code|cademy

Twitch Data

Analyze Data with SQL
Knox Maclean Bather
September 2023

Table of Contents

1) What are the **column names** for the Twitch data tables?

2) How many **rows** are there in the chat and the video_play tables?

3) How many **unique users** appear in both tables?

4) Show the **proportions** of only chat users compared to total chat users, and only video_play users compared to total video_play users.

5) Which **countries** make up the largest proportion of the viewership?

6) Which **games** are played? **Rank** them from most viewed to least.

7) Which **times** have peak and least **traffic**?

8) How does the **ranking** of game viewership **change** throughout the day? Choose the peak, lowest and two times between these times.

9a) How does League of Legend's viewership change over a 24-hour period?
9b) Does the top **locations** of League of Legend's viewers **change** over the course of the day?

10) How many channels are there, and which is the most popular? Also work out their subscribers.

Q1. What are the column names for the Twitch data tables?

```
CODE

select *
from video_play
limit 10;

select *
from chat
limit 10;
```

Video_play Table:
- Time
- device_id
- login
- channel
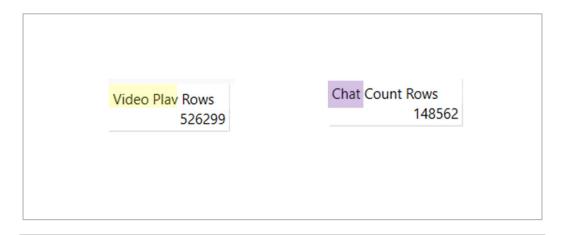- country
- player
- game
- stream_format
- subscriber

	time	device id	login	channel	country	player	game	stream format	subscriber
1	2015-01-01 18:33:52	40ffc2fa6534cf760...	085c1eb7b587bf...	frank	US	iphone_t	League of Legends	NULL	NULL
2	2015-01-01 23:35:33	9a8cc2b7162b99c0...	5ad49a7b408ce4...	george	US	site	DayZ	chunked	false
3	2015-01-01 04:39:38	2f9c3f9ee1033b71...	5b9a43e68f0191...	frank	US	site	League of Legends	chunked	false
4	2015-01-01 11:15:30	0cda8226ba258342...	02c7797faa4d8a3...	estelle	CH	site	Dota 2	high	false
5	2015-01-01 11:28:19	e3288ca5e3153aa8...	b920c228acbcbe...	morty	FR	site	Heroes of the Storm	medium	false
6	2015-01-01 23:27:36	343fe2bfd58595d5...	7814f661a54349f...	george	US	iphone_t	DayZ	NULL	NULL
7	2015-01-01 21:09:23	80a0c7d1abb6a5a0...	eb158cab3f606d...	frank	US	site	League of Legends	high	true
8	2015-01-01 19:14:27	1e342e5e4e228f61...	2aaf6a414bc3dc9...	frank	CA	site	League of Legends	high	false
9	2015-01-01 13:51:04	272cffbb1a9a33ad...	401716920e3435...	kramer	TR	site	Counter-Strike: Global Offensive	chunked	false
10	2015-01-01 22:00:14	593ed161c456eeeb...	ef24dc49ceb4bcd...	frank	US	site	League of Legends	medium	false

Chat Table:
- Time
- Device_id
- Login
- Channel
- Country
- Player
- Game

time	device id	login	channel	country	player	game
2015-01-01 18:45:50	70e2b95b5ac0d4c227e46966658d1b63e044996e	5c2f5c1f19a7738e16ed0be551d865e8a8fce71d	jerry	BY	NULL	Dota 2
2015-01-01 01:16:57	f2b9065b55fd80d6aa653ce989b489f4ec5198be	0d77740e4fb5ce77d94f9f6c8ef1f762990d0344	elaine	HK	NULL	Devil May Cry 4: Special Edition
2015-01-01 16:22:10	d448ba963d7e1023dd1b0a40b95d4f6611750692	77ab14c1fb815e1c369ba0cf7d4c56b4fe489997	frank	GB	iphone_t	League of Legends
2015-01-01 03:58:13	8d6823dc52b400b50aebf269bf1f03a36d19eeaa	91cb88c0743761589273fc5e800e7743ece46494	frank	US	iphone_t	League of Legends
2015-01-01 11:47:35	16c1e39594d62358d27ae604ad43a071f0d86bc4	51a9234f83d656607cfd7f26690c12d2ffbce353	estelle	DE	NULL	Dota 2
2015-01-01 17:59:51	6fcc75522de37833a0fb21fba4965aad3b63ea57	f628d1cb946ea2e8cffc0b327bc9d77775b8d3c0	jerry	RU	NULL	Dota 2
2015-01-01 02:24:33	dea94b3030025d837dd841fbfd479e775987f65d	9dbbcf6c7792074771c4c7284807041eac467ad5	elaine	TW	NULL	Gaming Talk Shows
2015-01-01 18:26:34	671bee0f3d66077876d9bc231990597292392cc2	51c286a41daa8e060275f622f2b8436bee9fab91	jerry	UA	NULL	Dota 2
2015-01-01 13:13:18	8b31d5ebd1f4f41d4365ae4a471c1686dd256745	06decad1d9565150791e183da017f47123433a4c	estelle	GB	ipad_t	Dota 2
2015-01-01 20:20:55	f2ebb129e6930e608f2ed3f5fb52bc4d533c4891	4679f8113aa157ba76fc6db5878d7ee625e88d55	frank	CA	NULL	League of Legends

Q2. How many rows are there in the chat and the video_play tables?

Video Play Rows	Chat Count Rows
526299	148562

CODE

```
select count(*) as 'Video Play Rows'
from video_play;

select count(*) as 'Chat Count Rows'
from chat;
```

Q3. How many unique users appear in both tables?

There is a large difference between the amount of people who are streaming on twitch and the number using the chat function. Only ~9.5% of people use the chat function (9.6% according to accounts used and 9.4% according to devices used).

There is a large cross over between the two tables with 24,250 common logins being recorded in both the video_play and chat tables. It is surprising that this number is smaller than the unique logins recorded in the chat table which suggests that a small number of people (~650 people) use the chat function without watching a stream.

Note of Interest: The difference between account and device numbers suggest that a small portion of users are logged in on multiple devices. This difference may be reduced by people with multiple accounts.

Accounts used in Chat	Devices used in Chat	Difference
24896	25265	369

Accounts used in Video	Devices used in Video	Difference
258313	267629	9316

common logins
24250

CODE
```
select count(distinct login) as 'Accounts used in Chat',
      count(distinct device_id) as 'Devices used in Chat',
      (count(distinct device_id) - count(distinct login)) as
'Difference'
from chat;


select count(distinct login) as 'Accounts used in Video',
      count(distinct device_id) as 'Devices used in Video',
      (count(distinct device_id) - count(distinct login)) as
'Difference'
from video_play ;


select count(distinct login) as 'common logins'
from video_play
where login in (
select login
from chat);
```
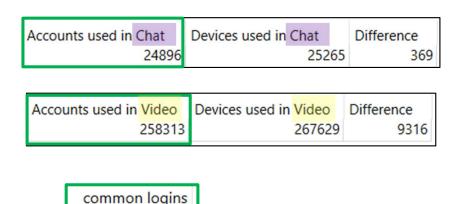
Q4. Show the proportions of only chat users compared to total chat users and only video_play users compared to total video_play users.

Using the previous figures (shown right) we can find the proportions of how people use twitch.

We will use the figures from 'Accounts' (highlighted yellow) for these calculations as the larger devices_id numbers suggests some users login on multiple devices making duplicate values.

- Only chat users / total chat users   = 2.59%
- Only chat users / total streamers   = 0.25%

- Non-chat streamers / total streamers   = 90.4%

Accounts used in Chat	Devices used in Chat	Difference
24896	25265	369

Accounts used in Video	Devices used in Video	Difference
258313	267629	9316

common logins
24250

Q5. Which countries make up the largest proportion of the viewership?

There are 167 unique countries in the video_play table [code 1].

By viewership, the **US** is the **largest country** by a significant margin for both the number of logins (**190,378**) and individuals (79,993) using the platform with 58.0% of users logging in multiple times.
The next five (**RU**, **TW**, **DE**, **CA**, and **GB** respectively) have similar in logins to each other ranging from **29,807 – 22,351**, very far from the US's numbers.

- The 'NULL' row is a large number of logins that have not been linked to a country. Given the extensive list of countries already identified we can assume these logins are mostly made up of countries on the list. It is likely that a large proportion of these 17,161 logins are from the US given the country's extensive numbers.

CODE

```
1) select count(distinct country)
   from video_play
   where country is not null
   order by country asc;

2) select country, count(login) as 'Login count', count(distinct
login) as 'Individuals',
count(login) - count(distinct login) as 'difference',
round(100.0*(count(login) - count(distinct login)) / count(login), 1)
as 'Multiple logins %'
from video_play
group by 1
order by 2 desc;
```

count(DISTINCT country)
167

	country	Login count	Individuals	difference	Multiple logins %
1	US	190378	79993	110385	58
2	RU	29807	15927	13880	46.6
3	TW	29489	16552	12937	43.9
4	DE	29262	16588	12674	43.3
5	CA	27350	12881	14469	52.9
6	GB	22351	11271	11080	49.6
7	NULL	17161	7640	9521	55.5
8	SE	12648	7638	5010	39.6
9	PL	10025	6363	3662	36.5
10	TR	9313	5420	3893	41.8
11	BR	7637	4149	3488	45.7
12	FR	7292	4220	3072	42.1
13	DK	6971	3997	2974	42.7
14	AU	6933	3394	3539	51
15	UA	6655	3498	3157	47.4
16	NL	6432	3568	2864	44.5

Q6. Which games are played? Rank them from most viewed to least.

League of Legends is by far the most streamed game with **193533** data rows linked to the game.
Dota 2 is the second most poular with 85,608 views
Counter-strike is third with 54,438.
- These top three games have been highlighted blue, orange and red, respectively.

CODE

```
select distinct game, count(*)
from video_play
group by 1
order by 2 desc;
```
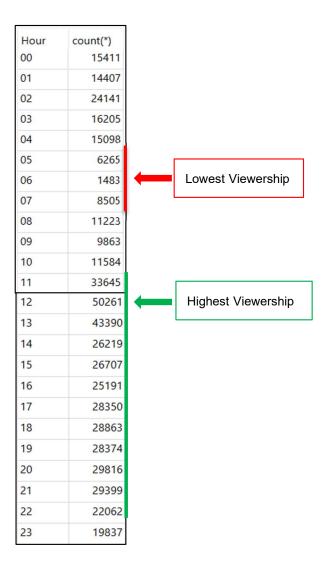
	game	count(*)
1	League of Legends	193533
2	Dota 2	85608
3	Counter-Strike: Global Offensive	54438
4	DayZ	38004
5	Heroes of the Storm	35310
6	The Binding of Isaac: Rebirth	29467
7	Gaming Talk Shows	28115
8	World of Tanks	15932
9	Hearthstone: Heroes of Warcraft	14399
10	Agar.io	11480
11	Rocket League	7087
12	ARK: Survival Evolved	4158
13	SpeedRunners	3367
14	NULL	3124
15	Duck Game	1063
16	Fallout 3	485
17	Devil May Cry 4: Special Edition	231
18	Breaking Point	161
19	Batman: Arkham Knight	117
20	Reign Of Kings	50

Q7. Which times have peak and least traffic?

Peak traffic time = **12:00** – 12:59 pm
Lowest traffic time = **06:00** – 06:59 am

Noon is the peak time for traffic with views scaling up quickly from 11am to 12pm before trailing down in viewership over the afternoon. This slow decline in viewership continues well into the night with a small peak at 2:00 before dropping suddenly at 5:00 and leading to the 6am trough.

CODE

```
select strftime('%H', time) as 'Hour', count(*)
from video_play
group by 1
order by 1
;
```

Hour	count(*)
00	15411
01	14407
02	24141
03	16205
04	15098
05	6265
06	1483
07	8505
08	11223
09	9863
10	11584
11	33645
12	50261
13	43390
14	26219
15	26707
16	25191
17	28350
18	28863
19	28374
20	29816
21	29399
22	22062
23	19837

← Lowest Viewership

← Highest Viewership

Q8. How does the ranking of game viewership change throughout the day? Choose the peak, lowest and two times between these times.

	game	count(*)	Peak
1	Dota 2	25249	12
2	Counter-Strike: Global Offensive	19595	12
3	Heroes of the Storm	2592	12
4	Hearthstone: Heroes of Warcraft	1102	12
5	World of Tanks	924	12
6	League of Legends	340	12
7	ARK: Survival Evolved	174	12
8	NULL	94	12
9	The Binding of Isaac: Rebirth	54	12
10	Agar.io	44	12
11	Fallout 3	36	12
12	Breaking Point	14	12
13	Batman: Arkham Knight	13	12
14	The Witcher 3: Wild Hunt	10	12
15	Gaming Talk Shows	10	12
16	Reign Of Kings	5	12
17	Super Mario Bros.	2	12
18	Block N Load	2	12
19	Devil May Cry 4: Special Edition	1	12

	game	count(*)	Lowest
1	World of Tanks	609	06
2	League of Legends	298	06
3	ARK: Survival Evolved	168	06
4	Counter-Strike: Global Offensive	106	06
5	Dota 2	92	06
6	Hearthstone: Heroes of Warcraft	49	06
7	The Binding of Isaac: Rebirth	44	06
8	Fallout 3	36	06
9	Agar.io	25	06
10	Heroes of the Storm	22	06
11	NULL	13	06
12	Gaming Talk Shows	8	06
13	Breaking Point	5	06
14	Reign Of Kings	4	06
15	Batman: Arkham Knight	2	06
16	Devil May Cry 4: Special Edition	1	06
17	Bridge Constructor Medieval	1	06

	game	count(*)	Time
1	League of Legends	12080	18
2	The Binding of Isaac: Rebirth	5896	18
3	Dota 2	5703	18
4	Heroes of the Storm	2233	18
5	Hearthstone: Heroes of Warcraft	1812	18
6	World of Tanks	314	18
7	ARK: Survival Evolved	231	18
8	NULL	148	18
9	Agar.io	141	18
10	Counter-Strike: Global Offensive	139	18
11	Rocket League	104	18
12	SpeedRunners	38	18
13	Gaming Talk Shows	10	18
14	Block N Load	4	18
15	Fallout 3	3	18
16	Risk of Rain	2	18
17	Breaking Point	2	18
18	Batman: Arkham Knight	2	18
19	The Witcher 3: Wild Hunt	1	18

	game	count(*)	Time
1	League of Legends	14016	00
2	World of Tanks	681	00
3	ARK: Survival Evolved	227	00
4	Counter-Strike: Global Offensive	137	00
5	Dota 2	97	00
6	NULL	96	00
7	Agar.io	35	00
8	Fallout 3	30	00
9	Heroes of the Storm	29	00
10	Hearthstone: Heroes of Warcraft	25	00
11	The Binding of Isaac: Rebirth	19	00
12	Breaking Point	8	00
13	Batman: Arkham Knight	5	00
14	Mortal Kombat X	2	00
15	The Witcher 3: Wild Hunt	1	00
16	The Last of Us	1	00
17	Reign Of Kings	1	00
18	Devil May Cry 4: Special Edition	1	00

Using peak time (**12pm**), lowest traffic time (**6am**) and two other times – **18:00** and **24:00**. A similar code (right) was used, varying the highlighted part for the numbers 12, 06, 18, 00 and appropriate descriptions.

Surprisingly, the most viewed game overall (LoL) does not have much of an audience during peak time. It's audience peaks in the afternoon and at night suggesting this may be to do with viewership locations/ work culture.

```
select distinct game, count(*),
strftime('%H', time) as 'Peak time'
from video_play
where strftime('%H', time) = '12'
group by 1
order by 2 desc;
```

Q9a. How does League of Legend's viewership change over a 24-hour period?

League of Legend viewership maintains a high count from the afternoon to the earlier hours of the morning [16:00 to 04:00]. This time range has been highlighted with the green lines.

There are two noticeable **peak** times, the highest of these is at **21:00** with 19,506 views (indicated with a green arrow) the other is just smaller at **02:00** with 18,306 views.

The times of **lowest** viewership are between 06:00 and 14:00, these have been highlighted with a red line, and the hour of least views is at **10:00** with just 259 views (indicated with red arrow).

These peak and least timings appear to correlate with standard working hours, perhaps the location of viewership will indicate further correlation.

CODE
```
select strftime('%H', time) as 'Time', game, count(*)
from video_play
where game like 'League%'
group by 1
order by 1;
```

Time	game	count(*)
00	League of Legends	14016
01	League of Legends	11106
02	League of Legends	18306
03	League of Legends	14987
04	League of Legends	13903
05	League of Legends	5118
06	League of Legends	298
07	League of Legends	316
08	League of Legends	282
09	League of Legends	286
10	League of Legends	259
11	League of Legends	315
12	League of Legends	340
13	League of Legends	310
14	League of Legends	322
15	League of Legends	7407
16	League of Legends	14062
17	League of Legends	14174
18	League of Legends	12080
19	League of Legends	10837
20	League of Legends	13192
21	League of Legends	19506
22	League of Legends	11723
23	League of Legends	10388

Lowest Viewership

Highest Viewership

Q9b. Does the top locations of League of Legend's viewers change over the course of the day?

Time	game	count(*)	country
21	League of Legends	11344	US
21	League of Legends	1906	CA
21	League of Legends	908	NULL
21	League of Legends	608	DE
21	League of Legends	424	TW
21	League of Legends	344	AU
21	League of Legends	329	TR
21	League of Legends	327	MX
21	League of Legends	311	BR
21	League of Legends	250	GB
21	League of Legends	162	SE
21	League of Legends	144	PT
21	League of Legends	138	CL
21	League of Legends	117	RU
21	League of Legends	107	PL
21	League of Legends	104	NO
21	League of Legends	98	RO
21	League of Legends	90	NZ
21	League of Legends	88	DK
21	League of Legends	87	HK

Time	game	count(*)	country
03	League of Legends	4099	US
03	League of Legends	1219	DE
03	League of Legends	705	GB
03	League of Legends	485	CA
03	League of Legends	472	NULL
03	League of Legends	445	TR
03	League of Legends	444	AU
03	League of Legends	437	NL
03	League of Legends	432	SE
03	League of Legends	385	GR
03	League of Legends	374	DK
03	League of Legends	372	RO
03	League of Legends	359	PL
03	League of Legends	332	IT
03	League of Legends	327	FR
03	League of Legends	268	TW
03	League of Legends	256	PT
03	League of Legends	245	BE
03	League of Legends	242	NO
03	League of Legends	198	CZ

Time	game	count(*)	country
09	League of Legends	132	US
09	League of Legends	27	CA
09	League of Legends	13	NULL
09	League of Legends	12	DE
09	League of Legends	10	SE
09	League of Legends	7	TW
09	League of Legends	7	GB
09	League of Legends	6	TR
09	League of Legends	6	GR
09	League of Legends	5	AU
09	League of Legends	4	RO
09	League of Legends	4	NL
09	League of Legends	4	MX
09	League of Legends	4	KW
09	League of Legends	4	DK
09	League of Legends	4	CH
09	League of Legends	3	SA
09	League of Legends	3	RU
09	League of Legends	3	PT
09	League of Legends	3	IT

Time	game	count(*)	country
15	League of Legends	2895	US
15	League of Legends	569	DE
15	League of Legends	442	GB
15	League of Legends	421	CA
15	League of Legends	266	NULL
15	League of Legends	188	NL
15	League of Legends	187	GR
15	League of Legends	186	SE
15	League of Legends	149	PT
15	League of Legends	141	DK
15	League of Legends	131	FR
15	League of Legends	122	PL
15	League of Legends	115	NO
15	League of Legends	102	TR
15	League of Legends	102	IT
15	League of Legends	99	ES
15	League of Legends	94	BR
15	League of Legends	92	RO
15	League of Legends	86	MX
15	League of Legends	85	BE

Taking the peak time and every 6 hours after that [**21:00, 03:00, 09:00** and **15:00**]
It can clearly be seen that the US has the greatest influence over viewership number, far out numbering any other country at all times of the day. The fact that viewership peaks outside of standard work hours and is lowest during them, lead us to the conclusion that the US working day has a large influence on the number of views a channel might get.

CODE
```
select strftime('%H', time) as 'Time', game, count(*),
country
from video_play
where game like 'League%' and strftime('%H', time) = '15'
group by 4
order by 3 desc
limit 20;
```

Q10. How many channels are there, and which is the most popular? Also work out their subscribers.

The largest channel is **Frank** with a massive **194300** viewers and **5830** subscribers.
The others are not spread as far apart in viewers with **Estell** second largest with **72484** viewers and then **Karmer** with **66756** viewers.
However, with the first and second queries we can see that more viewers does not necessarily translate to more subscribers. Frank is the only exception to this but his subscriber number is still very small relative to his viewership.

CODE

```
A) select channel, count(*)
   from video_play
   group by 1
   order by 2 desc;


B) select channel, count(*) as 'Subscribers'
   from video_play
   where subscriber = 'true'
   group by 1
   order by 2 desc;
```

A)

	channel	Viewership
1	frank	194300
2	estelle	72484
3	kramer	66756
4	george	43235
5	newman	41846
6	morty	35353
7	elaine	28390
8	susan	15958
9	helen	14769
10	jerry	13208

B)

	channel	Subscribers
1	frank	5830
2	newman	1133
3	george	719
4	elaine	692
5	kramer	690
6	morty	378
7	helen	309
8	susan	187
9	estelle	27
10	jerry	4