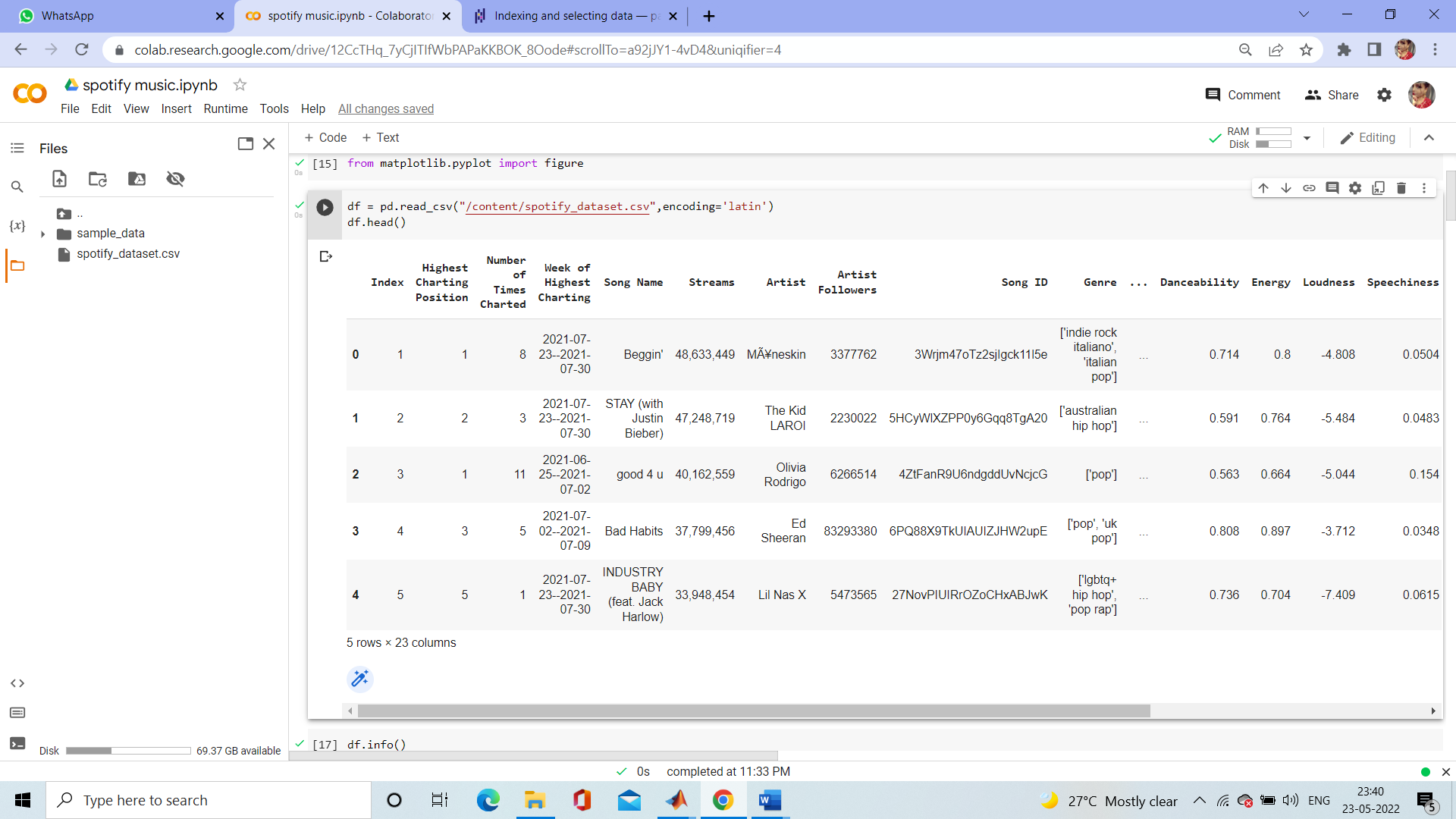
* Write and execute code in Python
* Document your code that supports mathematical equations
* Create/Upload/Share notebooks
* Import/Save notebooks from/to Google Drive
* Import/Publish notebooks from GitHub
* Import external datasets e.g. from Kaggle
* Integrate PyTorch, TensorFlow, Keras, OpenCV
*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

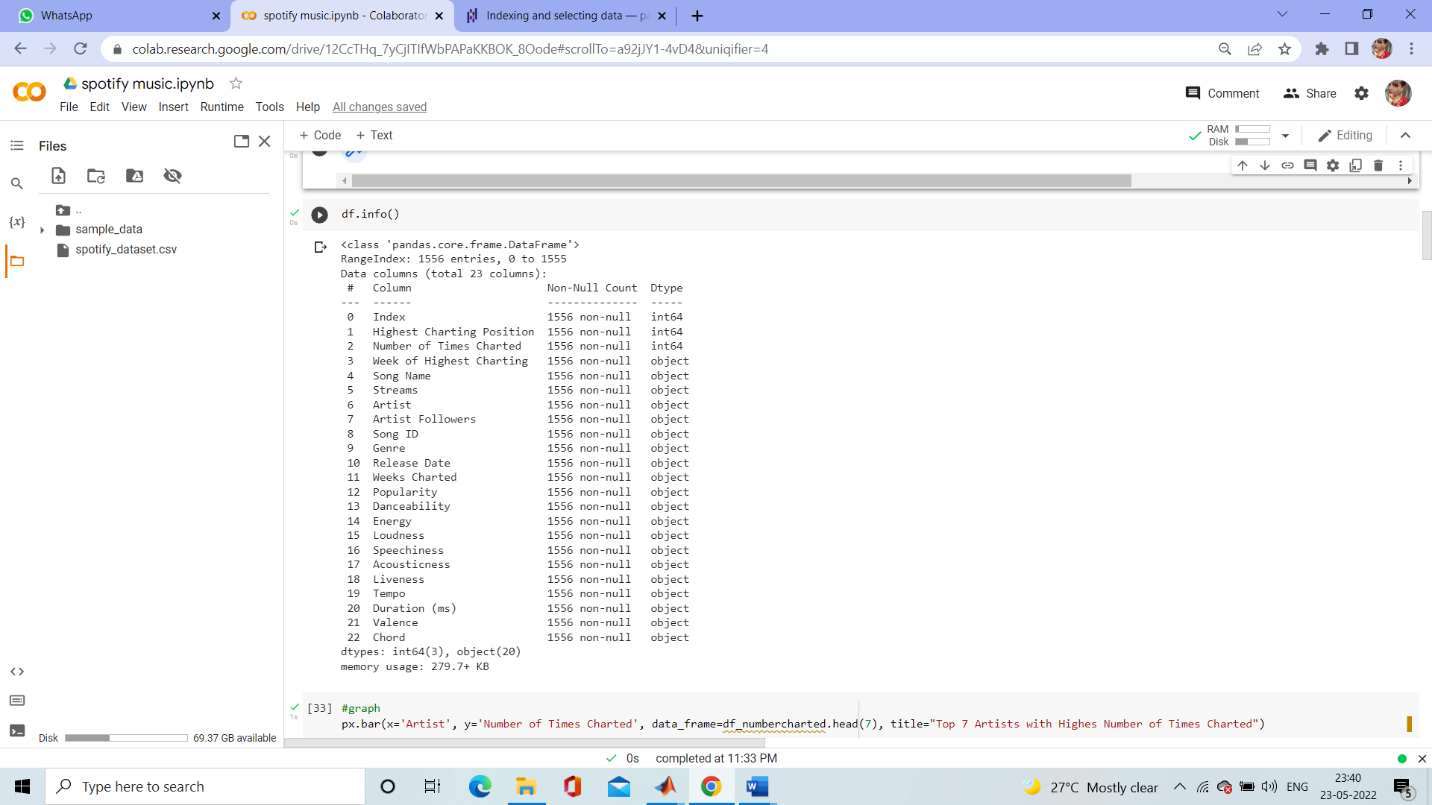
**Result with analysis**

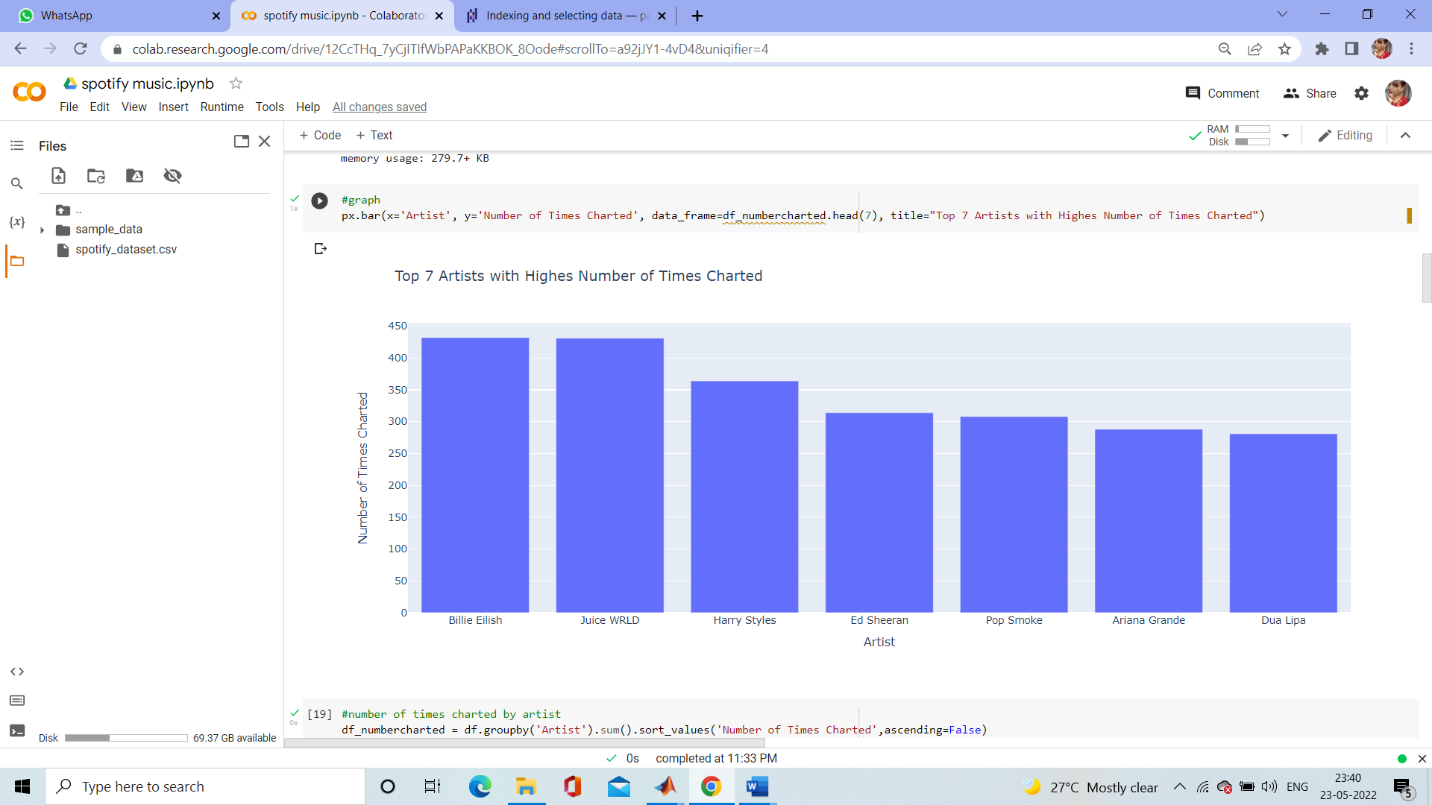
**Analysis of the code: -**

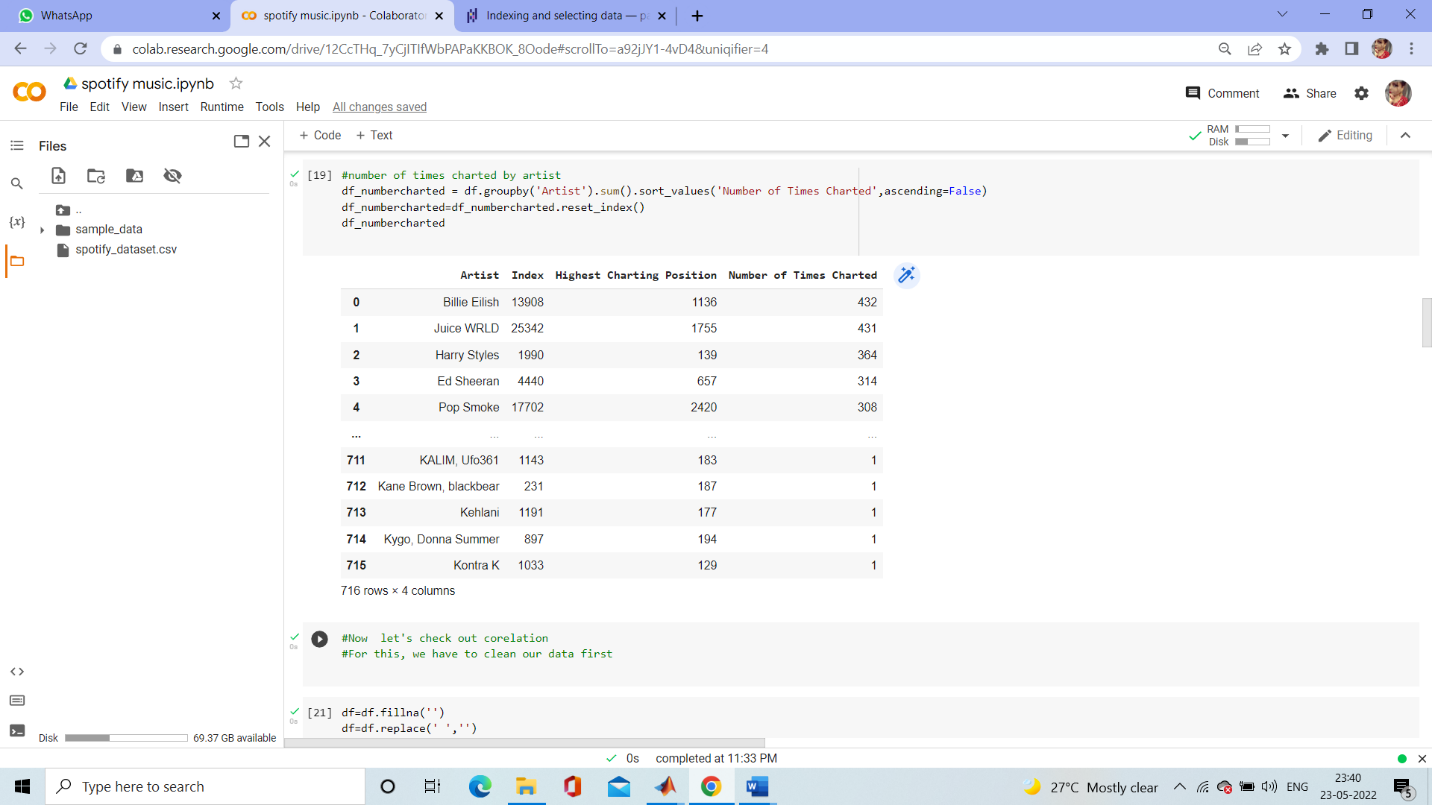
* First, we Import the libraries
* Secondly, Download the dataset and add that to the path to load the dataset. we use panda library and used head() function for displaying first five row of dataset.
* We get more information by using df.info().then for checking null values in dataset we used is null() 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.

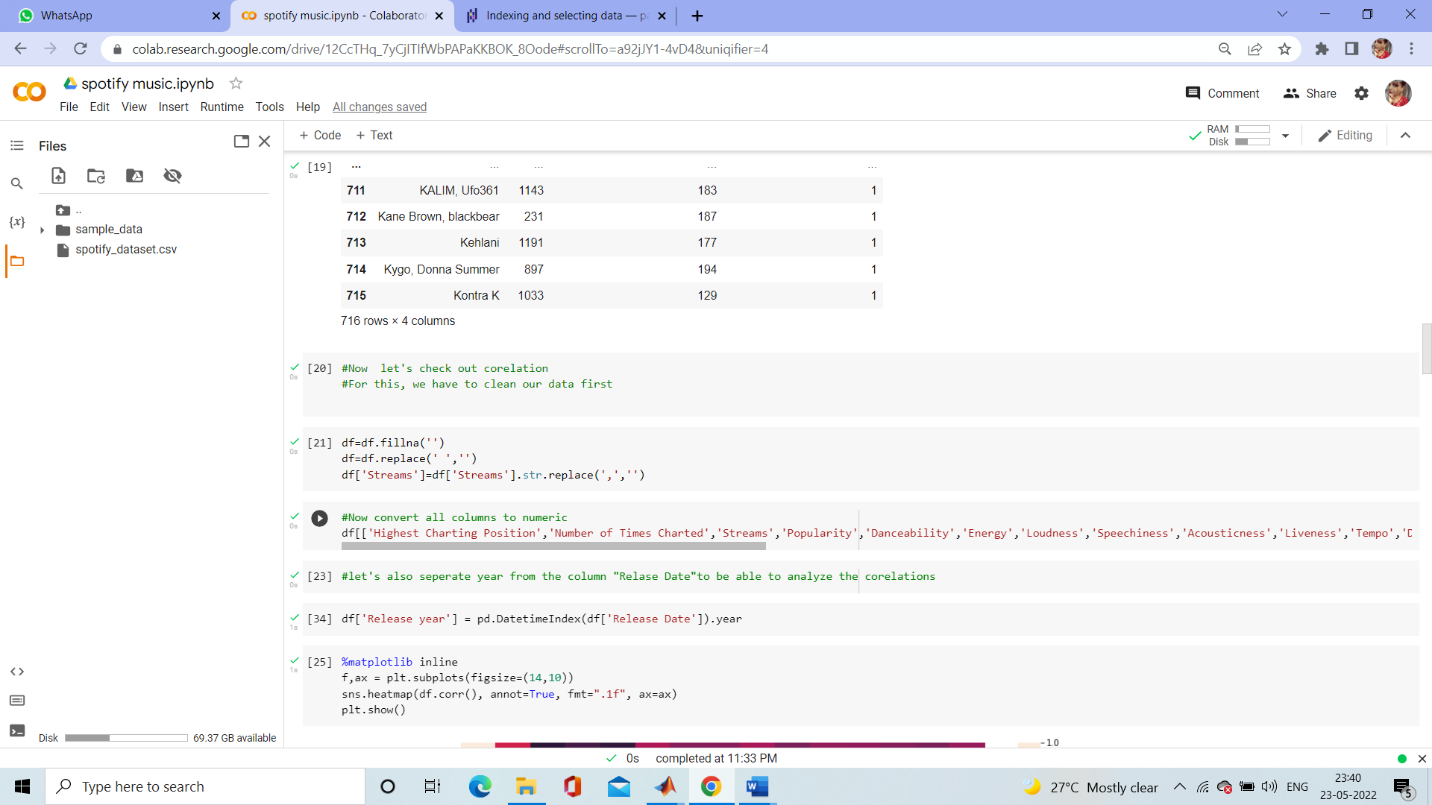


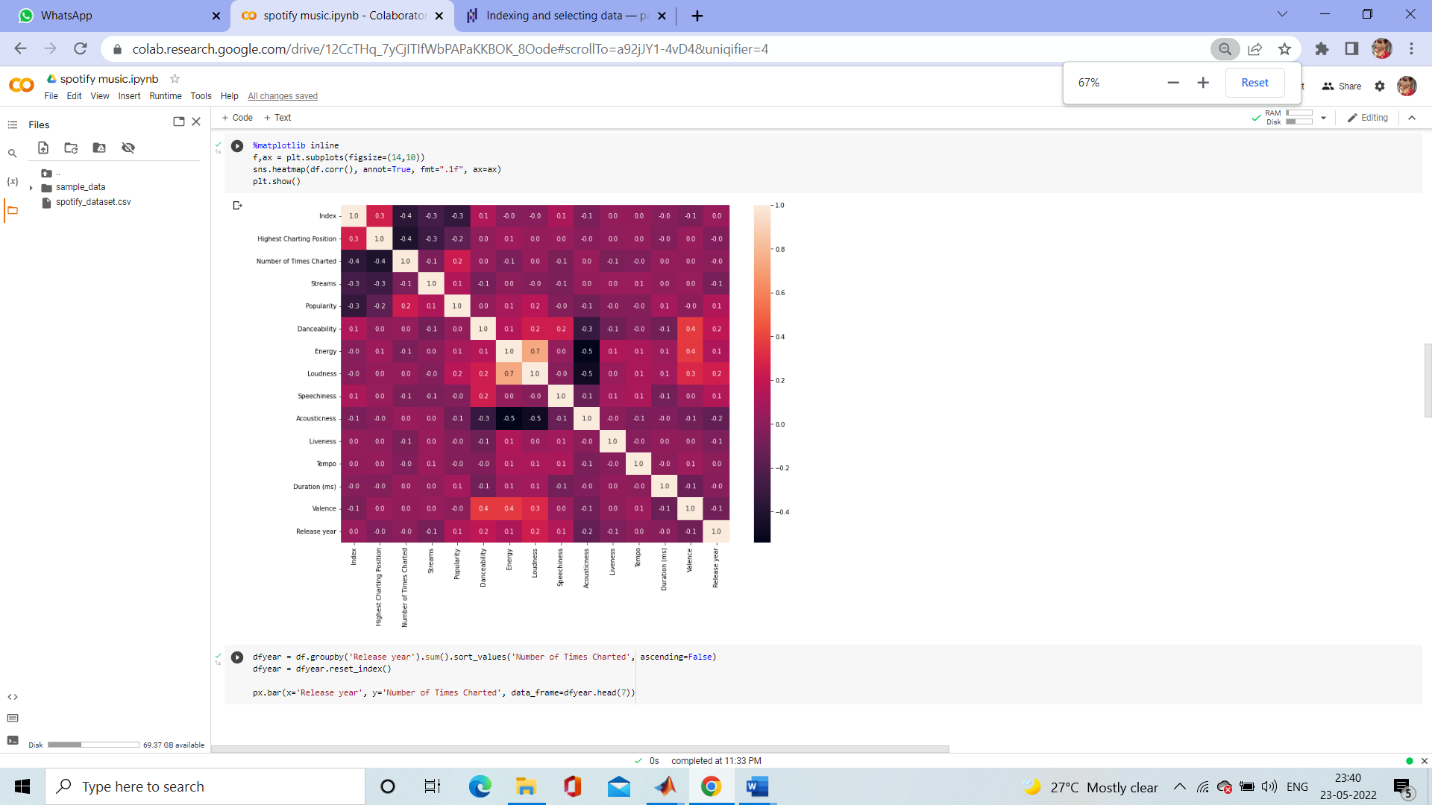


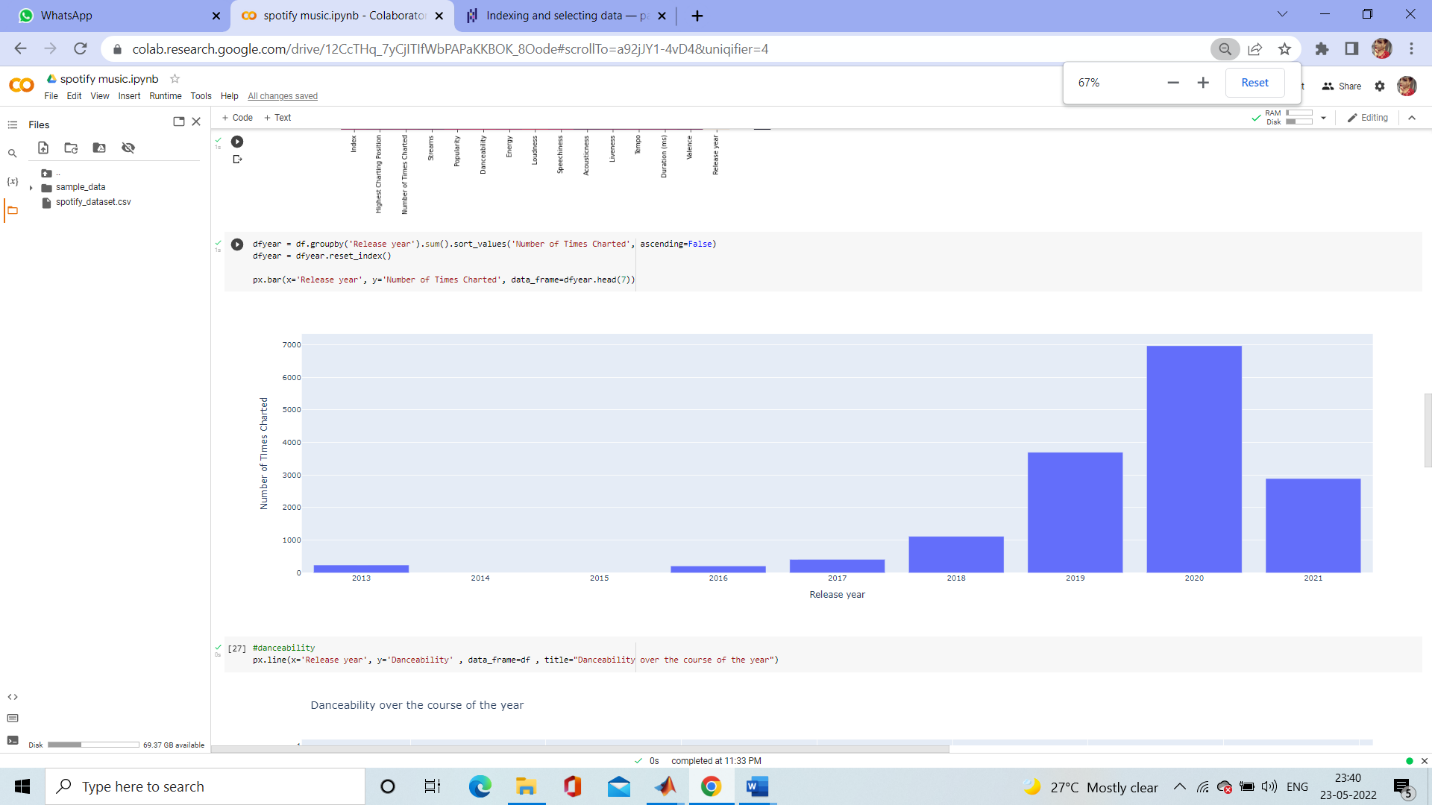


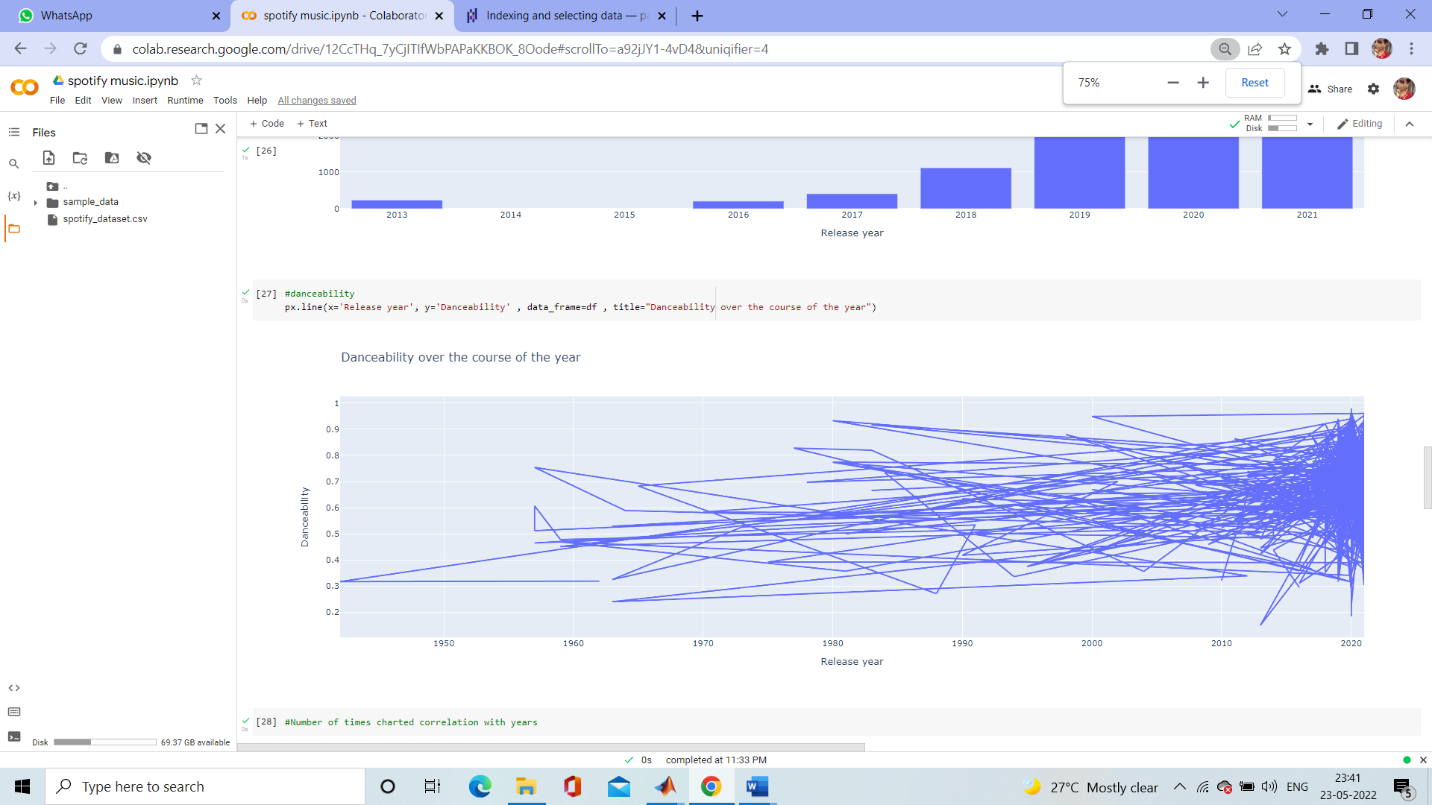


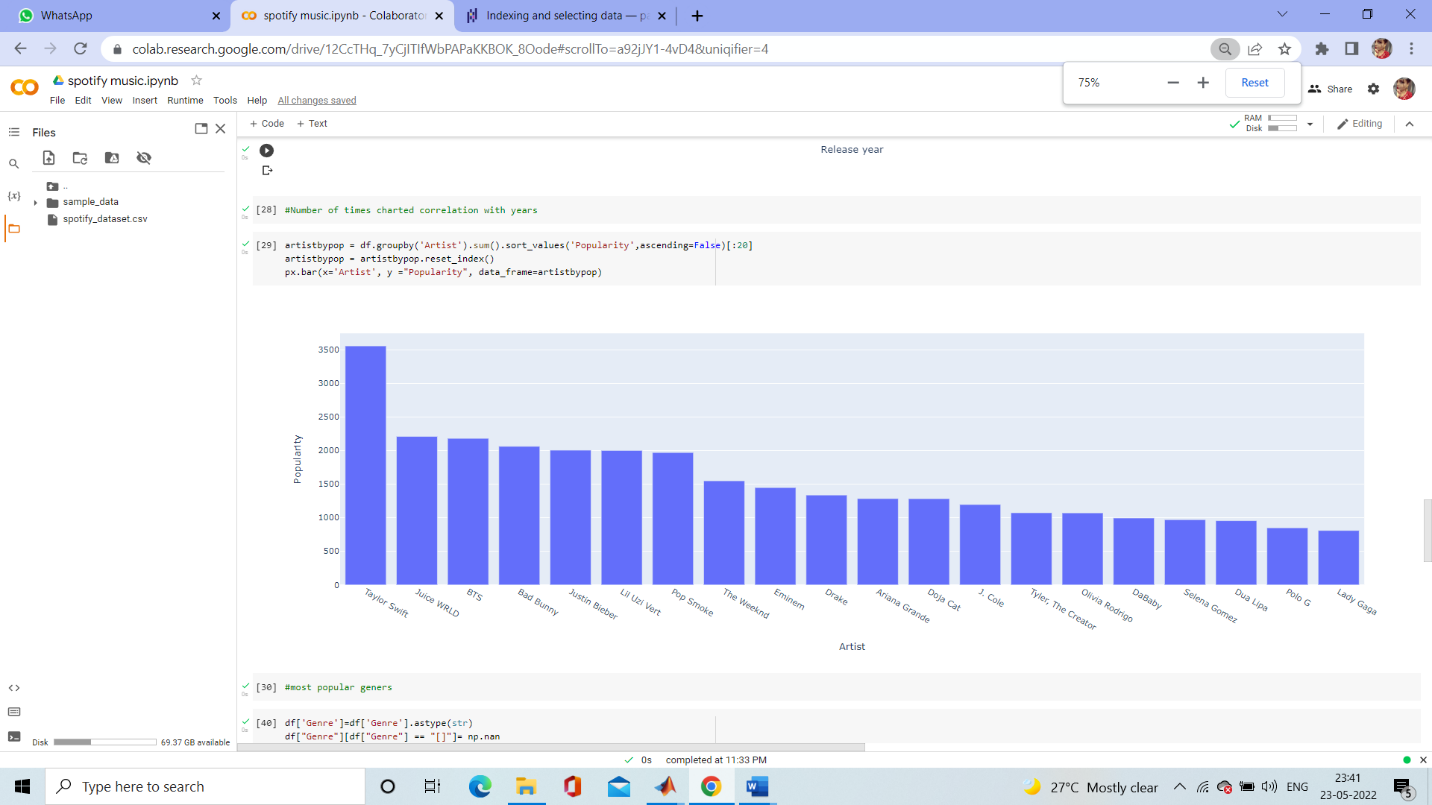
****

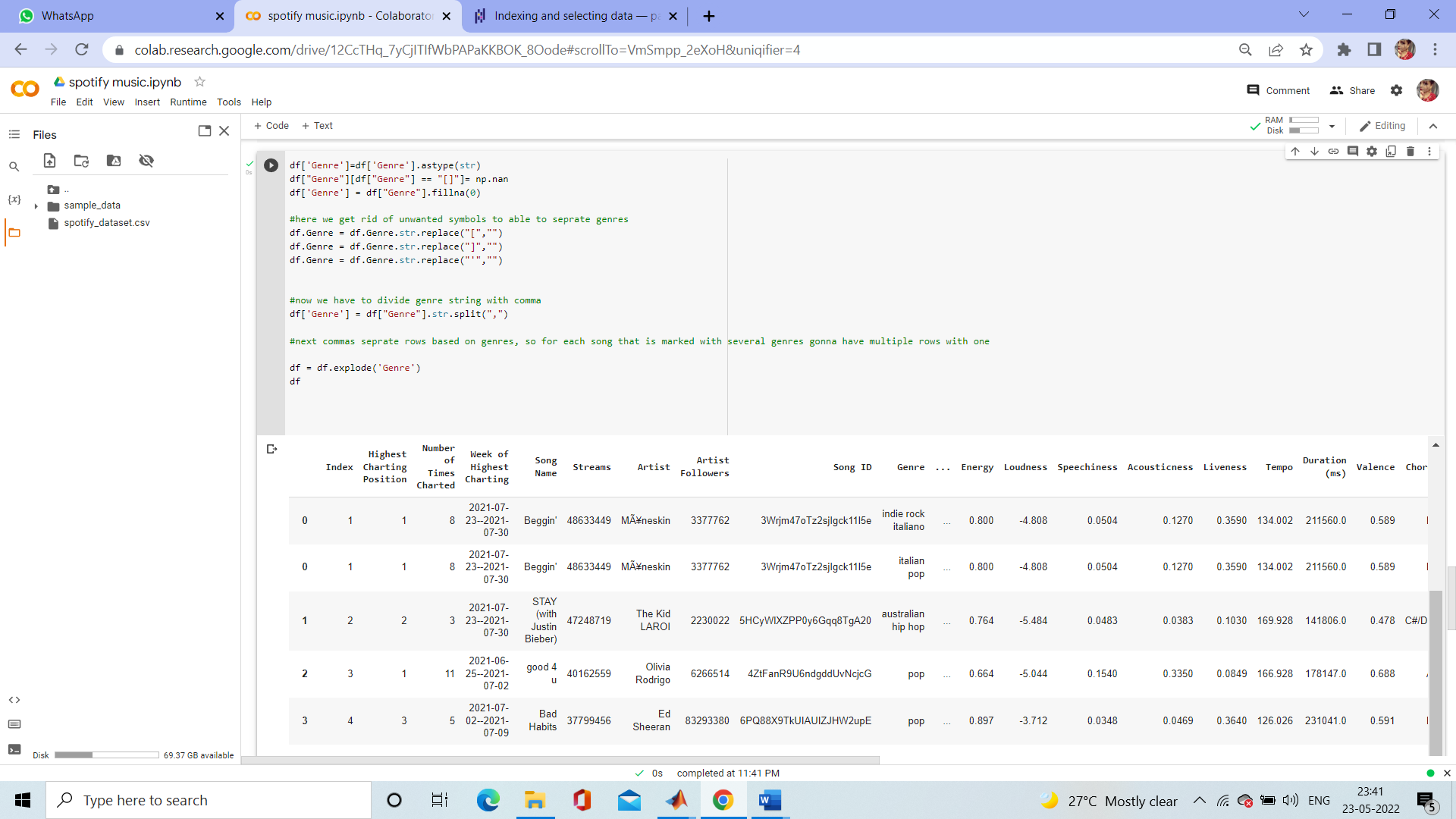
****

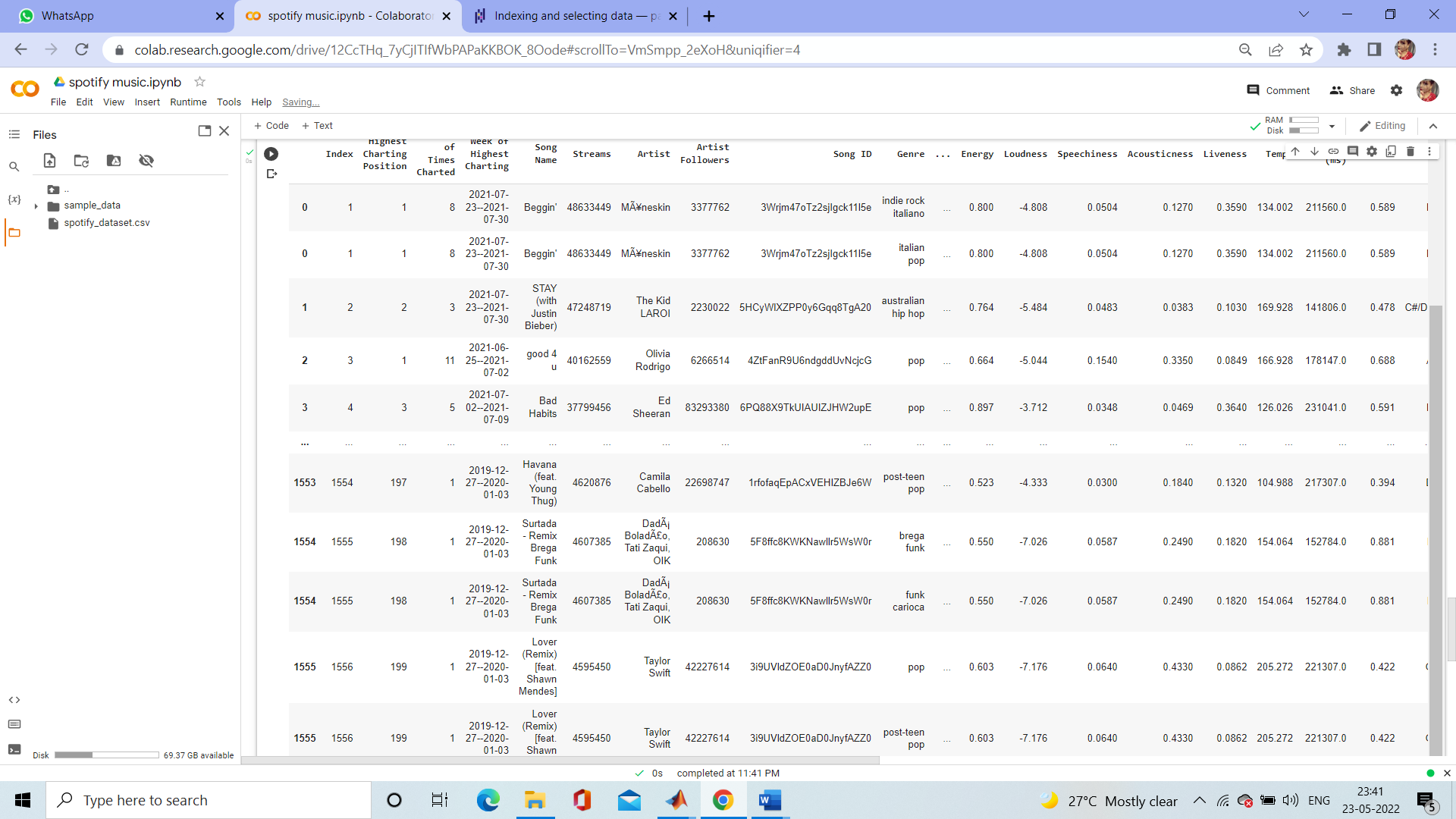
****

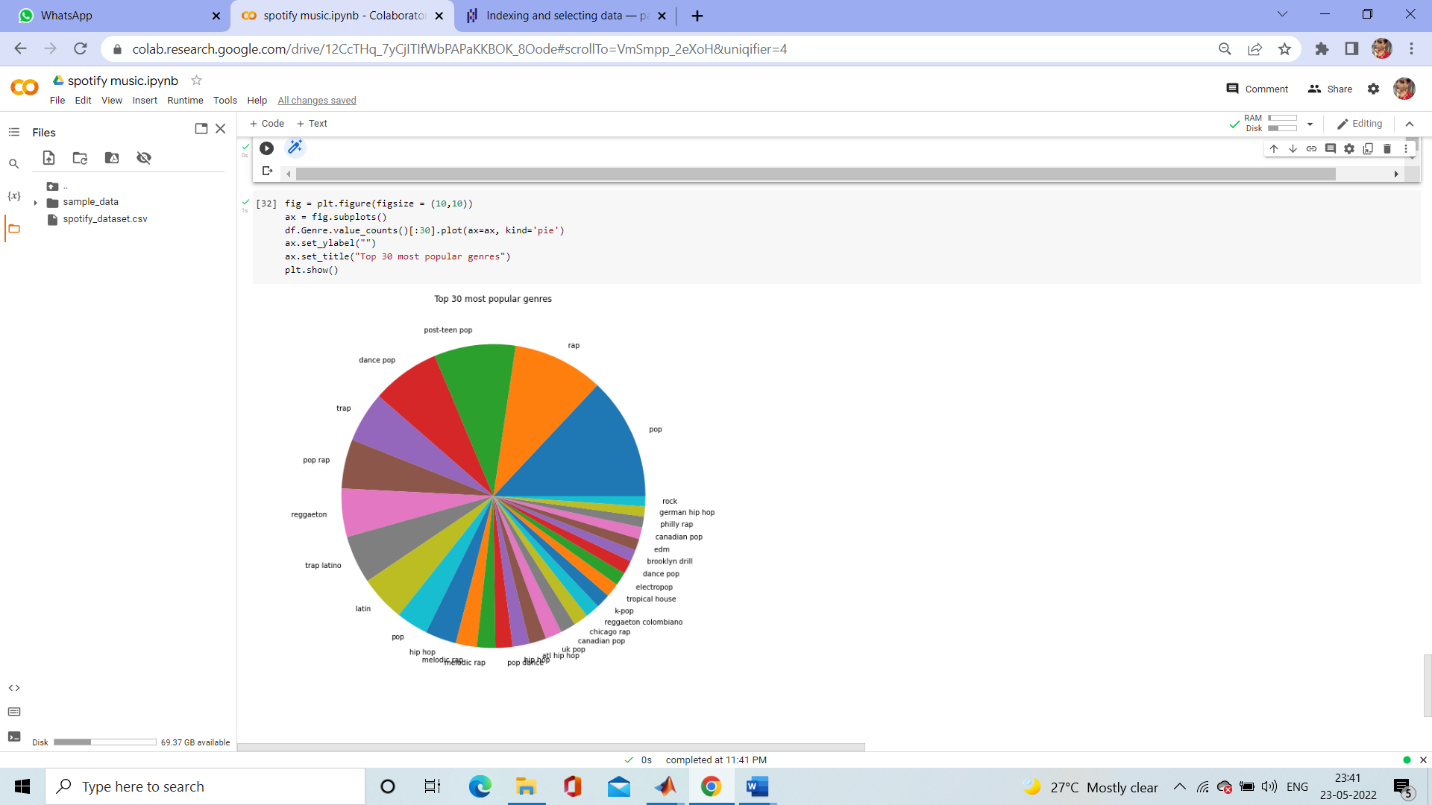
****

****

****

****

****

****

**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.

**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 .