

In [ ]:

```
import pandas as pd
import numpy as np
```

In [ ]:

```
from google.colab import drive
drive.mount('/content/drive')
```

Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect\\_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aoob&response\\_type=code&scope=email%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdocs.test%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdrive%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdrive.photos.readonly%20https%3a%2f%2fwww.googleapis.com%2fauth%2fpeopleapi.readonly](https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aoob&response_type=code&scope=email%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdocs.test%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdrive%20https%3a%2f%2fwww.googleapis.com%2fauth%2fdrive.photos.readonly%20https%3a%2f%2fwww.googleapis.com%2fauth%2fpeopleapi.readonly)

Enter your authorization code:  
.....  
Mounted at /content/drive

In [ ]:

```
!unzip "/content/drive/My Drive/SyncPC/Data Analytic/SpotifyData/data.csv.zip"
```

# Load dữ liệu

In [ ]:

```
data_df = pd.read_csv("data.csv")
data_df
```

Out[ ]:

	Position	Track Name	Artist	Streams	URL	Date	Region
0	1	Reggaetón Lento (Bailemos)	CNCO	19272	https://open.spotify.com/track/3AEZUABDXNtecAO...	2017-01-01	ec
1	2	Chantaje	Shakira	19270	https://open.spotify.com/track/6mICuAdrwEjh6Y6...	2017-01-01	ec
2	3	Otra Vez (feat. J Balvin)	Zion & Lennox	15761	https://open.spotify.com/track/3QwBODjSEzelZyV...	2017-01-01	ec
3	4	Vente Pa' Ca	Ricky Martin	14954	https://open.spotify.com/track/7DM4BPaS7uofFul...	2017-01-01	ec
4	5	Safari	J Balvin	14269	https://open.spotify.com/track/6rQSRBHf7HIZjtc...	2017-01-01	ec
...	...	...	...	...	...	...	...
3441192	196	More Than You Know	Axwell /\ Ingrosso	2126	https://open.spotify.com/track/71bBFbfn2OBK5Qw...	2018-01-09	hk
3441193	197	Issues	Julia Michaels	2117	https://open.spotify.com/track/7vu0JkJh0ldukEY...	2018-01-09	hk
3441194	198	Tuo Fei Lun	Eason Chan	2112	https://open.spotify.com/track/0pmGleExiQyd0Hx...	2018-01-09	hk
3441195	199	面具	Alfred Hui	2102	https://open.spotify.com/track/58SUsgamkW6vFFq...	2018-01-09	hk
3441196	200	Let Her Go	Passenger	2088	https://open.spotify.com/track/2jyjhRf6DVbMPU5...	2018-01-09	hk

3441197 rows x 7 columns

In [ ]:

```
feature_df = pd.read_csv("/content/drive/My Drive/SyncPC/Data Analytic/SpotifyData/feature_sdf.csv")
feature_df.head()
```

Out[ ]:

	id	name	artists	danceability	energy	key	loudness	mode	speechiness	acoustic
0	7qiZfU4dY1IWllzX7mPBI	Shape of You	Ed Sheeran	0.825	0.652	1.0	-3.183	0.0	0.0802	0.
1	5CtI0qwDJkDQGwXD1H1cL	Despacito - Remix	Luis Fonsi	0.694	0.815	2.0	-4.328	1.0	0.1200	0.
2	4aWmUDTfIPGksMNLV2rQP	Despacito (Featuring Daddy Yankee)	Luis Fonsi	0.660	0.786	2.0	-4.757	1.0	0.1700	0.
3	6RUKPb4LETWmmr3iAEQkt	Something Just Like This	The Chainsmokers	0.617	0.635	11.0	-6.769	0.0	0.0317	0.
4	3DXncPQOG4VBw3QHh3S81	I'm the One	DJ Khaled	0.609	0.668	7.0	-4.284	1.0	0.0367	0.

In [ ]:

```
feature_df["duration"] = np.round(feature_df["duration_ms"].values /1000)
feature_df
```

Out[ ]:

	id	name	artists	danceability	energy	key	loudness	mode	speechiness	acousti
0	7qiZfU4dY1IWllzX7mPBI	Shape of You	Ed Sheeran	0.825	0.652	1.0	-3.183	0.0	0.0802	(
1	5CtI0qwDJkDQGwXD1H1cL	Despacito - Remix	Luis Fonsi	0.694	0.815	2.0	-4.328	1.0	0.1200	(
2	4aWmUDTfIPGksMNLV2rQP	Despacito (Featuring Daddy Yankee)	Luis Fonsi	0.660	0.786	2.0	-4.757	1.0	0.1700	(
3	6RUKPb4LETWmmr3iAEQkt	Something Just Like This	The Chainsmokers	0.617	0.635	11.0	-6.769	0.0	0.0317	(
4	3DXncPQOG4VBw3QHh3S81	I'm the One	DJ Khaled	0.609	0.668	7.0	-4.284	1.0	0.0367	(
...	...	...	...	...	...	...	...	...	...	...
95	1PSBzsahR2AKwLJgx8ehB	Bad Things (with Camila Cabello)	Machine Gun Kelly	0.675	0.690	2.0	-4.761	1.0	0.1320	(
96	0QsvXlfqM0zZoerQfsI9l	Don't Let Me Down	The Chainsmokers	0.542	0.859	11.0	-5.651	1.0	0.1970	(
97	7mldq42yDuxiUNn08nvzH	Body Like A Back Road	Sam Hunt	0.731	0.469	5.0	-7.226	1.0	0.0326	(
98	7i2DJ88J7jQ8K7zqFX2fW	Now Or Never	Halsey	0.658	0.588	6.0	-4.902	0.0	0.0367	(
99	1j4kHkqpqZRBwE0A4CN4Y	Dusk Till Dawn - Radio Edit	ZAYN	0.258	0.437	11.0	-6.593	0.0	0.0390	(

In [ ]:

```
data_df[(data_df["Position"]==1) & (data_df["Date"]=="2017-05-27")]
```

Out[ ]:

	Position	Track Name	Artist	Streams	URL	Date	Region
29200	1	Felices los 4	Maluma	34469	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	ec
103400	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	232445	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	fr
177600	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	329911	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ar
251800	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	141129	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	fi
326000	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	386475	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	no
400200	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	235059	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	it
453002	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	6175	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	lt
491199	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	207000	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ph
565399	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	27440	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	tw
639599	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	81087	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	nz
691139	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	5795	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ee
726622	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	64110	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	tr
800822	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	2537245	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	us
858803	1	Felices los 4	Maluma	9135	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	sv
913915	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	33285	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	cr
988115	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	857625	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	de
1062315	1	Felices los 4	Maluma	189830	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	cl
1134914	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	19721	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	jp
1209114	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	799229	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	br

	Position	Track Name	Artist	Streams	URL	05-27 Date	Region
1269681	1	Felices los 4	Maluma	11274	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	hn
1320493	1	Felices los 4	Maluma	15378	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	gt
1393750	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	66822	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ch
1460044	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	22426	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	hu
1528122	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	438655	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ca
1602322	1	Felices los 4	Maluma	124375	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	pe
1676522	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	71667	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	be
1750722	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	76583	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	my
1824722	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	236659	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	dk
1882539	1	Felices los 4	Maluma	12328	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	bo
1938760	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	87054	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	pl
2012960	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	45505	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	at
2087160	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	47416	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	pt
2161360	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	589730	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	se
2235560	1	Felices los 4	Maluma	767697	https://open.spotify.com/track/0qYTZCo5Bwh1nsU...	2017-05-27	mx
2298591	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	12707	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	pa
2358362	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	19366	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	uy
2413027	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	19623	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	is
2465909	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	495063	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	es
2539000	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	17628	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	cz
2613200	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	94021	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	ie
2687400	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	381214	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	nl

	Position	Track Name	Artist	Streams	URL	Date	Region
2740003	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	7087	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	sk
2784197	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	76260	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	co
2858397	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	84318	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	sg
2932597	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	108235	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	id
2999885	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	15476	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	do
3040780	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	5311	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	lu
3071643	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	1005214	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	gb
3145443	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	11381520	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	global
3208783	1	Felices los 4	Maluma	18491	https://open.spotify.com/track/0qY TZCo5Bwh1nsU...	2017-05-27	py
3275440	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	309758	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	au
3328018	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	4885	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	lv
3349468	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	10809	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	gr
3396197	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	22533	https://open.spotify.com/track/4aWmUDTflPGksMN...	2017-05-27	hk

Bài tập 1: Tìm bài hát (track name) có stream lớn nhất.

In [ ]:

```
data_df.max()
```

Out[ ]:

Position 200  
Streams 11381520  
Date 2018-01-09  
Region uy  
dtype: object

In [ ]:

```
data_df["Streams"].max()
```

Out[ ]:

11381520

In [ ]:

```
data_df[data_df["Streams"]==data_df["Streams"].max()]
```

Out [ ]:

	Position	Track Name	Artist	Streams	URL	Date	Region
3145443	1	Despacito (Featuring Daddy Yankee)	Luis Fonsi	11381520	https://open.spotify.com/track/4aWmUDTfIPGksMN...	2017-05-27	global

## Bài tập 2: Tìm bài hát (track name) có stream lớn nhất theo ngày.

In [ ]:

```
data_df.groupby(["Date"])["Streams"].max()
```

In [ ]:

```
data_df.groupby(["Date"])["Streams"].idxmax()
```

Out [ ]:

Date  
2017-01-01      3116643  
2017-01-02      3116843  
2017-01-03      3117043  
2017-01-04      3117243  
2017-01-05      3117443  
...  
2018-01-05      3189443  
2018-01-06      3189643  
2018-01-07      3189843  
2018-01-08      3190043  
2018-01-09      3190243  
Name: Streams, Length: 371, dtype: int64

In [ ]:

```
data_df.loc[data_df.groupby(["Date"])["Streams"].idxmax()]
```

Out [ ]:

	Position	Track Name	Artist	Streams	URL	Date	Region
3116643	1	Starboy	The Weeknd	3135625	https://open.spotify.com/track/5aAx2yezTd8zXrk...	2017-01-01	global
3116843	1	Starboy	The Weeknd	3342769	https://open.spotify.com/track/5aAx2yezTd8zXrk...	2017-01-02	global
3117043	1	Starboy	The Weeknd	3563076	https://open.spotify.com/track/5aAx2yezTd8zXrk...	2017-01-03	global
3117243	1	Starboy	The Weeknd	3619247	https://open.spotify.com/track/5aAx2yezTd8zXrk...	2017-01-04	global
3117443	1	Starboy	The Weeknd	3579206	https://open.spotify.com/track/5aAx2yezTd8zXrk...	2017-01-05	global
...	...	...	...	...	...	...	...
3189443	1	rockstar	Post Malone	4346638	https://open.spotify.com/track/7wGoVu4Dady5GV0...	2018-01-05	global
3189643	1	rockstar	Post Malone	4208156	https://open.spotify.com/track/7wGoVu4Dady5GV0...	2018-01-06	global
...	...	...	...	...	...	...	...

3189843	1	rockstar	Post Malone	3814379	https://open.spotify.com/track/7wGoVu4Dady5GV0...	2018-01-07	global
	Position	Track Name	Artist	Streams	URL	Date	Region
3190043	1	rockstar	Post Malone	4062918	https://open.spotify.com/track/7wGoVu4Dady5GV0...	2018-01-08	global
3190243	1	rockstar	Post Malone	4139989	https://open.spotify.com/track/7wGoVu4Dady5GV0...	2018-01-09	global

371 rows x 7 columns

In [ ]:

```
data_df.groupby(["Date"])["Streams"].transform('max')
```

In [ ]:

In [ ]:

```
data_df[data_df["Streams"] == data_df.groupby(["Date"])["Streams"].transform('max')] .sort_values(by="Date")
```

In [ ]:

## Biểu đồ danh sách các nghệ sĩ có nhiều bài hát đạt top

In [ ]:

```
artist_count_df = feature_df.groupby(["artists"])["artists"].agg(['count'])
artist_count_df
```

In [ ]:

```
artist_count_df = artist_count_df[artist_count_df["count"] >=2].sort_values(by=["count"])
artist_count_df
```

In [ ]:

```
ax = artist_count_df.plot.barh(use_index=True, y='count')
```

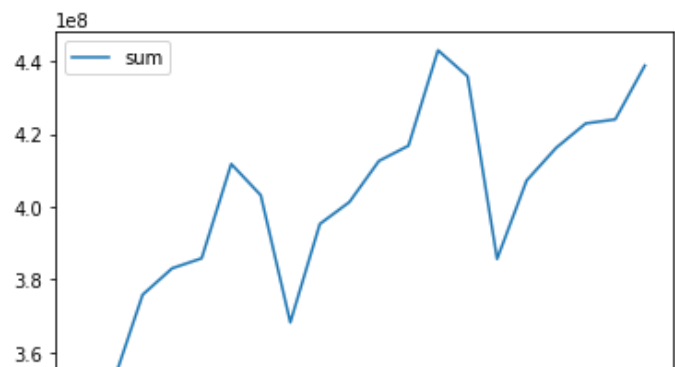
## Biểu đồ lượng streams của các top song theo ngày

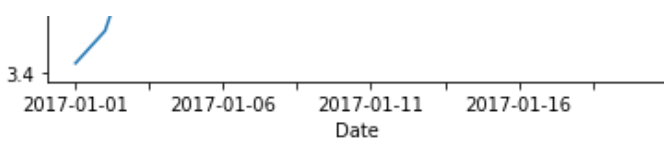
In [ ]:

```
#Biểu đồ theo ngày (Ví dụ: lấy 10 ngày đầu tiên)
total_stream_by_date_df = data_df.groupby(["Date"])["Streams"].agg(["sum"]).iloc[0:10]
```

In [ ]:

```
ax = total_stream_by_date_df.plot.line(use_index=True, y='sum')
```





## Bài tập 3: Biểu đồ lượng streams của các top song theo tháng

In [ ]:

```
data_df["month"] = data_df["Date"].apply(lambda x: x[0:7])
data_df
```

In [ ]:

```
total_stream_by_date_df = data_df.groupby(["month"])["Streams"].agg(["sum"])
ax = total_stream_by_date_df.plot.line(use_index=True, y='sum')
```

## Bài tập 4: Vẽ biểu đồ top 20 Tác giả có lượng play time cao nhất ở Mỹ (Region == "us")

Lưu ý: Play time = Streams \* duration

In [ ]:

```
#count stream
total_stream_df = data_df[data_df.Region == "us"].groupby(["Track Name"])["Streams"].agg(
    ["sum"])
total_stream_df.rename(columns={"sum": "total_streams"}, inplace=True)

playtime_df = feature_df.set_index(["name"]).join(total_stream_df, how="left")
playtime_df["play_time"] = (playtime_df["duration"].values * playtime_df["total_streams"]
    .values) / 86400
playtime_df
```

In [ ]:

```
top20_df = playtime_df.nlargest(20, ["play_time"])["play_time"]

play_time_ax = top20_df.sort_values(by=["play_time"], ascending=True).plot.barh(use_index
    =True, y='play_time')
```

## Bài tập 5 Số lần (số ngày) các bài hát của Ed Sheeran nằm trong top 100 ở Bảng xếp hạng Mỹ

In [ ]:

```
#How did Ed Sheeran's Songs perform in 2017 on daily basis
ed_sheeran_daily_df = data_df[(data_df.Region=="us") & (data_df.Artist == "Ed Sheeran")
    & (data_df.Position <= 100)]
count_ed_sheeran_daily_df = ed_sheeran_daily_df.groupby(["Track Name"])["Date"].agg(["count"])
count_ed_sheeran_daily_df.sort_values(by=["count"], ascending=False)
```

## Bài tập 6 Diễn biến xếp hạng của bài Shape of you của tác giả Ed Sheeran trong top 100 ở Mỹ. Mở rộng: diễn biến các bài hát của Ed Sheeran trong top 100 ở Mỹ

In [ ]:

```
ed_sheeran_daily_df[ed_sheeran_daily_df["Track Name"]=="Shape of You"]
```

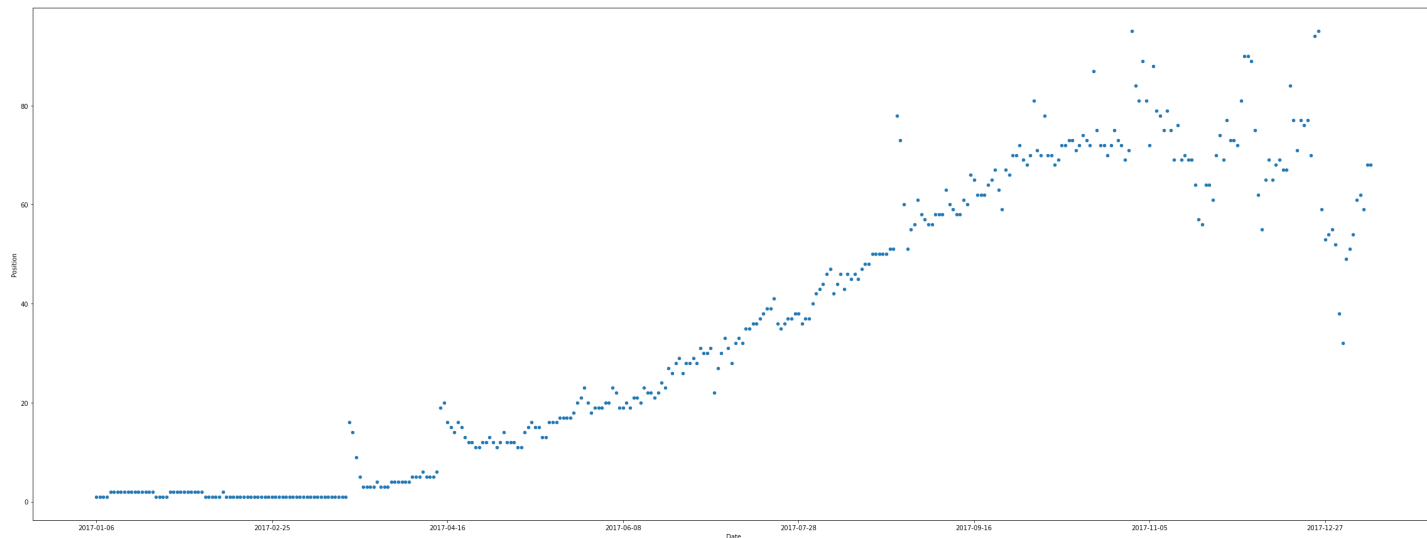


In [ ]:

```
ax = ed_sheeran_daily_df[ed_sheeran_daily_df["Track Name"]=="Shape of You"].plot.scatter(x="Date",y="Position",figsize=(40,15))
ax.get_xticks()
ax.set_xticks(ax.get_xticks()[::50])
```

Out[ ]:

```
[<matplotlib.axis.XTick at 0x7f74cf4e8f28>,
 <matplotlib.axis.XTick at 0x7f74cf4d7390>,
 <matplotlib.axis.XTick at 0x7f74cf495470>,
 <matplotlib.axis.XTick at 0x7f74cf495748>,
 <matplotlib.axis.XTick at 0x7f74cf495be0>,
 <matplotlib.axis.XTick at 0x7f74cf4ab128>,
 <matplotlib.axis.XTick at 0x7f74cf495898>,
 <matplotlib.axis.XTick at 0x7f74cf4ab278>]
```



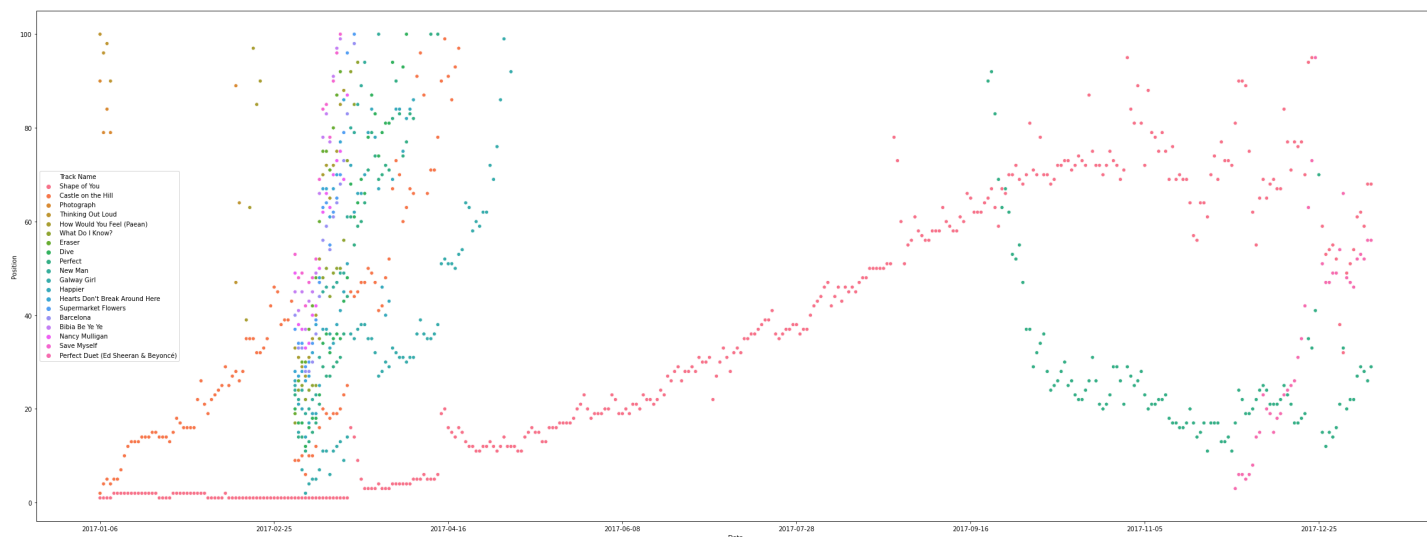
In [ ]:

```
import seaborn as sns
plt.figure(figsize=(40,15))

ax = sns.scatterplot(x="Date", y="Position",hue="Track Name", data=ed_sheeran_daily_df)
ax.get_xticks()
ax.set_xticks(ax.get_xticks()[::50])
ax.legend()
```

Out[ ]:

<matplotlib.legend.Legend at 0x7f74cff73e48>



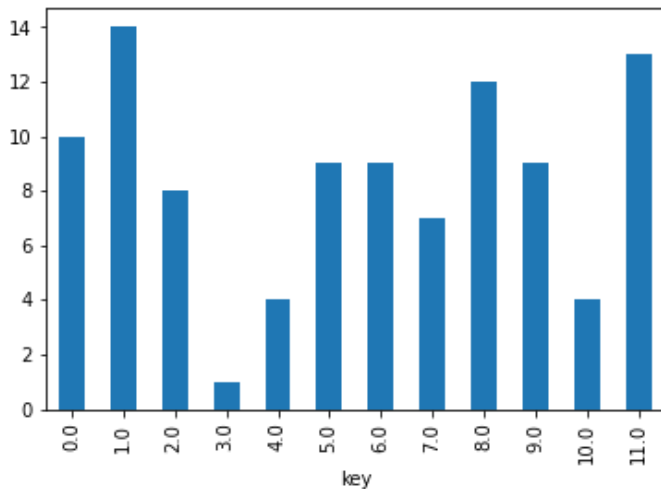
**Mối quan hệ giữa các đặc trưng của bài hát**

In [ ]:

```
#count key on Top Songs
feature_df.groupby(["key"])["id"].agg("count").plot.bar(use_index=True,y="count")
```

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f74cd9e2358>

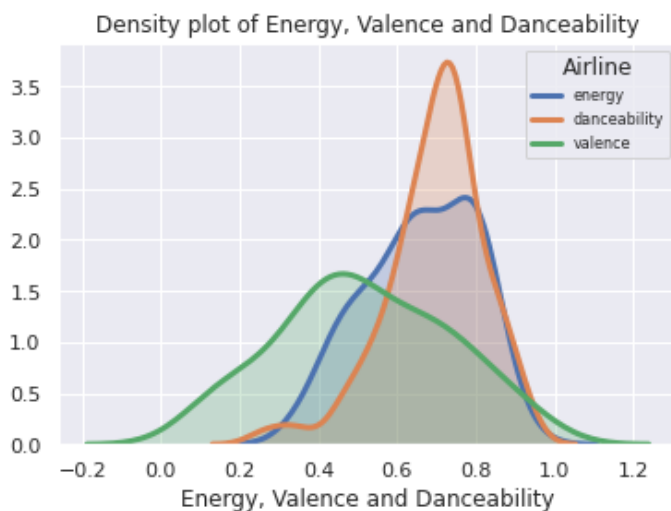


In [ ]:

```
sns.distplot(feature_df["energy"],hist=False, kde_kws = {'shade': True, 'linewidth': 3},
label = "energy")
sns.distplot(feature_df["danceability"],hist=False, kde_kws = {'shade': True, 'linewidth': 3},label = "danceability")
sns.distplot(feature_df["valence"],hist=False, kde_kws = {'shade': True, 'linewidth': 3}
,label = "valence")
plt.title('Density plot of Energy, Valence and Danceability')
plt.xlabel('Energy, Valence and Danceability')
plt.legend(prop={'size': 8}, title = 'Airline')
```

Out[ ]:

<matplotlib.legend.Legend at 0x7f74cbf39a90>



In [ ]:

```
# correlation
import matplotlib.pyplot as plt
corr = feature_df.corr()

mask = np.zeros_like(corr, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

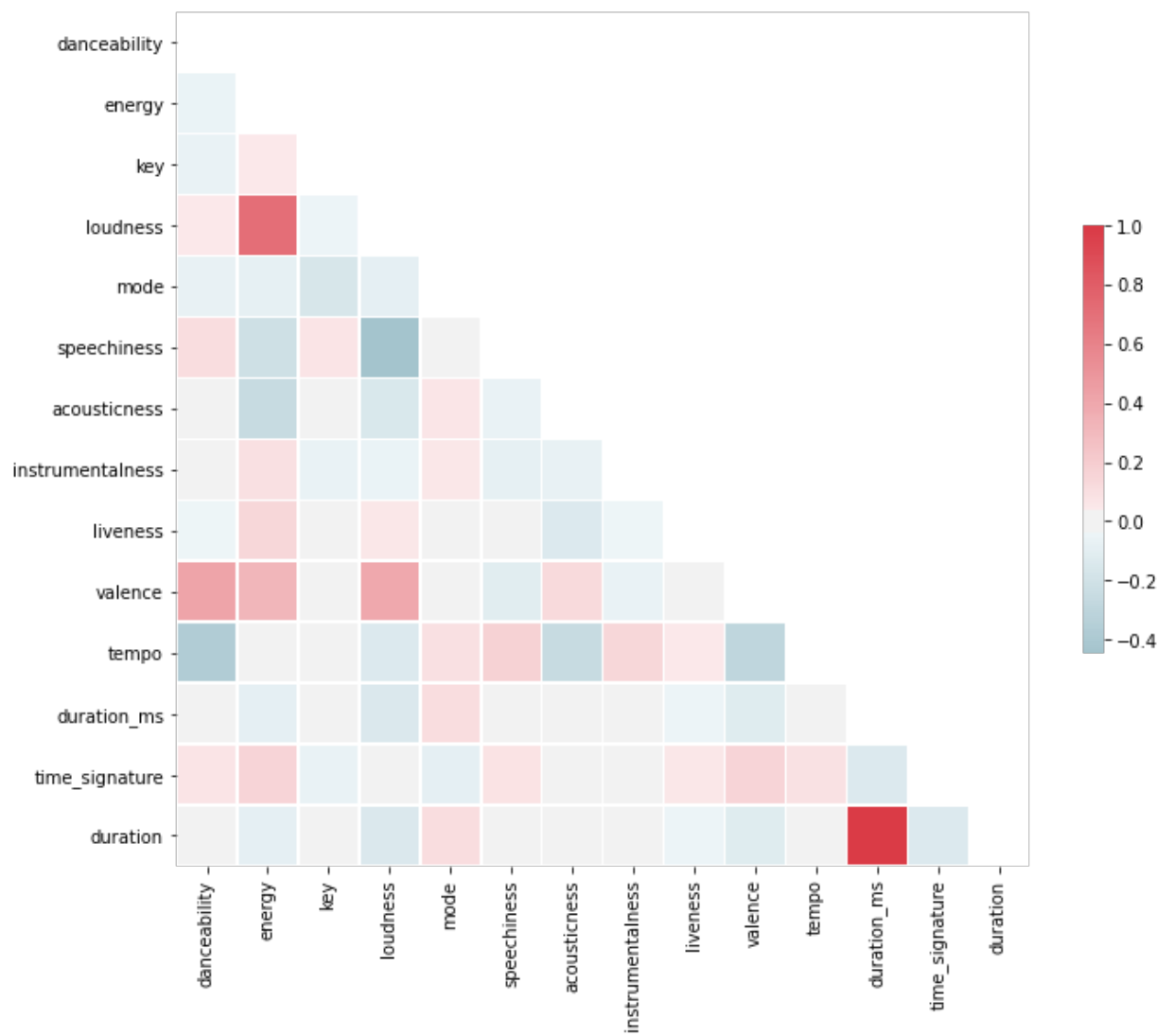
# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))
```

```
# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr, mask=mask, cmap=cmap, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": 0.5})
```

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f74ccf58cf8>



# Bài tập 7: Sử dụng PCA để biểu diễn các bài hát bằng biểu đồ scatter (dùng dữ liệu featuresdf)

```
In [ ]:
```

```
feature_df.head()
```

Out[ ]:

	id	name	artists	danceability	energy	key	loudness	mode	speechiness	acoustic
0	7qiZfU4dY1IWlIzX7mPBI	Shape of You	Ed Sheeran	0.825	0.652	1.0	-3.183	0.0	0.0802	0.
1	5CtI0qwDJkDQGwXD1H1cL	Despacito - Remix	Luis Fonsi	0.694	0.815	2.0	-4.328	1.0	0.1200	0.
2	4aWmUDTfIPGksMNLV2rQP	Despacito (Featuring Daddy Yankee)	Luis Fonsi	0.660	0.786	2.0	-4.757	1.0	0.1700	0.
		Something	The							

3	6RUKPb4LETWmmr3iAEQkt	Just Like This	Chainsmokers	danceability	energy	key	loudness	mode	speechiness	acoustic
4	3DXncPQOG4VBw3QHh3S81	I'm the One	DJ Khaled	0.609	0.668	7.0	-4.284	1.0	0.0367	0.

```
In [ ]:
feature_df[["danceability","energy","key","mode","speechiness","acousticness","instrumentalness","liveness","valence","tempo", "duration_ms","time_signature"]]
```

Out[ ]:

	danceability	energy	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	t
0	0.825	0.652	1.0	0.0	0.0802	0.5810	0.000000	0.0931	0.9310	95.977	233713.0	
1	0.694	0.815	2.0	1.0	0.1200	0.2290	0.000000	0.0924	0.8130	88.931	228827.0	
2	0.660	0.786	2.0	1.0	0.1700	0.2090	0.000000	0.1120	0.8460	177.833	228200.0	
3	0.617	0.635	11.0	0.0	0.0317	0.0498	0.000014	0.1640	0.4460	103.019	247160.0	
4	0.609	0.668	7.0	1.0	0.0367	0.0552	0.000000	0.1670	0.8110	80.924	288600.0	
...	...	...	...	...	...	...	...	...	...	...	...	...
95	0.675	0.690	2.0	1.0	0.1320	0.2100	0.000000	0.2870	0.2720	137.817	239293.0	
96	0.542	0.859	11.0	1.0	0.1970	0.1600	0.004660	0.1370	0.4030	159.797	208053.0	
97	0.731	0.469	5.0	1.0	0.0326	0.4630	0.000001	0.1030	0.6310	98.963	165387.0	
98	0.658	0.588	6.0	0.0	0.0367	0.1050	0.000001	0.1250	0.4340	110.075	214802.0	
99	0.258	0.437	11.0	0.0	0.0390	0.1010	0.000001	0.1060	0.0967	180.043	239000.0	

100 rows x 12 columns

```
In [ ]:
from sklearn.preprocessing import StandardScaler
x = feature_df[["danceability","energy","key","mode","speechiness","acousticness","instrumentalness","liveness","valence","tempo", "duration_ms","time_signature"]].values
x = StandardScaler().fit_transform(x)
```

```
In [ ]:
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents
                           , columns = ['principal component 1', 'principal component 2'])
principalDf
```

Out[ ]:

	principal component 1	principal component 2
0	-2.837690	1.232744
1	-1.499614	0.022366
2	0.313782	-0.785407
3	0.224982	-0.132229
4	-0.764038	0.329994
...	...	...
95	1.217487	-0.155512
96	1.704961	-1.074203
97	-1.199045	1.458176

98	0.048354	0.063101
	principal component 1	principal component 2
99	4.124497	0.296700

100 rows x 2 columns

In [ ]:

```
principalDf.plot.scatter(x="principal component 1",y="principal component 2")
```

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc1531aff60>

