

Business Problem

Quick Overview-

Battlegrounds is a player versus player shooter game in which up to one hundred players fight in a battle royale, a type of large-scale last man standing deathmatch where players fight to remain the last alive. Players can choose to enter the match solo, duo, or with a small team of up to four people.

Comprehensive statistical analysis and models for company and players:

- Give company a better insight of player behaviors to make future improvement on game design.
- Cluster and identify outliers in order to ban cheaters.
- Provide key insights for players to strategize to win the game.



Source

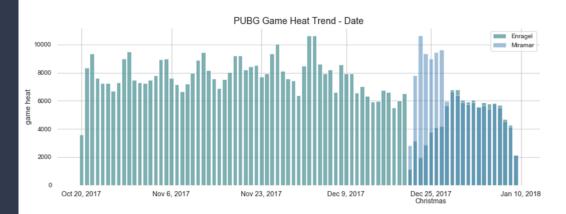
- 19GB PUBG Match dataset & Death dataset (Oct.2017-Jan.2018)
- 10 CSV files in total
- 65 million+ rows in each dataset
- Match: 15 features
 - Date, game size, match id, match type, party size, damage, assist, walk/ride distance, kills, survival time, player name, team id and team placement
- Death: 12 features
 - Match id, map, time, killer name, killer placement, killer position, victim name, victim placement, victim position and weapon
- Use Google Cloud Platform for storage and computing

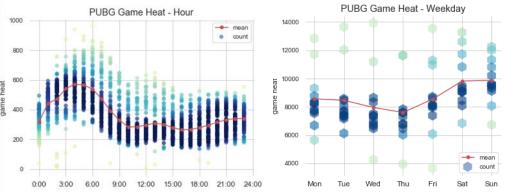


Game Time!

- A huge portion of players are more active at midnight and weekends
- New map launched on 2017-12-21. Great enthusiasm for new map but quickly gone

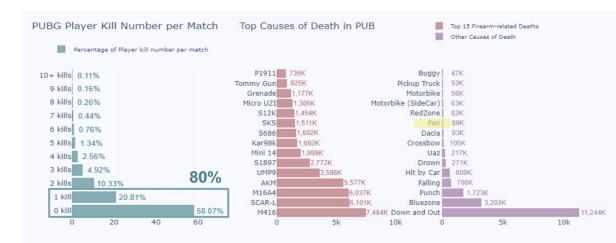


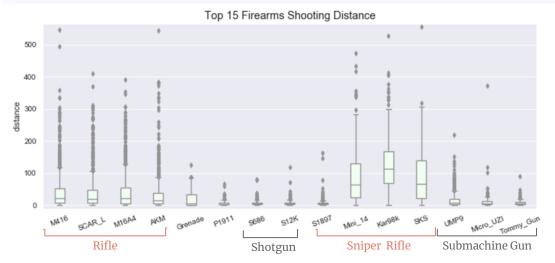




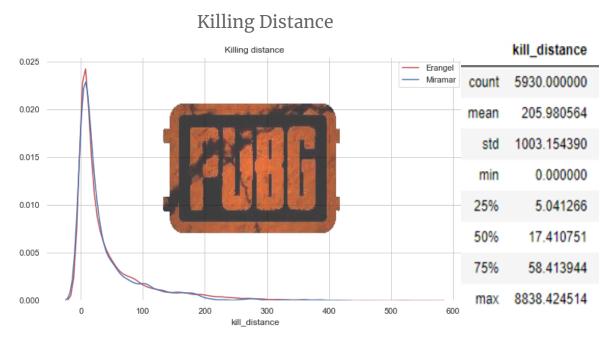
Kill Stats

- 80% of players have no more than 1 kill
- 80% deaths are caused by firearm attacks. A huge proportion (50%) of players are killed by listed rifles and sniper rifles
- Blue zone/ red zone only eliminate 5% of players, the best tactic early on is to avoid everyone and find the right gears





Killing Distance



- 1/10000 sampling to draw distribution
- Units in meter
- Median is only ~17 meters
- No long distance kills
 (sidenote: longest shoot range 650m)

Team Statistics

party	size=1 avg_survive_time	total_kill	total_assists	total_dbno	avg_ride_dist	avg_walk_dist
summary						
count	14039649	14039649	14039649	14039649	14039649	14039649
mean	145792	0	0	0	882	1164
stddev	96089538	1	0	0	1854	5308
min	0	0	0	0	0	0
25%	207	0	0	0	0	151
50%	578	0	0	0	0	688
75%	1178	1	0	0	1105	1819
max	63647875100	94	6	0	442536	1188375
party	size=2					
party	size=2 avg_survive_time	total_kill	total_assists	total_dbno	avg_ride_dist	avg_walk_dist
party summary		total_kill	total_assists	total_dbno	avg_ride_dist	avg_walk_dist
		total_kill	total_assists	total_dbno 10981317	avg_ride_dist	avg_walk_dist
summary	avg_survive_time		_			
summary	avg_survive_time	10981317	10981317	10981317	10981317	10981317
summary count mean	avg_survive_time 10981317 12354	10981317	10981317	10981317	10981317	10981317 1225
summary count mean stddev	10981317 12354 19206974	10981317	10981317 0	10981317 1	10981317 1131 1869	10981317 1225 3054
summary count mean stddev min	10981317 12354 19206974	10981317 1 2 0	10981317 0 0	10981317 1 1	10981317 1131 1869 0	10981317 1225 3054 0
count mean stddev min 25%	10981317 12354 19206974 0 255	10981317 1 2 0	10981317 0 0 0	10981317 1 1 0	10981317 1131 1869 0	10981317 1225 3054 0

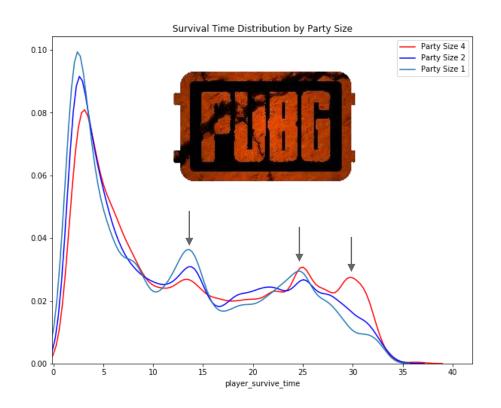
party	party size=4					
	avg_survive_time	total_kill	total_assists	total_dbno	avg_ride_dist	avg_walk_dist
summary						
count	9192801	9192801	9192801	9192801	9192801	9192801
mean	19280	3	1	2	1254	1340
stddev	26530186	3	1	2	1811	2234
min	0	0	0	0	0	0
25%	330	0	0	1	0	326
50%	724	2	0	2	216	1051
75%	1259	4	2	4	2112	2102
max	63648215000	71	28	152	163364	1006273

- Max: hackers (detect hackers!), but more likely bad data
- Squad usually survives longer
- Not many kills
- No assists



Survival Time

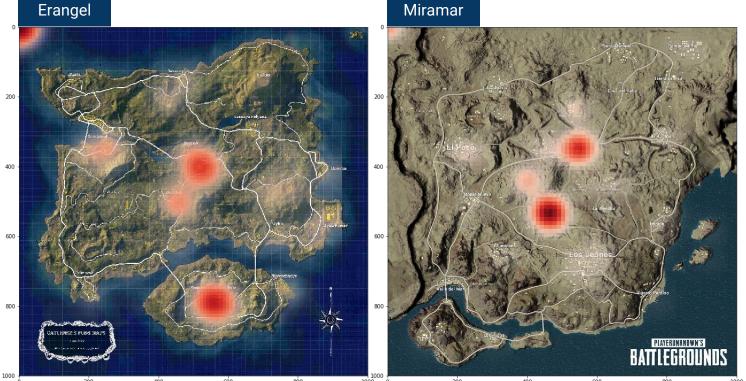
- Solo game is fiercer at the early game than duo and squad
- Peaks corresponding to the start of 2nd, 4th, and final Blue Zone (14 minutes, 25 minutes and 30 minutes)
- Wanna better game experience?
 Don't jump to those places....



Death HeatMap(<4min)



Death Heat Map within 4 min after skydiving

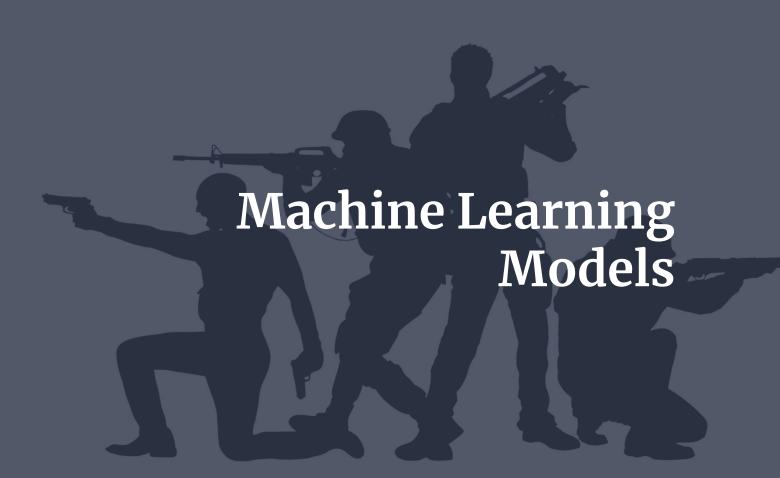


The Final Circle - Safe Zone

A DESCRIPTION OF MANY PARTY.

- Survival heatmap after the last blue zone starts shrinking
- A large portion on the map is never in the final circle
- A few land is more contested for final circle, primarily near very open areas and road intersections





Team Ranking Prediction by Performance

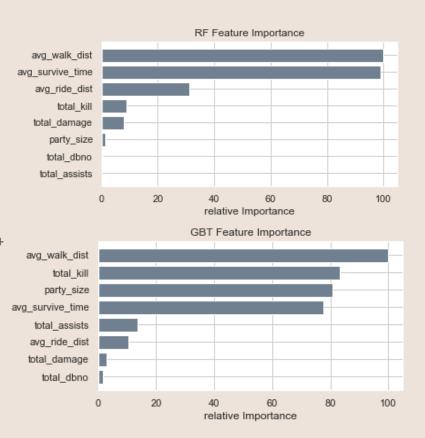
- Target Variable:
 - o Relative Ranking : Team placement / Game size
- Features:
 - o Party Size
 - o Team average survival time
 - Total kills
 - o Total assists
 - Total DBNO(down but not out)
 - o Total Damage
 - o Average walk distance
 - Average ride distance

Models:

- Linear Regression
- Random Forest
- Gradient Boosting Tree(GBT)

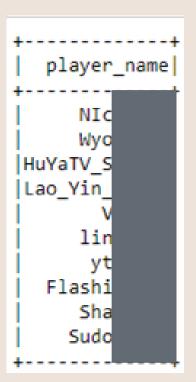
• Linear Regression:

- Linear Regression R2 on training data:
 0.454778
- Linear Regression R2 on test data = 0.458131
- Random Forest Regressor:
 - o 20 trees
 - Random Forest R2 on test data = 0.906374
 - Huge improve
 - Average survival time and average walk distance are the most important features
- Gradient Boosting Trees(GBT):
 - Maxiter 5, num of trees 5
 - \circ GBT R2 on test data = 0.917593
 - o Different feature importance



Hacker Detection - Clustering

- Use K-means to clustering players and try to identify hackers
- Find hackers have abnormal value in all three features: player_kills, player_dist_ride, and player_dist_walk
- Identified:
 - 574 players out of 7.6 million players
- Limitation: unable to distinguish with real talented players and more information required



Network Graph

```
id|degree|
feitengdedan|
               5460
   hzxiaobin|
               5346
   Slh Bunny
               5314
      JZalan
               5132
               4210
   yangyl123
    Sterealz
               4206
   jimmy0572
               4124
    virui887
               4106
BlueChick666
               3966
|Matthew wang|
               3940
```

- Identify Important Players:
 - Degree for each player
 - Send queries to those influencers in order to improve the game experience
- Friends / Teammates Recommendation:
 - Motif subgraph finding
 - Providing recommendation for players to improve game experience and game-player cohesion

a	b	c
[SMer_, babyt2]	[LeLocle0512, LeLocle0202] [babyt2, 333yinuo] [DouYu_GuLang, CNJianChen]	[333yinuo, babyt2]

Recommendation

Tips for Players:

- If you enjoy late-game combat, try to avoid those areas while skydiving
- If you exciting with the early-game fight,
 those areas are top choices
- Try to enter the next safe zone earlier to avoid the high death chances at the end each safe zone shrinkage

Insights for PUBG Corp.:

- Better algorithms to detect hackers in different perspectives such as aim hack, speed hack, and super sniper
- Could recommend possible known players in the squad mode based on the network graph
- Use triangle counting find influencers and send questionnaires to improve the game experience

Future Work



- Adding more features in the clustering model to detect hackers (higher dimensions)
- Get internal data such as players' report of hackers to improve the accuracy of hacker detection
- Larger memory for computing motif subgraph finding and providing recommendations for players while matching
- Develop a model to predict waiting time between two matches

