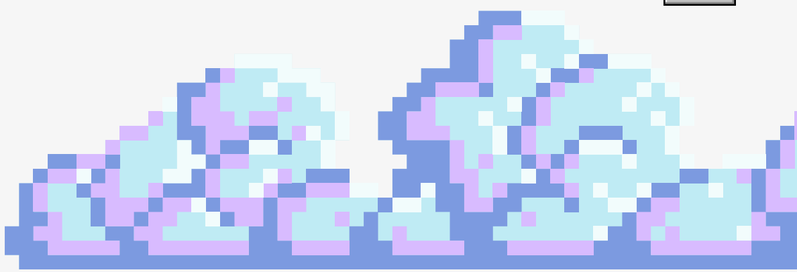
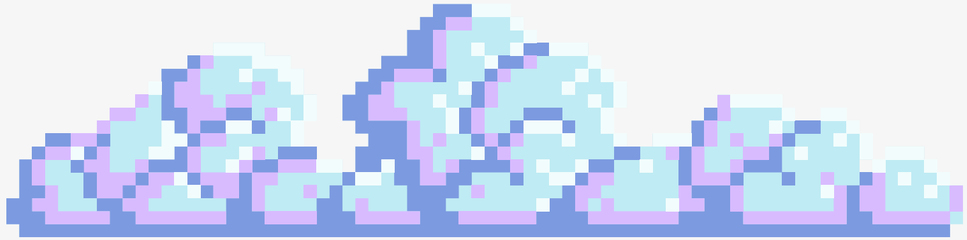
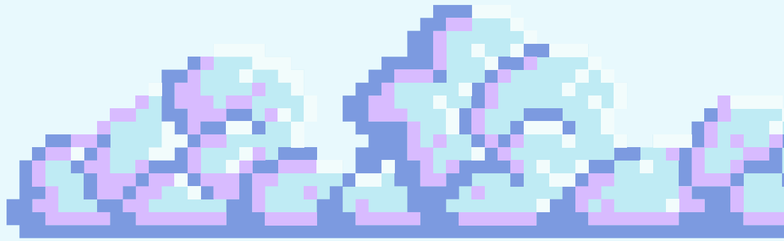
STEAM VIDEO GAME SALES AND... CS2



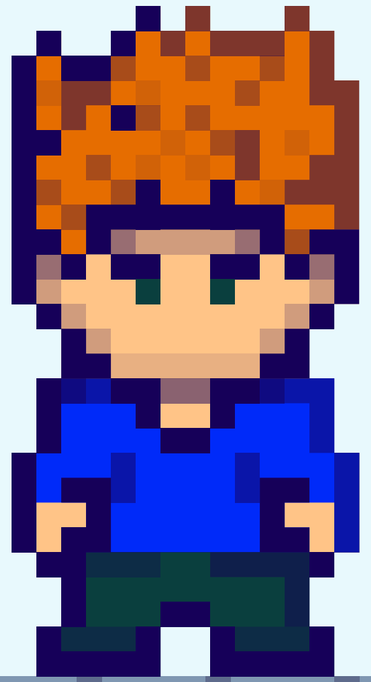


Author: Trần Tuấn Đạt

TABLE OF CONTENTS



DATA PREPARING



- Bước đầu tiên là thực hiện join và clean các bảng có chung column playerid thành một bảng final

Purchased_game

	playerid	library
0	76561198060698936	[60, 1670, 3830, 1600, 2900, 2910, 2920, 4800,...
1	76561198287452552	[10, 80, 100, 240, 2990, 6880, 6910, 6920, 698...
2	76561198040436563	[10, 80, 100, 300, 20, 30, 40, 50, 60, 70, 130...
3	76561198042412488	[300, 240, 220, 320, 360, 4300, 4800, 4000, 61...
4	76561198119605821	[47870, 108600, 550, 271590, 331470, 381210, 2...

Friends

	playerid	friends
0	76561198060422271	['76561198018120276', '76561198034545417', '76...
1	76561198113439786	['76561198047435192', '76561198059136488', '76...
2	76561198149851326	['76561197991555589', '76561198003513187', '76...
3	76561198296997371	NaN
4	76561198895573082	['76561197960300358', '76561197961330830', '76...

Players

	playerid	country	created
0	76561198287452552	Brazil	2016-03-02 06:14:20
1	76561198040436563	Israel	2011-04-10 17:10:06
2	76561198049686270	NaN	2011-09-28 21:43:59
3	76561198155814250	Kazakhstan	2014-09-24 19:52:47
4	76561198119605821	NaN	2013-12-26 00:25:50

Reviews_count

	playerid	review_count
0	76561197960265861	1
1	76561197960266039	7
2	76561197960266642	5
3	76561197960266945	5
4	76561197960268165	1

- Sau đó, ta sẽ có final table để thực hiện các phân tích

player_public

	playerid	country	created	library	friends	review_count	total_achievements	year_created	play_cs2	total_games	total_friends
0	76561198287452552	Brazil	2016-03-02 06:14:20	[10, 80, 100, 240, 2990, 6880, 6910, 6920, 698...	NaN	9.0	NaN	2016	1	476	0
1	76561198040436563	Israel	2011-04-10 17:10:06	[10, 80, 100, 300, 20, 30, 40, 50, 60, 70, 130...	['76561197961017729', '76561197963826101', '76...	31.0	NaN	2011	1	836	316
2	76561198049686270	NaN	2011-09-28 21:43:59	NaN	['76561197966947992', '76561197967022261', '76...	3.0	NaN	2011	0	0	718

Column name	Data type	Description
playerid	int64	ID của các player trong danh sách
country	object	Quốc gia của tài khoản steam
created	object	Ngày thành lập tài khoản steam
library	object	Danh sách các game có trong thư viện của tài khoản
friends	object	Danh sách bạn bè của tài khoản
review_count	float64	Tổng số review của người chơi
total_achievements	float64	Tổng số achievement của người chơi
year_created	object	Được trích từ "created", năm thành lập
play_cs2	int64	Giá trị Binary, 0 tức là không chơi cs2, 1 là có chơi cs2
total_friends	int64	Tổng số bạn bè trong tài khoản
total_games	int64	Tổng số game trong tài khoản

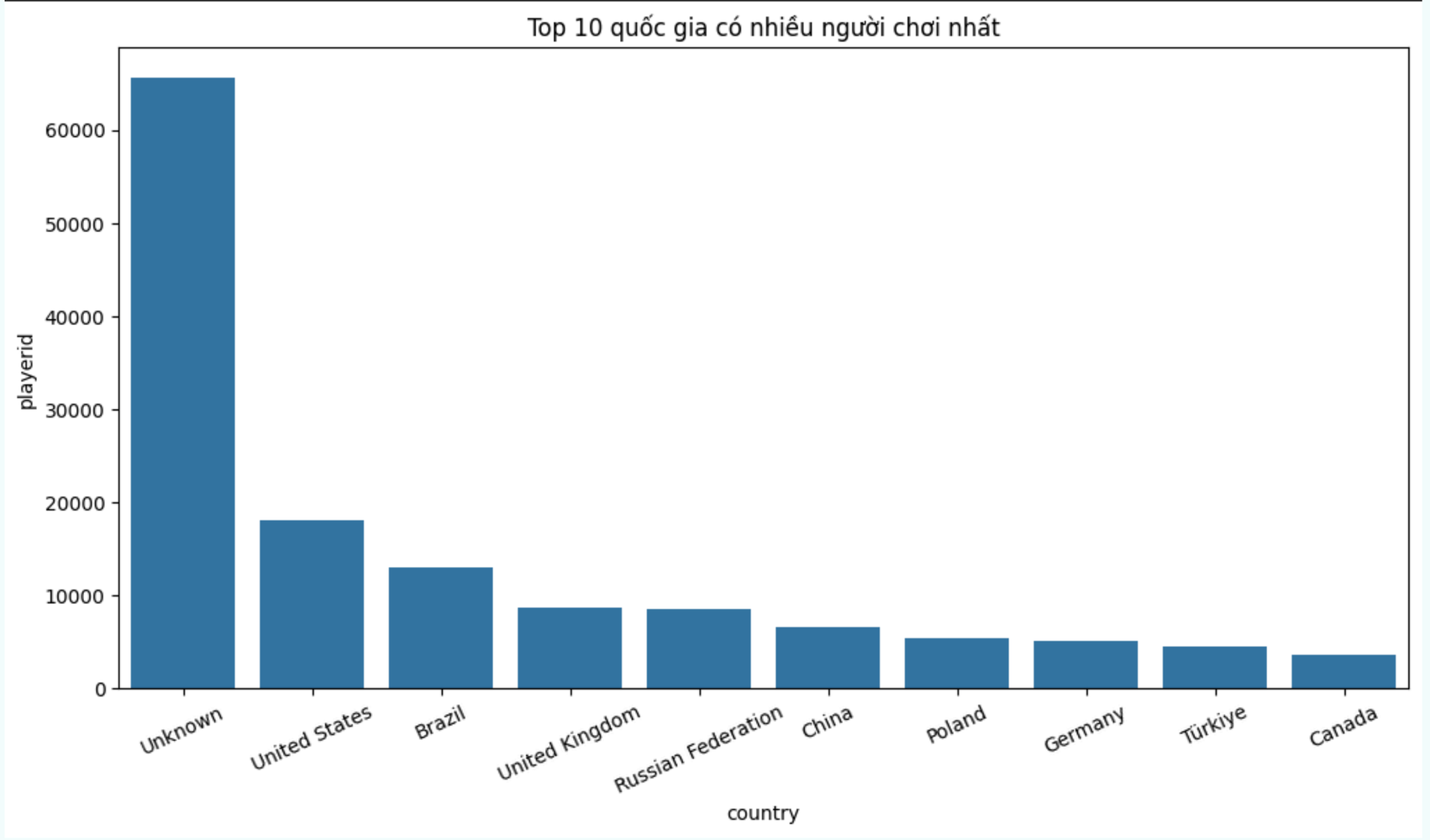
- Bên cạnh đó, ta có bảng thứ hai chứa thông tin các game và số lần xuất hiện trong thư viện người chơi

game_count

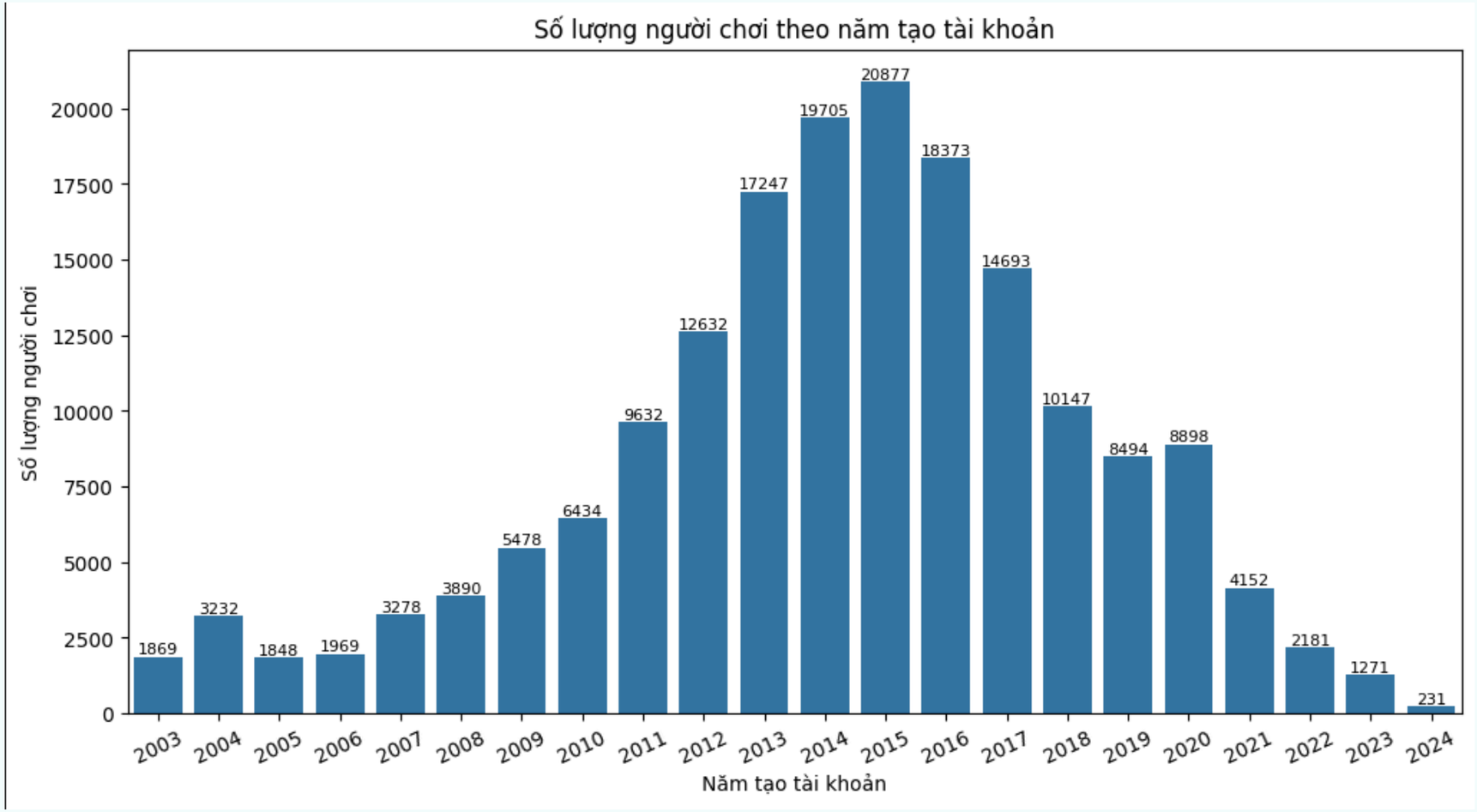
	gameid	count	title	developers	publishers	genres	supported_languages	release_date
0	730	28434	Counter-Strike 2	['Valve']	['Valve']	['Action', 'Free To Play']	['Czech', 'Danish', 'Dutch', 'English', 'Finni...	2012-08-21
1	578080	19690	PUBG: BATTLEGROUNDS	['PUBG Corporation']	['KRAFTON, Inc.']	['Action', 'Adventure', 'Massively Multiplayer...	['English', 'Korean', 'Simplified Chinese', 'F...	2017-12-21
2	550	18766	Left 4 Dead 2	['Valve']	['Valve']	['Action']	['Danish', 'Dutch', 'English', 'Finnish', 'Fre...	2009-11-16

Column name	Data type	Description
gameid	int64	Id của game trong dataset
count	int64	Số lần xuất hiện trong library của người chơi
title	object	Tên game
developers	object	Nhà phát triển
publishers	object	Nhà phát hành
genres	object	Thể loại game
supported_languages	object	Ngôn ngữ hỗ trợ
release_date	object	Ngày ra mắt

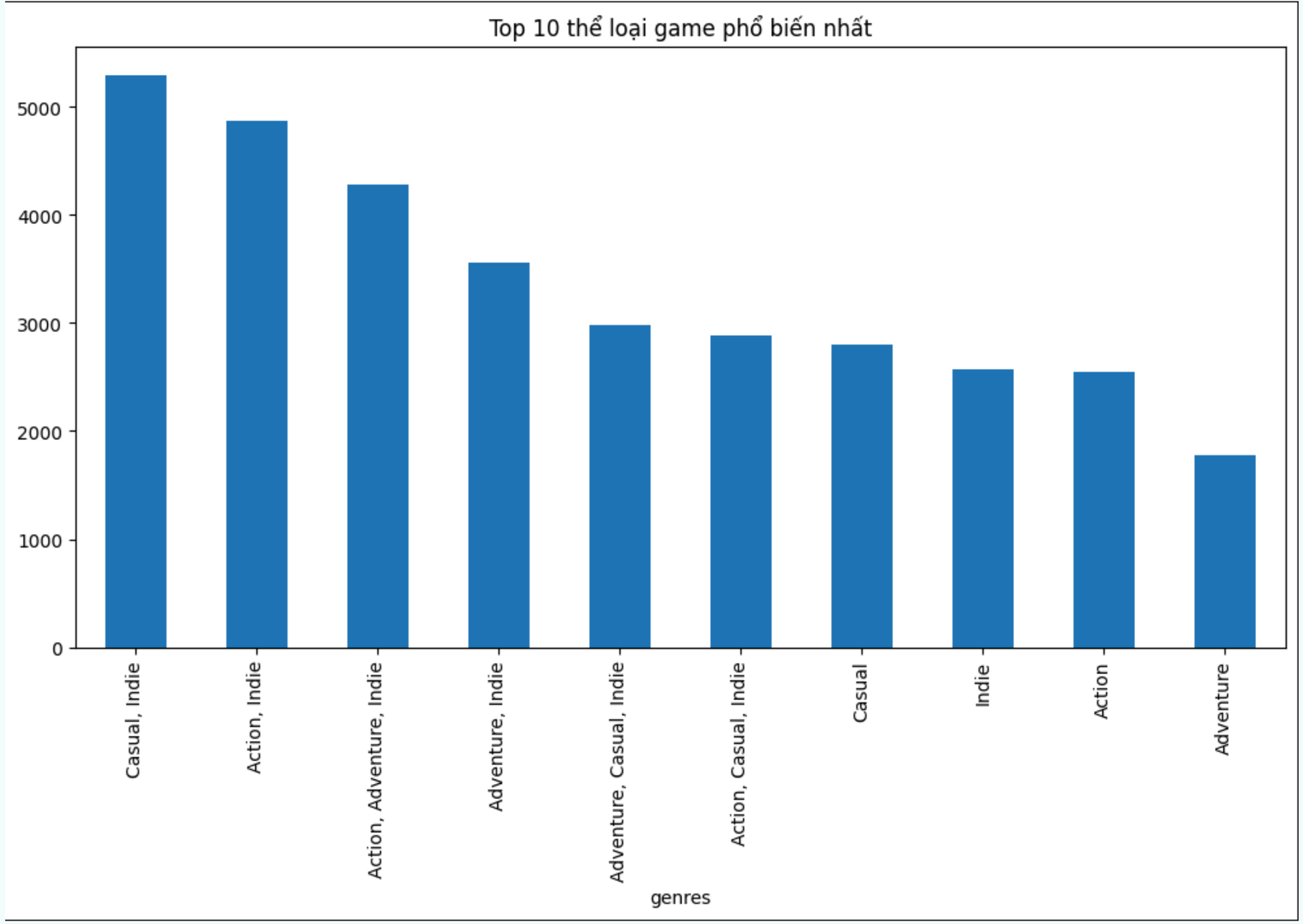
STEAM ANALYSIS



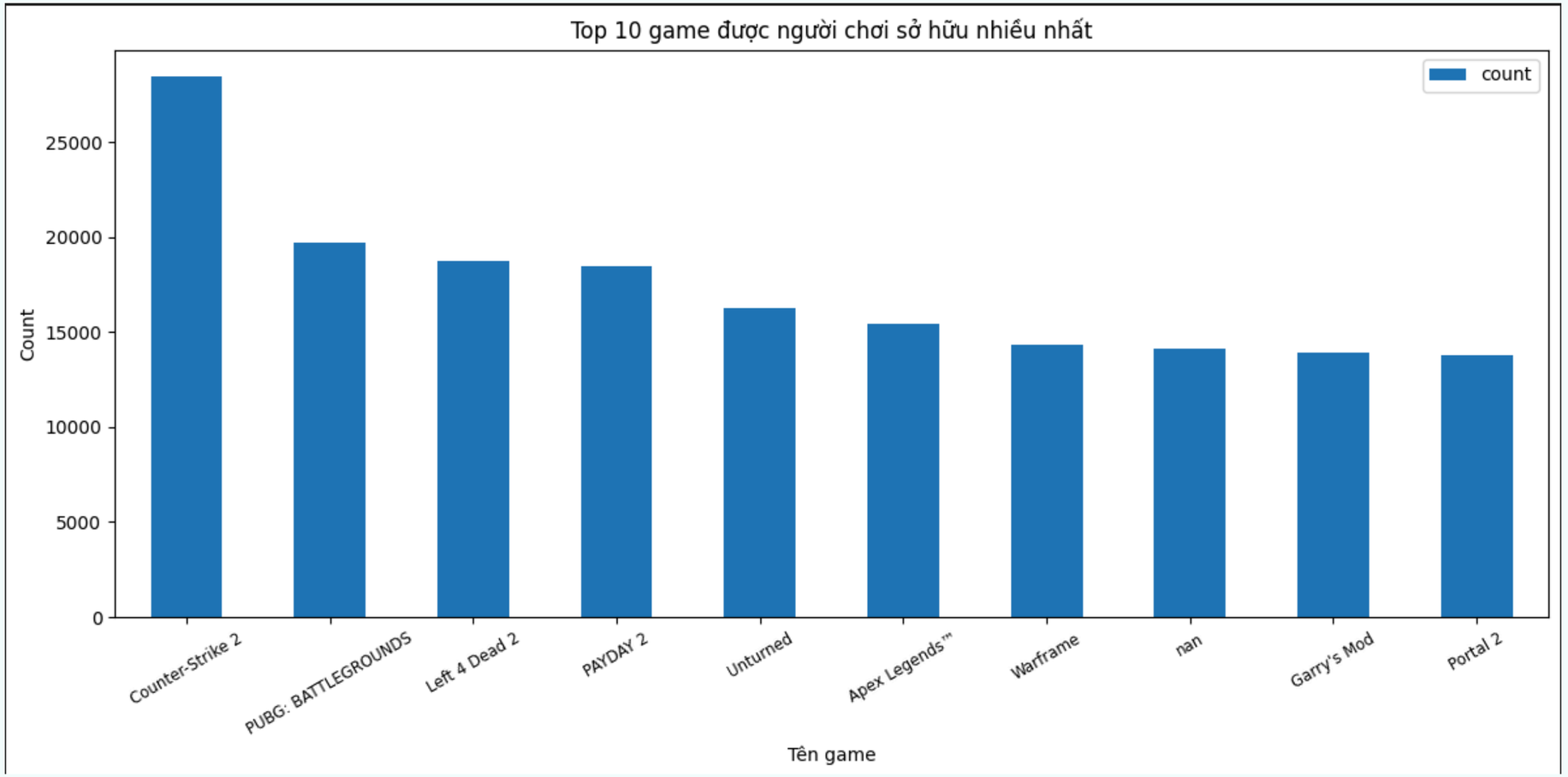
TRONG SỐ TOP MƯỜI QUỐC GIA CÓ NHIỀU NGƯỜI CÓ TÀI KHOẢN STEAM NHẤT, MỸ VÀ BRAZIL DẪN ĐẦU (NẾU KHÔNG TÍNH CÁC NƯỚC CHƯA XÁC ĐỊNH HOẶC NULL)



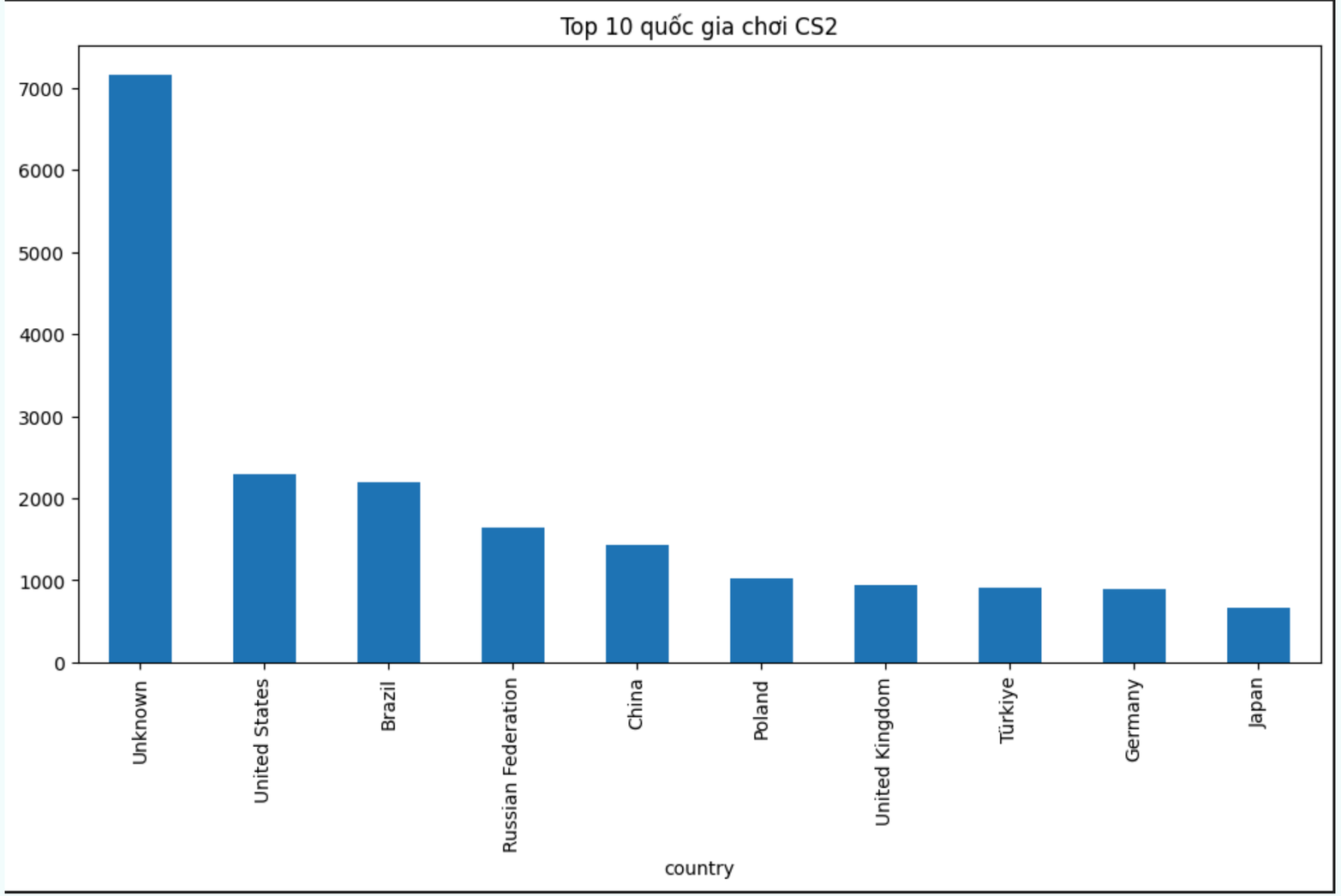
HẦU HẾT CÁC TÀI KHOẢN TRONG DATASET ĐƯỢC TẠO VÀO KHOẢNG 2015



NHỮNG TỰA GAME INDIE VẪN CHIẾM TỶ TRỌNG CAO NHẤT VÀ ÁP ĐẢO HOÀN TOÀN



TRONG DATASET NÀY, NHỮNG TỰA GAME PHỔ BIẾN NHẤT HẦU HẾT LÀ NHỮNG GAME FPS MIỄN PHÍ HOẶC GIÁ RẺ, NỔI BẬT NHẤT LÀ CS2



KHI SO SÁNH TOP 10 QUỐC GIA CHƠI CS2 VỚI TOP 10 QUỐC GIA CÓ NHIỀU NGƯỜI CHƠI NHẤT, THỨ TỰ KHÔNG THAY ĐỔI NHIỀU, CHỈ CÓ NGA VƯƠN LÊN HẠNG 4 VÀ UK TỤT XUỐNG HẠNG 7

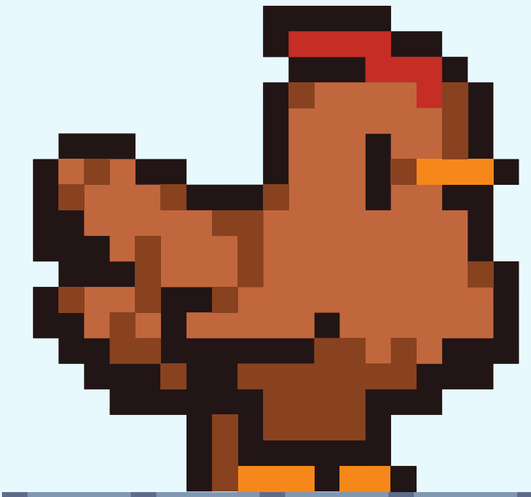
Một số quốc gia có tỉ lệ người chơi CS2 cao nhất là Lesotho, Senegal,.. Ngược lại, ở Nam Sudan, Timor-Leste, không có ai chơi CS2

country	
Saint Barthélemy	66.67
Lesotho	50.00
Senegal	36.36
Mali	33.33
Mauritania	33.33
	...
South Sudan	NaN
Svalbard and Jan Mayen	NaN
Tanzania, United Republic of	NaN
Timor-Leste	NaN
Vanuatu	NaN

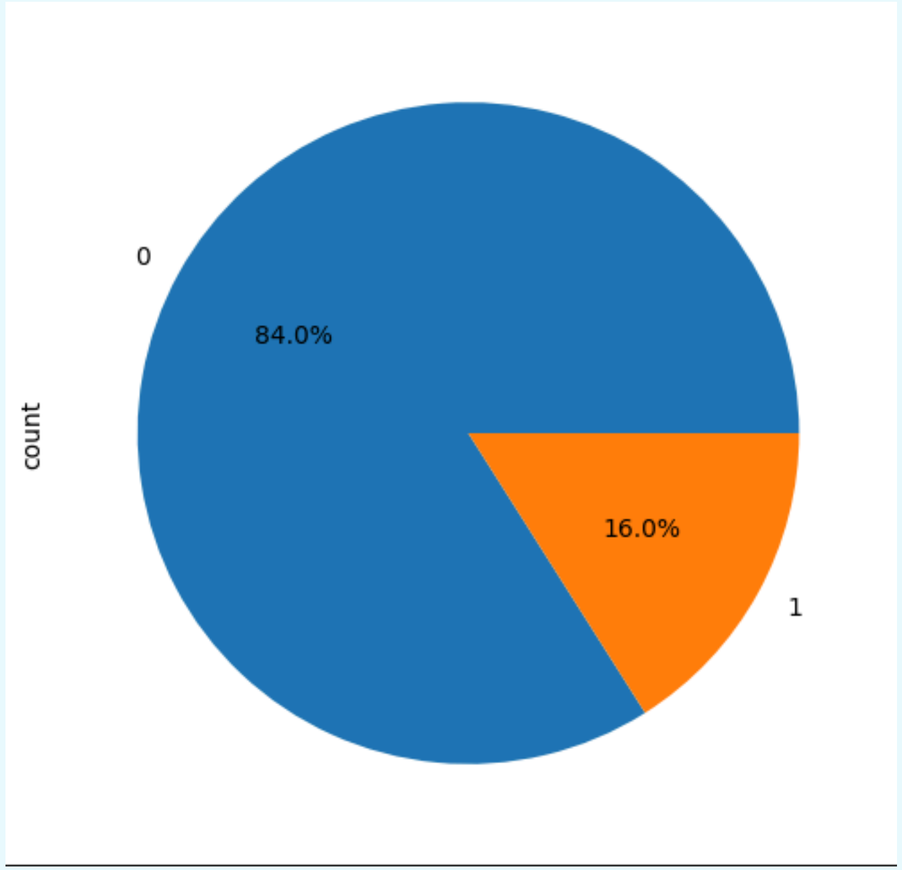
Những tài khoản tạo từ 2017-2021 có tỉ lệ chơi CS2 nhiều nhất, những năm càng xa hiện tại thì tỉ lệ này càng ít (Từ 2015 về trước)

year_created	
2019.0	19.71
2018.0	19.44
2020.0	19.26
2017.0	18.24
2021.0	18.04

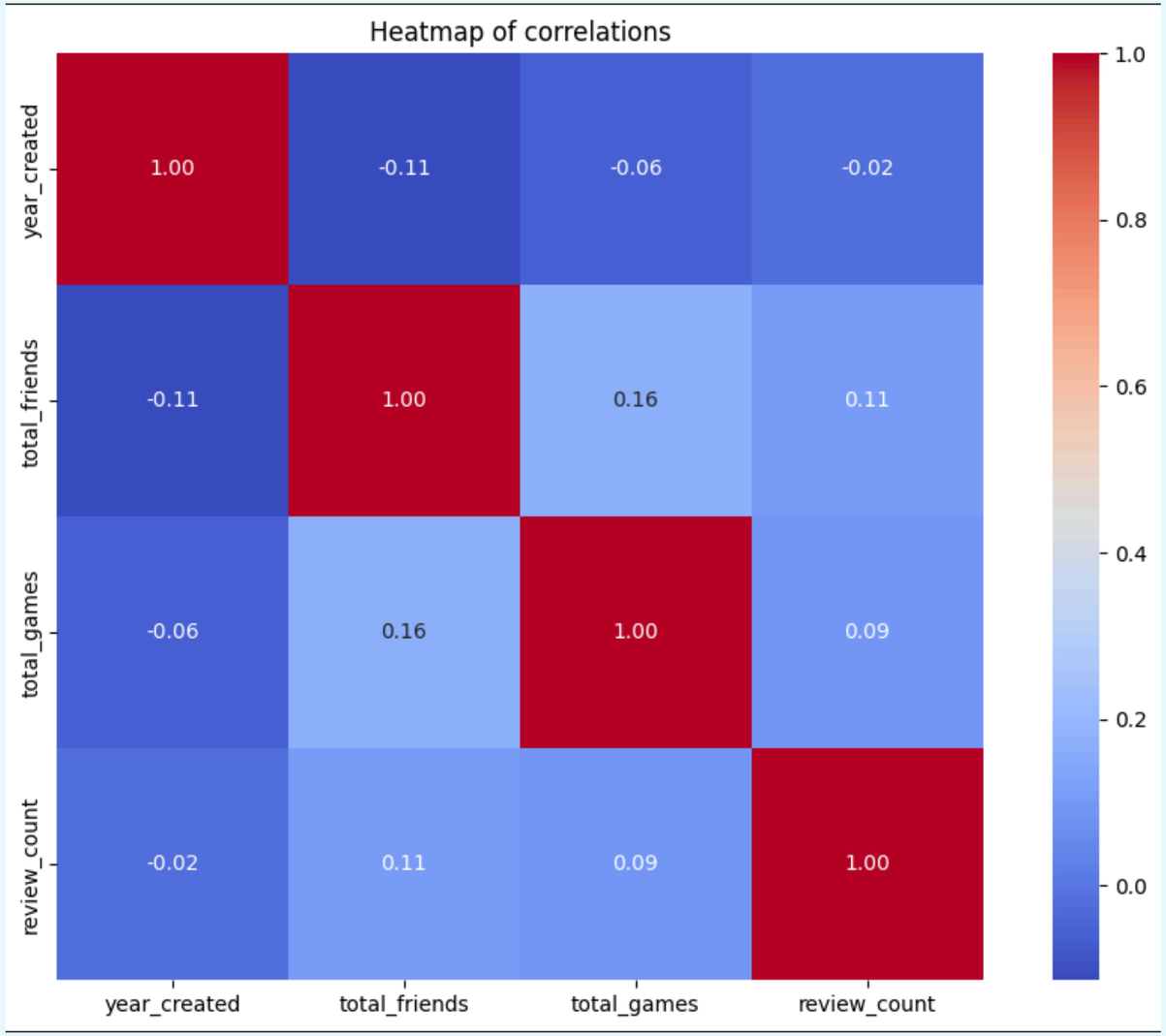
CS2 MODEL PREDICTION



- Để xây dựng mô hình dự đoán tài khoản có chơi cs2 hay không, cần biết training data có bị skewed hay không. Trong 176513 data, có khoảng 84% tài khoản không có cs2, vì vậy có thể cân nhắc sử dụng sampling

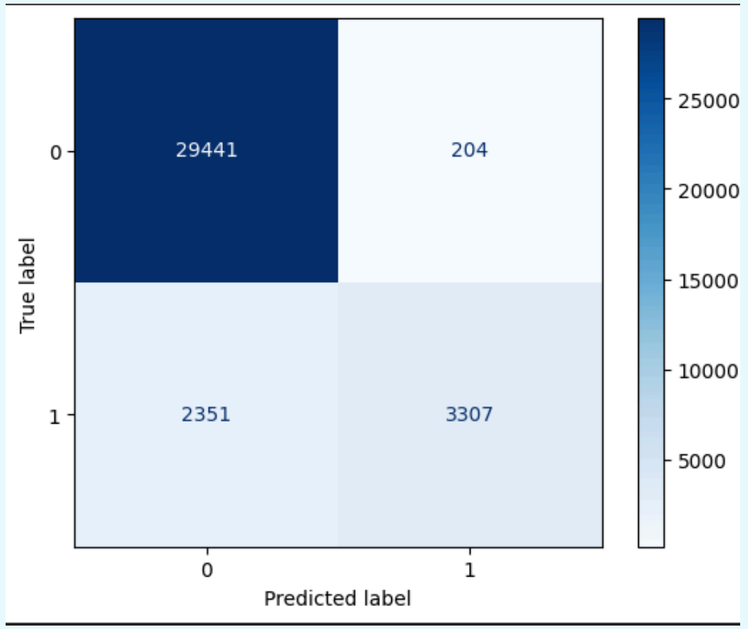


- Trong feature selection, có thể loại bỏ cột total_achievements do chứa quá nhiều giá trị null, các cột khác đều có thể sử dụng được vì không có multicollinearity

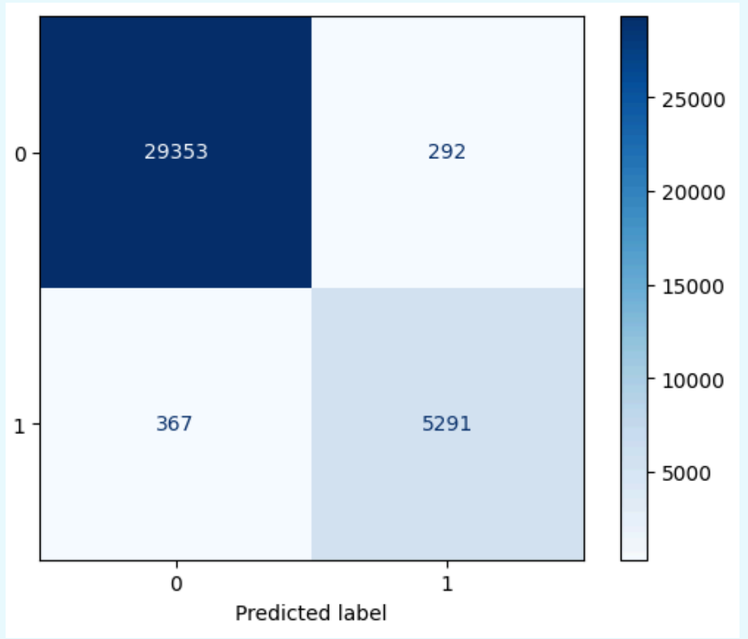


- TP: Dự đoán chơi, có chơi
- TN: Dự đoán không chơi, ko chơi
- FP: Dự đoán có chơi, nhưng không chơi
- FN: Dự đoán không chơi, nhưng có chơi

Logistic regression



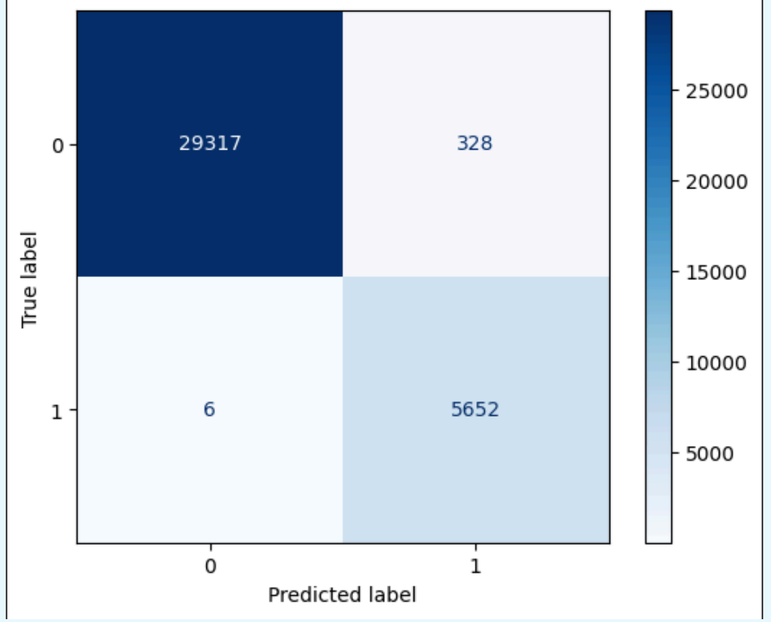
Decision tree



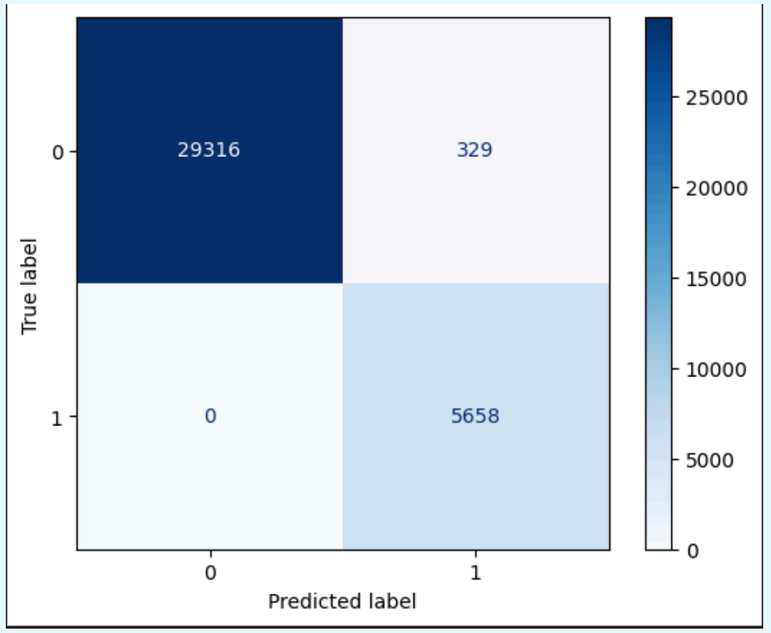
Tỉ lệ: 80%train, 20% test

- TP: Dự đoán chơi, có chơi
- TN: Dự đoán không chơi, ko chơi
- FP: Dự đoán có chơi, nhưng không chơi
- FN: Dự đoán không chơi, nhưng có chơi

XGBoost



Random Forest



Mức độ quan trọng của các feature:
Tổng số game là thước đo mạnh nhất, bỏ xa các chỉ số còn lại

	importance
total_games	0.959669
total_friends	0.019338
year_created	0.010552
review_count	0.010441

Bảng điểm của các model: Có thể thấy XGBoost là mô hình phù hợp nhất

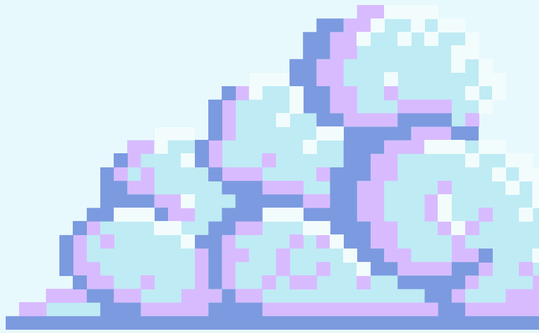
	Model	Accuracy	F1 Score	Precision	Recall
0	Logistic Regression	0.927627	0.721344	0.941897	0.584482
1	Decision Tree	0.981333	0.941375	0.947698	0.935136
2	XGBoost	0.990539	0.971301	0.945151	0.998940
3	Random Forest	0.971748	0.971748	0.945048	1.000000

HYPOTHESIS TESTING

Những người chơi CS2 có số lượng game trong thư viện lớn hơn những người không chơi CS2 hay không? Áp dụng one tail t-test

H_0: Số lượng game trong thư viện của người chơi CS2 và không chơi CS2 là bằng nhau.

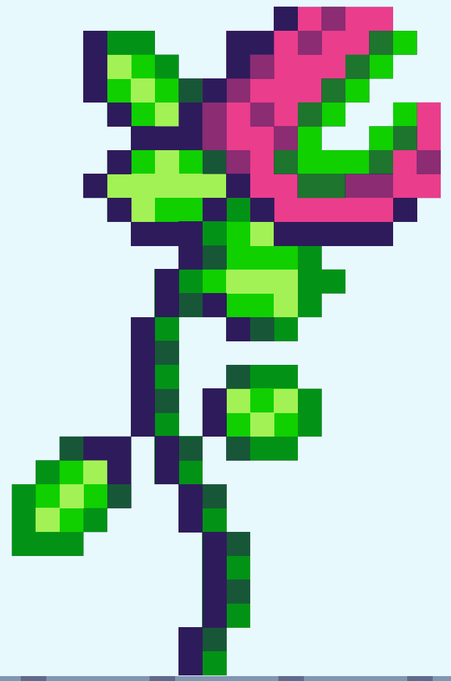
H_1: Số lượng game trong thư viện của người chơi CS2 lớn hơn những người không chơi CS2



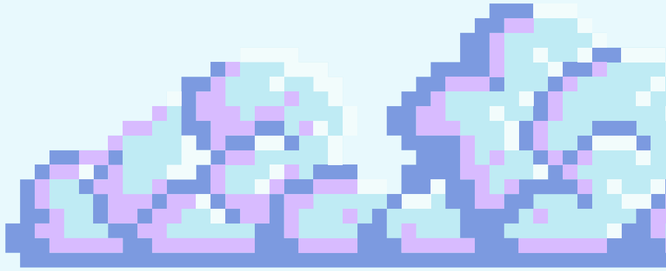
```python
t1= player_public[player_public["play_cs2"] == 1]["total_games"]
t2= player_public[player_public["play_cs2"] == 0]["total_games"]
t_stat, p_value_two_tailed = ttest_ind(t1, t2, equal_var=False)
if t_stat > 0:
    p_value_one_tailed = p_value_two_tailed / 2
else:
    p_value_one_tailed = 1 - p_value_two_tailed / 2

alpha = 0.05
if p_value_one_tailed < alpha:
    print("Bác bỏ H₀: Người chơi CS2 có số lượng game trong thư viện lớn hơn người không chơi CS2.")
    print (t_stat , p_value_one_tailed)
else:
    print("Không đủ bằng chứng để bác bỏ H₀.")
```

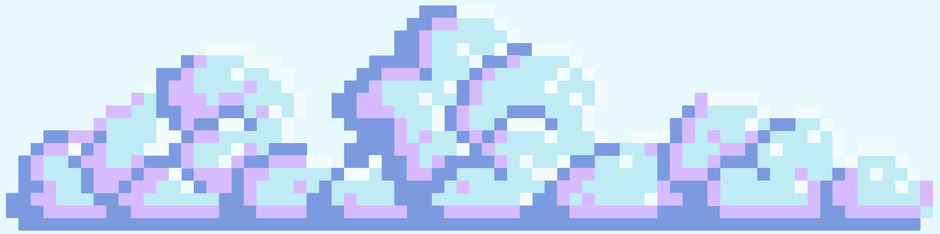
```
Bác bỏ H₀: Người chơi CS2 có số lượng game trong thư viện lớn hơn người không chơi CS2.
50.03999378358133 0.0
```

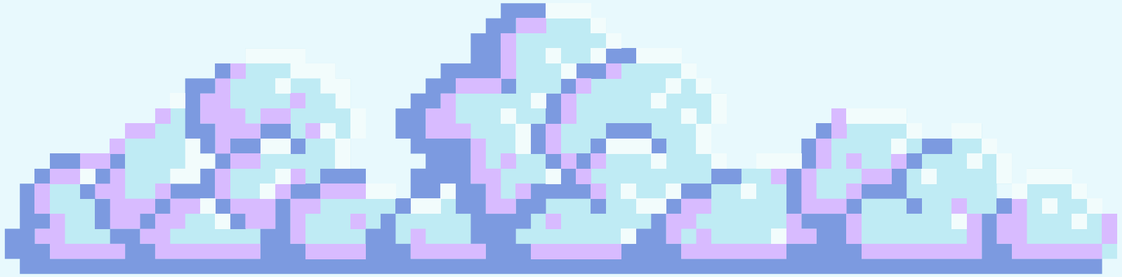
CONCLUSION



- Trong tất cả các model, model phù hợp nhất để sử dụng là XGBoost, tuy nhiên có thể cân nhắc sử dụng thêm các mô hình khác (SVM, Gradient Boosting) hoặc kết hợp model (Ensemble) để có kết quả tốt hơn
- Để biết một tài khoản có tựa game csgo trong thư viện hay không, dấu hiệu mạnh nhất là số lượng game trong tài khoản. Càng nhiều game trong thư viện người chơi thì khả năng rất cao sẽ có CS2 trong đó.
- Bên cạnh đó dựa vào hypothesis testing, những người chơi cs2 cũng sẽ có nhiều game trong thư viện hơn những người không chơi
- Valve cần có những động thái giữ chân người chơi ở các quốc gia có tỉ lệ chơi CS2 cao như mở thêm nhiều giải đấu tại đó hơn, ngược lại cần lôi kéo nhiều người chơi hơn ở các quốc gia ít người chơi bằng cách cân nhắc giảm giá các vật phẩm hoặc phối hợp cùng KOL
- Trong tất cả các tựa game, CS2 xuất hiện nhiều nhất, vì vậy việc giữ chân người chơi là vô cùng quan trọng, cần liên tục fix bug và ban cheater để nâng cao trải nghiệm chơi game







THANK YOU