code|cademy

Analyze Twitch Gaming Data

Analyze Data with SQL
Anna Zinkovich
Data Analyst

Project Description

This project was completed during Codecademy's Career Path "Learn Data Analysis for your Business":
https://www.codecademy.com/learn/paths/data-analyst-training-for-your-business-cfb

STACK: SQL

Twitch is the world's leading live streaming platform for gamers, with 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 worked with two tables that contain Twitch users' stream viewing data and chat room usage data.
Stream viewing data:
•stream table
Chat usage data:
•chat table

The Twitch Science Team provided this practice dataset (the **.csv** files with 800,000 rows)

To get a feel for the stream and the chat tables, select the first 10 rows from each of the two tables.

SELECT * FROM stream LIMIT 10;

time	device_id	login	channel	country	player	game	stream_format	subscriber
2015-01-01 18:33:52	40ffc2fa6534cf760becbdbf5311e31ad069e46e	085c1eb7b587bfe654f0df7b4ba7f4fc4013636c	frank	US	iphone_t	League of Legends	Null	Null
2015-01-01 23:35:33	9a8cc2b7162b99c0a0f501dc9a5ec4f68586a760	5ad49a7b408ce452140b180dd6efb57a9f4d22c7	george	US	site	DayZ	chunked	False
2015-01-01 04:39:38	2f9c3f9ee1033b71a3819564243f20ef3bec0183	5b9a43e68f019185f55615d0b83019dee4b5d06f	frank	US	site	League of Legends	chunked	False
2015-01-01 11:15:30	0cda8226ba2583424c80c3c1c22c1256b080ad17	02c7797faa4d8a3ff4b0c14ee1764b6817b53d0b	estelle	CH	site	Dota 2	high	False
2015-01-01 11:28:19	e3288ca5e3153aa85e32f64cdd994b7666968dcf	b920c228acbcbebee26d9c79f6eb73b73a9480c7	morty	FR	site	Heroes of the Storm	medium	False
2015-01-01 23:27:36	343fe2bfd58595d5c18602d420ecf6f9d694d5a8	7814f661a54349ff5eee84f9d6f476918c9b7270	george	US	iphone_t	DayZ	Null	Null
2015-01-01 21:09:23	80a0c7d1abb6a5a0060e18202b77bef831e08ca5	eb158cab3f606d3894a32e20bddbfd2d589095a9	frank	US	site	League of Legends	high	True
2015-01-01 19:14:27	1e342e5e4e228f617449029054b3bb19c5224528	2aaf6a414bc3dc923b04f986de7ba1b8101c6698	frank	CA	site	League of Legends	high	False
2015-01-01 13:51:04	272cffbb1a9a33ad3bb48a2ee9ae5cbcac5ca22e	401716920e3435b5e3eec9fc4ccd6a44c7af38f5	kramer	TR	site	Counter-Strike: Global Offensive	chunked	False
2015-01-01 22:00:14	593ed161c456eeeb9e18b8005786d42abc1a7373	ef24dc49ceb4bcd3cccb0fa862d8a150ecf935e1	frank	US	site	League of Legends	medium	False

To get a feel for the stream and the chat tables, select the first 10 rows from each of the two tables.

SELECT * FROM chat LIMIT 10;

time	device_id	login	channel	country	player	game	time	device_id
2015-01-01 18:45:50	70e2b95b5ac0d4c227e46966658d16b3e044996e	5c2f5c1f19a7738e16ed0be551d865e8a8fce71d	jerry	BY	Null	Dota 2	2015-01-01 18:45:50	70e2b95b5ac0d4c227e46966658d16b3e044996e
2015-01-01 01:16:57	f2b9065b55fd80d6aa653ce989b489f4ec5198be	0d77740e4fb5ce77d94f9f6c8ef1f762990d0344	elaine	HK	Null	Devil May Cry 4: Special Edition	2015-01-01 01:16:57	f2b9065b55fd80d6aa653ce989b489f4ec5198be
2015-01-01 16:22:10	d448ba963d7e1023dd1b0a40b95d4f6611750692	77ab14c1fb815e1c369ba0cf7d4c56b4fe489997	frank	GB	iphone_t	League of Legends	2015-01-01 16:22:10	d448ba963d7e1023dd1b0a40b95d4f6611750692
2015-01-01 03:58:13	8d6823dc52b400b50aebf269bf1f03a36d19eeaa	91cb88c0743761589273fc5e800e7743ece46494	frank	US	iphone_t	League of Legends	2015-01-01 03:58:13	8d6823dc52b400b50aebf269bf1f03a36d19eeaa
2015-01-01 11:47:35	16c1e39594d62358d27ae604ad43a071f0d86bc4	51a9234f83d656607cfd7f26690c12d2ffbce353	estelle	DE	Null	Dota 2	2015-01-01 11:47:35	16c1e39594d62358d27ae604ad43a071f0d86bc4
2015-01-01 17:59:51	6fcc75522de37833a0fb21fba4965aad3b63ea57	f628d1cb946ea2e8cffc0b327bc9d77775b8d3c0	jerry	RU	Null	Dota 2	2015-01-01 17:59:51	6fcc75522de37833a0fb21fba4965aad3b63ea57
2015-01-01 02:24:33	dea94b3030025d837dd841fbfd479e775987f65d	9dbbcf6c7792074771c4c7284807041eac467ad5	elaine	TW	Null	Gaming Talk Shows	2015-01-01 02:24:33	dea94b3030025d837dd841fbfd479e775987f65d
2015-01-01 18:26:34	671bee0f3d66077876d9bc231990597292392cc2	51c286a41daa8e060275f622f2b8436bee9fab91	jerry	UA	Null	Dota 2	2015-01-01 18:26:34	671bee0f3d66077876d9bc231990597292392cc2
2015-01-01 13:13:18	8b31d5ebd1f4f41d4365ae4a471c1686dd256745	06decad1d9565150791e183da017f47123433a4c	estelle	GB	ipad_t	Dota 2	2015-01-01 13:13:18	8b31d5ebd1f4f41d4365ae4a471c1686dd256745
2015-01-01 20:20:55	f2ebb129e6930e608f2ed3f5fb52bc4d533c4891	4679f8113aa157ba76fc6db5878d7ee625e88d55	frank	CA	Null	League of Legends	2015-01-01 20:20:55	f2ebb129e6930e608f2ed3f5fb52bc4d533c4891

What are the unique games in the stream table?

```
SELECT DISTINCT(game) FROM stream;
```

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

What are the unique channels in the stream table?

SELECT DISTINCT(channel) FROM stream;
channel
frank
george
estelle
morty
kramer
jerry
helen
newman
elaine
susan

What are the most popular games in the stream table?

```
SELECT game, COUNT(*) as number_of_viewers
FROM stream
GROUP BY game
ORDER BY number_of_viewers DESC;
```

game	number_of_viewers	game (continuation of the table)	number_of_viewers (continuation of the table)
League of Legends	1070	Rocket League	49
Dota 2	472	SpeedRunners	20
Counter-Strike: Global Offensive	302	ARK: Survival Evolved	19
DayZ	239	Null	15
Heroes of the Storm	210	Duck Game	5
The Binding of Isaac: Rebirth	171	Fallout 3	3
Gaming Talk Shows	170	Batman: Arkham Knight	3
Hearthstone: Heroes of Warcraft	90	Breaking Point	2
World of Tanks	86	Block N Load	2
Agar.io	71	Devil May Cry 4: Special Edition	1

Where are League of Legends (LoL/lol) stream viewers located?

```
SELECT country, COUNT(*) as lol_viewers
FROM stream
WHERE game = "League of Legends"
GROUP BY country
ORDER BY 2 DESC LIMIT 10;
```

country	lol_viewers
US	447
DE	66
CA	64
Null	49
GB	45
TR	28
BR	25
DK	20
PL	19
NL	17

How many viewers are using each of Twitch's stream sources to watch streams?

```
SELECT player AS stream_source, COUNT(*) AS viewers_count
FROM stream
GROUP BY player
ORDER BY viewers_count DESC;
```

stream_source	viewers_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

Create a genre column, group games by their genres and count viewers.

```
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(*) as viewers_count
FROM stream GROUP BY game ORDER BY 3 DESC;
```

game	genre	viewers_count	game (continuation of the table)	genre (continuation of the table)	viewers_count (continuation of the table)
League of Legends	MOBA	1070	Rocket League	OTHER	49
Dota 2	MOBA	472	SpeedRunners	OTHER	20
Counter-Strike: Global Offensive	FPS	302	ARK: Survival Evolved	Survival	19
DayZ	Survival	239	Null	OTHER	15
Heroes of the Storm	MOBA	210	Duck Game	OTHER	5
The Binding of Isaac: Rebirth	OTHER	171	Fallout 3	OTHER	3
Gaming Talk Shows	OTHER	170	Batman: Arkham Knight	OTHER	3
Hearthstone: Heroes of Warcraft	OTHER	90	Breaking Point	OTHER	2
World of Tanks	OTHER	86	Block N Load	OTHER	2
Agar.io	OTHER	71	Devil May Cry 4: Special Edition	OTHER	1

Write a query that returns the hours of the time column, the view count for each hour and filter the result with only users in your country.

```
SELECT strftime('%H', time) as hours, COUNT(*)
FROM stream
WHERE country = 'DE'
GROUP BY hours;
```

hours	view_count	hours (continuation of the table)	view_count (continuation of the table)
00	1	13	24
01	3	14	18
02	7	15	10
03	9	16	11
04	12	17	7
05	5	18	8
08	1	19	5
09	2	20	3
10	2	21	3
11	12	22	6
12	15	23	1

Write a query that returns the hours of the time column, the view count for each hour and filter the result with only users in your country.

```
SELECT strftime('%H', time) as hours, COUNT(*)
FROM stream
WHERE country = 'BY'
GROUP BY hours;
```

hours	view_count
01	1
08	1
12	2
16	1
18	1

Thank you.
Do you have any questions?