

```
In [16]: import pandas as pd
import numpy as np
data = pd.read_csv("C:\\datasci\\spotify-2023.csv",encoding='latin-1')
data
```

Out[16]:

	track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_s
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	
1	LALA	Myke Towers	1	2023	3	23	
2	vampire	Olivia Rodrigo	1	2023	6	30	
3	Cruel Summer	Taylor Swift	1	2019	8	23	
4	WHERE SHE GOES	Bad Bunny	1	2023	5	18	
...
948	My Mind & Me	Selena Gomez	1	2022	11	3	
949	Bigger Than The Whole Sky	Taylor Swift	1	2022	10	21	
950	A Veces (feat. Feid)	Feid, Paulo Londra	2	2022	11	3	
951	En La De Ella	Feid, Sech, Jhayco	3	2022	10	20	
952	Alone	Burna Boy	1	2022	11	4	

953 rows × 24 columns



In [3]: data.head(6)

Out[3]:

(s)_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify
atto, Jung Kook	2	2023	7	14		553
ke Towers	1	2023	3	23		1474
a Rodrigo	1	2023	6	30		1397
aylor Swift	1	2019	8	23		7858
ad Bunny	1	2023	5	18		3133
e, Central Cee	2	2023	6	1		2186

In [5]: data.shape

Out[5]: (953, 24)

In [6]: data.isnull()

Out[6]:

	track_name	artist(s)_name	artist_count	released_year	released_month	released_day	ir
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	
948	False	False	False	False	False	False	
949	False	False	False	False	False	False	
950	False	False	False	False	False	False	
951	False	False	False	False	False	False	
952	False	False	False	False	False	False	

```
In [7]: data.dropna()
```

```
Out[7]:
```

	track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_s
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	
1	LALA	Myke Towers	1	2023	3	23	
2	vampire	Olivia Rodrigo	1	2023	6	30	
3	Cruel Summer	Taylor Swift	1	2019	8	23	
4	WHERE SHE GOES	Bad Bunny	1	2023	5	18	
...
948	My Mind & Me	Selena Gomez	1	2022	11	3	
949	Bigger Than The Whole Sky	Taylor Swift	1	2022	10	21	
950	A Veces (feat. Feid)	Feid, Paulo Londra	2	2022	11	3	
951	En La De Ella	Feid, Sech, Jhayco	3	2022	10	20	
952	Alone	Burna Boy	1	2022	11	4	

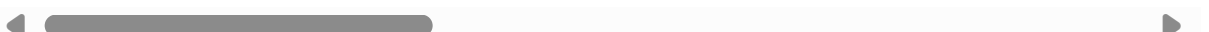
817 rows × 24 columns



```
In [8]: data.describe()
```

```
Out[8]:
```

	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify_
count	953.000000	953.000000	953.000000	953.000000	953.000000	953.000000
mean	1.556139	2018.238195	6.033578	13.930745	5200.124869	12.000000
std	0.893044	11.116218	3.566435	9.201949	7897.608990	19.500000
min	1.000000	1930.000000	1.000000	1.000000	31.000000	0.000000
25%	1.000000	2020.000000	3.000000	6.000000	875.000000	0.000000
50%	1.000000	2022.000000	6.000000	13.000000	2224.000000	3.000000
75%	2.000000	2022.000000	9.000000	22.000000	5542.000000	16.000000
max	8.000000	2023.000000	12.000000	31.000000	52898.000000	147.000000



```
In [9]: dict = {'name':["aparna", "pankaj", "sudhir", "Geeku"],
               'degree': ["MBA", "BCA", "M.Tech", "MBA"],
               'score':[90, 40, 80, 98]}

df = pd.DataFrame(dict, index = [True, False, True, False])

print(df)
```

	name	degree	score
True	aparna	MBA	90
False	pankaj	BCA	40
True	sudhir	M.Tech	80
False	Geeku	MBA	98

```
In [12]: spotify_data = {
        'track_name': ['seven', 'haegeum', 'like crazy', 'arson'],
        'artist(s)_name': ['jungkook', 'suga', 'jimin', 'jhope'],
        'duration_ms': [200000, 180000, 220000, 240000],
        'popularity': [80, 60, 70, 90]
    }

spotify_df = pd.DataFrame(spotify_data)

popularity_threshold = 50

filtered_spotify_df = spotify_df[spotify_df['popularity'] > popularity_thresho

print(filtered_spotify_df)
```

	track_name	artist(s)_name	duration_ms	popularity
0	seven	jungkook	200000	80
1	haegeum	suga	180000	60
2	like crazy	jimin	220000	70
3	arson	jhope	240000	90

```
In [22]: data['in_apple_playlists'].mean()
```

```
Out[22]: 67.8121720881427
```