ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO ĐỒ ÁN THỰC HÀNH

Học phần: NHẬP MÔN KHOA HỌC DỮ LIỆU

Mã lớp : 20_21

Nhóm sinh viên thực hiện: Lê Nguyễn Thanh Hoàng - 20120088
Trần Xuân Hòa - 20120085
Nguyễn Thế Đạt - 20120055
Lê Hoàng Huy - 20120105

TP.HCM, tháng 10 năm 2022

Mục lục

Danh mục hình

1. Tổng hợp quá trình thực hiện đồ án

1.1. Những khó khăn khi thực hiện đồ án

- Một vài thành viên lần đầu sử dụng Git và Github dẫn đến khó khăn trong việc lưu trữ, sửa đổi code, branch.
- Các file notebook có cấu trúc json khác với cấu trúc file thực thi (.py, .cpp) dẫn đến nhiều khó khăn trong việc merge và resolve conflict giữa các branch.

1.2. Những kiến thức học được

- Cách sử dụng các thư viện Pandas và Numpy để phân tích dữ liệu
- Cách sử dụng các thư viện Matplotlib và Seaborn để trực quan hóa dữ liệu bằng các loại biểu đồ phù hợp.
- Cách làm việc với Git và Github.

1.3. Phân công công việc

Link Notion bảng phân công công việc: tại đây

Quy trình	Công việc	Phân công
Chuẩn bị dữ liệu	Tất cả thành viên tải file kill_match_stats_final0.csv về máy local của mình	Tất cả thành viên
	Tải tất cả asset cần thiết	Hoàng
Tiền xử lí dữ liệu	- Lọc bỏ một số dòng/cột không cần thiết - Đổi tên các cột tọa độ thành tên ngắn hơn để thuận tiện cho phân tích, truy xuất - Lưu kết quả vào file kill_match_stats_v1.csv	Đạt
	- Xử lí missing values - Xử lí duplicated values - Xử lí outlier	Hòa

	- Nhất quán các giá trị của cột killed_by - Chuẩn hóa các giá trị tọa độ về khoảng từ 0 tới 8000 - Phân nhóm cột killed_by thành từng nhóm nguyên nhân ở cột type - Phân nhóm cột time thành các phase của trận đấu ở cột phase - Tính khoảng cách từ killer tới victim dựa trên các cột tọa độ và lưu vào cột dis - Xử lí các giá trị không hợp lệ ở cột dis và phase - Kết quả lưu vào fle kill_match_stats_v2.csv	Huy
Khám phá dữ liệu	- Tổng quan kích thước, kiểu dữ liệu của dữ liệu - Phân bố dữ liệu của các cột tọa độ và cột dis - Phân bố dữ liệu của cột time và phase	Hòa
	- Phân bố dữ liệu của cột killed_by và type - Phân bố cột killed_by theo type - Phân bố cột type theo time	Đạt
	- Phân bố dis theo time - Phân bố dữ liệu tọa độ 2 chiều	Hoàng

	bằng scatter plot - Phân bố dữ liệu tọa độ 2 chiều bằng heatmap - Phân bố dữ liệu tọa độ theo thời gian	
Phân tích dữ liệu	Loại vũ khí nào hiệu quả nhất?	Huy
	Những vị trí nào giao tranh nhiều?	Hoàng
	Những vị trí nào thuận lợi để gây kill từ xa?	Hoàng
	Những vị trí thuận lợi gây kill từ xa có thực sự an toàn?	Hòa
	Thời điểm thích hợp để qua cầu	Đạt
Báo cáo	Các thành viên viết báo cáo cho phần mình được phân công	Tất cả thành viên, Đạt kiểm tra lại (về định dạng, để mục,…)
Slide		Hoàng

Figure 1

2. Nội dung báo cáo

2.1. Thu thập dữ liệu

Trong đồ án này, nhóm thu thập và phân tích dataset về game PUBG được lấy từ Kaggle.

2.1.1. Về PUBG

PUBG là game thể loại đấu súng battle royale góc nhìn thứ nhất/thứ ba. Mỗi trận đấu trong game, 90 người chơi sẽ cùng xuất hiện trên một hòn đảo lớn và chiến đấu với nhau cho đến khi chỉ còn một người chơi duy nhất. Người chơi sẽ được thả lên đảo từ một chiếc máy bay, đáp xuống những thị trấn và tòa nhà bỏ hoang để thu thập vũ khí, trang bị. Người chơi sau đó sẽ quyết định chiến đấu với các người chơi khác hoặc lẩn trốn để đạt được cái đích cuối cùng: người duy nhất sống sót. Một Bluezone sẽ xuất hiện sau vài phút trận đấu bắt đầu, gây sát thương cho bất kì ai ở trong vùng này. Bluezone sẽ liên tục

thu nhỏ vùng an toàn (không bị sát thương) để ép các người chơi nằm trong vùng an toàn sát lại gần nhau hơn.

2.1.2. Về dataset

Dataset dược chia làm 2 tập: aggregate và deaths

- Trong tập deaths: các file ghi nhận các sự kiện có người chơi tử trận trong vòng 720.000 trận đấu. Mỗi dòng dữ liệu mà một sự kiện người chơi tử trận
- Trong tập aggregate: chứa các thông tin tổng hợp về người chơi và các trận đấu. Tập này cung cấp nhiều thông tin tổng hợp về số lượng kill, sát thương hay quãng đường đã di chuyển,… cũng như các thông tin tổng hợp về trận đấu như ngày diễn ra, chế độ, kích thước hàng chờ của trận đấu,…

2.1.3. Phương pháp thu thập dữ liệu

Dữ liệu về các trận đấu được thu thập từ website pubg.op.gg.

- Người thu thập dữ liệu này (tạm gọi là tác giả) bắt đầu thu thập các chỉ số với một "người chơi hạt giống", tác giả tự lấy tài khoản game của mình để làm người chơi hạt giống này.
- Sau đó tác giả thu thập tất cả người chơi mà người chơi hạt giống đã gặp trong tất cả các trận đấu.
- Tác giả tiếp tục lấy mẫu 5000 người chơi trong tập người chơi đã có phía trên, sau đó lấy thông tin của tất cả trận đấu mà tập 5000 người chơi này đã tham gia, từ đó thu thập được tập dữ liệu cuối cùng.

2.1.4. License

Dữ liệu này được thu thập tuân thủ giấy phép về bản quyền CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, cho phép người khác có quyền chia sẻ, sử dụng và xây dựng, thậm chí cho cả mục đích thương mại mà không cần tác giả cho phép, dựa trên tác phẩm mà tác giả tạo ra, trong trường hợp này là dữ liệu chơi game của người chơi.

2.1.5. Một số thuật ngữ về game sử dụng trong đồ án

- kill: một sự kiện người chơi tử trận, tương ứng với một dòng trong tập dữ liệu
- player: người chơi
- killer: người chơi gây ra kill

- victim: người chơi tử trận ở kill
- bluezone: khu vực gây sát thương của game, đã đề cập ở trên

2.2. Tiền xử lí dữ liệu

Trong đồ án này, vì kích thước tập dữ liệu gốc quá lớn, nhóm chỉ sử dụng một file csv trong tập deaths, gồm: 13 triệu dòng, mỗi dòng là các thông tin về một kill

2.2.1. Mô tả các cột trong dữ liệu

Mỗi kill xảy ra sẽ được ghi nhận thông tin về tọa độ của killer và victim trên hệ tọa độ 2 chiều với gốc tọa độ ở góc bản đồ, thời gian diễn ra kill và các thông tin liên quan

- killed_by: Nguyên nhân gây ra kill.
- killer_name: tên player gây ra kill (từ nay về sau gọi ngắn gọi là killer), với một số nguyên nhân cụ thể sẽ trùng với victim_name.
- killer_placement: thứ hạng của killer. Thứ hạng, hay "top" là số player còn lại của trận đấu ngay trước khi killer tử trận. Ví dụ placement = 10 nghĩa player tử trận khi trận chỉ còn 10 player.
- killer_position_x: tọa độ x của killer tại thời điểm gây ra kill.
- killer_position_y: tọa độ y của killer tại thời điểm gây ra kill.
- map: tên map của trận đấu.
- match_id: id của trận đấu.
- time: thời điểm kill xảy ra tính từ lúc bắt đầu trận đấu.
- victim_name: tên player tử trận (từ nay gọi về sau tạm gọi là victim), với một số nguyên nhân cụ thể sẽ trùng với killer_name.
- victim_placement: thứ hạng của victim.
- victim_position_x: tọa độ x của victim tại thời điểm kill xảy ra.
- victim_position_y: tọa độ y của victim tại thời điểm kill xảy ra.

Trong đồ án này, nhóm chủ yếu tập trung vào thời điểm và vị trí các kill diễn ra, không quá chú trọng vào tên player và match id hay các yếu tố khác. Do đó nhóm sẽ loại bỏ các cột killer_name, victim_name, killer_placement, victim_placement, match_id. Nhóm cũng sẽ tập trung vào phân tích diễn biến trận đấu trên map Erangel, do đó nhóm cũng sẽ

lọc bỏ các dòng ghi nhận trên map Mirama. Sau khi tiến hành lọc bỏ các dòng có map Mirama, nhóm cũng xóa bỏ cột map.

Để thuận tiện cho việc truy xuất cũng như phân tích về sau, nhóm sẽ đổi tên các cột như sau:

- killer_position_x thành kx
- killer_position_y thành ky
- victim_position_x thành vx
- victim_position_y thành vy

Sau khi thực hiện lọc bỏ một số cột và đổi tên, dữ liệu còn lại được mô tả gồm các cột

- killed_by: Nguyên nhân gây ra kill.
- kx: tọa độ x của killer tại thời điểm gây ra kill
- ky: tọa độ y của killer tại thời điểm gây ra kill
- time: thời điểm kill diễn ra
- vx: tọa độ x của victim tại thời điểm gây ra kill
- vy: tọa độ y của victim tại thời điểm gây ra kill

2.2.2. Làm sạch dữ liệu – Data Cleaning

2.2.2.1. Dữ liệu khuyết – Missing values

Nhóm tiến hành kiểm tra dữ liệu khuyết trên tất cả các cột, phát hiện có 2 cột có cùng số dữ liệu khuyết là kx và ky với 741,597 dòng mỗi cột. Khi kiểm tra số lượng dòng của dữ liệu, nhóm biết được dữ liệu có chính xác 10,865,476 dòng, nghĩa là các dòng dữ liệu khuyết chỉ chiếm khoảng 7% dữ liệu, ta có thể đơn giản xóa đi các dòng này. Nhưng khi quan sát kĩ hơn các dòng dữ liệu khuyết này, nhóm nhận thấy ở cột killed_by có chứa các giá trị sau:

Nguyên nhân	**Số lượng**
Bluezone	468994
Down and Out	141913
Falling	61571
Drown	41150
RedZone	13683
Uaz	4451

Dacia	3002
Buggy	2091
Hit by Car	1719
Motorbike	994
Motorbike (SideCar)	895
Punch	514
Grenade	330
death.Buff_FireDOT_C	145
death.RedZoneBomb_C	90
Boat	28
death.ProjMolotov_DamageField_C	24
S686	1
SCAR-L	1
Aquarail	1

Nhóm nhận thấy chủ yếu các nguyên nhân này thuộc vào loại Zone (các vùng gây sát thương), các loại xe, các loại vũ khí ném và các nguyên nhân tự thân (Falling: Ngã, Drown: Đuối nước). Đối vơi các nguyên nhân này, không có killer, do đó không có tọa độ của killer. Ta có thể tạm gán tọa độ kx và ky của các dòng này bằng với tọa độ của victim vx và vy.

2.2.2.2. Giá trị lặp – Duplicated values

Quan sát dữ liệu, nhóm nhận thấy có khoảng hơn 250,000 dòng dữ liệu bị trùng, tuy nhiên ta có thể quan sát kĩ hơn các dòng này

kx	**ky**	**vx**	**vy**	**Số lượng**
0.0	0.0	0.0	0.0	257127
399045.9	300804.1	0.0	0.0	2
482652.2	446492.6	0.0	0.0	2
446353.9	629764.4	446300.3	629816.6	1
444775.8	623105.9	444581.1	623158.9	1
				…
348630.2	312316.8	350118.1	318378.7	1

349536.5	563234.7	349230.2	564633.3	1
349832.6	565963.8	350066.0	566620.0	1
349866.4	563824.9	354611.3	570966.2	1
738463.6	434072.6	738463.6	434072.6	1

Quan sát các giá trị khác nhau của các cột tọa độ, ta nhận thấy đa số các cột trùng nhau có tọa độ killer và victim nằm ở (0, 0), nghĩa là ở góc bản đồ, nơi không có địa hình (sẽ được minh họa kỹ hơn ở phần trực quan hóa dữ liệu). Đây có thể là lỗi trong quá trình game vận hành, số lượng các dòng bị lỗi cũng chỉ khoảng 2,5% dữ liệu, ta có thể đơn giản xóa đi các dòng này.

2.2.2.3. Dữ liệu ngoại lai – Outliers

Hiện tại dữ liệu có 1 cột phân loại (categorical) là killed_by và 5 cột còn lại là dữ liệu số (numerical). Đối với cột categorical, nhóm sử dụng hàm đếm đơn giản để xác định các outliers, còn đối với cột numerical, nhóm sử dụng boxplot, kết hợp với thông tin về bản đồ của game để xác định và xử lí các outliers.

2.2.2.3.1. Cột Categorical

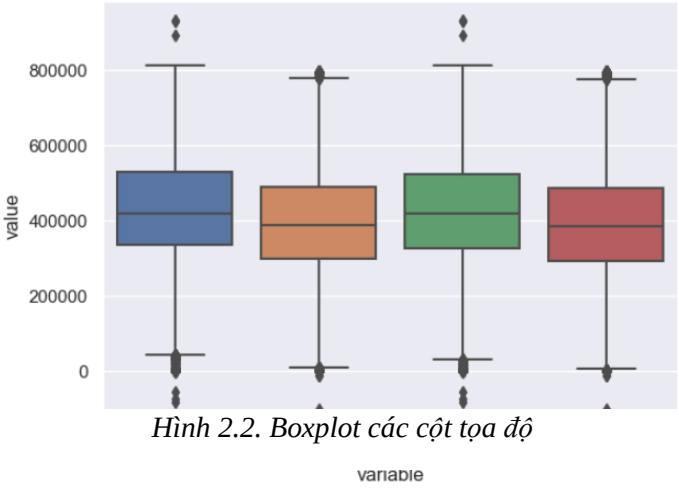
Quan sát cột killed_by, nhóm nhận thấy một vài outlier là death.PlayerMale_A_C, death.RedZoneBomb_C, Aquarial, death.ProjMolotov, Boat, death.Buf_FireDOT_C. Tuy là outliers nhưng các nguyên nhân gây kill này vẫn tồn tại trong game, do ít được sử dụng nên được ghi nhận ít. Nhóm có tra cứu thì nhận thấy ngoài death.PlayerMale_A_C và death.Buff_FireDOT_C, các nguyên nhân còn lại đều hợp lệ, chỉ bị sai tên, sẽ được sửa ở phần Data Quality bên dưới. Do dó ở cột killed_by, nhóm chỉ thực hiện xóa các dòng có death.PlayerMale_A_C và death.Buff_FireDOT_C của cột killed_by

```
death.PlayerMale_A_C                1
Aquarail                           21
death.RedZoneBomb_C                90
death.ProjMolotov_C               251
Boat                             1742
death.Buff_FireDOT_C             2711
Sickle                           4338
Crowbar                          4704
```

Hình 2.1. Một phần bảng đếm giá trị cột killed_by

2.2.2.3.2. Cột Numerical

Đối với 5 cột numerical thì chỉ có cột time là khác ý nghĩa với 4 cột tọa độ còn lại, ta gom nhóm 4 cột tọa độ vào cùng một boxplot.

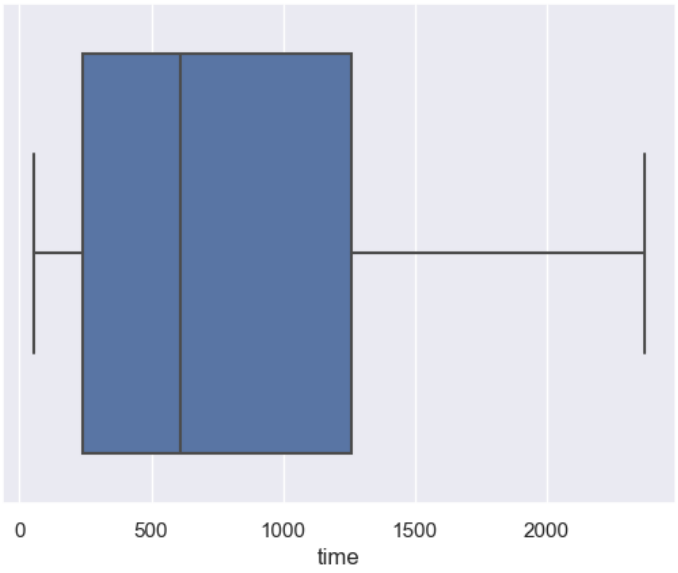


Hình 2.2. Boxplot các cột tọa độ

Quan sát các giá trị tọa độ, ta nhận ra đa số các giá trị đều nằm trong khoảng từ 0 tới 800,000. Đây cũng là khoảng dữ liệu hợp lệ đã được đề cập trong mô tả dữ liệu của tác giả.

Các giá trị ngoại lai lớn hơn 0 và nhỏ hơn 800,000 cũng xuất hiện không nhiều, do đó ta có thể dễ dàng xóa các giá trị này. Do đó ở đây nhóm chỉ loại bỏ các giá trị tọa độ lớn hơn 800,000 hoặc nhỏ hơn 0.

Đối với cột time, nhóm cũng tiến hành vẽ boxplot, tuy nhiên tại thời điểm này chưa có dữ liệu ngoại lai nên nhóm không xử lí dữ liệu trên cột time



Hình 2.3. Boxplot cột time

2.2.3. Đánh giá chất lượng dữ liệu

2.2.3.1. Cột killed_by

Như đã đề cập ở phần 1.2.2.3.1, cột killed_by có một vài giá trị có vẻ khác với các giá trị còn lại. Nhóm đã tiến hành tra cứu và dựa vào hiểu biết của mình về game để chỉnh sửa các giá trị này cho phù hợp.

Qua tìm hiểu, nhóm ghi nhận được:

- death.RedZoneBomb_C là cách thể hiện khác của RedZone
- death.ProjMolotov_C và death.ProjMolotov_DamageField_C đều là kill do bom xăng Molotov gây ra. Tuy nhiên điểm khác biệt nằm ở chỗ nếu victim tử trận khi bị đốt cháy trong đám lửa của xăng, killed_by sẽ là death.ProjMolotov_C. Nếu người chơi tử trận do lửa còn đốt cháy trên người sau khi đã thoát khỏi đám cháy của Molotov, killed_by sẽ được ghi nhận là death.ProjMolotov_DamageField_C

Do đó nhóm quyết định đổi các giá trị death.RedZoneBomb_C thành RedZone, đổi cả 2 giá trị death.ProjMolotov_C và death.ProjMolotov_DamageField_C thành Molotov

Hình 2.5. death.ProjMolotov_DamageField_C

Hình 2.4. death.ProjMolotov_C

Nhóm cũng nhận thấy Bluezone và RedZone có sự khác biệt về chữ z và Z, dù cả 2 đối tượng này đều chỉ các zone. Để tránh nhầm lẫn trong lập trình, nhóm cũng sẽ replace Bluezone thành BlueZone.

2.2.3.2. Các cột tọa độ

Như đã đề cập, khoảng hợp lệ của các tọa độ là từ 0 tới 800,000. Hiện tại trong dữ liệu các cột tọa độ là số thực. Tuy nhiên trong game, khoảng cách giữa 2 cạnh bản đồ là 8,000m, do đó nhóm nhận thấy có thể quay tọa độ từ khoảng 0 tới 800,000 thành 0 tới

8,000, đồng thời làm tròn các tọa độ thành số nguyên (vì sự khác biệt ở sau dấu phẩy thập phân không có ý nghĩa khi khoảng dữ liệu trải dài từ 0 tới 800,000)

2.2.3.3. Thêm một số cột khác

Để thuận tiện cho quá trình phân tích dữ liệu, nhóm tiến hành thêm một số cột khác vào dữ liệu từ các cột có sẵn

2.2.3.3.1. Distance

Một tiêu chí đánh giá quan trọng của kill là khoảng cách giữa killer và victim, phần nào cung cấp thêm thông tin về giao tranh đã xảy ra. Khoảng cách ở đây được tính theo công thức Euclid.

$$dis = \sqrt{(kx - vx)^2 + (ky - vy)^2}$$

Ở đây nhóm cũng sẽ làm tròn khoảng cách thành số nguyên để tiết kiệm chi phí lưu trữ.

2.2.3.3.2. Phase

Ngoài thời gian time, ta có thể dùng phase để xác định khoảng thời gian diễn ra giao tranh. Mỗi trận đấu có 9 phase, bắt đầu từ phase 1 tới 9. Mỗi phase sẽ diễn ra trong một khoảng thời gian nhất định, trong khoảng thời gian này, bluezone sẽ đứng yên. Tới cuối mỗi phase, bluezone sẽ thu hẹp lại, ép người chơi di chuyển vào trung tâm, và chuyển sang phase tiếp theo. Thời gian dành cho mỗi phase sẽ giảm dần cùng với kích thước bluezone tại phase đó. Theo tra cứu của nhóm tại http://battlegrounds.party/circle/, thời gian cho từng phase sẽ như sau:

Phase	Bắt đầu	Kết thúc
1	121	720
2	721	1060
3	1061	1300
4	1301	1480
5	1481	1650
6	1651	1760
7	1761	1880
8	1881	1970

9	1971	2150

Từ giây thứ 1 (ngay sau khi lên máy bay) tới giây thứ 120s, bluezone chưa xuất hiện nên không được tính là phase 1. Tuy nhiên để thuận tiện, nhóm sẽ gom khoảng thời gian này vào phase 1

2.2.3.3.3. Type

Quan sát cột killed_by, nhóm nhận thấy có quá nhiều giá trị categorical. Nhóm quyết định sẽ gom nhóm nguyên nhân đã được quy ước sẵn trong game:

- AR: Assualt Riffle - Súng trường hay trung liên
- DMR: Designated Marksman Riffl - Súng trường thiện xạ
- SR: Sniper Riffle - Súng ngắm
- SMG: Sub Machine Gun - Tiểu liên
- LMG: Light Machine Gun - Súng máy hạng nhẹ
- Shotgun: Súng săn hay súng hoa cải
- Handgun: Súng ngắn, nhóm cũng sẽ gom Crossbow (nỏ) vào nhóm này
- Melee: Vũ khí cầm tay, nhóm quyết định gom Punch (nắm đấm) vào nhóm này
- Zone: Các vùng gây sát thương, gồm BlueZone và RedZone
- Throwable: Các loại vũ khí ném gồm Grenade (lựu đạn) và Molotov (bom xăng)
- Vehicle: phương tiện di chuyển
- Self: các nguyên nhân tự thân gồm Falling (té ngã) và Drown (đuối nước)

Tuy nhiên có 2 giá trị cần lưu ý ở đây:

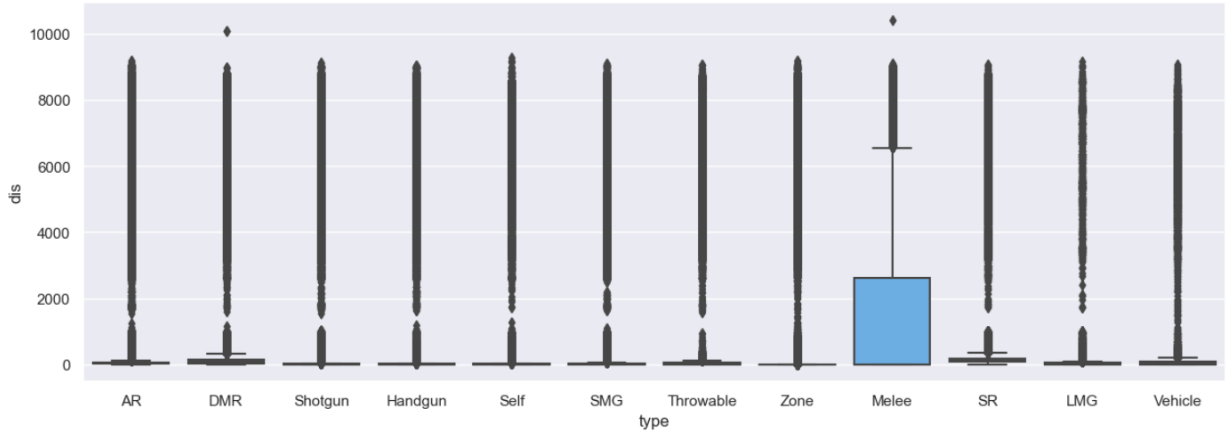
- Down and Out: khi người chơi tham gia trận đấu thể thức nhóm 2, 4 người thì khi mức máu trở về 0, người chơi sẽ không tử trận ngay và rơi vào trạng thái 'knock' (không thể sử dụng vũ khí, hồi máu, tốc độ di chuyển rất chậm, không thể di chuyển trên xe, máu liên tục giảm). Nếu đồng đội của người chơi này không 'cứu' kịp thời, người chơi này sẽ tử trận và được ghi nhận là một sự kiện 'Down and Out'. Do không rõ nguyên nhân gây kill ban đầu là gì, nhóm quyết định loại bỏ các dòng này.

- Hit by Car: bị đụng xe. Bởi vì đã có các giá trị khác nêu rõ loại xe gây tai nạn (Dacia, Uaz, ...) nên nhóm cũng không rõ Hit by Car có ý nghĩa gì khác. Thêm vào đó là số lượng các dòng Hit by Car chỉ chiếm 1% dữ liệu nên nhóm quyết định loại bỏ các dòng này

2.2.4. Tiếp tục làm sạch dữ liệu

Sau khi thêm một số cột mới, nhóm tiến hành làm sạch dữ liệu trên các cột này.

2.2.4.1. Dữ liệu ngoại lai ở cột dis



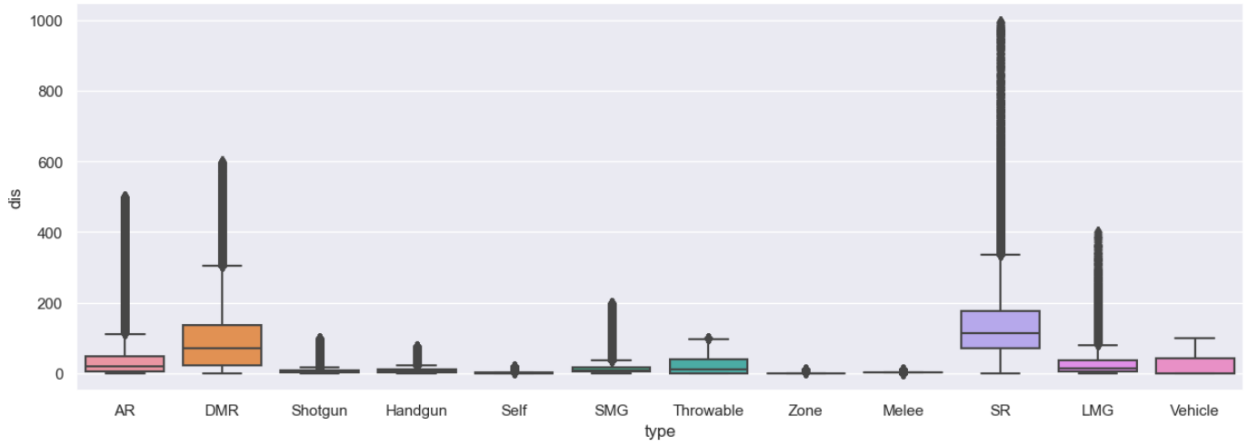
Hình 2.6. Boxplot của dis và type

Nhóm kiểm tra dữ liệu ngoại lai ở cột dis theo cột type bằng boxplot dưới đây

Quan sát boxplot, nhóm nhận ra có khá nhiều dữ liệu khoảng cách bị lỗi (có thể do sai sót của hệ thống hoặc do cheater). Dựa vào kinh nghiệm chơi game cũng như tra cứu về các loại vũ khí, nhóm quyết định lọc khoảng cách theo các loại vũ khí như bên dưới.

type	Giá trị dis tối đa
Self	20
AR	500
DMR	600
SR	1000
Shotgun	100
Handgun	75
SMG	200

LMG	400
Throwable	100
Zone	10
Melee	10
Vehicle	100



Hình 2.7. Boxplot của dis và type sau khi loại bỏ outliers

2.2.4.2. Giá trị khuyết và giá trị ngoại lai cột Phase

Như đã đề cập, phase 9 của trận đấu kết thúc vào giây thứ 2150, nghĩa là tại giây 2150, bluezone sẽ thua nhỏm đến hết cỡ, gây sát thương lên toàn bản đồ và thường vào lúc này cũng chỉ còn nhiều nhất 1 player sống sót, do đó nhóm sẽ xóa các dòng có time lớn hơn 2160, khoảng tầm 10s sau khi vòng bo thu vào hết cỡ.

Ngoài ra khi gán phase cho cột time, nhóm cũng không rõ vì sao có một số lượng rất nhỏ các hàng không gán được phase dù có giá trị time hợp lệ. Vì số lượng các dòng này quá nhỏ, nhóm quyết định xóa toàn bộ các dòng này

2.3. Khám phá dữ liệu

Trên thực tế việc khám phá dữ liệu đã được nhóm thực hiện một phần ở phần 1.2. Tiền xử lí dữ liệu. Ở phần này, nhóm chủ yếu tìm hiểu sự phân bố của dữ liệu ở các cột.

2.3.1. Tổng quan kích thước, kiểu dữ liệu

Sau khi tiến hành tiền xử lí, dữ liệu thu được dùng để phân tích có 8,253,143 dòng cà 9 cột. Mỗi dòng mô tả một kill diễn ra trong trận đấu, gồm:

Mỗi dòng mô tả một kill diễn ra trong trận đấu, gồm:

- killed_by: nguyên nhân gây kill
- kx và ky: tọa của killer trên bản đồ
- time: thời điểm kill xảy ra (tính từ đầu trận đấu)
- vx và vy: tọa của victim trên bản đồ
- dis: khoảng cách giữa killer và victim
- phase: phase diễn ra kill
- type: phân loại nguyên nhân gây kill

Về kiểu dữ liệu Dữ liệu của các cột:

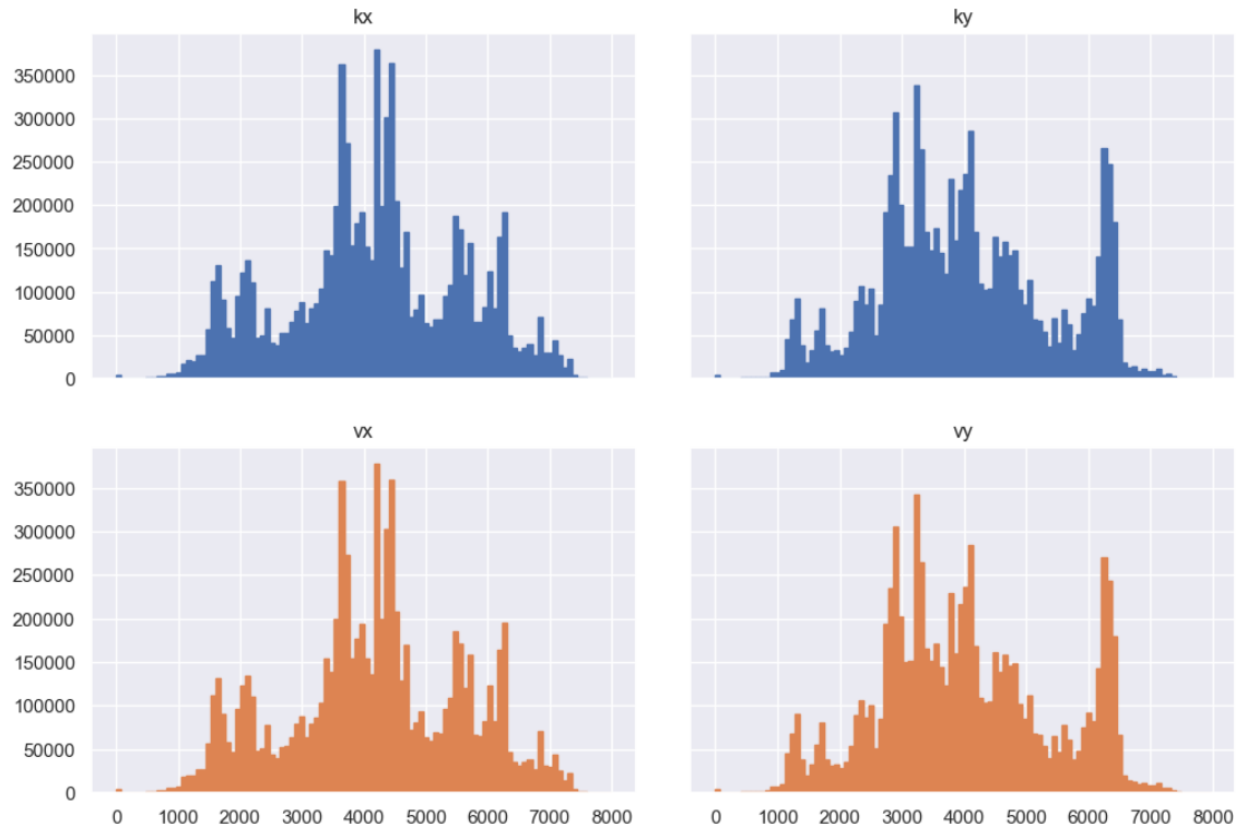
- killed_by và type: nominal
- kx, ky, vx, vy, dis: numerical
- time và phase: 2 cột này mang ý nghĩa thời điểm, nên nhóm xếp 2 cột này vào nhóm dữ liệu thời gian

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8286873 entries, 0 to 8286872
Data columns (total 9 columns):
 #   Column     Dtype
---  ------     -----
 0   killed_by  object
 1   type       object
 2   time       int64
 3   phase      int64
 4   dis        int64
 5   kx         int64
 6   ky         int64
 7   vx         int64
 8   vy         int64
dtypes: int64(7), object(2)
memory usage: 569.0+ MB
```

Hình 2.8. Bảng kiểu dữ liệu của các cột

2.3.2. Phân bố của dữ liệu trong từng cột

2.3.2.1. Các cột tọa độ



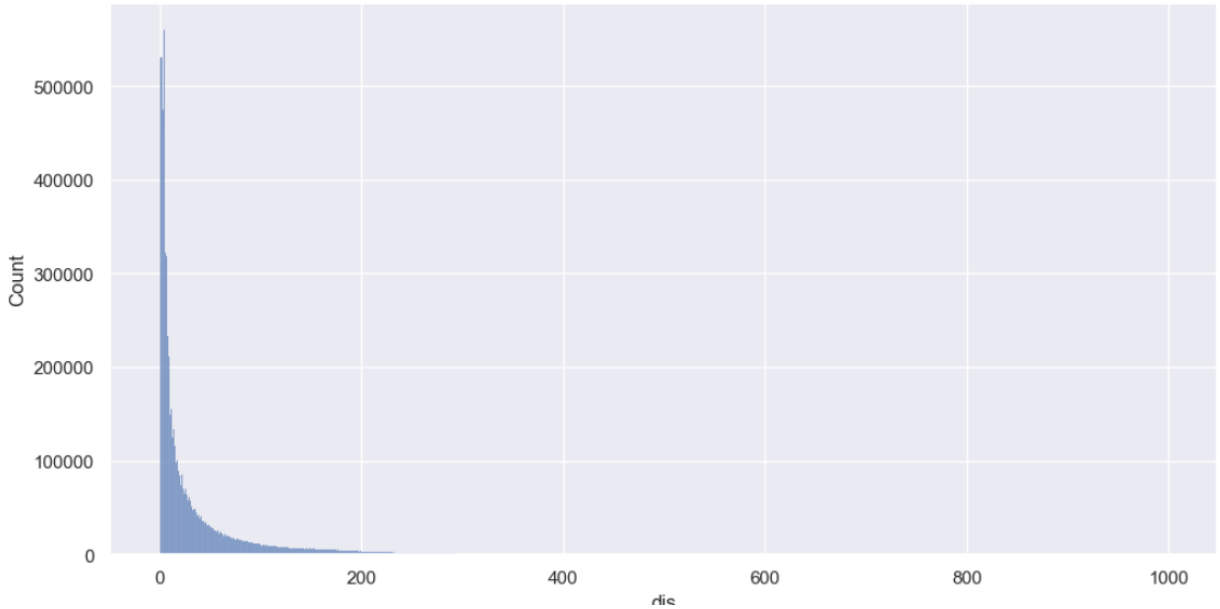
Hình 2.9. Histogram các cột tọa độ

Ở đây nhóm sử dụng histogram để trực quan hóa phân bố của các cột tọa độ kx, ky, vx, vy.

Cả 4 cột tọa độ đều phân bố trong khoảng từ 0 tới 8000:

- Tọa độ tập trung của killer và victim nhìn chung khá giống nhau
- Đối với các tọa độ x, có khoảng 3 cụm đỉnh lần lượt ở 1500 - hơn 2000, 3500 - 4500 và 5500 - 6500
- Đối với các tọa độ y, có một cụm đỉnh lớn trong khoảng 3000 - 4500 và một đỉnh riêng lẻ ở khoảng 6500
- Vị trí của các cụm sẽ được giải thích ở phần sau, khi nhóm sử dụng thêm các biểu đồ trực quan hóa khác.

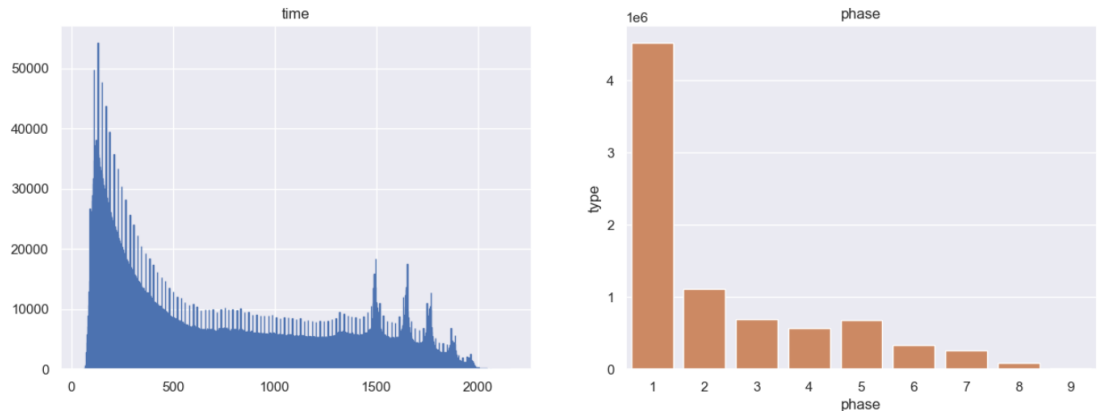
2.3.2.2. Cột dis



Hình 2.10. Histogram cột dis

Cột dis phân bố từ 0 tới 1000 (do nhóm đã thực hiện lọc ở phần Data Preprocessing), tập trung cao nhất ở mức 0, số lượng kill giảm dần khi dis tăng lên (có vẻ là theo hàm log)

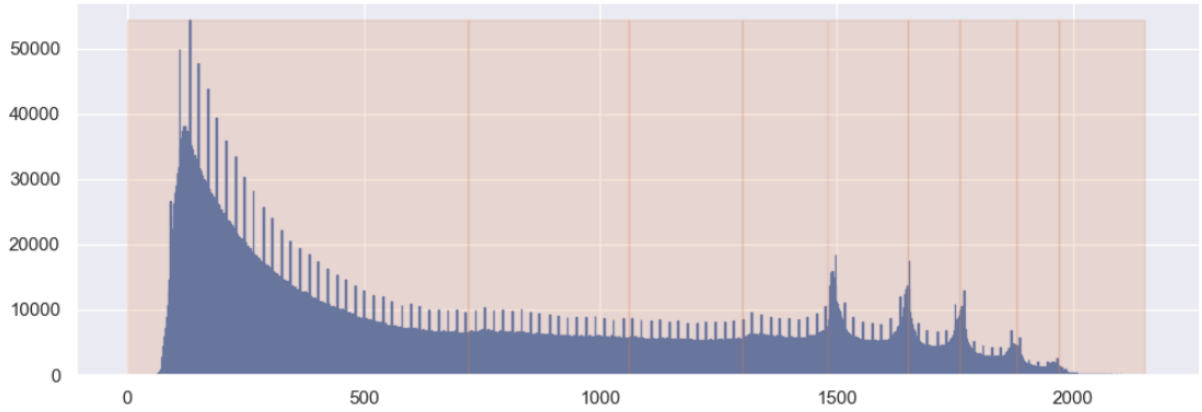
2.3.2.3. Cột time và phase



Hình 2.11. Histogram của time và barplot tương ứng của phase

- Đa số kill diễn ra ở phase 1, giảm dần ở các phase tiếp theo. Có thể do ở phase 1, các player vừa nhảy dù xuống, nhiều player sẽ nhảy dù vào cùng những địa điểm thuận lợi, dẫn tới giao tranh xảy ra vào phase 1 là cao nhất. Sau đó nhiều player tử trận, giao tranh giảm xuống, số lượng kill cũng giảm theo

- Khi quan sát histogram của time, ta nhận thấy có một số đỉnh cao đột ngột từ khoảng 1500s trở về sau, khi xem xét bảng thời gian bắt đầu và kết thúc của các phase dưới đây, ta nhận ra thời điểm kill tăng lên là ở các khoảng thời gian chuyển tiếp từ phase này sang phase kia, khi vòng bluezone thu vào, buộc người chơi phải di chuyển khỏi các nơi trú ẩn. Thêm vào đó là số lượng kill gây ra bởi sát thương từ bluezone, dẫn tới số lượng kill tăng cao. Nhóm kết hợp các đường khoảng đánh dấu phase ở histogram bên dưới để làm nổi bật điều này



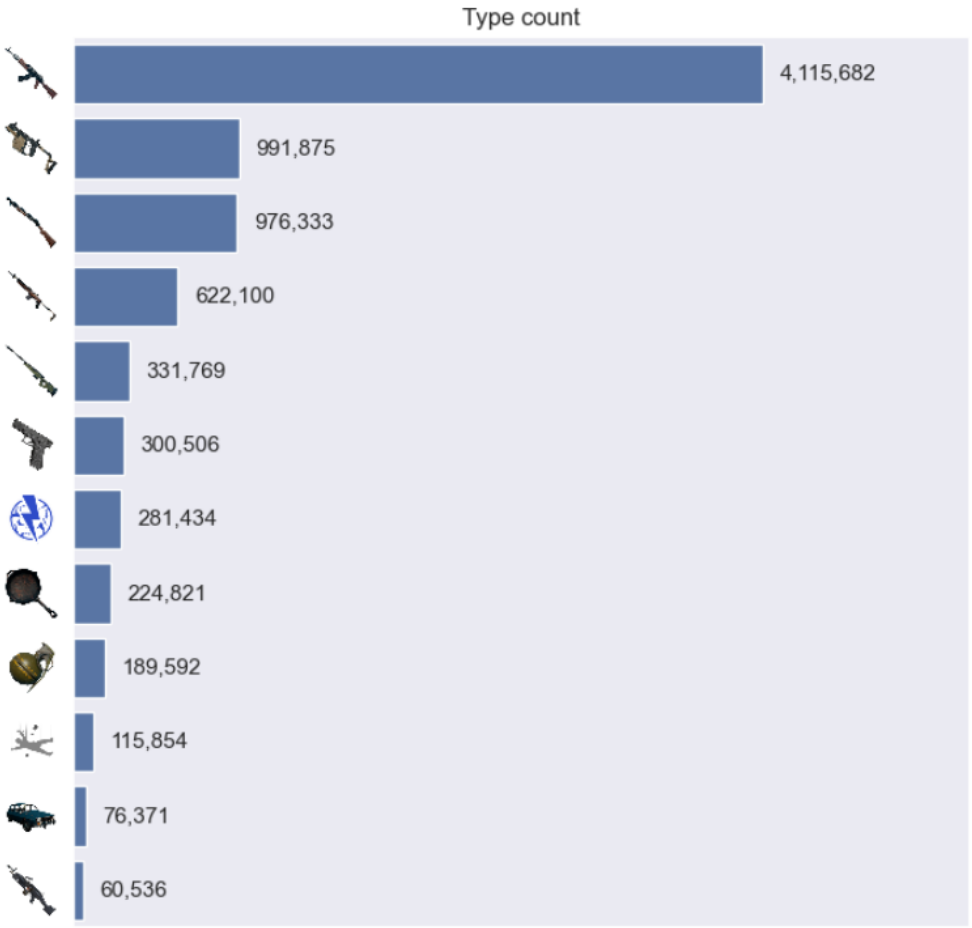
Hình 2.12. Histogram của time với các đường phân chia phase

2.3.2.4. Cột killed_by và type

Cột killed_by có số lượng phần tử khá nhiều, do đó thay vì đánh giá từng nguyên nhân gây kill khác nhau, ta có thể đánh giá theo từng nhóm nguyên nhân. Ở đây thay vì sử dụng tên các type làm nhãn cho biểu đồ, nhóm sử dụng các kí hiệu vũ khí đặc trưng của nhóm để làm nhãn cho biểu đồ.

- Đa số kill được gây ra bởi các súng AR với hơn 4 triệu kill, theo sau đó là SMG và Shotgun với số lượng kill xấp xỉ nhau khoảng gần 1 triệu kill
- các loại súng nhắm DMR và SR xếp hạng 4 và 5 với lần lượt hơn 600,000 và hơn 300,000 kill
- Đáng chú ý số lượng kill gây bởi Handgun nhiều hơn cả Bluezone và Redzone cộng lại
- Vũ khí cận chiến gây hơn 200,000 kill trong khi vũ khí ném chỉ gây gần 190,000 kill

- Số player tử trận bởi đuối nước và té ngã nhiều hơn cả số player bị tông bởi các loại xe
- Súng máy hạng nhẹ LMG gây ra ít kill nhất, khoảng 60,000 kill



Hình 2.13. Phân bố của cột type

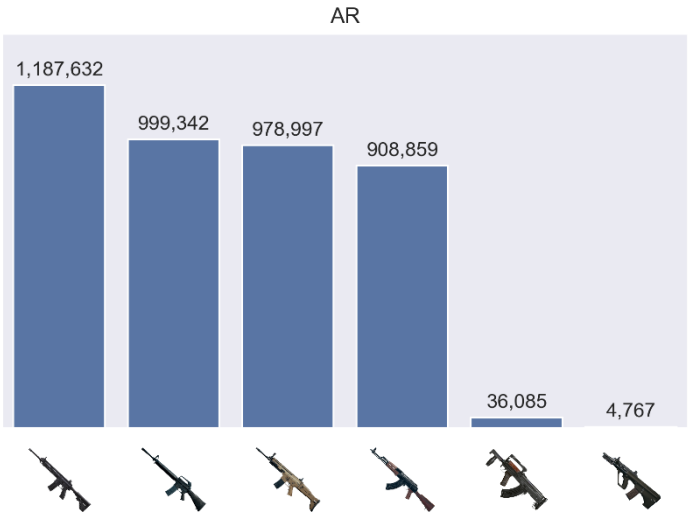
2.3.3. Phân bố của dữ liệu giữa các cột với nhau

2.3.3.1. Đếm killed_by theo từng type

2.3.3.1.1. AR

- Số lượng kill của AR tập trung chủ yếu vào 4 loại súng M416, M16A4, SCAR-L và AKM. Trong đó cao nhất là M416 với hơn 1 triệu kill, có thể là do M416 là loại súng gây sát thương trung bình cao, tốc độ đạn trung bình cao và độ ổn định tốt hơn so với các loại súng khác.

- Groza và AUG có số lượng kill khá thấp (khoảng 36000 và 4500) vì đây là các loại súng xuất hiện trong hòm cứu trợ, có sát thương cao, tốc độ bắn nhanh nhưng số lượng xuất hiện trong trận đấu rất ít.

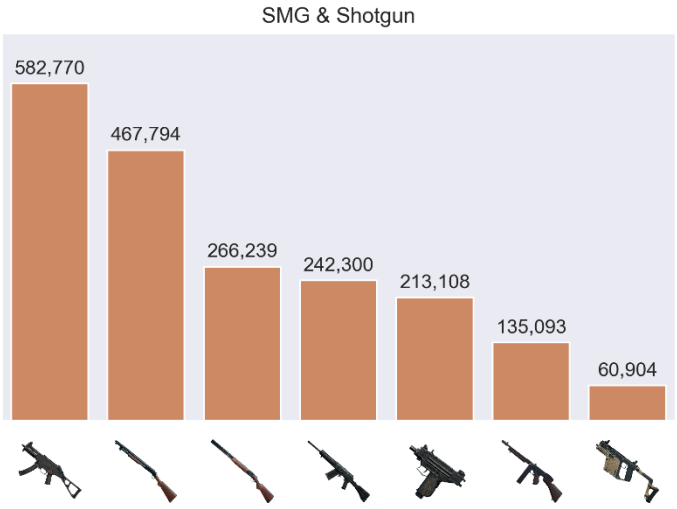


Hình 2.14. Các loại súng AR

2.3.3.1.2. SMG và Shotgun

Đều là các loại súng gây kill tầm ngắn nên nhóm tiến hành phân tích 2 loại súng này trên cùng một biểu đồ

- UMP45 có số lượng kill cao nhất, gần 600,000 kill
- Theo sau đó là 3 loại Shotgun S686, S1897 và S12K
- 3 loại SMG còn lại gây khá ít kill, thấp nhất là Vector với khoảng 60,000 kill



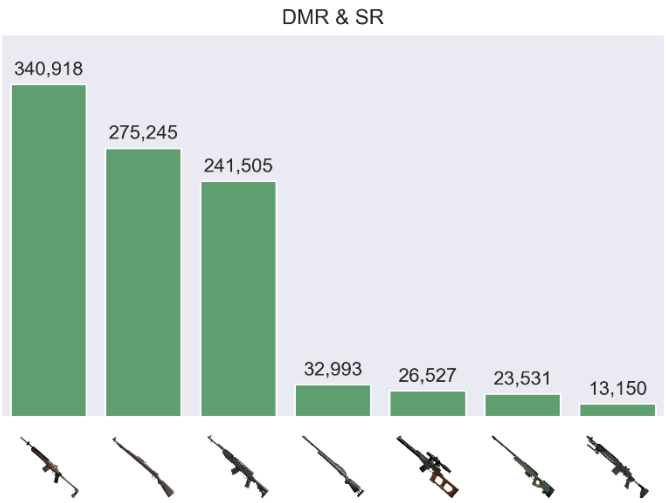
Hình 2.15. Các loại súng SMG

2.3.3.1.3. DMR và SR

Đều là các loại súng gây kill tầm xa nên nhóm tiến hành phân tích 2 loại súng này trên cùng một biểu đồ

- Top 1 là một DMR: Mini 14 với khoảng 340,000 kill
- Theo sau đó là Kar98k, súng nhắm được nhiều player ưa thích với 275,000 kill
- DMR SKS cũng có số lượng kill tương đối cao: 241,000 kill

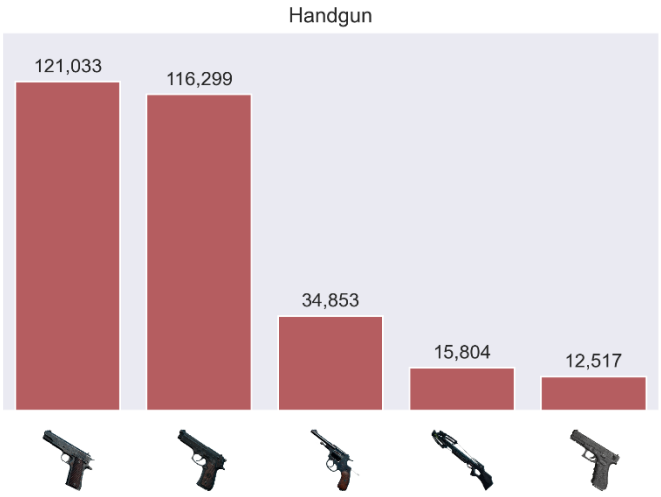
- Trong 4 loại súng top cuối thì M24, AWM và MK14 đều là những khẩu súng trong hòm cứu trợ, gây sát thương cao nhưng xuất hiện rất ít trong trận đấu
- VSS do sát thương khá ít, gây tương đối ít kill so với các loại súng top đầu, chỉ khoảng 26,000 kill



Hình 2.16. Các loại súng DMR và SR

2.3.3.1.4. Handgun

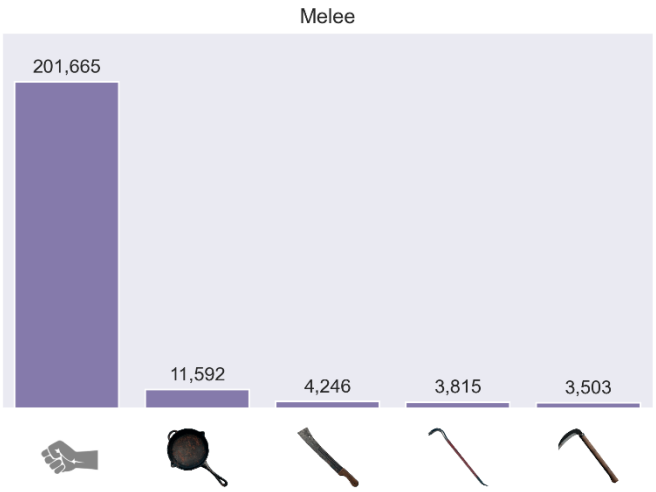
- P1911 gây hơn 120,000 kill, cùng với P92 gây hơn 116,000 kill là 2 loại súng lục hiệu quả nhất với cơ chế bắn liên thanh thay vì bắn từng viên
- R1895 đứng thứ 3, nỏ với sát thương lớn "one hit one kill" đứng thứ 4 và P18C đứng hạng cuối



Hình 2.17. Các loại súng lục và nỏ

2.3.3.1.5. Melee

- Nắm đấm lại chiếm nhiều kill nhất so với nhóm vũ khí cận chiến, đạt 200,000 kill, bỏ xa chảo ở vị trí thứ 2 với 11,600 kill
- Các loại vũ khí còn lại gây khá ít kill, khoảng dưới 5000



Hình 2.18. Các loại Melee

2.3.3.1.6. Các nhóm còn lại

- Không quá bất ngờ khi Bluezone gây nhiều kill nhất trong các nguyên nhân còn lại, khoảng 270,000 kill
- Grenade ở hạng 2 với khoang 180,000 kill
- Số lượng player tử trận do té ngã là g còn nhiều hơn cả số lượng kill của DP-28 và M249 cộng lại

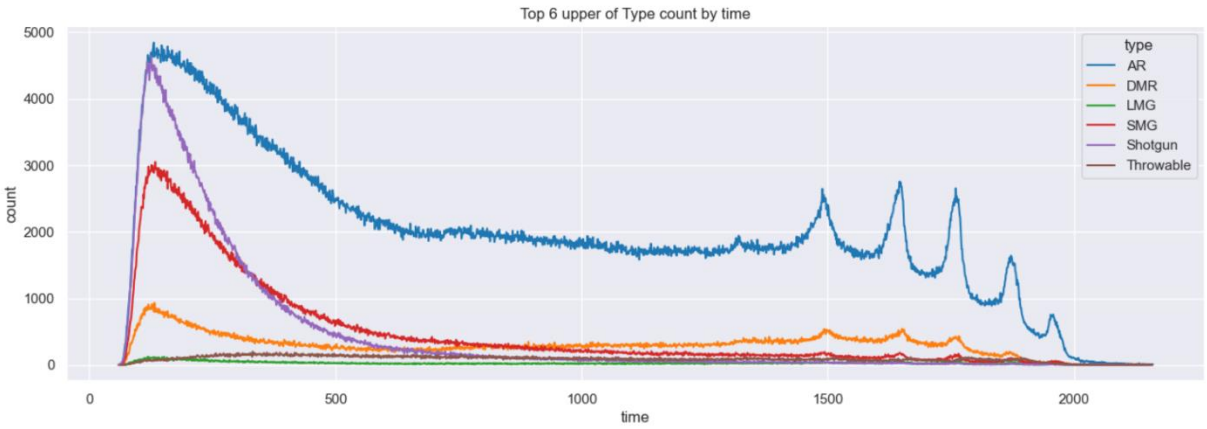
Hình 2.19. Các loại nguyên nhân còn lại

- Số kill của Molotov là ít nhất, khoảng 7000 kill

2.3.3.2. Đếm type theo time

Ở phần này nhóm trực quan hóa số lượng kill bởi từng loại nguyên nhân theo thời gian trận đấu

- Đa phần các type có số lượng tuân theo tổng số kill theo thời gian: cao ở giai đoạn đầu, giảm dần qua thời gian và tăng cao đột ngột ở các giai đoạn chuyển phase.
- Thứ hạng của các type hầu như giữ nguyên theo thời gian, trừ:



Hình 2.20. Số lượng kill của top 6 loại vũ khí gây nhiều kill nhất theo thời gian trận đấu

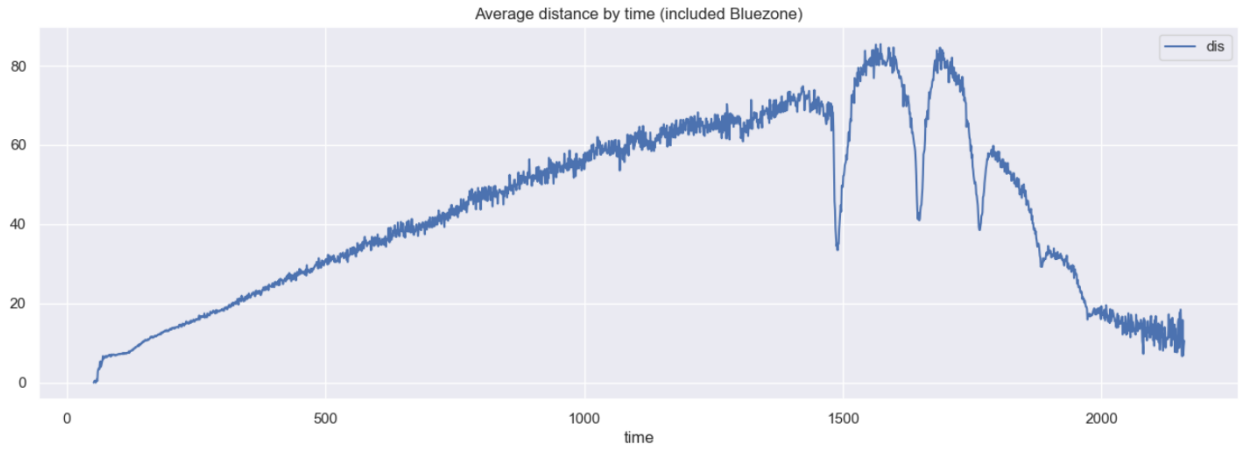
- Shotgun sau khi giữ top 2 trong khoảng 750s đầu trận rơi dần xuống top 6 ở giai đoạn cuối trận. Giải thích: Shotgun là gây sát thương rất cao nhưng tầm bắn rất ngắn, phù hợp với giai đoạn đầu trận khi giao tranh xảy ra ở các công trình, khoảng cách giao tranh nhỏ, số lượng player đông, cần phải xử lí nhanh. Khi trận đấu dần ổn định, các player giữ khoảng cách xa, shotgun gần như không có tác dụng.
- SMG nắm giữ top 3, sau đó vươn lên top 2 khi Shotgun rơi xuống top 3, sau đó tiếp tục rơi xuống top 3 khi DMR vươn lên top 2. Nguyên nhân cũng giống Shotgun, SMG là súng tiểu tiên, tốc độ ra đạn cực kì nhanh, có khả năng gây sát thương lớn trong khoảng thời gian ngắn ở khoảng cách gần, được sử dụng nhiều ở giai đoạn đầu trận. Ở các giai đoạn sau SMG mất ưu thế khi khoảng các của các player tăng lên, tuy nhiên SMG vẫn hiệu quả hơn Shotgun (tầm bắn xa hơn, liên thanh) nên vẫn nắm giữ top 3.
- DMR giữ top 4 ở giai đoạn đầu trận, sau đó từ phase 2 trở đi dần dần leo lên top 2. Giải thích: DMR là súng bắt tỉa theo cơ chế semi-sniper (bắn từng viên), tuy gây sát thương tương đối cao nhưng không thích hợp giao tranh tầm gần, rất gần (do cần gây sát thương lên kẻ thù với tốc độ nhanh) nên khó gây nhiều kill ở giai đoạn đầu. Ở giai đoạn sau, khi các player ổn định vị trí ở các khoảng cách tương đối, DMR phát huy ưu thế hơn so với SMG và Shotgun.

Hình 2.21.Hình 1.20. Số lượng kill của các loại vũ khí còn lại theo thời gian trận đấu

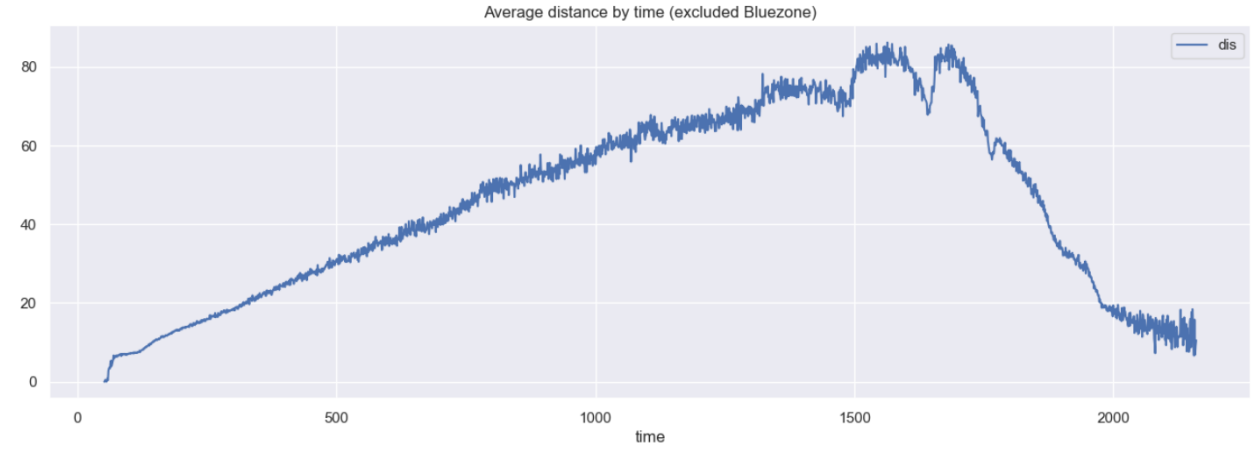
- Tương tự với SMG và Shotgun, ở giai đoạn đầu trận, Handgun và Melee nắm giữ top đầu gây kill (do các player vừa đáp dù xuống mặt đất, bất kì thứ vũ khí gì có trong tay đều được sử dụng để tiêu diệt đối thủ, kể cả các cũ khí cầm tay và đấm tay không). Tất nhiên sau khi đã có súng, các player chỉ dùng súng để gây kill
- Seft có một đỉnh cao nhẹ ở giai đoạn đầu và ở giai đoạn 1500s (chuyển phase) có thể do ở giai đoạn đầu lầ giai đoạn đáp dù (dễ đáp dù lỗi, vướng vào các tòa nhà, cây cao) và di chuyển giữa các kiến trúc nhiều (nhảy qua lại giữa các tòa nhà, nhảy từ tầng cao xuống để giao tranh/tránh giao tranh) dẫn đến nhiều player tử trận vì té ngã. Ở giai đoạn chuyển phase 5, nhiều player trong quá trình di chuyển có thể bị té ngã từ các vách núi, công trình hoặc đuối nước trong lúc né tránh kẻ thù khi đang bơi.
- Zone đặt biệt cao ở các giai đoạn chuyển phase, khi Bluezone bắt đầu thu vào, nhiều player mắc kẹt ở ngoài vòng xanh, bị gây sát thương dẫn đới tử trận.

2.3.3.3. Trung bình khoảng cách theo thời gian

Ở phần này nhóm tính trung bình của cột dis theo thời gian diễn ra trong trận đấu



Hình 2.22. Trung bình khoảng cách theo thời gian (Có tính Bluezone)

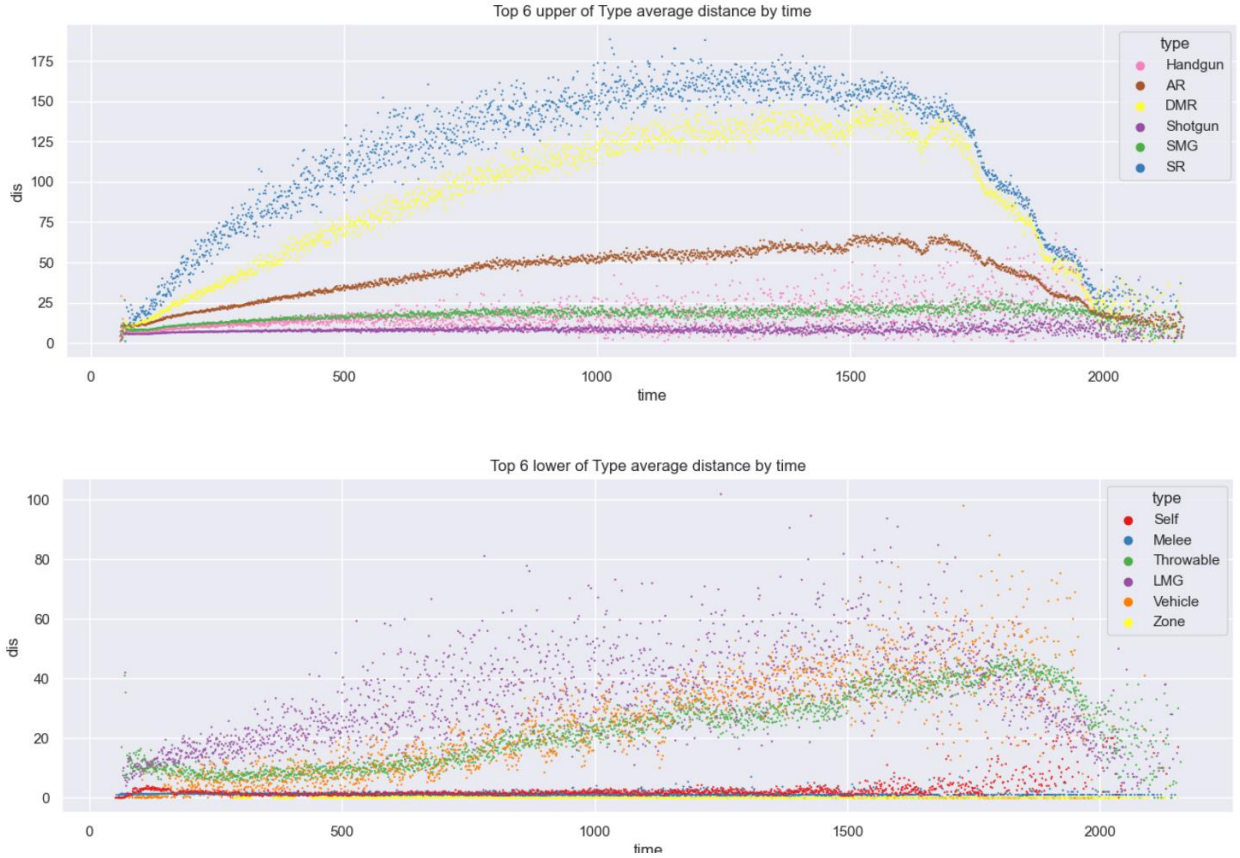


Hình 2.23. Trung bình khoảng cách theo thời gian (không tính Bluezone)

- Khoảng cách trung bình của các kill gây ra thấp ở đầu trận, khi các player vừa đáp dù xuống đất, va chạm với nhau ở các khu vực công trình thuận lợi, vị trí có xe. Thêm vào đó là ở giai đoạn này các player chưa thu thập được nhiều trang bị như ống ngắm, giảm giật nên việc gây kill ở khoảng cách xa cũng rất ít. Do đó khoảng cách trung bình ở giai đoạn này thấp
- Sau đó khi trận đấu ổn định, các player giữ khoảng cách với nhau, chuyển sang sử dụng AR, DMR và SR nhiều hơn thay vì các vũ khí tầm gần như SMG và Shotgun, do đó khoảng cách trung bình của các kill tăng lên.
- Ta có thể thấy một vài đỉnh lõm xuống bất thường giống với các đỉnh tăng lên bất thường của số lượng kill theo thời gian ở các giai đoạn chuyển phase. Cùng lí do rằng khi chuyển phase, vòng bluezone sẽ ép các tuyển thủ di chuyển, trong nhiều

trường hợp là ra khỏi các công trình kiến trúc và ca chạm nhau trong khoảng cách hẹp, do đó làm giảm khoảng cách trung bình của kill. Sau đó ở giai đoạn vòng bluezone đứng yên, các tuyển thủ lại tiếp tục gây kill ở khoảng cách xa. Thêm vào đó, các kill do bluezone gây ra ở giai đoạn này tăng vọt lên, mà khoảng cách kill của bluezone là bằng 0, do đó trung bình khoảng cách sẽ bị kéo xuống nhiều.

- Khi trận đấu tiến về cuối, bluezone thu hẹp lại nên khoảng cách gây kill cũng giảm dần theo
- Tuy nhiên ta có thể thấy sự bất thường khi khoảng dữ liệu từ giây thứ 2000 trở đi lại giãn rộng ra

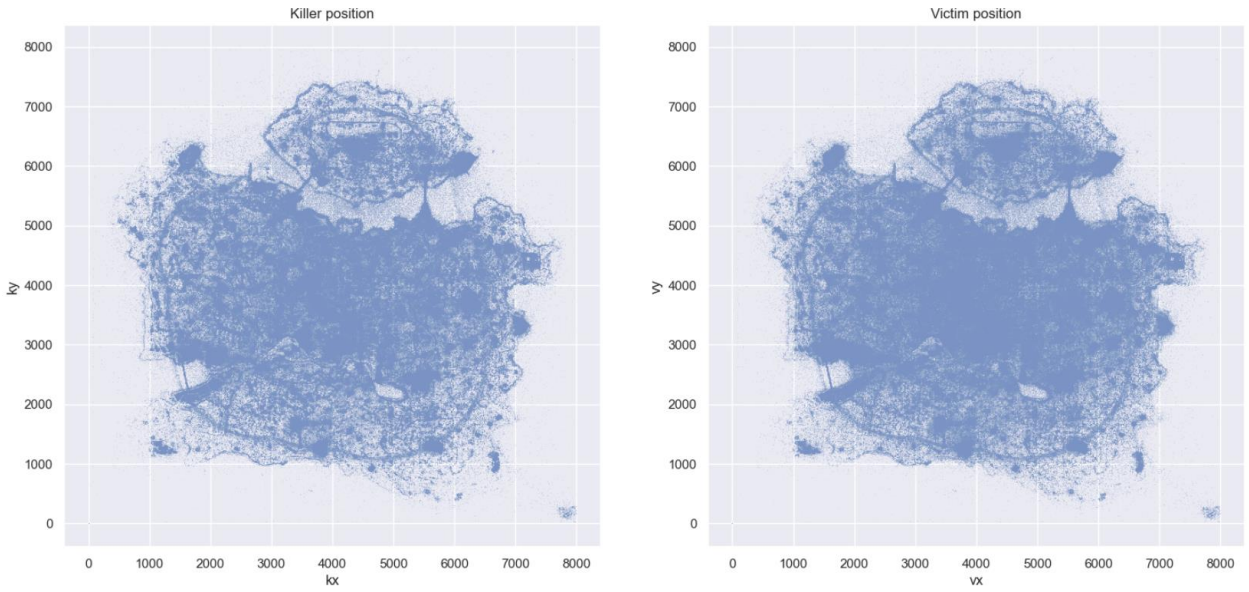


Hình 2.24. Khoảng cách trung bình của từng nhóm nguyên nhân theo thời gian

- Khoảng cách trung bình của từng nhóm nguyên nhân gây kill theo thời gian cũng tuân theo quy luật tăng giảm như quy luật đã nêu ở trên

2.3.3.4. Phân bố dữ liệu tọa độ trên mặt phẳng 2 chiều

Ở phần trước, nhóm đã biểu diễn phân bố của các cột tọa độ trên 1 chiều (trục ngang). Ở phần này nhóm sẽ biểu diễn phân bố các cột tọa độ theo 2 chiều.

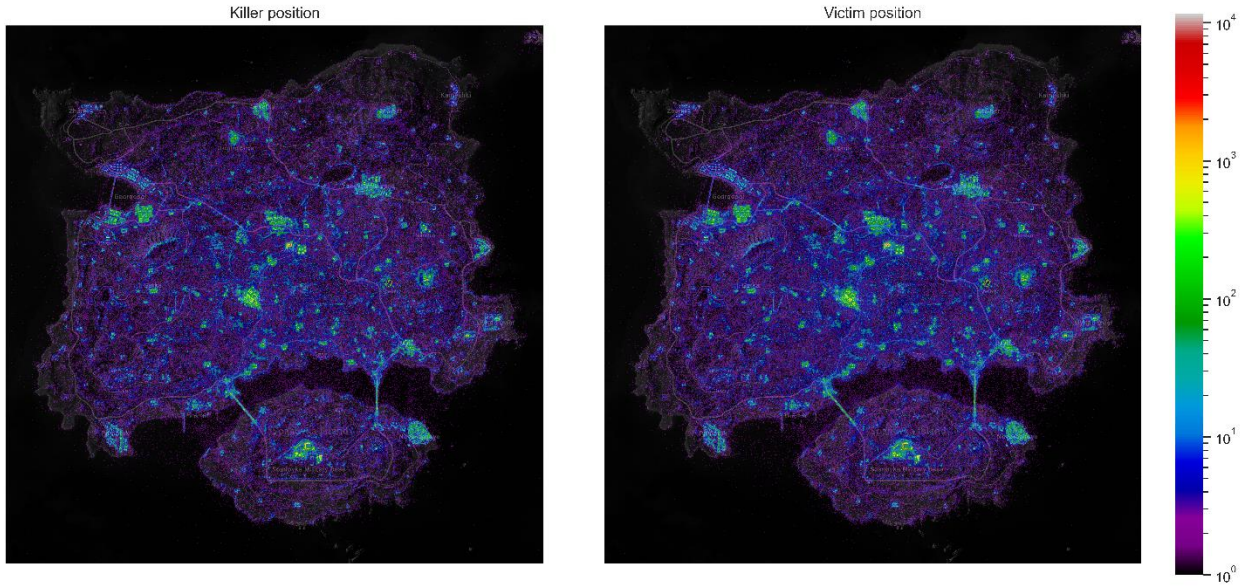


Hình 2.25. Phân bố 2 chiều của killer và victim

Nhìn chung ta thấy phân bố của killer có vẻ phân bố rõ ràng các vùng đặc, rỗng hơn so với phân bố của victim. Có thể giải thích do các killer có xu hướng tập trung trong các công trình để gây kill còn victim thường có xu hướng tử trận ở đều các vị trí hơn

Quan sát phân bố của vị trí của killer và position, ta dễ dàng nhận ra dữ liệu phân bố theo hình dạng của bản đồ ERANGEL. Để trực quan hơn, nhóm có thể visualize đè lên bản đồ ERANGEL.

Hình 2.26. Bản đồ Erangel



Hình 2.27. Heatmap phân bố của killer và victim

Trên đây là heatmap biểu diễn số lượng kill ở các khu vực với thang đo màu bên phải biểu hiện cho log_{10} của số lượng kill tại 1 điểm. Ví dụ những điểm màu xanh dương sẽ có $10^1 = 10$ kill còn những điểm có màu đỏ sẽ có $10^{3.5} \approx 3000$ kill. Mỗi điểm ở đây là 1 khu vực hình vuông 5m*5m. Sỡ dĩ nhóm không chia thành hình vuông 1m*1m là vì khi đó số lượng điểm trên bản đồ sẽ là $64 * 10^6$, bộ nhớ RAM của thành viên trong nhóm chỉ có 8GB, không thể xử lí tác vụ này.

Quan sát sơ bộ, ta có thể thấy lại điều đã nêu ở trên, vị trí của victim có xu hướng phủ rộng hơn (các vùng màu xanh có vẻ trải rộng hơn). Khi nhìn vào các khu vực nhỏ như Severny ở phía bắc, School ở trung tâm hay thành phố Georgopol ở phía tây bắc ta có thể thấy rõ điều này.

2.4. Một số câu hỏi có ý nghĩa

2.4.1. Loại vũ khí nào hiệu quả nhất?

2.4.1.1. Ý nghĩa

Khi tìm được loại vũ khí hiệu quả nhất, các player sẽ có sự lựa chọn loại vũ khí hợp lí cho chiến thuật của mình (vì mỗi player chỉ được trang bị 2 vũ khí chính, 1 Handgun, 1 Melee và 1 vũ khí ném (Throwable)), đặc biệt là với những player mới chơi, đang làm

quen với game. Ở phần này nhóm chỉ phân tích độ hiệu quả của các loại vũ khí chính (trừ Handgun, Melee, Throwable, Zone, Self, Vehicle)

2.4.1.2. Phân tích

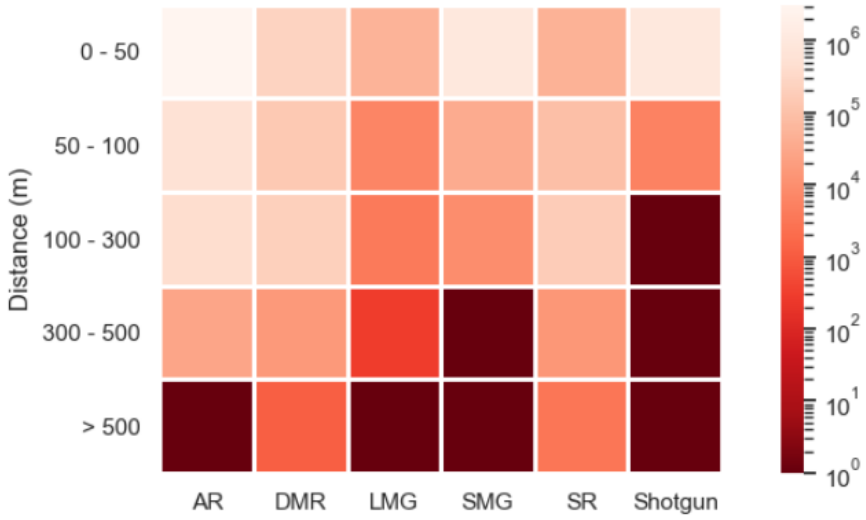
Để trả lời cho câu hỏi này, ta cần một thang đo độ hiệu quả. Khoảng cách và số lượng kill là 2 tiêu chí quan trọng để đánh giá độ hiệu quả của vũ khí.

- Khoảng cách kill ngoài phụ thuộc vào vũ khí còn phụ thuộc vào trình độ của player cũng như vài yếu tố máy mắn khác, do đó ta sẽ sử dụng khoảng cách trung bình để đánh giá
- Số lượng kill ngoài cho biết loại vũ khí đó gây kill nhiều hay ít còn cho biết loại vũ khí nào được sử dụng nhiều hơn (nên gây kill nhiều hơn, vì nếu không gây kill hiệu quả thì các player sẽ đổi sang loại vũ khí hiệu quả hơn)
- Ngoài ra còn một số yếu tố khác như sát thương, độ giật. Tuy nhiên với dữ liệu nhóm đang có hiện tại, ta sẽ tạm phân tích dựa trên khoảng cách và số lượng kill

Vì vậy nhóm sẽ sử dụng heatmap để trực quan hóa độ hiệu quả của vũ khí như sau:

- Các dòng là các khoảng cách khác nhau:
    - 0m - 50m
    - 50m - 100m
    - 100 - 300m
    - 300 - 500m
    - Trên 500m
- Các cột là các loại vũ khí
- Mỗi cell giao của dòng với cột là số lượng kill của loại vũ khí ở khoảng cách tương ứng

2.4.1.3. Trực quan hóa và trả lời



Hình 2.28. Heatmap số lượng kill theo khoảng cách và loại súng

Trong heatmap trên, những cell càng sáng nghĩa là loại vũ khí đó càng gây được nhiều kill ở khoảng cách tương ứng

- Đối với khoảng cách rất ngắn dưới 50m thì AR, SMG và Shotgun là các loại vũ khi hiệu quả
- Khi khoảng cách tăng dần lên thì SMG và Shotgun dần mất lợi thế, lên đến mức 100 – 300m thì số lượng kill của SR và DMR đã vượt hẳn lên, đến mức trên 500m thì SR và DMR thống trị bảng kill
- LMG dường như không lọt top hiệu quả nhất ở bất kì khoảng cách nào, chỉ nên lựa chọn khi không còn sự lựa chọn khác

2.4.2. Những vị trí nào giao tranh nhiều ở khoảng đầu trận?

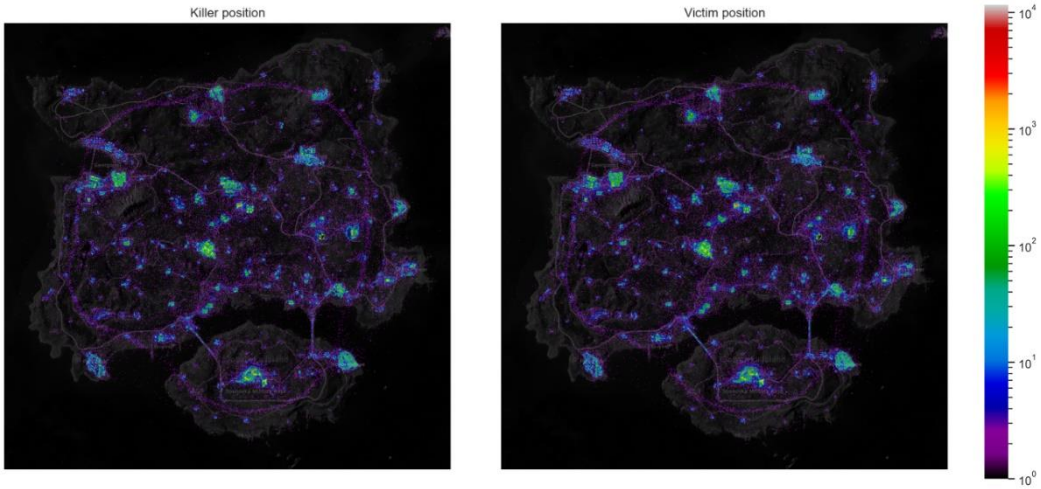
2.4.2.1. Ý nghĩa

Trên bản đồ sẽ có những khu vực giao tranh nhộn nhịp hơn những khu vực khác, biết được các khu vực này giúp player chọn nơi nhảy dù, trú ẩn phù hợp với lối chơi của mình

- Ví dụ player thích thử thách, nâng cao kỹ năng, có thể chọn các khu vực đông đúng, giao tranh xảy ra liên tục
- Nếu người chơi chọn lối chơi sinh tồn có thể chọn các khu vực yên ắng, ít người

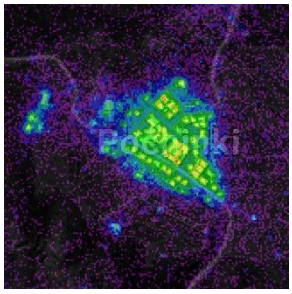
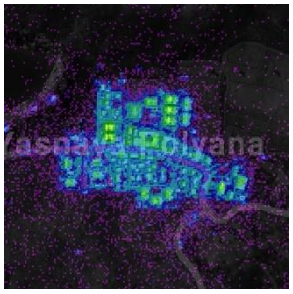
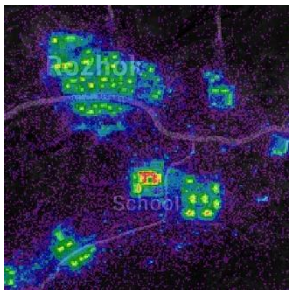
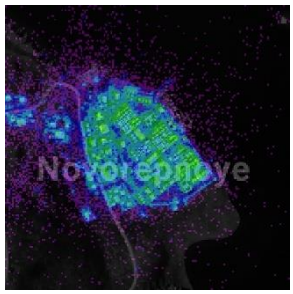
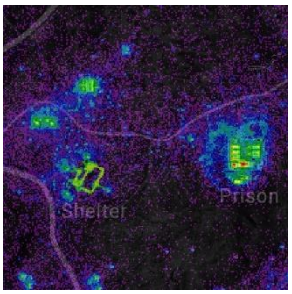
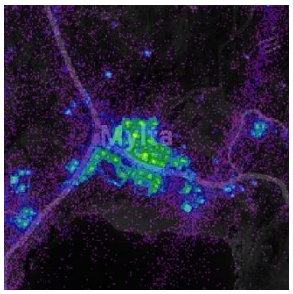
2.4.2.2. Phân tích

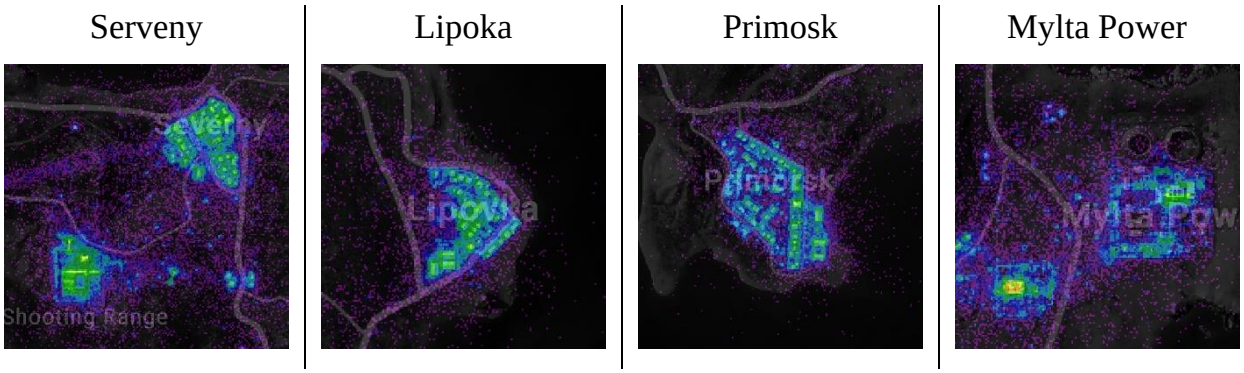
Câu hỏi này có thể trả lời dễ dàng bằng phân bố tọa độ 2 chiều đã làm ở Data Understanding nhưng ở đây ta sẽ giới hạn thời gian lại ở khoảng đầu trận, cụ thể là từ lúc bắt đầu tới giây thứ 600



2.4.2.3. Trực quan hóa và trả lời

Ta có thể các kiến trúc, các khu vực thành phố là những khu vực giao tranh gay gắt giai đoạn đầu trận

Pochinki	Military Base	Georgopol	Yasnaya Polyana
			
Rozhok & School	**Novorepnoy**	**Prison & Shelter**	**Mylta**
			



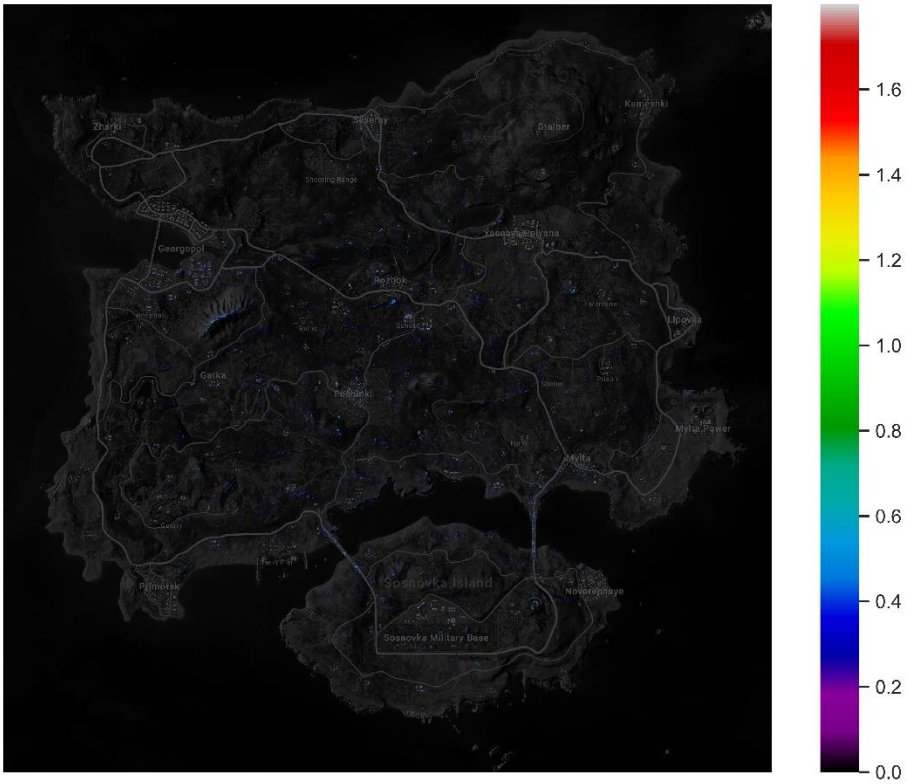
2.4.3. Những vị trí nào thuận lợi để gây kill từ xa?

2.4.3.1. Ý nghĩa

Trong đa số giai đoạn của trận đấu thì khả năng gây kill từ xa sẽ mang lại lợi thế rất lớn cho các player. SR và DMR là các vũ khí ưa chuộng dùng để gây kill từ xa. Biết được những vị trí thuận lợi để gây kill từ xa sẽ giúp các player có thêm những sự lựa chọn trong chiến lược của mình.

2.4.3.2. Phân tích

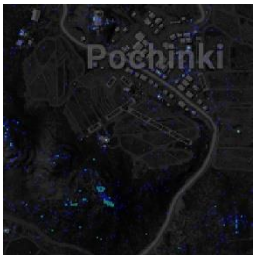
Ta có thể làm tương tự cách biểu diễn heatmap số lượng kill đã làm ở trên



Hình 2.29. Các vị trí dễ gây kill từ xa

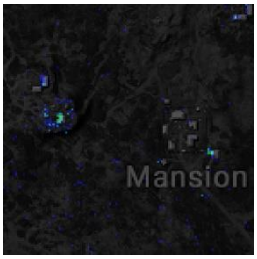
2.4.3.3. Trực quan hóa và trả lời

Đồi gần Pochinki



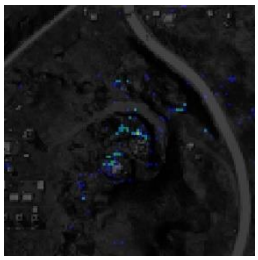
Núi xương cá gần cảng Georgopol

Nhà trên vách núi gần Mansion



School và Rozhok

Núi cao giữa gần cảng Novorepnoye



Ngoài các vị trí trên còn 2 vị trí cầu nối đảo lớn và đảo nhỏ của bản đồ

2.4.4. Những vị trí gây kill thuận lợi gây kill từ xa có thực sự an toàn?

2.4.4.1. Ý nghĩa

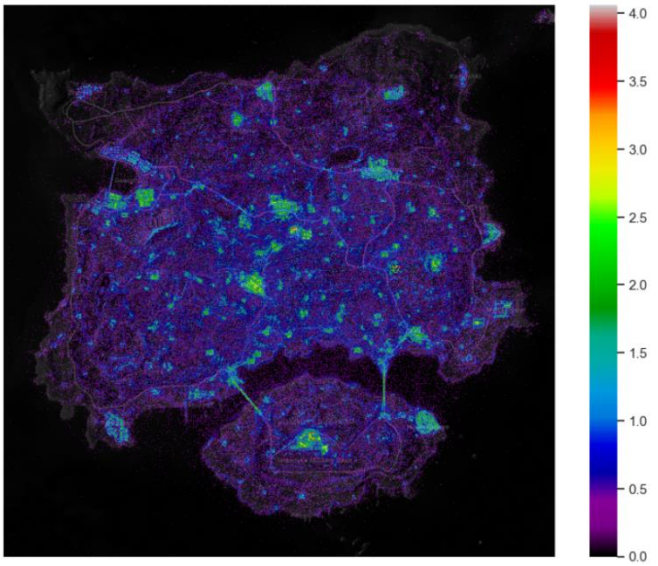
Tuy việc biết được các vị trí thuận lợi gây kill từ xa sẽ mang lại lợi thế nhưng các player cũng cần biết tình hình giao tranh tại các khu vực này.

2.4.4.2. Phân tích

Ta có thể biết được phần nào tình hình giao tranh dựa trên phân bố tọa độ của victim 2 chiều như đã làm ở phần trước. Bên cạnh đó ta cũng có thể nhìn sâu hơn, đây là các khu vực thuận lợi gây kill từ xa nên cũng có thể bị gây kill từ xa

2.4.4.3. Trả lời

Quan sát bản đồ, ta có thể thấy các địa điểm dễ dây kill từ xa đã đề cập ở câu 3 cũng là những địa điểm nhiều player tử trận nhất. Ngoài việc dễ tử trận, tức là các vị trí này diễn ra nhiều giao tranh, các vị trí này cũng dễ bị gây kill từ xa.



Quan sát kĩ hơn các hướng đạn bắn vào các vị trí này, ta nhận thấy các vị trí này dễ bị bắn ở mọi hướng. Do đó nếu muốn chiếm các vị trí này để gây kill từ xa, player cũng cần phải cẩn thận quan sát nhiều hướng để đề phòng

2.4.5. Thời điểm thích hợp để qua cầu?

2.4.5.1. Ý nghĩa

Bản đồ gồm 2 đảo lớn và nhỏ, được nối bởi 2 cầu lớn. Trong trận đấu, các player thường phải đi qua cầu để sang đảo kia (để tìm kiếm trang bị, do vòng bluezone thu nhỏ). Tuy nhiên như đã trực quan hóa ở các phần trước, 2 khu vực cầu này là nơi tử trận của rất nhiều player, do đó ta cần cân nhắc thời điểm qua cầu phù hợp. Hoặc ngược lại, các

player có thể chọn lối chơi "camp cầu", thay vì đi qua cầu, các player sẽ dừng ở 2 đầu cầu để canh các player khác đi qua và bắn.

2.4.5.2. Phân tích

Ở các phần trước, nhóm đã trực quan hóa tọa độ của killer và victim trên mặt phẳng tọa độ 2 chiều. Để trả lời câu hỏi này, nhóm sẽ tiến hành animate các tọa độ này theo thời gian để tìm ra những khoảng thời gian phù hợp để qua cầu.

2.4.5.3. Trực quan hóa và trả lời

Do bản báo cáo ở dạng tĩnh nên nhóm đã đăng video ở link này.

Quan sát video trên, ta nhận thấy khoảng thời gian thuận lợi nhất để qua cầu là khoảng trước 800s, nghĩa là ở phase 1 và đầu phase 2. Tuy nhiên ở giai đoạn này vòng bluezone chưa thực sự thu về đảo lớn hay đảo nhỏ, do đó bên cạnh việc sử dụng dữ liệu để quyết định thời gian qua cầu, các player cần phải quyết đoán đưa ra phán đoán có nên qua cầu hay không. Ngược lại nếu các player muốn camp cầu có thể bắt đầu canh từ đầu phase 2 đến cuối phase 5 tùy vào tình hình vòng bluezone vì đây là khoảng thời gian nhiều player qua cầu nhiều (để né bluezone) và tử trận tại đây.

3. Tài liệu tham khảo

Trong quá trình thực hiện đồ án, nhóm có tham khảo cũng như tra cứu nhiều trên các nguồn online. Do quá nhiều lần tra cứu không thể liệt kê hết ở đây, nhóm chỉ trình bày các nguồn tham khảo chính ở đây

[1] Matplotlib Cheatsheet, [Online]. https://matplotlib.org/cheatsheets/.

[2] Matplotlib Documentation, [Online]. https://matplotlib.org/stable/index.html.

[3] Seaborn Gallery. [Online]. https://seaborn.pydata.org/examples/index.html.

[4] Stack Overflow, [Online]. https://stackoverflow.com/.

[5] Geeks for Geeks. [Online]. https://www.geeksforgeeks.org.