code|cademy

Gaming Data with Twitch

Analyze Data with SQL
Dillon Medd
11-15-2021

1. Getting Started

1.1 Twitch Stream Dataset

- The practice dataset for this activity was provided by the Twitch Science Team
- It includes columns for time, device_id, login, channel, country, player, game , stream_format and subscriber
- There are 3000 rows, which all contain data from January 1st, 2015

Let's take a look at our response and the query result to the prompt,
"What are the unique games in the stream table?"

```
3
4   SELECT DISTINCT game FROM stream;
5
```

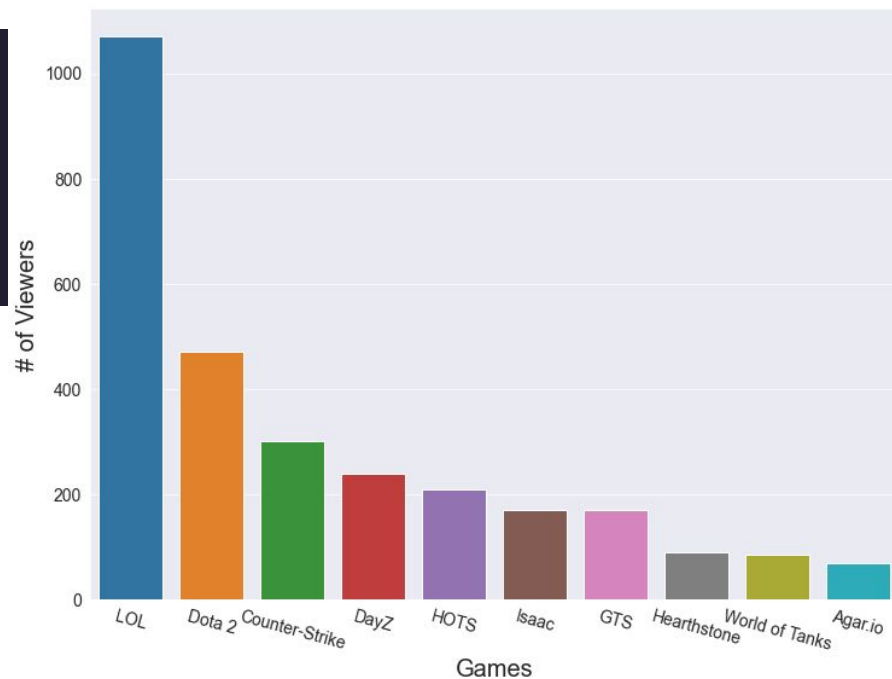| game |
| --- |
| League of Legends |
| DayZ |
| Dota 2 |
| Heroes of the Storm |
| Counter-Strike: Global Offensive |
| Hearthstone: Heroes of Warcraft |
| The Binding of Isaac: Rebirth |
| Agar.io |
| Gaming Talk Shows |
| Ø |
| Rocket League |
| World of Tanks |
| ARK: Survival Evolved |
| SpeedRunners |
| Breaking Point |
| Duck Game |
| Devil May Cry 4: Special Edition |
| Block N Load |
| Fallout 3 |
| Batman: Arkham Knight |

2. Aggregate Functions

2.1 Most Popular Games

Lets now looks at the ten most popular games in the stream table.
- This will include using Group By and Limit Clauses
- We will also include a barplot visualization created with the data visualization library, Seaborn

```
7
8  ▼ SELECT
9      game AS 'Game', COUNT(*) AS 'Number of Viewers'
10   FROM stream
11   GROUP BY game
12   ORDER BY COUNT(*) DESC
13   LIMIT 10;
14
```

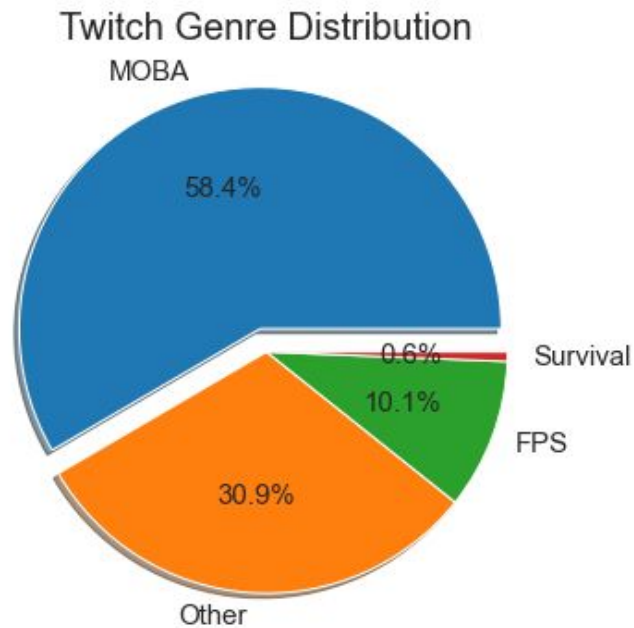| Game	Number of Viewers
League of Legends	1070
Dota 2	472
Counter-Strike: Global Offensive	302
DayZ	239
Heroes of the Storm	210
The Binding of Isaac: Rebirth	171
Gaming Talk Shows	170
Hearthstone: Heroes of Warcraft	90
World of Tanks	86
Agar.io	71

2.2 League of Legends by Country

- These are some big numbers from the game League of Legends (also known as LoL).
- Let take a look at where the stream viewer for LOL are located using Where, Group By and Order By clauses.

Country	Number of Viewers
US	447
DE	66
CA	64
GB	45
TR	28
BR	25
DK	20
PL	19
NL	17
BE	17
SE	16
RO	16
MX	16
PT	15
FR	15

```
15  ▼ SELECT
16      country as 'Country', COUNT(*) 'Number of Viewers'
17  ▼ FROM
18      stream
19    WHERE game = 'League of Legends'
20    AND country IS NOT NULL
21    GROUP BY 1
22    ORDER BY 2 DESC
23    LIMIT 15;
24
```

2.3 New Column for Genre

- Let's create a new column to look at the Genre distribution of the games in this dataset
- We can do this with a Case Statement
- In the top right, the distribution is displayed with a pie chart created with the Matplotlib visualization library

Twitch Genre Distribution


```
31
32  ▼ SELECT
33  ▼    CASE
34         WHEN game IN ('League of Legends', 'Dota 2', 'Heroes of the Storm'
35  ▼     )   THEN 'MOBA'
36  ▼       WHEN game = 'Counter-Strike: Global Offensive'
37           THEN 'FPS'
38  ▼       WHEN game IN ('Dayz', 'ARK: Survival Evolved')
39           THEN 'Survival'
40         ELSE 'Other'
41         END AS 'Genre',
42         COUNT(*) AS 'Number of Streams'
43  ▼ FROM
44       stream
45    GROUP BY 1
46    ORDER BY 2 DESC;
47
```

Genre	Number of Streams
MOBA	1752
Other	927
FPS	302
Survival	19

3. Strftime Function

3.1 Viewer Count by Hour

- Using the strftime function, let's take a look the viewers organized by the hour of the day

```
▼ SELECT
    strftime('%H', time) as 'Hour',
    COUNT(*) as 'View Count'
FROM stream
WHERE country = 'US'
GROUP BY 1;
```

Hour	View 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