

# Data Visualization Python

## Spotify Dataset

The dataset for this tutorial tracks global daily streams on the music streaming service Spotify. We focus on five popular songs from 2017 and 2018:

1. "Shape of You", by Ed Sheeran ([link](#))
2. "Despacito", by Luis Fonsi ([link](#))
3. "Something Just Like This", by The Chainsmokers and Coldplay ([link](#))
4. "HUMBLE.", by Kendrick Lamar ([link](#))
5. "Unforgettable", by French Montana ([link](#))

## Set up notebook

```
In [2]: # Imports
import pandas as pd
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
print("Setup Complete")
```

Setup Complete

## Load and Examine Data

```
In [6]: # Path of the file to read
spotify_filepath = "./spotify.csv"

# Read the file into a variable spotify_data
spotify_data = pd.read_csv(spotify_filepath, index_col="Date", parse_dates=True)

# Print the first 5 rows of the data
spotify_data.head()
```

Out[6]:

	Shape of You	Despacito	Something Just Like This	HUMBLE.	Unforgettable
--	--------------	-----------	--------------------------	---------	---------------

Date					
2017-01-06	12287078	NaN	NaN	NaN	NaN
2017-01-07	13190270	NaN	NaN	NaN	NaN
2017-01-08	13099919	NaN	NaN	NaN	NaN
2017-01-09	14506351	NaN	NaN	NaN	NaN
2017-01-10	14275628	NaN	NaN	NaN	NaN

```
In [7]: # Print the last five rows of the data
spotify_data.tail()
```

```
Out[7]:
```

	Shape of You	Despacito	Something Just Like This	HUMBLE.	Unforgettable
<b>Date</b>					
<b>2018-01-05</b>	4492978	3450315.0	2408365.0	2685857.0	2869783.0
<b>2018-01-06</b>	4416476	3394284.0	2188035.0	2559044.0	2743748.0
<b>2018-01-07</b>	4009104	3020789.0	1908129.0	2350985.0	2441045.0
<b>2018-01-08</b>	4135505	2755266.0	2023251.0	2523265.0	2622693.0
<b>2018-01-09</b>	4168506	2791601.0	2058016.0	2727678.0	2627334.0

Notice that the first date that appears is January 6, 2017, corresponding to the release date of "The Shape of You", by Ed Sheeran. And, using the table, you can see that "The Shape of You" was streamed 12,287,078 times globally on the day of its release. Notice that the other songs have missing values in the first row, because they weren't released until later!

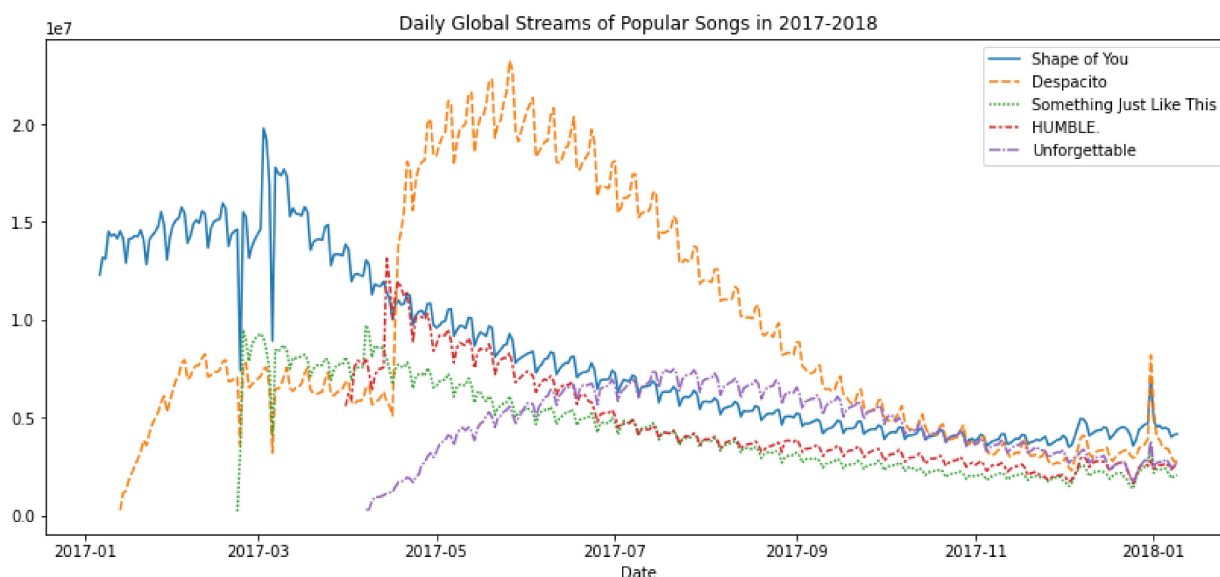
## Plot the data using lineplot

```
In [9]: # Set the width and height of the figure
plt.figure(figsize=(14,6))

# Add title
plt.title("Daily Global Streams of Popular Songs in 2017-2018")

# Line chart showing daily global streams of each song
sns.lineplot(data=spotify_data)
```

```
Out[9]: <AxesSubplot:title={'center':'Daily Global Streams of Popular Songs in 2017-2018'}, x
label='Date'>
```



## Analysis

From the above graph, the songs follow a bell curve. That is the streams rapidly grows after the songs is released, reaches a peak and then gradually falls as time passes. Additionally although different songs might have different streams at the begining or the peak, they generally fall to about the same streams in the long run.

In [ ]: