A

PROJECT BASED LEARNING REPORT

On

### “Create a Spotify Music Analysis visualization using Python pandas”

Submitted in the partial fulfilment of the requirements for the

project based learning (PBL) in Semester-IV of Subject: Essentials of Data Science

in

Electronics & Communication Engineering By

|  |  |
| --- | --- |
| SAHIL PAWAR (96) | 2314110774 |
| MANISH SHIVAKOTI (97) | 2314110842 |
| AHAMAD TAMBOLI (93) | 2314110772 |

Under the guidance of Course In-Charge Prof. V.P.KADUSKAR



Department of Electronics & Communication Engineering Bharati Vidyapeeth

(Deemed to be University)

College of Engineering, Pune-411004

Academic Year: 2023-24

Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune – 411043

DEPARTMENT OF ELECTRONICS & COMMUNICATION

ENGINEERING

# CERTIFICATE

Certified that the Project Based Learning report **“Create a Spotify Music Analysis visualization using Python pandas”**

is work done by

|  |  |
| --- | --- |
| SAHIL PAWAR (96) | 2314110774 |
| MANISH SHIVAKOTI (97) | 2314110842 |
| AHAMAD TAMBOLI (93) | 2314110772 |
|  |  |

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 Department of Electronics and Communication Engineering.

Date:

Prof. V.P. Kaduskar

Dr. Arundhati A. Shinde

Course In-charge Professor & Head

# ACKNOWLEDGEMENT

We extend our deepest gratitude to all those who have contributed to the successful completion of this ―Create a Spotify Music Analysis visualization using Python pandas‖.

First and foremost, I am deeply thankful to our project supervisor, Mr.V.P. Kaduskar, for his invaluable guidance, encouragement, and support throughout the duration of this project. His expertise, constructive feedback, and dedication played a crucial role in shaping the project's direction and ensuring its successful completion.

Furthermore, we would like to acknowledge the support and assistance provided by our classmates, friends, and fellow students who contributed to the project in various capacities, whether through brainstorming sessions, feedback, or moral support.

We are also grateful to the faculty and staff of Bharati Vidyapeeth College of Engineering for their administrative support, access to resources, and encouragement throughout the project.

Lastly, we would like to express my appreciation to the authors of relevant literature, research papers, and open-source resources that served as valuable references and inspiration for our project.

In conclusion, we are deeply indebted to everyone mentioned above for their contributions, support, and encouragement, which have been instrumental in the successful completion of our college project.

Sahil Pawar Manish Shivakoti Ahamad Tamboli

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr. No | Title | age No. |
| 1. | Introduction | 1 |
| 2. | Problem statement | 2-3 |
| 3. | About Spotify | 4-5 |
| 4. | Software Used | 6-7 |
| 5. | Result | 8-14 |
| 6. | Outcome & Conclusion | 15 |

# Introduction:

In the age of digital streaming, music consumption has become not only widespread but also meticulously tracked. Services like Spotify provide a treasure trove of data, offering insights into listeners' preferences, trends, and behaviors. Harnessing the power of Python and the data manipulation capabilities of pandas, we embark on a journey to dissect Spotify's vast musical landscape.

This tutorial delves into the process of extracting, analyzing, and visualizing Spotify's music data using Python's popular data manipulation library, pandas. By leveraging pandas' intuitive syntax and powerful functionality, we can unlock fascinating insights into genres, artist popularity, song attributes, and more. Whether you're a data enthusiast, music aficionado, or simply curious about the intersection of technology and music, this tutorial offers a comprehensive guide to unraveling Spotify's musical tapestry. Let's dive in and uncover the rhythms of the digital age.

However, behind the seamless user experience lies a treasure trove of data waiting to be explored. By tapping into this wealth of information, we can uncover valuable insights into listener preferences, emerging trends, and the ever-evolving nature of popular music. This is where Python and the versatile data analysis library, pandas, come into play.

In this tutorial, we embark on a journey to dissect Spotify's vast musical ecosystem using Python and pandas. We'll delve into the process of extracting data from the Spotify API, wrangling it into a usable format, and unleashing the analytical power of pandas to gain meaningful insights. From exploring the popularity of different genres to analyzing the acoustic attributes of top hits, we'll cover a range of topics to provide a comprehensive understanding of Spotify's music landscape.

Whether you're a data enthusiast looking to sharpen your skills, a music aficionado curious about the inner workings of streaming platforms, or simply someone intrigued by the intersection of technology and culture, this tutorial offers something for everyone. So, grab your headphones, fire up your Python environment, and let's embark on a journey to uncover the rhythms of the digital age.

# Problem Statement :

Create a Spotify Music Analysis visualization using Python pandas.

# Solution:

## What is Music Analysis?

Music analysis is the process of examining and understanding the structure, content, and characteristics of a piece of music. It involves breaking down music into its constituent elements such as melody, harmony, rhythm, form, instrumentation, and dynamics, and studying how these elements interact to create the overall musical experience. Music analysis can be performed for various purposes, including:

1. **Understanding**: Analyzing music helps musicians, composers, and listeners gain a deeper understanding of how music works, including its compositional techniques, stylistic features, and expressive qualities.
2. **Interpretation**: Analysis can aid in interpreting the meaning, emotion, and intentions behind a piece of music, providing insights into the composer's creative process and artistic choices.
3. **Performance**: Musicians often analyze music to inform their performance decisions, such as phrasing, articulation, and expression, to convey the intended musical expression effectively.
4. **Composition**: Composers use analysis to study existing music as inspiration and to learn from established techniques and conventions, helping them develop their own compositional style.
5. **Research**: Musicologists and scholars analyze music to conduct research on variousaspects of musical theory, history, culture, and aesthetics, contributing to the academicunderstanding of music.

## What is Music Visualization?

usic visualization refers to the creation of visual representations or animations that accompany and enhance the experience of listening to music. It involves synchronizing visualelements, such as graphics, animations, or light patterns, with the audio content of the music in real-time or pre-defined sequences. Music visualization can take various forms, including:

1. **Audio-Driven Graphics**: Visual elements react to the amplitude, frequency, and dynamics of the music, creating patterns, shapes, and movements that correspond to the sound.
2. **Spectral Analysis**: Visualizing the frequency spectrum of the music, often using spectrograms or frequency histograms, to display the distribution of energy across different frequency bands.
3. **Beat Detection**: Detecting rhythmic patterns and beats in the music and synchronizing visual effects, such as pulsating lights or moving objects, with the rhythm to enhance the sense of groove and tempo.
4. **Interactive Visualizations**: Allowing users to interact with the visual elements in real- time, for example, by controlling parameters like color, shape, or movement using input devices or gestures.
5. **Immersive Environments**: Creating immersive experiences, such as 3D virtual environments or augmented reality scenes, where the music becomes the soundtrack for the visual world, enhancing the sense of immersion and engagement.

## What is Python Pandas?

Python pandas is a powerful open-source library used for data manipulation and analysis. It provides data structures and functions designed to make working with structured data, such as tabular or time series data, more intuitive and efficient. Some key features of pandas include:

1. **DataFrame**: The DataFrame is the primary data structure in pandas, resembling a table with rows and columns. It allows for easy indexing, selection, and manipulation of data.
2. **Series**: A Series is a one-dimensional labeled array capable of holding data of any type. It is often used to represent a single column or row of data in a DataFrame.
3. **Data Manipulation**: Pandas provides a wide range of functions for data manipulation, including merging and joining datasets, reshaping data, handling missing values, and performing operations across rows and columns.
4. **Data Input/Output**: Pandas supports reading and writing data from various file formats, such as CSV, Excel, SQL databases, JSON, and HDF5, making it easy to work with data from different sources.
5. **Data Analysis**: Pandas offers powerful tools for data analysis, including descriptive statistics, group-by operations, time series analysis, and data visualization integration with libraries like Matplotlib and Seaborn.
6. **Indexing and Selection**: Pandas provides flexible indexing and selection methods, allowing users to access data by label, integer position, boolean masks, or through advanced indexing techniques.

# Spotify Music:



**Spotify** [audio streaming](https://en.wikipedia.org/wiki/Streaming_media) and media services provider founded on 23 April 2006 by [Daniel](https://en.wikipedia.org/wiki/Daniel_Ek) [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](https://en.wikipedia.org/wiki/Record_label) [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*](https://en.wikipedia.org/wiki/The_New_York_Times)[*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 Jupyter Notebook

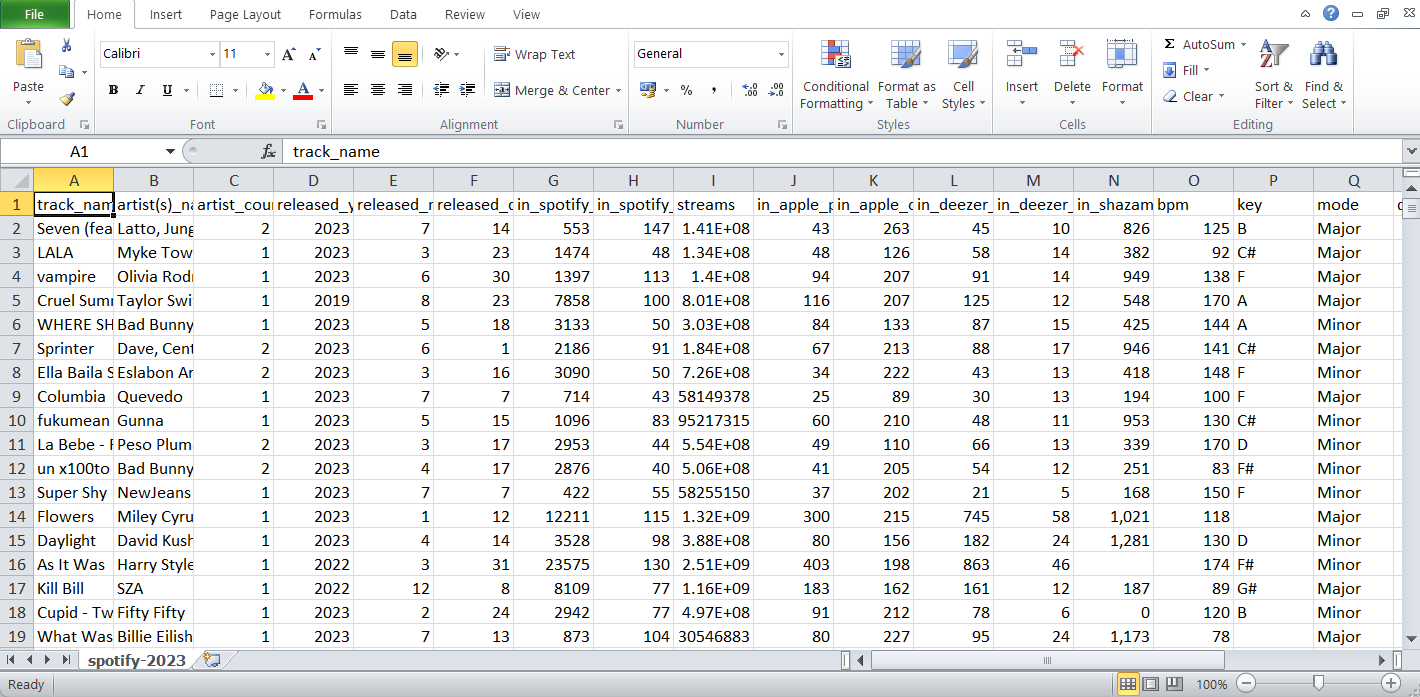


Fig. (1) Data set in Excel

## 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 isualization tools such as Dash and Chart Studio. It has also developed open source graphing Application Programming Interface (API) libraries for Python.

## Software used Jupyter Notebook



I assume you meant “Jupyter software.” Jupyter is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, including Python, R, Julia, and Scala, but Python is the most commonly used.

## Here are some key components and features of Jupyter:

1. Notebooks: Jupyter Notebooks are interactive documents that contain code cells, text cells, and multimedia elements. They allow you to write and execute code in real-time, view the output, and include explanatory text using Markdown formatting.
2. Support for Multiple Kernels: Jupyter supports multiple programming languages through its kernel system. Each kernel allows you to execute code written in a specific language within the same notebook interface.
3. Rich Output: Jupyter Notebooks can display rich output, including HTML, images, videos, LaTeX equations, and interactive visualizations created using libraries like Matplotlib, Seaborn, Plotly, and Bokeh.
4. Collaboration and Sharing: Jupyter Notebooks can be easily shared with others via email, Dropbox, GitHub, or through the Jupyter Notebook Viewer. They can also be converted to various formats, such as HTML, PDF, and slideshows, for easier sharing and collaboration.

## What Jupyter Notebook Offer Us?

### As a programmer, Jupyter software offers several capabilities and tasks you can perform:

1. Data Exploration and Analysis: You can use Jupyter Notebooks to explore datasets, analyze data using libraries like pandas, NumPy, and SciPy, and visualize the results with Matplotlib, Seaborn, Plotly, or other plotting libraries.
2. Prototyping and Testing Code: Jupyter Notebooks provide an interactive environment for prototyping and testing code snippets or algorithms. You can quickly iterate on ideas, experiment with different approaches, and see immediate results.
3. Machine Learning and Data Science: Jupyter is widely used in machine learning and data science workflows. You can train machine learning models, evaluate their performance, and tune hyperparameters—all within the same notebook interface. Libraries like scikit-learn, TensorFlow, and PyTorch are commonly used for these tasks.
4. Scientific Computing: Jupyter is popular in scientific computing for solving mathematical and scientific problems. You can perform numerical simulations, solve differential equations, and conduct computational experiments using libraries like SymPy and SciPy.
5. Education and Teaching: Jupyter Notebooks are valuable for teaching programming concepts and data analysis techniques. You can create interactive tutorials, exercises, and demonstrations, allowing students to engage with the material actively.
6. Documentation and Reports: Jupyter Notebooks serve as a platform for documenting your code, analyses, and research findings. You can combine code, text, visualizations, and explanations in a single document, making it easier to communicate your work to others.
7. Web Development and Deployment: Jupyter can be used for web development tasks, such as creating interactive dashboards, prototypes, or data-driven applications. You can convert notebooks to HTML or deploy them as interactive web applications using tools like Voilà.
8. Collaboration and Sharing: Jupyter facilitates collaboration among team members or the broader community. You can share your notebooks with others, receive feedback, and collaborate on projects in real-time using platforms like GitHub or JupyterHub.

## 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.



Fig. (1) Data Loading and Input :

You can use libraries like Pandas to read data from CSV, Excel, or other file formats. You can use libraries like SQLAlchemy to connect to databases and retrieve data

Exploring Spotify music data can involve various aspects, including the analysis of features provided by Spotify for each track, understanding user preferences, and identifying trends in the music industry.

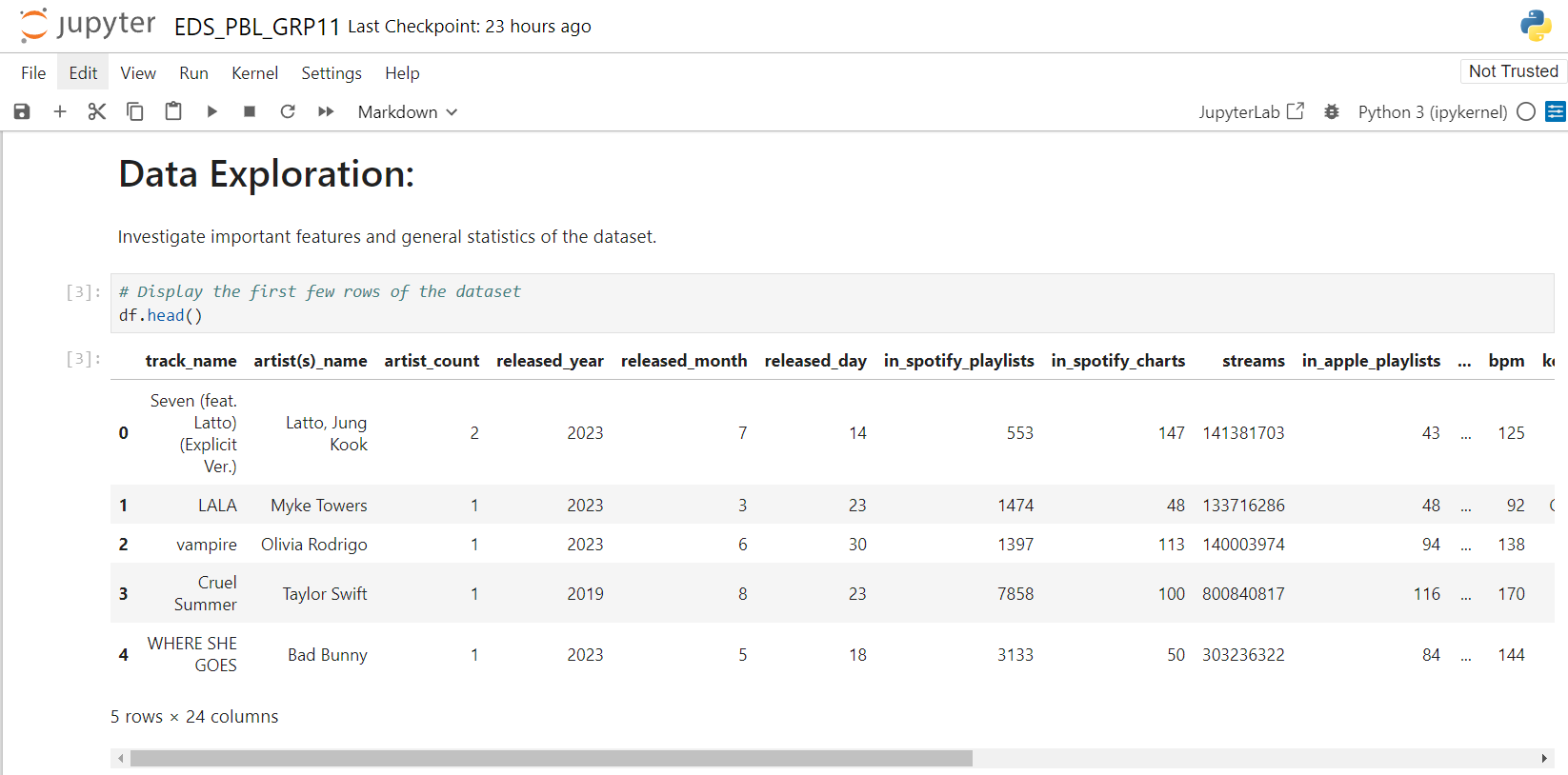
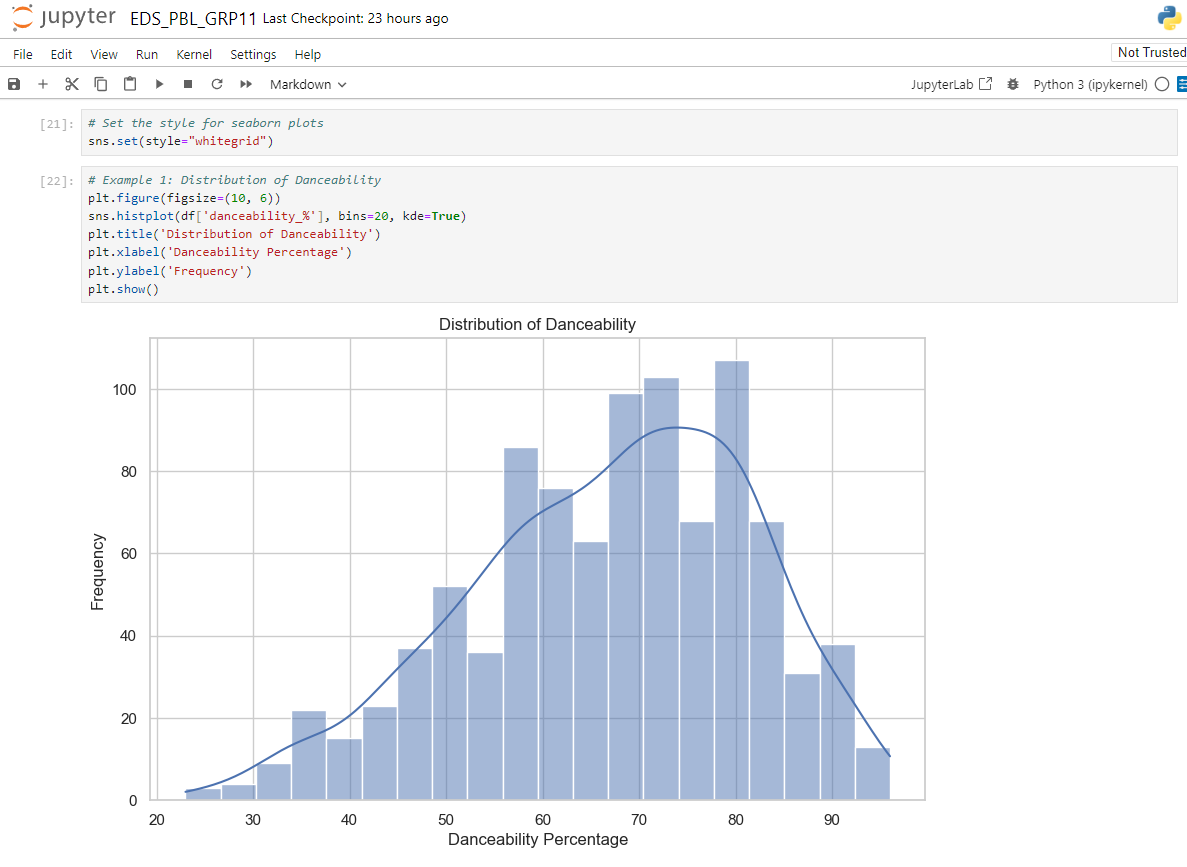


Fig. (2) Data Exploration :



**Fig. (3) Distribution of Danceability**

Analyzing the distribution of danceability on Spotify can provide insights into the general trends and preferences of listeners regarding the rhythmic qualities of music. Here's how you might approach this analysis:

**Data Collection**: Obtain a dataset containing information about songs on Spotify, including their danceability scores. You can use the Spotify API or third-party datasets available online.

**Exploratory Data Analysis (EDA)**:

Start by examining summary statistics of the danceability feature, such as mean, median, minimum, maximum, and standard deviation.

Create visualizations like histograms or box plots to visualize the distribution of danceability scores

.

**Interpretation**:

Identify any patterns or trends in the distribution of danceability scores.

**Visualization and Reporting**:

Present your findings visually using plots and charts. This could include bar charts comparing danceability across genres or time periods, scatter plots showing the relationship between danceability and other variables, or heatmaps illustrating correlations between danceability and other song features.

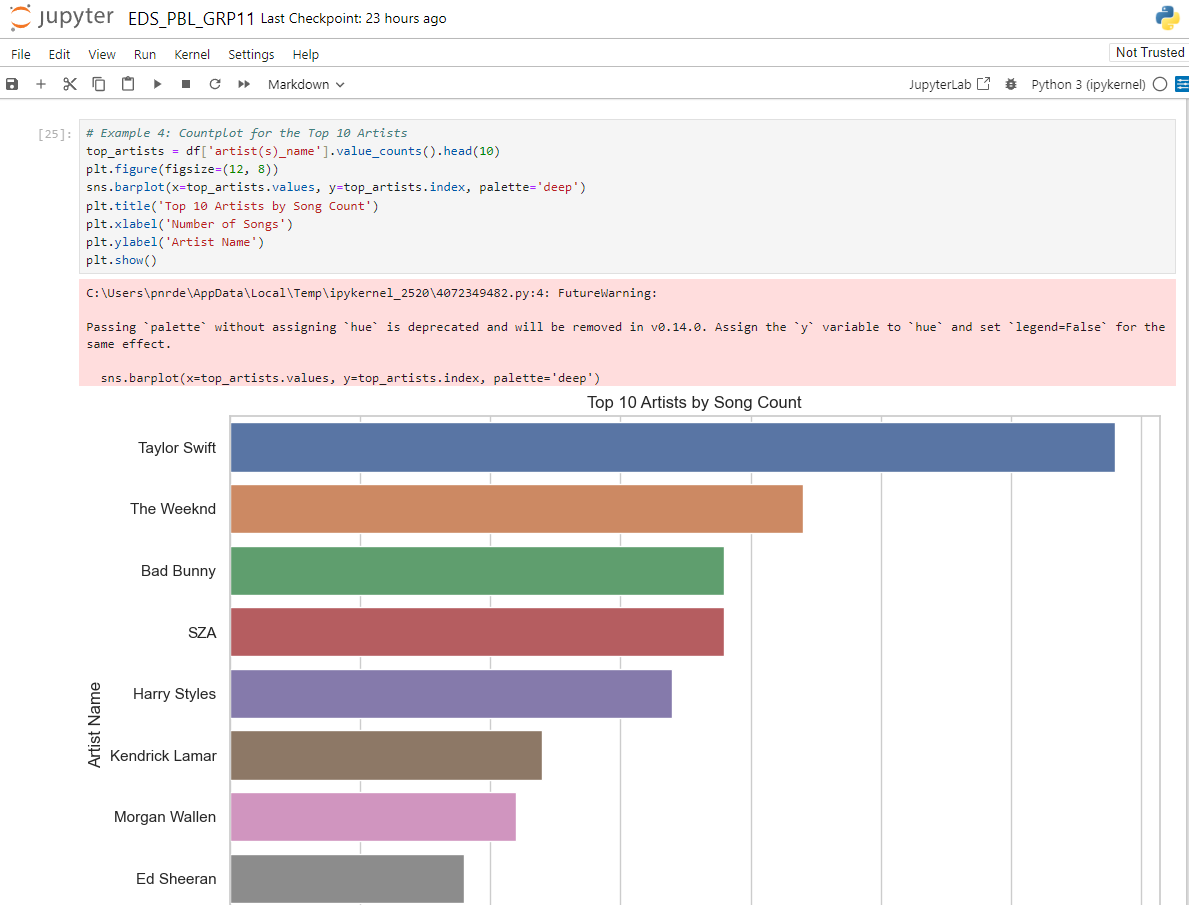
****

**Fig. (4) Correlation Heatmap:**

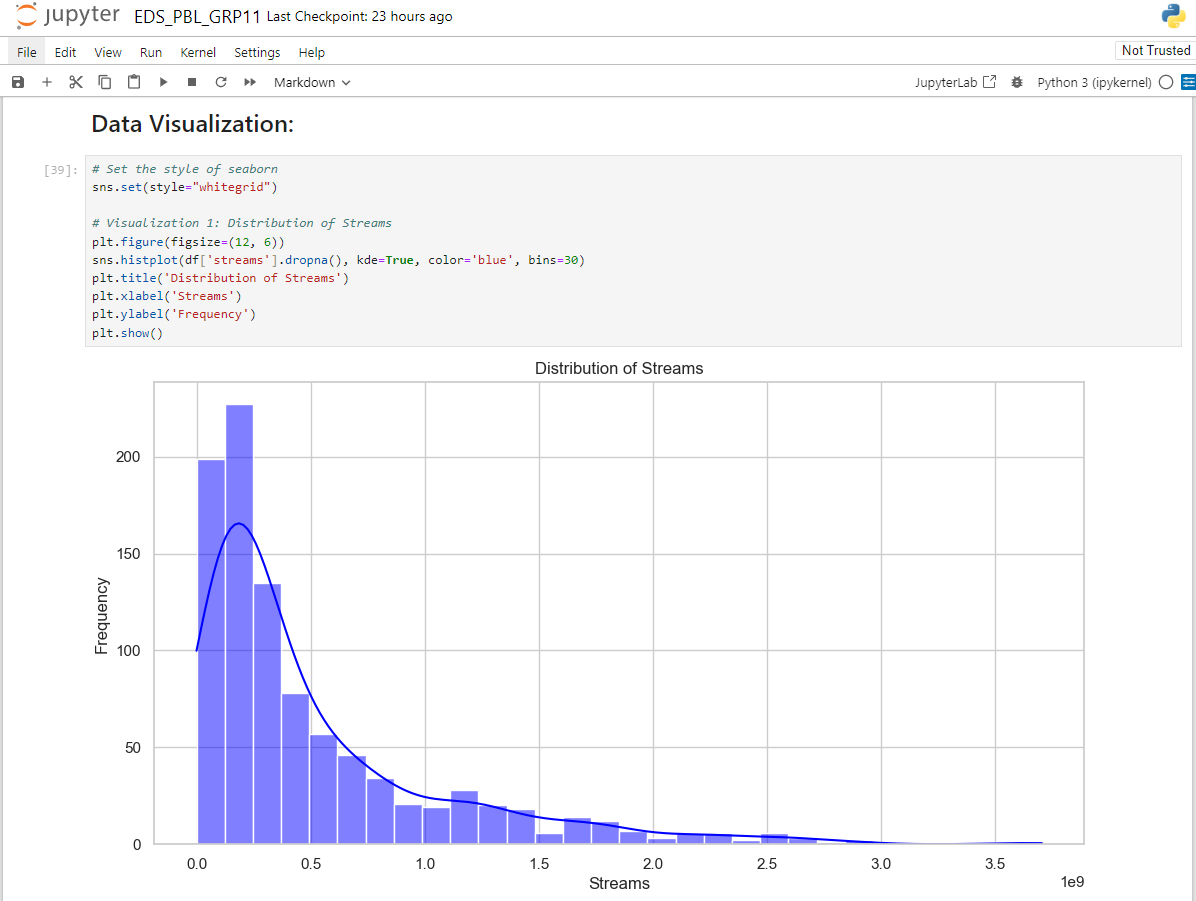
Creating a correlation heatmap is a great way to visualize the relationships between different variables in your dataset, including danceability and other features of songs on Spotify. Here's how you can do it using Python with libraries like Pandas, Matplotlib, and Seaborn.



**Fig. (5) Streams vs Energy with Color by Speechiness**



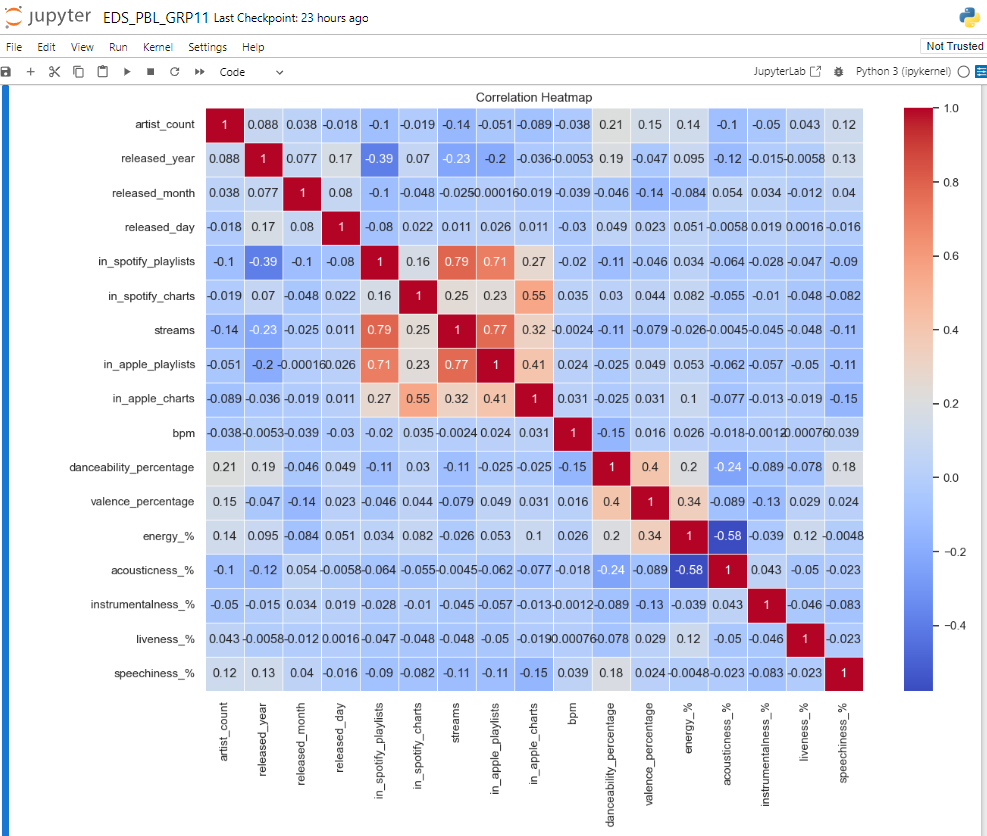
**Fig. (6) Top 10 Artist By Song Count**



**Fig. (7) Data Visualization**

Data visualization is the graphical representation of data and information. It uses visual elements such as charts, graphs, and maps to communicate insights and patterns in the data. The primary goal of data visualization is to make complex datasets more understandable, accessible, and actionable.

* Understanding Complex Data
* Effective Communication
* Effective Communication
* Storytelling
* Decision Support



**Fig. (8) Correlation Heatmap**



**Fig. (9) Scatter Plot of Danceability Percentage vs. Percentage**

## 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 Python Pandas used to analysis and visulization results to the solution approach followed and assessing the solution approach.

## Project Conclusion:

From this project, we gained the knowledge of software – Python Pandas(Jupiter Notebook). We learnt to analyze the datasets and afterwards, visualizing them. We learnt about various plots .