IMPORTING NECESSARY LIBRARIES FOR THIS PROJECT

In [ ]:

```
import pandas as pd
import numpy as
import matplotlib.pyplot as plt
%matplotlib inne
```

In [72]:

```
pd.read_csv?
```

In [73]:

```
spotify_df = pd.read_csv('Spotify_final_dataset.csv')
```

In [74]:

```
spotify_df
```

Out[74]:

	Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	T Strea
0	1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	(x29)	2118242	883369
1	2	Juice WRLD	Lucid Dreams	1673	178.0	1	(x20)	2127668	864832
2	3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	(x4)	1660502	781153
3	4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	734857
4	5	Post Malone	rockstar	1223	186.0	1	(x124)	2905678	718865
...	...	...	...	...	...	...	...	...	
11079	11080	The Band Perry	If I Die Young	1	0.0	184	0	51321	51
11080	11081	Justin Timberlake	Not a Bad Thing	1	0.0	196	0	49512	49
11081	11082	Mike WiLL Made	It 23	1	0.0	167	0	46547	46
11082	11083	The Vamps	Somebody To You	1	0.0	200	0	44962	44
11083	11084	JAY	Z Holy Grail	1	0.0	184	0	44323	44

11084 rows × 9 columns

GETTING THE DATA INFORMATION BEFORE PREPARATION

In [75]:

```
spotify_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11084 entries, 0 to 11083
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Position                11084 non-null  int64  
 1   Artist Name             11084 non-null  object 
 2   Song Name               11080 non-null  object 
 3   Days                    11084 non-null  int64  
 4   Top 10 (xTimes)         11084 non-null  float64
 5   Peak Position           11084 non-null  int64  
 6   Peak Position (xTimes)  11084 non-null  object 
 7   Peak Streams            11084 non-null  int64  
 8   Total Streams           11084 non-null  int64  
dtypes: float64(1), int64(5), object(3)
memory usage: 779.5+ KB
```

In [76]:

```
spotify_df.isnull().sum()
```

Out[76]:

```
Position                  0
Artist Name               0
Song Name                 4
Days                      0
Top 10 (xTimes)           0
Peak Position             0
Peak Position (xTimes)    0
Peak Streams              0
Total Streams             0
dtype: int64
```

DATA CLEANING/PREPARATION FOR VISUALISATION

In [77]:

```
# DROPPING NULL VALLUES
spotify_df=spotify_df.dropna(how='any')
```

In [78]:

```
spotify_df.isnull().sum()
```

Out[78]:

```
Position                 0
Artist Name              0
Song Name                0
Days                     0
Top 10 (xTimes)          0
Peak Position            0
Peak Position (xTimes)   0
Peak Streams             0
Total Streams            0
dtype: int64
```

In [79]:

```
# SETTING UP MY INDEXING
spotify_df.set_index('Position', inplace=True)
```

In [80]:

```
spotify_df.head(20)
```

Out[80]:

Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	Stre
1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	(x29)	2118242	88336
2	Juice WRLD	Lucid Dreams	1673	178.0	1	(x20)	2127668	86483
3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	(x4)	1660502	78115
4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	73485
5	Post Malone	rockstar	1223	186.0	1	(x124)	2905678	71886
6	Travis Scott	goosebumps	1995	4.0	8	0	977275	67297
7	The Weeknd	Blinding Lights	1100	233.0	1	(x11)	2355059	64428
8	XXXTENTACION	Jocelyn Flores	1673	44.0	2	(x1)	3175206	62445
9	XXXTENTACION	SAD!	1217	133.0	1	(x6)	4437612	61987
10	Juice WRLD	All Girls Are The Same	1681	2.0	5	0	1239152	61387
11	Kendrick Lamar	HUMBLE.	1175	152.0	1	(x67)	4060034	60630
12	Post Malone	Circles	1172	163.0	1	(x38)	3441328	59852
13	Travis Scott	SICKO MODE	1046	190.0	1	(x19)	4129691	58663
14	Lil Baby	Drip Too Hard (Lil Baby & Gunna)	1453	122.0	1	(x6)	1759294	58344
15	Post Malone	Congratulations	1215	136.0	5	0	1263208	54603
16	Post Malone	I Fall Apart	1278	132.0	2	(x14)	1275320	54314
17	Glass Animals	Heat Waves	763	248.0	1	(x34)	1230046	54075
18	Drake	God's Plan	634	139.0	1	(x70)	4739798	53499
19	Roddy Ricch	The Box	650	197.0	1	(x85)	3781452	50421
20	J. Cole	MIDDLE CHILD	1119	106.0	1	(x16)	3800251	48198

In [ ]:

```
#REMOVING UNWANTED VALLUE FROM DATASET
spotify_df['Peak Position (xTimes)']=spotify_df['Peak Position (xTimes)'].str.replace('x'
```

In [ ]:

```
spotify_df['Peak Position (xTimes)']=spotify_df['Peak Position (xTimes)'].str.replace('('
```

In [ ]:

```
spotify_df['Peak Position (xTimes)']=spotify_df['Peak Position (xTimes)'].str.replace(')'
```

In [87]:

```
spotify_df
```

Out[87]:

Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	Total Streams
1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	29	2118242	883369738
2	Juice WRLD	Lucid Dreams	1673	178.0	1	20	2127668	864832399
3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	4	1660502	781153024
4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	734857487
5	Post Malone	rockstar	1223	186.0	1	124	2905678	718865961
...	...	...	...	...	...	...	...	...
11080	The Band Perry	If I Die Young	1	0.0	184	0	51321	51321
11081	Justin Timberlake	Not a Bad Thing	1	0.0	196	0	49512	49512
11082	Mike WiLL Made	It 23	1	0.0	167	0	46547	46547
11083	The Vamps	Somebody To You	1	0.0	200	0	44962	44962
11084	JAY	Z Holy Grail	1	0.0	184	0	44323	44323

11080 rows × 8 columns

In [88]:

```
# MAKING A COPY OF DATASET SO I CAN ADD MORE COLUMNS
spotify_df_copy = spotify_df.copy()
```

In [89]:

```
# CREATING A TOTAL AMOUNT OF STREAMS COLUMN
spotify_df_copy ['TAOS'] = spotify_df_copy['Total Streams'] * 0.005
```

In [90]:

```
spotify_df_copy
```

Out[90]:

Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	Total Streams	
1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	29	2118242	883369738	44
2	Juice WRLD	Lucid Dreams	1673	178.0	1	20	2127668	864832399	43
3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	4	1660502	781153024	39
4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	734857487	36
5	Post Malone	rockstar	1223	186.0	1	124	2905678	718865961	35
...	...	...	...	...	...	...	...	...	
11080	The Band Perry	If I Die Young	1	0.0	184	0	51321	51321	
11081	Justin Timberlake	Not a Bad Thing	1	0.0	196	0	49512	49512	
11082	Mike WiLL Made	It 23	1	0.0	167	0	46547	46547	
11083	The Vamps	Somebody To You	1	0.0	200	0	44962	44962	
11084	JAY	Z Holy Grail	1	0.0	184	0	44323	44323	

11080 rows × 9 columns

In [91]:

```
spotify_df_copy=spotify_df_copy.rename(columns={'TAOS':'Total Amount Of Stream'})
```

In [92]:

```
spotify_df_copy
```

Out[92]:

Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	Total Streams	A
1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	29	2118242	883369738	44
2	Juice WRLD	Lucid Dreams	1673	178.0	1	20	2127668	864832399	43
3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	4	1660502	781153024	39
4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	734857487	36
5	Post Malone	rockstar	1223	186.0	1	124	2905678	718865961	35
...	...	...	...	...	...	...	...	...	
11080	The Band Perry	If I Die Young	1	0.0	184	0	51321	51321	
11081	Justin Timberlake	Not a Bad Thing	1	0.0	196	0	49512	49512	
11082	Mike WiLL Made	It 23	1	0.0	167	0	46547	46547	
11083	The Vamps	Somebody To You	1	0.0	200	0	44962	44962	
11084	JAY	Z Holy Grail	1	0.0	184	0	44323	44323	

11080 rows × 9 columns

TOP 10 BEST SELLING SONG

In [93]:

```
Best_Song = 'Song Name'
```

In [94]:

```
sorted_song_df=spotify_df_copy.sort_values(by='Total Streams', ascending=False)
```

In [95]:

```
Top_10_songs=sorted_song_df.head(10)
Top_10_songs
```

Out[95]:

Position	Artist Name	Song Name	Days	Top 10 (xTimes)	Peak Position	Peak Position (xTimes)	Peak Streams	Tot Strean
1	Post Malone	Sunflower SpiderMan: Into the SpiderVerse	1506	302.0	1	29	2118242	8833697
2	Juice WRLD	Lucid Dreams	1673	178.0	1	20	2127668	8648323
3	Lil Uzi Vert	XO TOUR Llif3	1853	212.0	1	4	1660502	7811530
4	J. Cole	No Role Modelz	2547	6.0	7	0	659366	7348574
5	Post Malone	rockstar	1223	186.0	1	124	2905678	7188659
6	Travis Scott	goosebumps	1995	4.0	8	0	977275	6729727
7	The Weeknd	Blinding Lights	1100	233.0	1	11	2355059	6442879
8	XXXTENTACION	Jocelyn Flores	1673	44.0	2	1	3175206	6244571
9	XXXTENTACION	SAD!	1217	133.0	1	6	4437612	6198792
10	Juice WRLD	All Girls Are The Same	1681	2.0	5	0	1239152	6138723

In [45]:

```
# CHANGING THE DATATYPE OF TAOS(TOTAL AMOUNT OF STREAM)
spotify_df_copy['TAOS']=spotify_df_copy['TAOS'].astype(int)
```

TOP 10 BEST SELLING ARTIST BY TOTAL AMOUNT PER STREAMS

In [46]:

```
artist_total_stream = spotify_df_copy.groupby('Artist Name')['TAOS'].sum()
sorted_artist_streams= artist_total_stream.sort_values(ascending=False)
Top_10_artist=sorted_artist_streams.head(10)
Top_10_artist
```

Out[46]:

```
Artist Name
Drake             46144535
Post Malone       37000963
Juice WRLD        24289941
The Weeknd        21335004
XXXTENTACION      18448821
Taylor Swift      16114773
Ariana Grande     15843811
Billie Eilish     15597469
Lil Uzi Vert      14575602
Bad Bunny         13948835
Name: TAOS, dtype: int32
```

TOP 10 BEST SELLING ARTIST BY TOTAL STREAMS

In [47]:

```
artist_total_stream = spotify_df_copy.groupby('Artist Name')['Total Streams'].sum()
sorted_artist_streams= artist_total_stream.sort_values(ascending=False)
Top_10_artist=sorted_artist_streams.head(10)
Top_10_artist
```

Out[47]:

```
Artist Name
Drake             9228928893
Post Malone       7400199432
Juice WRLD        4857996474
The Weeknd        4267010379
XXXTENTACION      3689771047
Taylor Swift      3222972346
Ariana Grande     3168771349
Billie Eilish     3119498389
Lil Uzi Vert      2915127161
Bad Bunny         2789775409
Name: Total Streams, dtype: int64
```

In [52]:

```
# CONVERTING THE DATATYPE OF Peak Position (xTimes) TO INT
spotify_df_copy['Peak Position (xTimes)']=spotify_df_copy['Peak Position (xTimes)'].astyp
```

TOP 10 ARTISTS WITH THE MOST PEAK STREAMS

In [99]:

```
artist_Peak_Position = spotify_df_copy.groupby('Artist Name')['Peak Streams'].sum()
sorted_Peak_Position= artist_Peak_Position.sort_values(ascending=False)
Top_10_artist=sorted_Peak_Position.head(10)
Top_10_artist
```

Out[99]:

```
Artist Name
Drake              369854586
Taylor Swift       305560909
Post Malone        115411593
Kendrick Lamar     110191432
Juice WRLD         110143807
Kanye West         108962605
J. Cole            107891220
Future             104696355
Ariana Grande      100983832
The Weeknd          94998935
Name: Peak Streams, dtype: int64
```

EXPORTING DATASET FOR VISUALISATION IN POWER BI

I couldn't make some visualisation here with PYTHON cause my JUPYTER NOTEBOOK cause some functions are not functioning especially Seaborn and Matplotlib not showing some graph information so i decided to take all visualisation to MICROSOFT POWER BI.

In [ ]:

```
spotify_df_copy.to_csv('C:\\Users\\J6IX\\Desktop\\spotify_df_copy.csv')
```