

Minor Project 1– EDA

Given by Rinex

Name- Ritam Sarkar

College- Kalyani Government Engineering College

Year- 3rd

Course- Machine Learning with Python

Exploratory Data Analysis with Spotify Dataset

The screenshot shows the EDA.ipynb notebook interface. The code cell contains the following Python code:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

sns.set_style("darkgrid") #setting the style

df = pd.read_csv("~/content/dataset.csv") #loading the dataset

df
```

The output of the `df` command is a preview of the dataset, showing columns: genre, artist_name, track_name, track_id, popularity, acousticness, danceability, duration_ms, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time_signature, and valence. The preview shows rows 0 through 4064 and 4065.

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKKJfDqeFgVWV	0	0.611	0.389	99373	0.910	0.0000	C#	0.3460	-1.828	Major	0.0525	166.969	4/4	
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BJC1NfoEOOusryehmNudP	1	0.246	0.590	137373	0.737	0.0000	F#	0.1510	-5.559	Minor	0.0868	174.003	4/4	
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	170267	0.131	0.0000	C	0.1030	-13.879	Minor	0.0362	99.488	5/4	
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tivf	0	0.703	0.240	152427	0.326	0.0000	C#	0.0985	-12.178	Major	0.0395	171.758	4/4	
4	Movie	Fabien Natat	Ouverture	0lusiXpMROHdEPvSI1fTQK	4	0.950	0.331	82625	0.225	0.1230	F	0.2020	-21.150	Major	0.0456	140.576	4/4	
...
4064	Country	Whiskey Myers	Good Ole Days (feat. Brent Cobb)	0BJj79eZoWPD9Cg8pr04m	40	0.706	0.591	237791	0.538	0.0000	A#	0.1370	-8.099	Major	0.0513	135.797	4/4	
4065	Country	My Morning Jacket	Rocket Man	3q6BJcc0i7UyM8oWn6ntRQ	53	0.534	0.565	299827	0.270	0.0362	D#	0.1070	-9.591	Major	0.0302	120.308	4/4	

The screenshot shows the EDA.ipynb notebook interface. The code cell contains the following Python code:

```
df.isna().sum() #checking for missing values
```

The output of the `df.isna().sum()` command is a series of zeros for all columns, indicating no missing values.

```
genre      0
artist_name 0
track_name 0
track_id   0
popularity 0
acousticness 0
danceability 0
duration_ms 0
energy      0
instrumentalness 0
key         0
liveness    0
loudness    0
mode        0
speechiness 0
tempo       0
time_signature 0
valence     0
dtype: int64
```

The code cell also contains the following Python code:

```
df.info()
```

The output of the `df.info()` command is the following information:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4069 entries, 0 to 4068
Data columns (total 18 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   genre               4069 non-null  object
 1   artist_name         4069 non-null  object
 2   track_name          4069 non-null  object
 3   track_id            4069 non-null  object
 4   popularity           4069 non-null  int64
 5   acousticness        4069 non-null  float64
 6   danceability         4069 non-null  float64
 7   duration_ms         4069 non-null  int64
 8   energy              4069 non-null  float64
 9   instrumentalness     4069 non-null  float64
10   key                 4069 non-null  object
11   liveness            4069 non-null  float64
12   loudness            4069 non-null  float64
13   mode                4069 non-null  object
```

EDA.ipynb - Colaboratory — Mozilla Firefox

EDA.ipynb - Colaboratory x EDA.ipynb - Colaboratory x New Tab x +

https://colab.research.google.com/github/ritamsarkar2000/rinex-training-and-internship/blob/main/minor_project1/EDA.ipynb

Getting Started | Inbox | ritam.formal... | happy new year - Go...

EDA.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

Connect Editing

```
[ ] df.shape
```

```
(4869, 18)
```

```
[ ] df.columns
```

```
Index(['genre', 'artist_name', 'track_name', 'track_id', 'popularity',
       'acousticness', 'danceability', 'duration_ms', 'energy',
       'instrumentalness', 'key', 'liveness', 'loudness', 'mode',
       'speechiness', 'tempo', 'time_signature', 'valence'],
      dtype='object')
```

```
[ ] len(df.columns) #number of columns
```

```
18
```

```
[ ] df.describe()
```

	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence
count	4069.000000	4069.000000	4069.000000	4.069000e+03	4069.000000	4069.000000	4069.000000	4069.000000	4069.000000	4069.000000	4069.000000
mean	46.310641	0.271402	0.560013	2.277946e+05	0.632455	0.030355	0.189895	-7.557801	0.080694	120.562065	0.493221
std	19.349293	0.306988	0.144688	1.327858e+05	0.232826	0.128343	0.148311	3.730657	0.107474	30.376112	0.231483
min	0.000000	0.000001	0.061700	1.880000e+04	0.001540	0.000000	0.021800	-29.368000	0.022400	32.244000	0.000000
25%	41.000000	0.012900	0.470000	1.881480e+05	0.466000	0.000000	0.098100	-9.338000	0.032500	95.223000	0.312000
50%	52.000000	0.129000	0.566000	2.162340e+05	0.664000	0.000006	0.131000	-6.605000	0.043200	118.989000	0.483000
75%	60.000000	0.476000	0.657000	2.482210e+05	0.827000	0.000528	0.241000	-4.938000	0.074200	141.514000	0.672000
max	83.000000	0.995000	0.971000	3.631469e+06	0.998000	0.984000	0.976000	-0.521000	0.961000	216.029000	0.983000

```
[ ] #Analysis
```

```
#top 5 artist
```

```
top5a = df.groupby("artist_name").count().sort_values(by = "track_name", ascending = False)["track_name"][:5]
```

EDA.ipynb - Colaboratory — Mozilla Firefox

EDA.ipynb - Colaboratory x EDA.ipynb - Colaboratory x New Tab x +

https://colab.research.google.com/github/ritamsarkar2000/rinex-training-and-internship/blob/main/minor_project1/EDA.ipynb

Getting Started | Inbox | ritam.formal... | happy new year - Go...

EDA.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

Connect Editing

```
[ ] #Analysis
```

```
#top 5 artist
```

```
top5a = df.groupby("artist_name").count().sort_values(by = "track_name", ascending = False)["track_name"][:5]
```

```
top5a
```

```
artist_name
```

```
Chorus      102
```

```
Henri Salvador    88
```

```
The Singers Unlimited    36
```

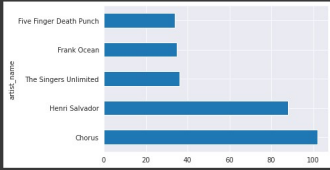
```
Frank Ocean      35
```

```
Five Finger Death Punch    34
```

```
Name: track_name, dtype: int64
```

```
[ ] top5a.plot.barh() #plotting the data
```

```
plt.show()
```

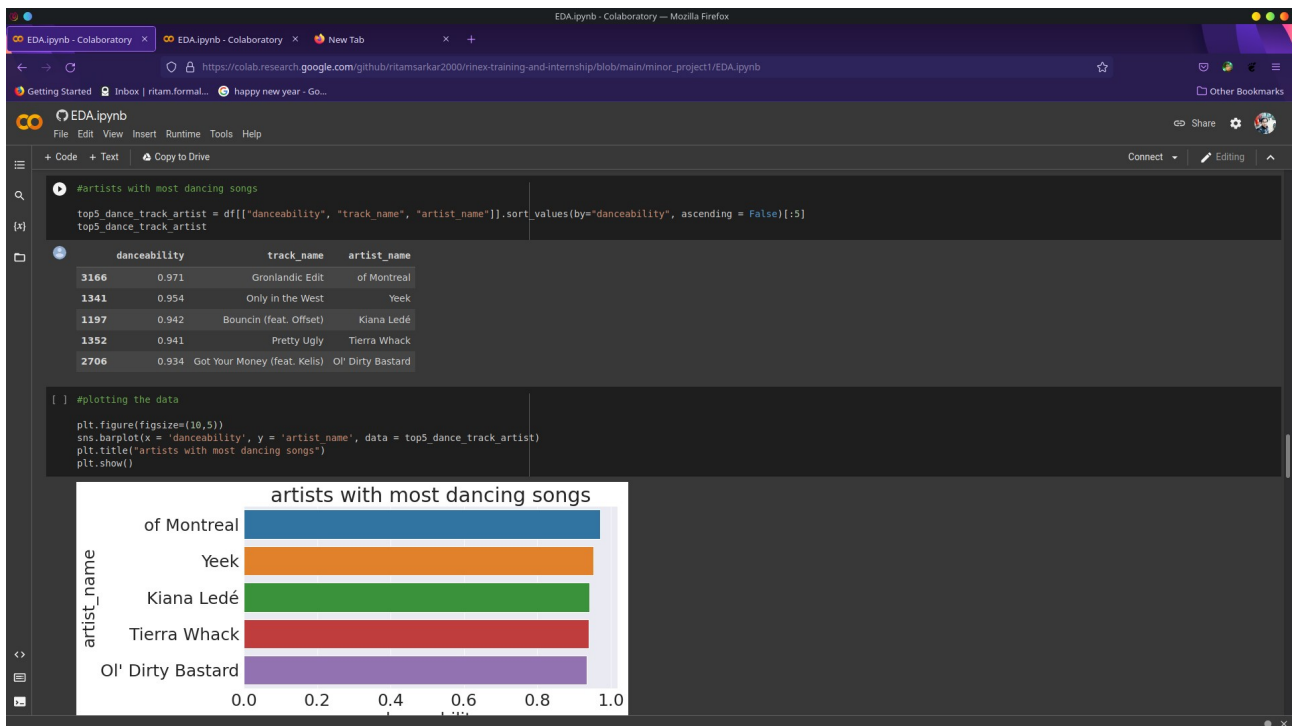
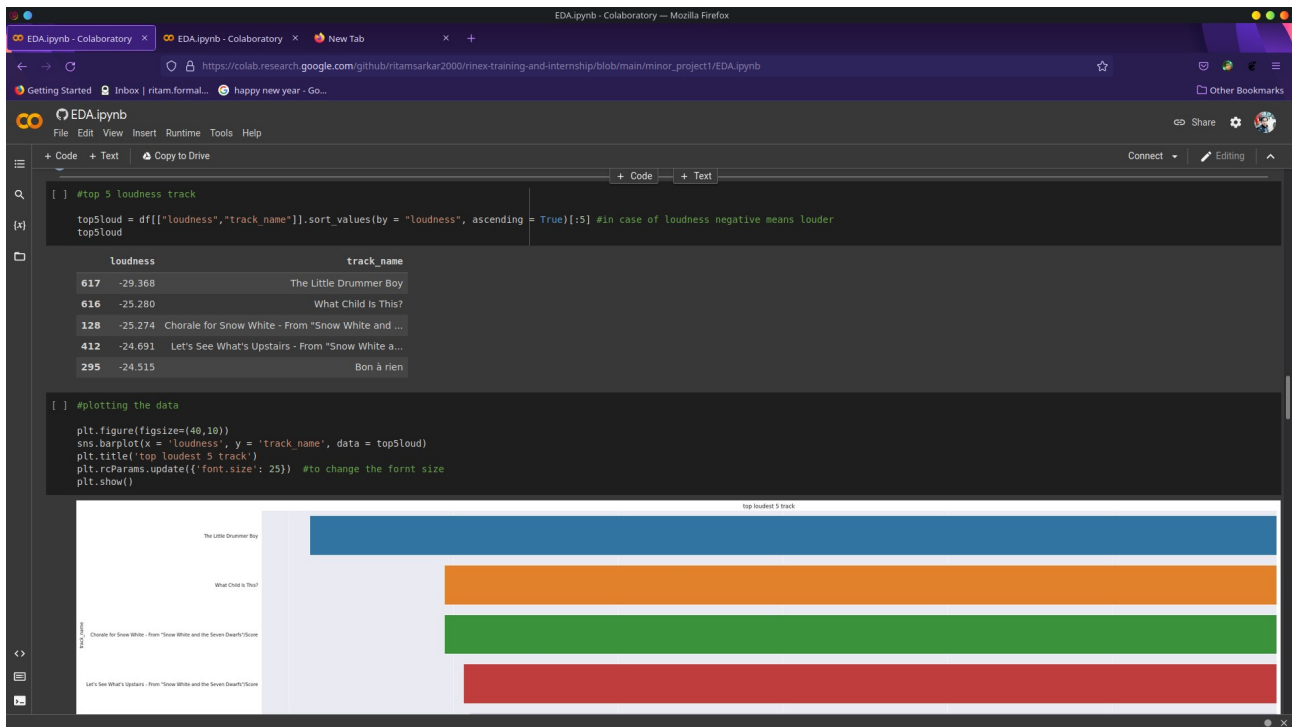


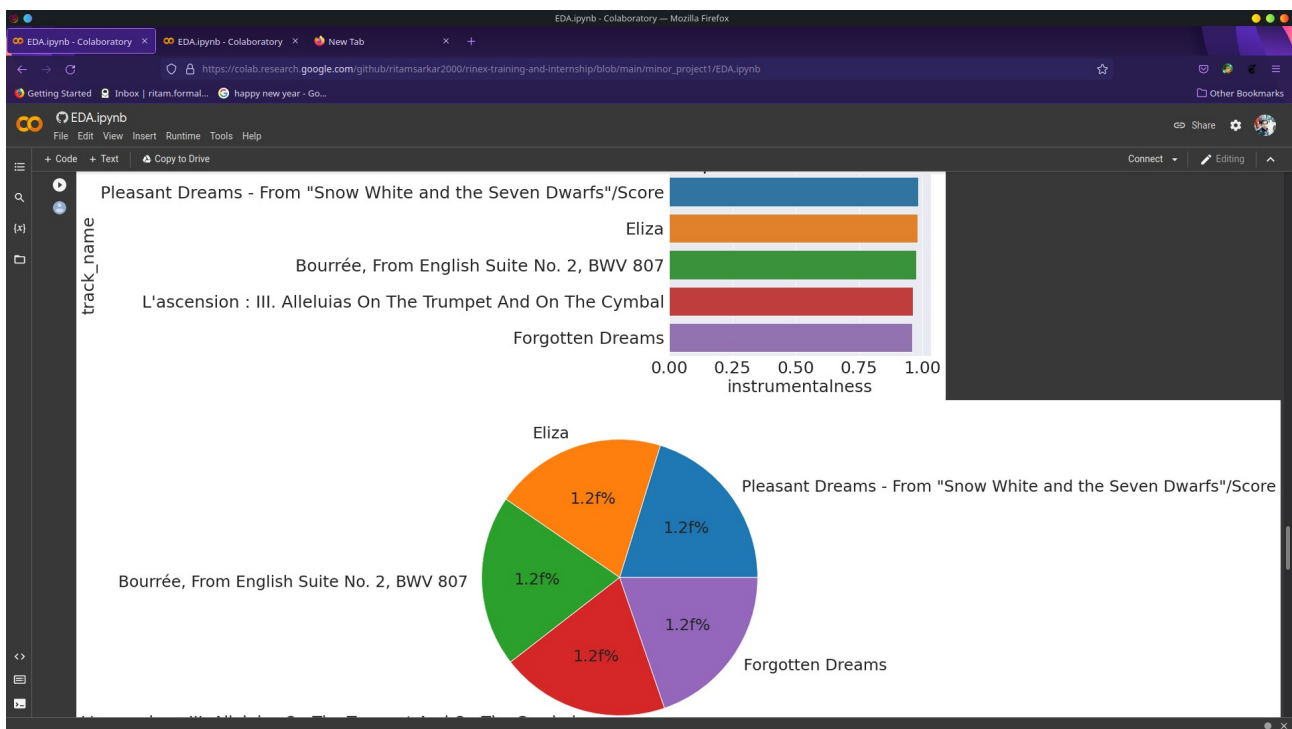
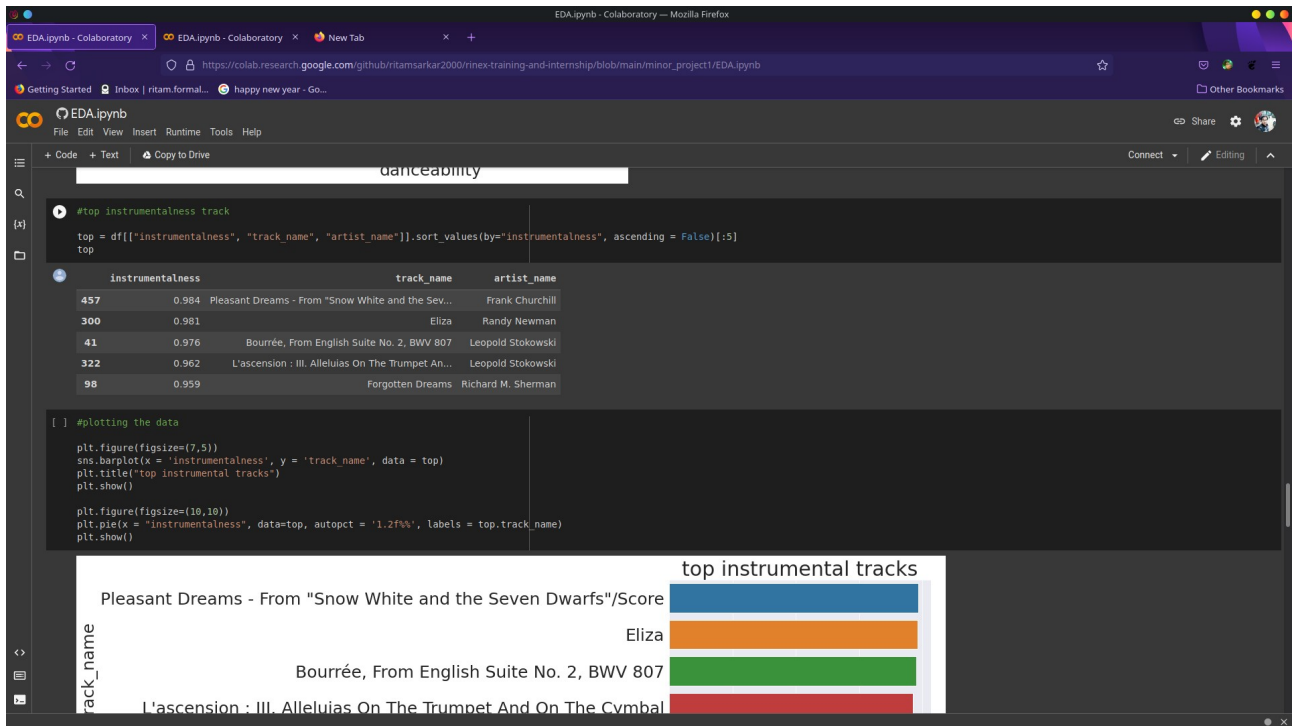
```
[ ] #top 5 loudness track
```

```
top5loud = df[["loudness", "track_name"]].sort_values(by = "loudness", ascending = True)[:5] #in case of loudness negative means louder
```

```
top5loud
```

loudness	track_name	
617	-29.368	The Little Drummer Boy
616	-26.760	What Child Is This?





EDA.ipynb - Colaboratory — Mozilla Firefox

EDA.ipynb - Colaboratory x EDA.ipynb - Colaboratory x New Tab x +

← → https://colab.research.google.com/github/ritamsarkar2000/rinex-training-and-internship/blob/main/minor_project1/EDA.ipynb ☆

Getting Started | Inbox | ritam.formal... | happy new year - Go... | Other Bookmarks

EDA.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive Connect Editing

Ascension : III. Alleluia On The Trumpet And On The Cymbal

```
[ ]
#longest tracks
top = df[['duration_ms', 'track_name', 'artist_name']].sort_values(by='duration_ms', ascending = False)[:5]
top
```

	duration_ms	track_name	artist_name
470	3631469	Pedda Gedda	Chorus
450	3475594	Sri Sankashta Chathurthi Vrata	Chorus
96	3435625	Koti Chennaya(Vol.1 & 2)	Chorus
15	2447870	Les aventures d'Arsène Lupin: 813 - Partie 6	Michel Roux
114	2359123	Les aventures d'Arsène Lupin: 813 - Partie 5	Michel Roux

```
[ ] #most common keys
top = df['key'].value_counts()[:5].index.tolist()
top
```

```
['G', 'C', 'D', 'A', 'C#']
```

Conclusion

1. top artists are:

```
Chorus,
Henri Salvador,
The Singers Unlimited,
Frank Ocean,
Five Finger Death Punch
```

EDA.ipynb - Colaboratory — Mozilla Firefox

EDA.ipynb - Colaboratory x EDA.ipynb - Colaboratory x New Tab x +

← → https://colab.research.google.com/github/ritamsarkar2000/rinex-training-and-internship/blob/main/minor_project1/EDA.ipynb ☆

Getting Started | Inbox | ritam.formal... | happy new year - Go... | Other Bookmarks

EDA.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive Connect Editing

Conclusion

1. top artists are:

```
Chorus,
Henri Salvador,
The Singers Unlimited,
Frank Ocean,
Five Finger Death Punch
```

2. Top 5 loudest tracks are:

```
The Little Drummer Boy,
What Child Is This?,
Chorale for Snow White - From "Snow White and ....
Let's See What's Upstairs - From "Snow White a...,
Bon à rien
```

3. Artists with most danceable songs are :

```
off Montreal,
Yeek,
Kiana Ledé,
Tierra Whack,
Ol' Dirty Bastard,
```

4. Most common keys are:

```
G, C, D, A, C#
```

```
Conclusion

1. top artists are:

Chorus,
Henri Salvador,
The Singers Unlimited,
Frank Ocean,
Five Finger Death Punch

2. Top 5 loudest tracks are:

The Little Drummer Boy,
What Child Is This?,
Chorus for Snow White - From "Snow White and ....",
Let's See What's Upstairs - From "Snow White a....",
Bon à rien

3. Artists with most danceable songs are :

off Montreal,
Yeek,
Kiana Ledé,
Tierra Whack,
Ol' Dirty Bastard,

4. Most common keys are:

G, C, D, A, C#
```

```
4. Most common keys are:

G, C, D, A, C#

5. longest tracks are:

Pedda Gedda,
Sri Sankashta Chaturthi Vrattha,
Koti Chennai(Vol.1 6amp; 2),
Les aventures d'Arsène Lupin: 813 - Partie 6,
Les aventures d'Arsène Lupin: 813 - Partie 5,

6. Most instrumental songs or songs with most instrumentality are:

Pleasant Dreams - From "Snow White and the Sev...",
Eliza,
Bourrée, From English Suite No. 2, BWV 807,
L'ascension - III. Alleluia On The Trumpet An...,
Forgotten Dreams,

Double-click (or enter) to edit

[ ]
```

Github link: https://github.com/ritamsarkar2000/rinex-training-and-internship/tree/main/minor_project1

link of the dataset: <https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db>

Conclusions:

1. top artists are:

Chorus,
Henri Salvador,
The Singers Unlimited,
Frank Ocean,
Five Finger Death Punch

2. Top 5 loudest tracks are:

The Little Drummer Boy,
What Child Is This?,
Chorale for Snow White - From "Snow White and ...",
Let's See What's Upstairs - From "Snow White a...",
Bon à rien

3. Artists with most danceable songs are :

off Montreal,
Yeek,
Kiana Ledé,
Tierra Whack,
Ol' Dirty Bastard,

4. Most common keys are:

G, C, D, A, C#

5. longest tracks are:

Pedda Gedda,
Sri Sankashta Chathurthi Vratha,
Koti Chennaya(Vol.1 & 2),
Les aventures d'Arsène Lupin: 813 - Partie 6,

Les aventures d'Arsène Lupin: 813 - Partie 5,

6. Most instrumental songs or songs with most instrumentality are:

Pleasant Dreams - From "Snow White and the Sev...",

Eliza,

Bourrée, From English Suite No. 2, BWV 807,

L'ascension : III. Alleluias On The Trumpet An...,

Forgotten Dreams,