

Project Overview

I conducted an extensive data analysis project using raw JSON data from an online game environment, analyzing over 180 days' worth of interaction logs of 2 factions (Tattaglia and Cuneo) totaling around 230 players. This exercise focused on turning complex, unstructured data into meaningful insights through data cleaning, transformation, and visualization techniques. The analysis was done using R and Microsoft Excel.

Raw JSON File

```
{
  "success": true,
  "warInfo": {
    "war_id": 40,
    "prize": 500000000,
    "date_started": "2024-01-17T18:00:05.000Z",
    "date_end": "2024-06-28T20:19:20.000Z",
    "started_by": 70409,
    "winner_team_id": 1,
    "factions": [
      {
        "team_id": 0,
        "faction_id": 2,
        "kills": 22263,
        "deaths": 28253,
        "money_lost": 3390250000,
        "faction": "Tattaglia",
        "f_name": "Tattaglia",
        "f_abbr": "Tattaglia"
      },
      {
        "team_id": 1,
        "faction_id": 4,
        "kills": 25033,
        "deaths": 26060,
        "money_lost": 3102320000,
        "faction": "Cuneo",
        "f_name": "Cuneo",
        "f_abbr": "Cuneo"
      }
    ],
    "kills": [
      {
        "killerName": "Angel Dinerio",
        "killerId": 123222300,
        "killerFactionId": 2,
        "killedName": "Pablo Tattaglia",
        "killedId": 218973,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:14.000Z",
        "reason": 34,
        "x": 1077.742554,
        "y": 2080.465820,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Scar Salerno",
        "killerId": 174322,
        "killerFactionId": 4,
        "killedName": "Jason McKamey",
        "killedId": 123156712,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:12.000Z",
        "reason": 31,
        "x": 2155.784424,
        "y": -2297.185791,
        "z": 13.447949,
        "interior": 0,
        "virtualworld": 0,
        "killerName": "Elita Cuneo",
        "killerId": 123133717,
        "killerFactionId": 4,
        "killedName": "Blade Wilson",
        "killedId": 27290,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:08.000Z",
        "reason": 24,
        "x": 2156.273193,
        "y": -2299.179932,
        "z": 13.466101,
        "interior": 0,
        "virtualworld": 0,
        "killerName": "Pablo Tattaglia",
        "killerId": 218973,
        "killerFactionId": 2,
        "killedName": "Quiss Tattaglia",
        "killedId": 123135580,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:05.000Z",
        "reason": 24,
        "x": 1072.408447,
        "y": 2083.945312,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Elita Cuneo",
        "killerId": 123133717,
        "killerFactionId": 4,
        "killedName": "Fred Woods",
        "killedId": 180194,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:02.000Z",
        "reason": 24,
        "x": 1076.285034,
        "y": 2082.132080,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": null,
        "killerId": 65535,
        "killerFactionId": 0,
        "killedName": "Clayde Cuneo",
        "killedId": 123131553,
        "killedFactionId": 4,
        "date_killed": "2024-06-28T20:19:00.000Z",
        "reason": 51,
        "x": 2150.230225,
        "y": -2287.146729,
        "z": 14.773438,
        "interior": 0,
        "virtualworld": 0,
        "killerName": "Kevin Salerno",
        "killerId": 218222,
        "killerFactionId": 4,
        "killedName": "Calvin Dinerio",
        "killedId": 123216105,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:19:00.000Z",
        "reason": 24,
        "x": 2158.214111,
        "y": -2297.697998,
        "z": 13.478495,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": null,
        "killerId": 65535,
        "killerFactionId": 0,
        "killedName": "Boris Bagarella",
        "killedId": 123243772,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:54.000Z",
        "reason": 51,
        "x": 1057.937134,
        "y": 2086.608887,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": null,
        "killerId": 65535,
        "killerFactionId": 0,
        "killedName": "Paco Pakito",
        "killedId": 148754,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:52.000Z",
        "reason": 51,
        "x": 1057.727539,
        "y": 2082.587646,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Paco Pakito",
        "killerId": 148754,
        "killerFactionId": 2,
        "killedName": "Norman Sheppard",
        "killedId": 158146,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:49.000Z",
        "reason": 35,
        "x": 1071.846680,
        "y": 2087.914551,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Fred Woods",
        "killerId": 180194,
        "killerFactionId": 2,
        "killedName": "Bugsy King",
        "killedId": 192424,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:44.000Z",
        "reason": 37,
        "x": 1058.968262,
        "y": 2084.150391,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Agazio Bernardi",
        "killerId": 123241095,
        "killerFactionId": 4,
        "killedName": "Ricky Gonzalez",
        "killedId": 119810,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:41.000Z",
        "reason": 34,
        "x": 2104.723877,
        "y": -2259.230713,
        "z": 13.396226,
        "interior": 0,
        "virtualworld": 0,
        "killerName": "Norman Sheppard",
        "killerId": 158146,
        "killerFactionId": 2,
        "killedName": "Martino Gonzalez",
        "killedId": 135845,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:34.000Z",
        "reason": 24,
        "x": 1058.577148,
        "y": 2093.225586,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Norman Sheppard",
        "killerId": 158146,
        "killerFactionId": 2,
        "killedName": "Cruz Tattaglia",
        "killedId": 117139,
        "killedFactionId": 2,
        "date_killed": "2024-06-28T20:18:29.000Z",
        "reason": 24,
        "x": 1059.500854,
        "y": 2086.244141,
        "z": 10.820312,
        "interior": 0,
        "virtualworld": 10002,
        "killerName": "Cruz Tattaglia",
        "killerId": 117139,
        "killerFactionId": 2,
        "killedName": "Quiss Tattaglia",
        "killedId": 123135580,
        "killedFactionId": 2,
        "date_killed": "2024-06-
```

I worked with a JSON file that provides detailed data from the event, focusing primarily on a comprehensive log of every kill that occurred during the event. The war information includes the war ID, start and end dates, prize for the winning team, and the initiator of the war. The primary section, which logs over 54,000 individual kills, contains detailed entries for each kill. Each entry records the names and IDs of the killer and the killed, their factions, the time and date of the kill, as well as precise location coordinates. Additional metadata includes the reason for the kill (weapon used) and virtual world identifiers. The detailed kill log offers a deep view into player interactions during the war, which I used to extract useful insights through data analysis and data visualization techniques.

Initial Variables in the Kill Logs

- Killer Name
- Killer ID
- Killer Faction ID (2 for Tattaglia, 4 for Cuneo)
- Killed Name
- Killed ID
- Killed Faction ID (2 for Tattaglia, 4 for Cuneo)
- Date-Time Killed
- Reason (ID of weapon used)
- X coordinate
- Y coordinate
- Z coordinate

- Interior ID (0 means the kill occurred in the open world)
- Virtual World ID (0 means the kill occurred in the common world with all other players)

Death Classification

As every row represents a death, we can only differentiate between the types of deaths (regular death to another player, suicide, NPC death). Therefore, we will set a few criteria to classify deaths:

- If the Killer Name is "N/A", the death is either a suicide (death to the environment) or an NPC death (death occurred when the player left the game, NPC state leaves a static player model for 5 minutes if the player quits the game)
- I used the Reason variable to differentiate between environment deaths (suicide) and NPC deaths:
 - o The Reason ID for NPC deaths is always a weapon ID
 - o The Reason ID for suicides is always one of a few environment or splash damage IDs

I used the "subset" function with all weapon IDs to select the NPC deaths, and the negation of those weapon IDs to select the suicide deaths.

```
npc_deaths <- subset(mydata, killedFactionId == 4 & is.na(killerName)
  & reason %in% c(0, 1, 2, 5, 6, 7, 8, 10, 11, 12, 14, 15, 22, 24, 25, 26, 28, 29, 32,
    30, 31, 33, 34, 41, 42))
```

```
suicides <- subset(mydata, killedFactionId == 4 & is.na(killerName)
  & !(reason %in% c(0, 1, 2, 5, 6, 7, 8, 10, 11, 12, 14, 15, 22, 24, 25, 26, 28, 29, 32,
    30, 31, 33, 34, 41, 42)))
```

Data Transformation Process

The dataset was processed using R and Excel.

1. Converted the JSON file to a data frame table in R using the "jsonlite" library
2. Imported and merged a separately calculated "Hours" data frame into the main data frame
3. Cleaned up the formatting and variable names of certain variables
4. Extracted hour, week, day, and month info from the date and time variable using "as.POSIX" and "cut" functions
5. Imported a map of the game world to be used as background for a scatter plot of the coordinates using the "png" library
6. Added the tiers of the killer and killed players using known names of the tier 0 and tier 1 players and a series of "ifelse" statements to work out the rest (tier 2 and tier 3 players are merged into tier 2 as they are treated the same as far as the purposes of this analysis are concerned)
7. Added death fee paid by the killed player for every kill, depending on their tier
 - a. Tier 0: \$150,000 (double if the death is a suicide or an NPC death)
 - b. Tier 1: \$120,000

c. Tier 2: \$100,000

8. Calculated the distance to an important coordinate in the map using the Euclidean distance between two points in a 2D space (using only the X and Y coordinates):

Point A (x1, y1) – Static coordinates on the map of the important location

Point B (x2, y2) – Variable, depends on every death's XY coordinates

$$\text{Euclidean distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

```
mydata$distance <- round((sqrt((mydata$x - 2189.5590)^2 + (mydata$y - 2262.3320)^2)), digits = 0)
```

Frequency Analysis

This was primarily done using filters using the “subset” function to select kills, deaths, suicides, and NPC kills for specific factions or players or using specific weapons or between specific dates, etc.

Those were then transformed into frequency tables by counting instances using the “table” function and sorted using the “sort” function.

a. Basic frequencies

Counting the number of kills, deaths, suicides, and NPC deaths

b. Hourly Activity

Identifying peak periods of activity to better adapt the times daily offensives to the hours with the best performance.

The images below show how I visualized the activity and KDR for every hour of the day. The images show the end results after 6 months of events, but I would do these weekly and monthly or for notable events.

Actual numbers:

CUNEO	Kills			Deaths			Suicides			NPC deaths		
	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3
0	43	101	465	32	79	564	2	7	75	1	1	4
1	11	13	60	3	6	59	1	3	16		1	3
2		8	9	1	1	6						
3			3			7			2			
4			2									
5			10			4			2			
6	1		25	1		29			4			1
7		5	41		2	18			4			1
8	3	11	167		5	116			25			4
9	7	54	377	8	23	213		4	45			1
10	2	62	350		25	209		7	47		1	8
11	6	71	496	3	44	447	1	5	81		1	8
12	12	64	521	6	47	452		6	79			14
13	28	154	664	19	101	698		8	96		1	14
14	51	165	873	23	112	851	3	12	120		4	21
15	29	102	700	17	61	596	3	11	130			14
16	60	197	1248	52	127	1117		21	151		4	34
17	73	276	1791	48	150	1773	1	19	252		8	41
18	61	270	1739	31	158	1624		15	321		1	34
19	51	231	1416	22	148	1273	1	12	208	1	2	40
20	128	622	3031	74	346	3058		46	463		3	53
21	204	556	2933	98	326	2895	3	30	402	2	6	38
22	178	421	1909	89	256	1956	7	29	314	1	2	26
23	120	250	1105	66	147	1221	5	12	191	3	1	14
TOT	1068	3633	19935	593	2164	19186	28	247	3028	8	36	373
		24636			21943			3303			417	

KDR		
Tier 0	Tier 1	Tier 2+3
1.23	1.16	0.72
2.75	1.30	0.77
	8.00	1.50
		0.33
		1.67
1.00		0.74
	2.50	1.78
	2.20	1.15
0.88	2.00	1.46
	1.88	1.33
1.50	1.42	0.93
2.00	1.21	0.96
1.47	1.40	0.82
1.96	1.29	0.88
1.45	1.42	0.95
1.15	1.30	0.96
1.49	1.56	0.87
1.97	1.55	0.88
2.13	1.43	0.93
1.71	1.57	0.85
1.98	1.54	0.88
1.84	1.47	0.83
1.62	1.56	0.77
TOT	1.70	1.48
		0.88

Percentages:

[illegible]

TATTA	Kills			Deaths			Suicides			NPC deaths		
Hour	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3	Tier 0	Tier 1	Tier 2+3
0	3.6%	2.7%	3.1%	2.8%	2.0%	2.5%	3.9%	1.4%	1.8%	4.3%		1.7%
1	0.4%	0.3%	0.3%	0.5%	0.3%	0.3%	0.6%	0.3%	0.1%			0.5%
2	0.1%		0.0%			0.1%	0.6%		0.4%			
3		0.1%	0.0%		0.1%							
4						0.0%			0.0%			
5			0.0%			0.1%			0.1%			
6		0.2%	0.1%		0.1%	0.1%			0.3%			
7	0.1%	0.0%	0.1%	0.1%	0.1%	0.2%			0.5%			0.7%
8	0.7%	0.3%	0.6%	0.4%	0.4%	0.8%	0.6%	0.3%	1.3%			2.2%
9	1.3%	0.4%	1.3%	1.1%	0.3%	2.1%	1.7%	1.4%	2.0%			2.7%
10	0.3%	0.7%	1.4%	0.7%	0.8%	1.9%	0.6%	1.4%	2.1%			2.7%
11	2.8%	2.1%	2.2%	2.4%	1.8%	2.4%	1.7%	3.1%	2.7%	4.3%	6.8%	4.1%
12	2.2%	1.5%	2.6%	2.0%	1.2%	2.7%	4.4%	3.1%	2.8%			3.6%
13	3.7%	3.5%	3.8%	3.6%	3.5%	3.4%	5.0%	3.1%	3.3%	6.4%	2.3%	5.3%
14	4.4%	3.6%	4.8%	4.5%	2.9%	4.6%	4.4%	5.1%	4.4%	8.5%	6.8%	3.4%
15	3.4%	3.2%	2.9%	3.8%	2.8%	3.4%	5.5%	1.7%	2.2%	4.3%	4.5%	4.3%
16	6.0%	6.2%	5.8%	6.8%	6.2%	6.0%	6.1%	8.2%	4.3%	6.4%	6.8%	6.8%
17	8.1%	9.7%	9.0%	8.0%	10.6%	8.5%	9.4%	10.2%	9.1%	8.5%	6.8%	8.0%
18	7.3%	8.7%	8.4%	8.3%	9.0%	8.3%	5.5%	8.8%	10.0%	2.1%	4.5%	8.0%
19	6.6%	7.5%	6.2%	7.1%	7.7%	6.7%	4.4%	10.2%	7.3%	14.9%	6.8%	7.5%
20	16.8%	15.9%	15.6%	15.7%	16.2%	15.2%	12.2%	14.3%	15.3%	14.9%	13.6%	12.3%
21	14.4%	16.9%	14.7%	14.2%	18.6%	14.5%	14.4%	15.6%	13.6%	17.0%	25.0%	10.1%
22	11.2%	10.6%	10.3%	11.2%	10.6%	10.0%	10.5%	8.2%	9.5%	2.1%	13.6%	10.1%
23	6.5%	5.7%	6.8%	6.7%	4.9%	6.1%	8.8%	3.7%	6.9%	6.4%	2.3%	6.0%
TOT	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Although these frequencies are too long-term and wouldn't be ideal to be able to derive any meaningful insights, some recommendations could be:

- For Cuneo to play between 17:00 and 22:00
- For Tattaglia to play between 22:00 and 01:00

c. Sniper Hunters

Sniper hunter kills are the logs that match these filters:

- Killed Tier = 0 or 1
- Killer Tier ≠ 0 or 1 (To exclude sniper vs sniper kills)
- Killer Faction ID ≠ Killed Faction ID (To exclude team kills)
- Killer Name ≠ "N/A" (To exclude suicides and NPC deaths)

Name	Sniper Hunter kills	Faction
Renji_Corzetti	189	Cuneo
Kevin_Salerno	174	Cuneo
Calvin_Dinero	164	Tattaglia
Ralph_Toretto	150	Cuneo
Clayde_Cuneo	146	Cuneo
Lorenzo_Koizumi	131	Cuneo
Tyson_Salerno	121	Cuneo
Bjorgi_Cuneo	120	Cuneo
Brian_Morgan	115	Cuneo
Hakimi_Caruzo	114	Tattaglia
Jack_Kawasaki	112	Cuneo
Neo_Morgan	111	Cuneo
Kaspar_Vitali	109	Cuneo
Vinzy_Corzetti	108	Cuneo
Sharon_Riccio	99	Cuneo
Kyle_Cuneo	96	Cuneo
Ace_Alvarez	93	Cuneo
Kristo_Bucci	91	Cuneo
Bernd_Baustella	91	Tattaglia
Zirga_Galva	91	Tattaglia
Luca_Smoke	83	Cuneo
Kaan_Corzetti	82	Cuneo

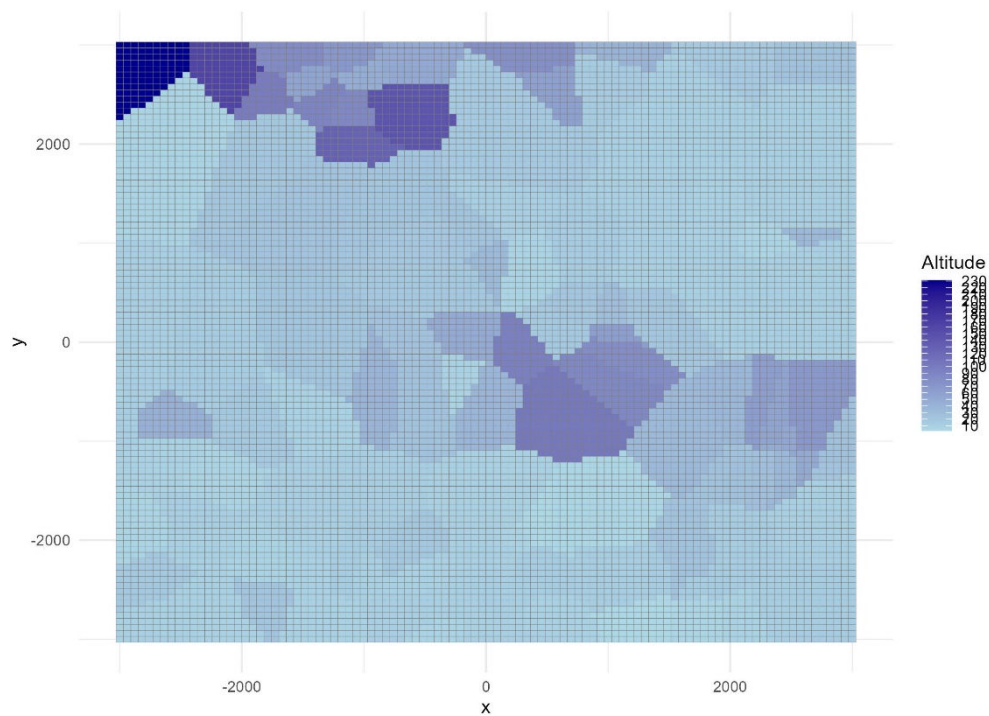
Faction	Total Sniper Hunter Kills	Percentage
Cuneo	3479	69%
Tattaglia	1590	31%
Total	5069	100%

We can see that Cuneo players make up most of the Sniper Hunter leaderboard, and Cuneo in general have more than twice the number of Sniper Hunter kills.

d. Air deaths (deaths while flying, they are almost always registered as suicides)

In order to count the number of air deaths, I had to create a terrain elevation map to determine the Z coordinate value that separates the ground from the air. I added a buffer of 15 units to account for high-rise buildings and for errors in the calculation of the ground level.

Elevation map:



Air deaths table:

Name	Air deaths	Hours	Air deaths per hour
Calvin_Dinero	104	209.61	0.50
Renji_Corzetti	97	303.35	0.32
Kevin_Salerno	76	258.8	0.29
Ralph_Toretto	75	196.78	0.38
Martino_Gonzalez	59	133.3	0.44
Jason_McKamey	52	129.74	0.40
Ernesto_Dinero	52	190.72	0.27
Cruz_Tattaglia	43	334.06	0.13
Adam_Morgan	42	316.13	0.13
Thomas_Dinero	39	222.78	0.18
Clayde_Cuneo	37	169.6	0.22
Yami_Giovacchini	36	84.29	0.43
Kenji_Salerno	36	228.5	0.16
Ace_Alvarez	34	147.82	0.23
Bakar_Koizumi	32	324.06	0.10
Sharon_Riccio	28	190.94	0.15
Agazio_Bernardi	28	211.92	0.13
Sergio_Cacchione	28	238.7	0.12
Lorenzo_Koizumi	28	263.07	0.11
Kaan_Corzetti	27	160.91	0.17
Angel_Dinero	27	215.9	0.13
Elita_Cuneo	25	305.3	0.08
Zacaris_Marcellus	24	459.23	0.05
Bjorgi_Cuneo	24	168.64	0.14
Exor_Elarma	22	67.58	0.33
Lorenzo_Scarpelli	22	147.44	0.15
Pablo_Tattaglia	22	312.33	0.07
Brian_Morgan	21	88.38	0.24
Ron_Tattaglia	20	203.48	0.10
Boris_Bagarella	20	203.71	0.10

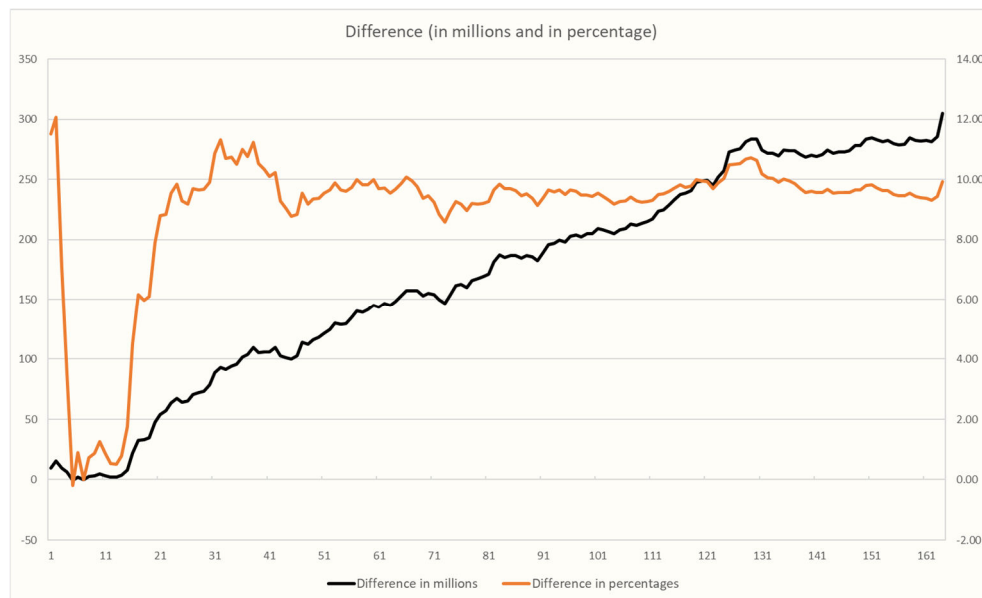
I combined the count of air deaths with the hours data to get a better idea of which players actually die in the air more relative to their time played.

Temporal Trends Analysis

I measured several metrics across time periods (usually days, weeks, or months) in order to identify trends over time.

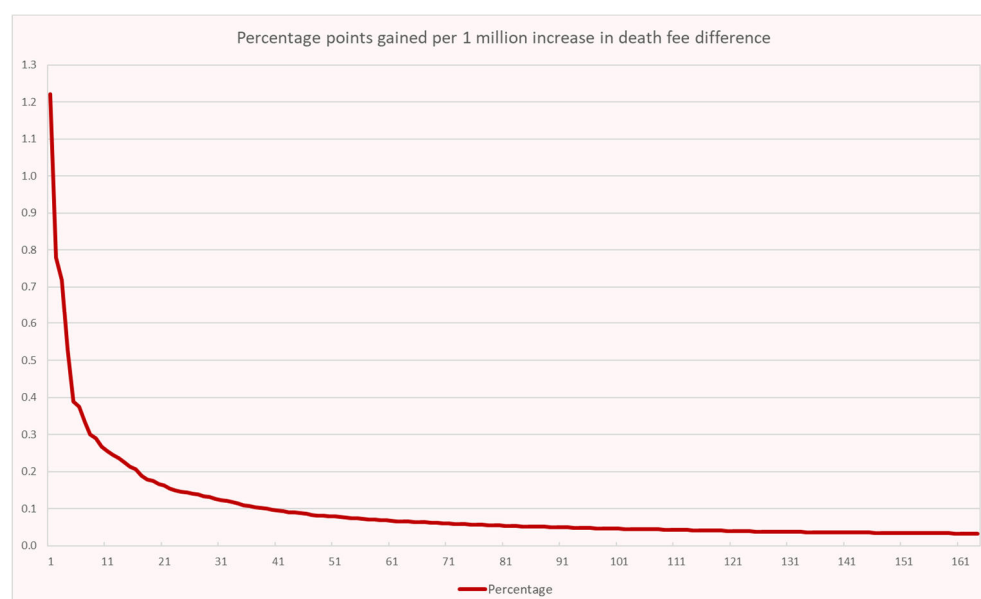
a. Analysis of the Difference in Death Fees

I tracked the difference in death fees for every day of the event. We can see the interplay between the difference in the actual values (dollars) and the percentage difference between the two factions.



We can see that as the difference in millions increases, it becomes much harder to increase the percentage. Therefore, we would advise the factions to try to increase the difference in the early days of the event and avoid accumulating death fees as much as possible while also making sure the other side does.

We can observe this dynamic better in the following graph, where we can see it gets much harder to gain percentage as time goes on.



b. Daily and Weekly Kills, Deaths, Suicides, NPC deaths, and Distance to HQ

A

c. Daily Use of AK and M4 Weapons

Since AKs can be bought using in-game cash and M4s can only be bought using faction-points that require time and effort to make. I tracked their used for both factions to see how they evolved over the course of the event and see what conclusions could be made.

d. Comprehensive Weekly KDSN for All Players

A