**Spotify Data Analysis using Python**

**Project Overview**

**Spotify is a leading audio streaming service with over 381 million monthly active users, including 172 million paid subscribers. This project focuses on exploring and analyzing Spotify’s dataset to gain insights into music trends and features.**

**Objectives**

* **Data Cleaning & Preparation: Clean and preprocess the dataset for analysis.**
* **Exploratory Data Analysis (EDA): Conduct EDA to understand the dataset’s structure and key characteristics.**
* **Data Visualization: Create visualizations to represent data insights effectively.**
* **Feature Analysis: Investigate relationships between audio features and lyric sentiments.**
* **Pattern Identification: Identify and interpret patterns and correlations in the data.**

**Tools & Technologies**

* **Python Libraries: Pandas, NumPy, Matplotlib, Seaborn**
* **Environment: Jupyter Notebook or any Python IDE**

**Dataset**

**The analysis utilizes secondary data from Spotify, including:**

* **Audio Features: Attributes like danceability, energy, acousticness, etc.**
* **Lyrics Sentiment: Positive or negative sentiment scores of song lyrics.**

**Methodology**

1. **Data Collection: Source data from Spotify’s dataset.**
2. **Data Cleaning: Handle missing values, outliers, and inconsistencies.**
3. **Exploratory Data Analysis (EDA):**
   * **Analyze summary statistics.**
   * **Explore distributions and correlations.**
4. **Data Visualization:**
   * **Create charts (e.g., histograms, scatter plots, heatmaps) to visualize data distributions and relationships.**
5. **Feature Analysis:**
   * **Examine how audio features correlate with lyric sentiment.**
   * **Use sentiment analysis tools to quantify lyric emotions.**
6. **Pattern Identification:**
   * **Discover trends and patterns in audio features and sentiment.**
   * **Generate insights about music trends and listener preferences.**

* **Spotify Dataset: [Link to dataset or source]**
* **Python Libraries Documentation: Links to Pandas, NumPy, Matplotlib, Seaborn documentation**

**Prerequisite:** [Data Analyst Roadmap](https://github.com/mrankitgupta/Data-Analyst-Roadmap) ⌛ , [Python Lessons](https://github.com/mrankitgupta/PythonLessons) 📑 & [Python Libraries for Data Science](https://github.com/mrankitgupta/PythonLibraries) 🗂️

**Technologies used ⚙️**

* [Python](https://github.com/mrankitgupta/Python-Lessons)
* [Statistics](https://github.com/mrankitgupta/Statistics-for-Data-Science-using-Python)

**Python Libraries :**

* [Pandas](https://github.com/mrankitgupta/Kaggle-Pandas-Solved-Exercises)| [NumPy](https://numpy.org/)| [Matplotlib](https://matplotlib.org/)| [Seaborn](https://seaborn.pydata.org/)

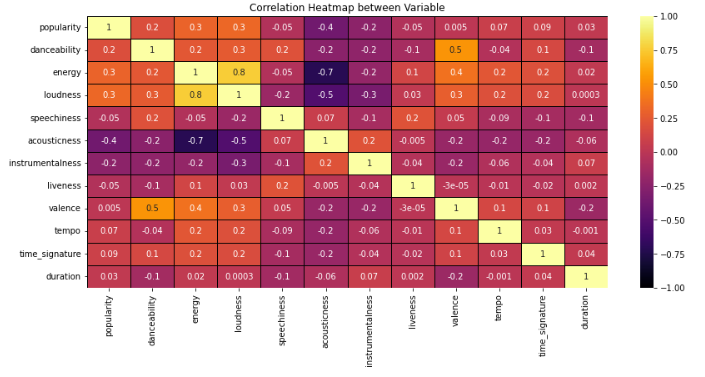
**Project - Spotify Data Analysis using Python**

[**Kaggle Project: Spotify Data Analysis**](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python) [🔗](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

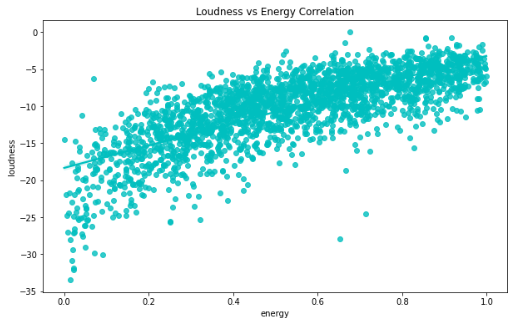
[Kaggle Spotify Datasets:](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python-project/data) [Spotify Tracks](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python/data?select=tracks.csv) & [Spotify Features](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python/data?select=SpotifyFeatures.csv)

**Objective**

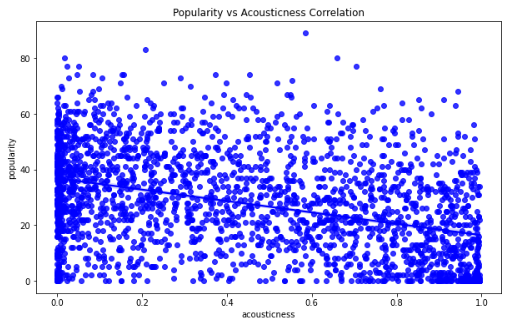
1. **Top 10 most popular songs on Spotify**
2. **Top 10 least popular songs on Spotify**
3. **Correlation Heatmap between Variable**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

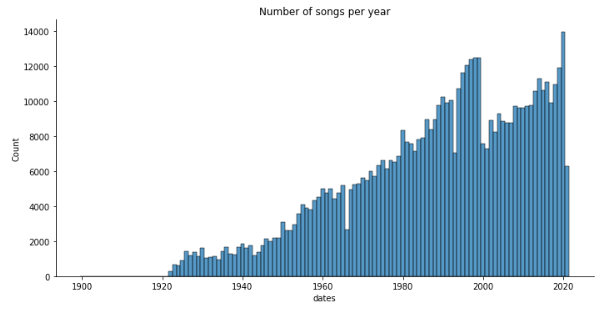
1. **Regression plot - Correlation between Loudness and Energy**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

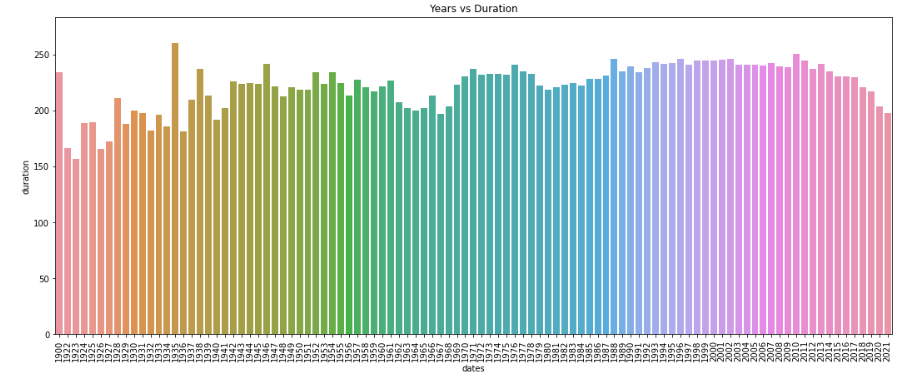
1. **Regression plot - Correlation between Popularity and Acousticness**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

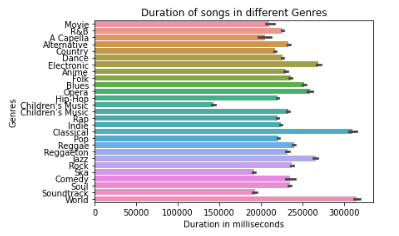
1. **Distibution plot - Visualize total number of songs on Spotify since 1992**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

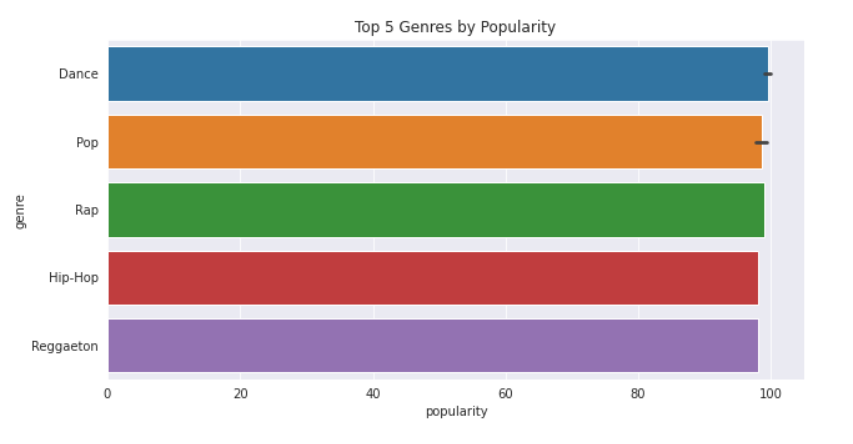
1. **Change in Duration of songs wrt Years**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

1. **Duration of songs in different Genres**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

1. **Top 5 Genres by Popularity**

[](https://www.kaggle.com/code/mrankitgupta/spotify-data-analysis-using-python)

**Results**

* **Key Findings: Summarize the main insights from the data, such as notable correlations or trends.**
* **Visualizations: Include relevant charts and graphs that support the findings.**

**Conclusion**

**The analysis reveals significant patterns in Spotify’s dataset, providing insights into how audio features and lyrics sentiment influence music trends. This project enhances skills in data cleaning, analysis, and interpretation.**

**Future Work**

* **Further Analysis: Explore additional datasets or deeper analysis into specific music genres.**
* **Advanced Techniques: Apply machine learning models for predictive analytics.**

**Related Projects:question: 👨‍💻 🛰️**

[Data Analyst Roadmap](https://github.com/mrankitgupta/Data-Analyst-Roadmap) ⌛

[Sales Insights - Data Analysis using Tableau & SQL](https://github.com/mrankitgupta/Sales-Insights-Data-Analysis-using-Tableau-and-SQL) 📊

[Statistics for Data Science using Python](https://github.com/mrankitgupta/Statistics-for-Data-Science-using-Python) 📊

[Kaggle - Pandas Solved Exercises](https://github.com/mrankitgupta/Kaggle-Pandas-Solved-Exercises) 📊