

Analysing and Visualizing Gaming Trends with Twitch Data

Contents

1. Project Overview	1
2. Understanding the database schema.....	1
3. Querying data.....	2
4. Visualizing Twitch Data using Python and Matplotlib	6
5. Conclusion	9

1. Project Overview

Twitch is a popular live streaming platform that allows users to watch and broadcast video streams. It is mainly focused on video game streaming, but also offers music, art, cooking and other types of entertainment streams.

Twitch has over 15 million daily active users. Using data to understand its users and products is one of the main responsibilities of the Twitch Data Science Team. In this project, I will be extracting data using SQL and then visualizing it with Matplotlib using Python.

2. Understanding the database schema

In this project, I will be working with 2 tables- “stream” and “chat”. Their schema is given below:

Stream Table

time
device_id
login
channel
country
player
game
stream_format
subscriber

Chat table

time
device_id

login
channel
country
player
game

3. Querying data.

3.1. Most popular games in stream table

```
SELECT game, COUNT(*)  
FROM stream  
GROUP BY 1  
ORDER BY 2 DESC;
```

Query Results	
game	COUNT(*)
League of Legends	1070
Dota 2	472
Counter-Strike: Global Offensive	302
DayZ	239
Heroes of the Storm	210
The Binding of Isaac: Rebirth	171
Gaming Talk Shows	170
Hearthstone: Heroes of Warcraft	90
World of Tanks	86
Agar.io	71
Rocket League	49
SpeedRunners	20
ARK: Survival Evolved	19
Ø	15
Duck Game	5
Fallout 3	3

3.2. Viewers by country (for League of Legends)

```
SELECT country, COUNT(*) FROM stream  
WHERE game='League of Legends'  
GROUP BY 1  
ORDER BY 2 DESC;
```

Query Results	
country	COUNT(*)
US	447
DE	66
CA	64
Ø	49
GB	45
TR	28
BR	25
DK	20
PL	19
NL	17
BE	17
SE	16

3.3. Source of Viewers

--SOURCE OF VIEWERS

SELECT player, COUNT(*) FROM stream

GROUP BY 1

ORDER BY 2 DESC;

Query Results	
player	COUNT(*)
site	1365
iphone_t	622
android	547
ipad_t	290
embed	125
xbox_one	22
home	16
amazon	6
frontpage	4
xbox360	1
roku	1
chromecast	1

3.4. Categorising games by Genre

--CREATING GENRE COLUMN AND CATEGORISING GAMES

SELECT game,

CASE

WHEN game='League of Legends'

THEN 'MOBA'

```
WHEN game='Dota 2'
  THEN 'MOBA'
WHEN game='Heroes of the Storm'
  THEN 'MOBA'
WHEN game='Counter-Strike: Global Offensive'
  THEN 'FPS'
WHEN game='DayZ'
  THEN 'Survival'
WHEN game='ARK: Survival Evolved'
  THEN 'Survival'
ELSE 'Other'
END AS 'genre',
COUNT(*)
FROM stream
GROUP BY 1
ORDER BY 3 DESC;
```

Query Results		
game	genre	COUNT(*)
League of Legends	MOBA	1070
Dota 2	MOBA	472
Counter-Strike: Global Offensive	FPS	302
DayZ	Survival	239
Heroes of the Storm	MOBA	210
The Binding of Isaac: Rebirth	Other	171
Gaming Talk Shows	Other	170
Hearthstone: Heroes of Warcraft	Other	90
World of Tanks	Other	86
Agar.io	Other	71
Rocket League	Other	49
SpeedRunners	Other	20
ARK: Survival Evolved	Survival	19
Ø	Other	15
Duck Game	Other	5
Fallout 3	Other	3

3.5. Change in view count during a day (USA)

-- CHANGE IN VIEW COUNT IN A DAY

```
SELECT strftime("%H",time),COUNT(*)
```

```
FROM stream
```

```
WHERE country='US'
```

```
GROUP BY 1;
```

strftime("%H",time)	COUNT(*)
00	30
01	17
02	34
03	29
04	19
05	14
06	3
07	2
08	4
09	9
10	5
11	48
12	62
13	58
14	40
15	51
16	69
17	55
18	76
19	81
20	102
21	120
22	71

23	63
----	----

3.6. Joining stream and chat tables.

--JOINING STREAM AND CHAT TABLES

SELECT * FROM stream

JOIN chat ON

stream.device_id=chat.device_id;

time	device_id	login	channel	country	player	game	stream_format	subscriber	time	device_id	login	channel	country	player	game
01-01-2015 15:48	bab2163d35dd88e6ec804436e8cbb1902d1655520277ce		frank		iphone_t	League of Legends			01-01-2015 22:37	163d35dd88e6ec804436e8cbb1902d1655520277ce		frank		iphone_t	League of Legends
01-01-2015 13:42	b28971fa094091fbc052a2c8bd9b12c81be4e5c0e6eebc69f		helen	US	site	Hearthstone: Heroes of Warcraft	medium	TRUE	01-01-2015 12:43	1fa094091fbc052a2c8bd9b12c81be4e5c0e6eebc69f		helen	US		Hearthstone: Heroes of Warcraft
01-01-2015 14:47	6d60d8ebf5dc343f601b880ad14ee6d1ba50a467f2e92e		morty	TW	android	Heroes of the Storm			01-01-2015 13:26	8ebf5dc343f601b880ad14ee6d1ba50a467f2e92e		morty	TW	android	Heroes of the Storm
01-01-2015 02:08	453dd66d5be0167ad9b56223c2eda57c84e5063500898f2		susan	RU	ipad_t	World of Tanks			01-01-2015 00:17	66d5be0167ad9b56223c2eda57c84e5063500898f2		susan	RU	ipad_t	World of Tanks
01-01-2015 03:16	4135dd9393e3d357b0a80c36bd3aa517e3bfac508bb0f		frank		iphone_t	League of Legends			01-01-2015 04:04	dd9393e3d357b0a80c36bd3aa517e3bfac508bb0f		frank		iphone_t	League of Legends

4. Visualizing Twitch Data using Python and Matplotlib

```
import codecademylib3_seaborn
```

```
from matplotlib import pyplot as plt
```

```
import numpy as np
```

```
import pandas as pd
```

```
# Bar Graph: Featured Games
```

```
games = ["LoL", "Dota 2", "CS:GO", "DayZ", "HOS", "Isaac", "Shows", "Hearth", "WoT", "Agar.io"]
```

```
viewers = [1070, 472, 302, 239, 210, 171, 170, 90, 86, 71]
```

```
#Bar chart to graph viewers by game
```

```
ax=plt.subplot()
```

```
ax.set_xticks(range(len(games)))
```

```
ax.set_xticklabels(games, rotation=30)
```

```
plt.bar(range(len(games)),viewers,color="purple")
```

```
plt.xlabel("Games")
```

```
plt.ylabel("Viewers")
```

```
plt.legend(["Twitch"])
plt.title("Top 10 trending games")
plt.show()
plt.clf()
```

Pie Chart: League of Legends Viewers' Whereabouts

```
labels = ["US", "DE", "CA", "N/A", "GB", "TR", "BR", "DK", "PL", "BE", "NL", "Others"]
```

```
countries = [447, 66, 64, 49, 45, 28, 25, 20, 19, 17, 17, 279]
```

```
colors = ['lightskyblue', 'gold', 'lightcoral', 'gainsboro', 'royalblue', 'lightpink',
'darkseagreen', 'sienna', 'khaki', 'gold', 'violet', 'yellowgreen']
```

```
explode = (0.1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
```

```
plt.pie(countries, explode=explode, colors=colors, shadow=True, startangle=345,
autopct='%1.0f%%', pctdistance=1.15)
```

```
plt.legend(labels, loc="right")
```

```
plt.title("League of Legends Viewers' Whereabouts")
```

```
plt.show()
```

```
plt.clf()
```

Line Graph: Time Series Analysis

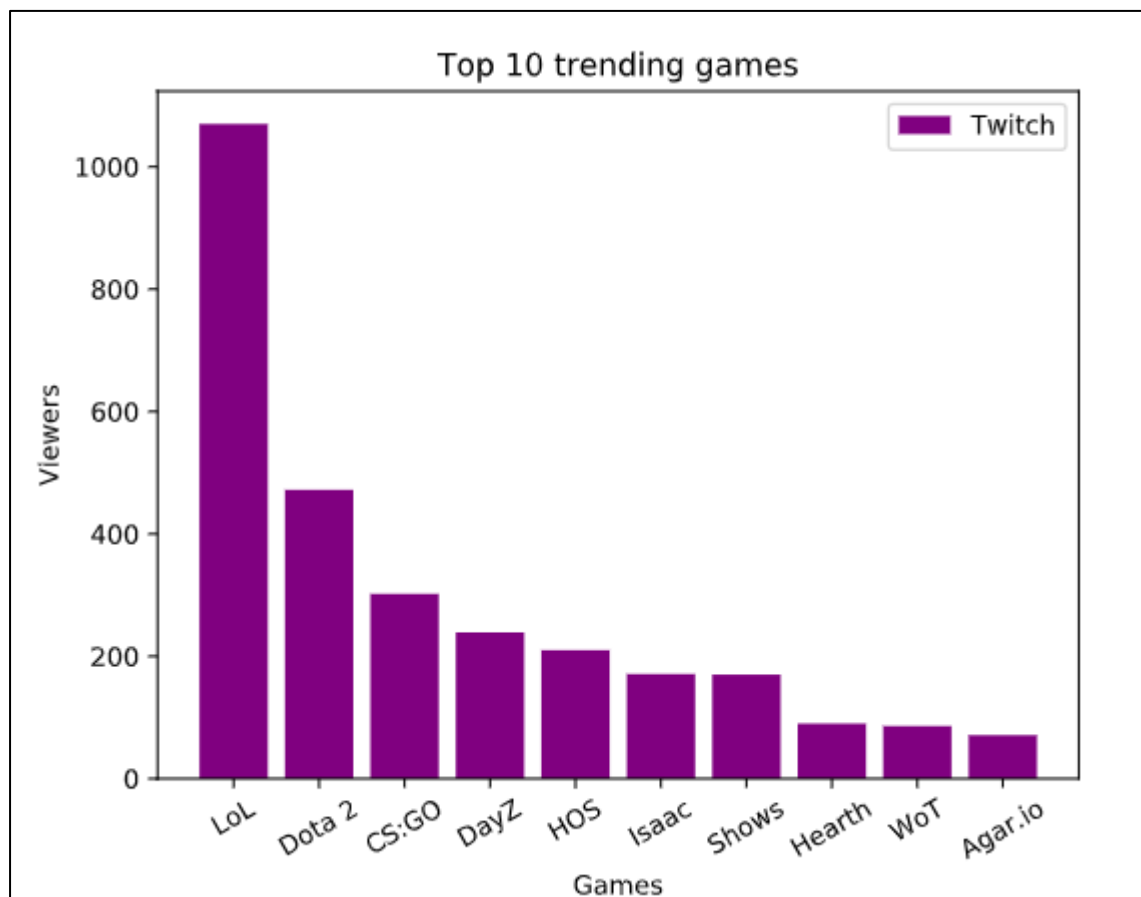
```
hour = range(24)
```

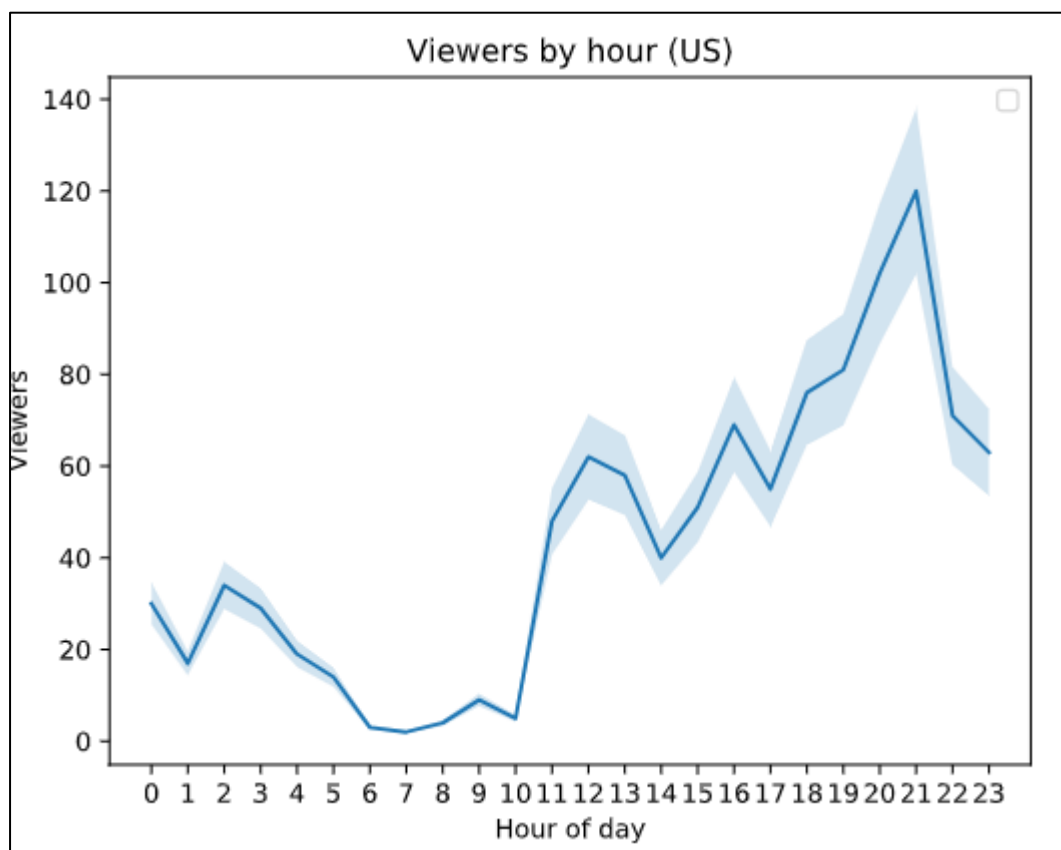
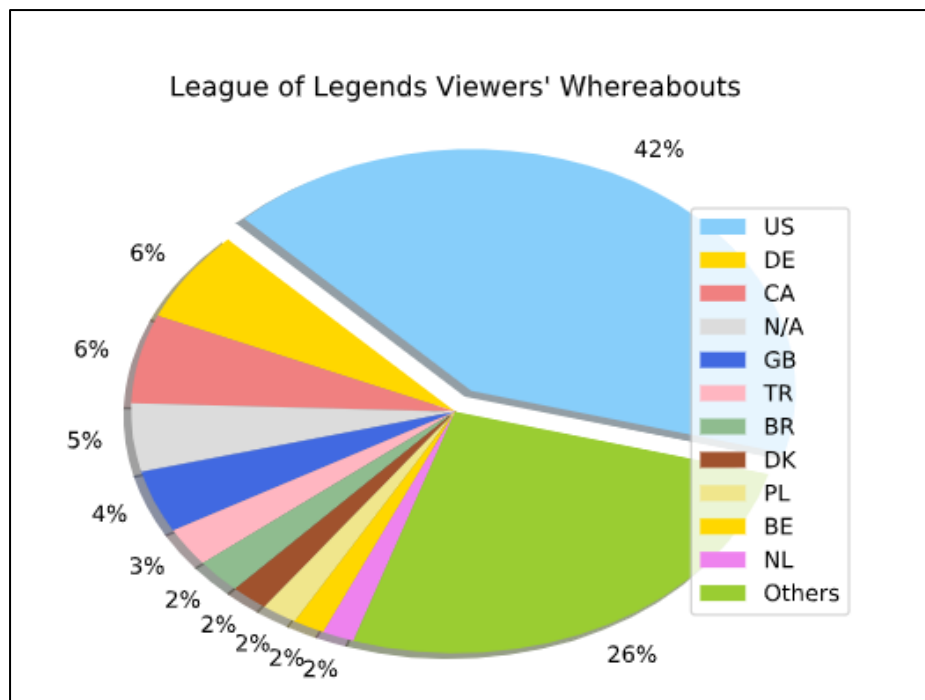
```
viewers_hour = [30, 17, 34, 29, 19, 14, 3, 2, 4, 9, 5, 48, 62, 58, 40, 51, 69, 55, 76,
81, 102, 120, 71, 63]
```

```
y_upper=[i*1.15 for i in viewers_hour]
```

```
y_lower=[0.85*i for i in viewers_hour]
```

```
ax=plt.subplot()
ax.set_xticks(hour)
ax.set_xticklabels(hour)
plt.legend("Viewers")
plt.xlabel("Hour of day")
plt.ylabel("Viewers")
plt.title("Viewers by hour (US)")
plt.legend("2015-01-01")
plt.fill_between(hour,y_lower,y_upper, alpha=0.2)
plt.plot(hour,viewers_hour)
plt.show()
plt.clf()
```





5. Conclusion

5.1. I was able to extract specific datasets using SQL to answer targeted questions. This data formed the base data for the visualization part of the project.

- 5.2. League of legends, DOTA 2 and CS:GO are the 3 most popular video games by number of viewers. 2 of the top 3 games by viewership are “MOBA” (multiplayer online battle arena) type while the third is a “FPS” (first-person shooter) type game.
- 5.3. An average user from USA tends to be most active on the platform between 8pm-9pm local time.
- 5.4. The lion’s share (>40%) of viewers are from USA.