Pubg Death Data Analysis

By Riddhi Padte and Varun Sriram
GSND 6350

Introduction

- Data set acquired from Kaggle
- Two sets of Data: Aggregate and Death
- We used the Death Data as it has the killers' and victims' placement and the mode of kill.
- Feature Engineering
- We aim to identify the areas where players have a higher chance of killing other players.
- Develop a model which predicts a gun based on the distance and elevation of the victim, player and game time .

PUBG MAP - Erangel

- Urban Structures:Yellow
- White Lines: Highways

Data Cleaning

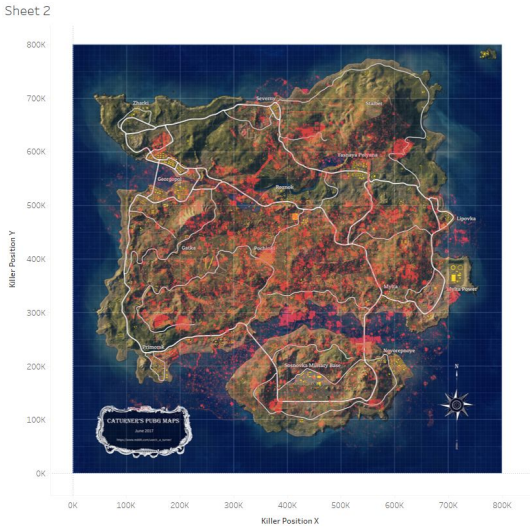
- Original Data set:
  - 12 columns
  - Approx 60,000,000 rows
- Redundant Data:
  - Killer_name
  - Match_id
  - Map
  - Victim_name
  - killer_Placement
  - victim_placement

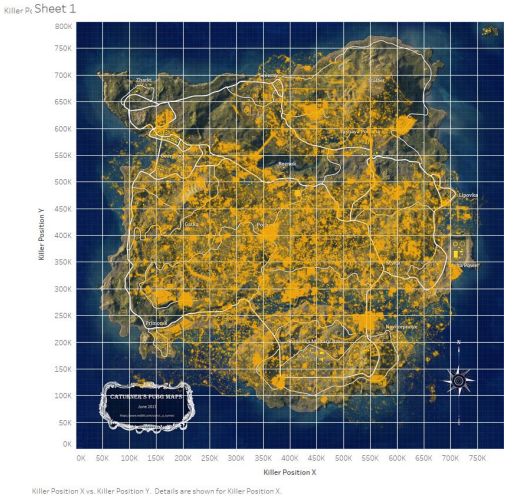
Columns

- killed_by
- killer_name
- killer_placement
- killer_position_x
- killer_position_y
- map
- match_id
- time
- victim_name
- victim_placement
- victim_position_x
- victim_position_y

OUTLIERS

- Outliers could only be present in killer and victim (x,y)
- Upon visualization, there were no points that were going out of map





Data Imputation

- N/A values present in values falling under columns whose corresponding rows were removed:
  - Killed_by (Gun)
  - victim x, y position
  - Killer x, y position
- Many positional values have 0.0 and they were removed as well.
- “Killed_by” columns had some unidentified values and they were removed.
  - death.None
  - death.PlayerMale_A_C
  - death.Buff_FireDOT_C

Feature derivation

- Derived 4 features:
  - Gun Type: Classified guns according to their class
    - Rifle = "Groza", "M16A4", "Mk14", "SCAR-L", "M416", "Mini 14", "AKM", "AUG"
    - MG = "M249", "DP-28"
    - Sniper = "AWM", "Kar98k", "M24", "Crossbow", "SKS"
  - Killer and Victim Elevation:
    - Identified hills and mountains from the map
    - Tagged killer/victim positions as “Elevated” if they were on a hill
    - Killer_position_x > 462500 & killer_position_x < 500000 & killer_position_y > 110000 & killer_position_x < 131500 = "Elevated"
  - Kill Distance: Used distance formula to measure the trajectory of kill
    - sqrt(((killer_position_x - victim_position_x)^2) + ((killer_position_y - victim_position_y)^2)))

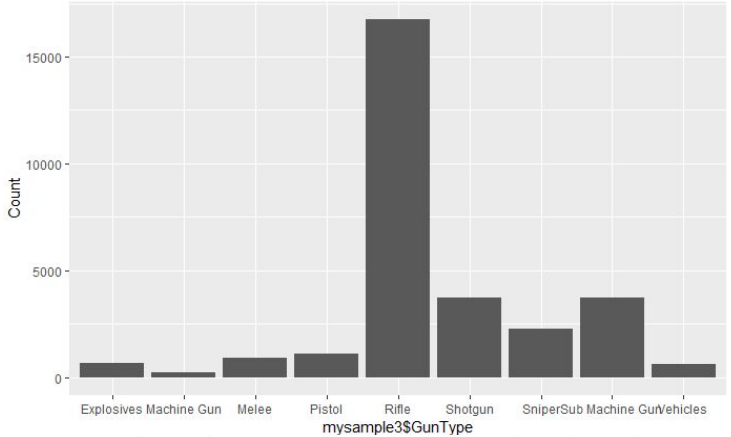
Approach

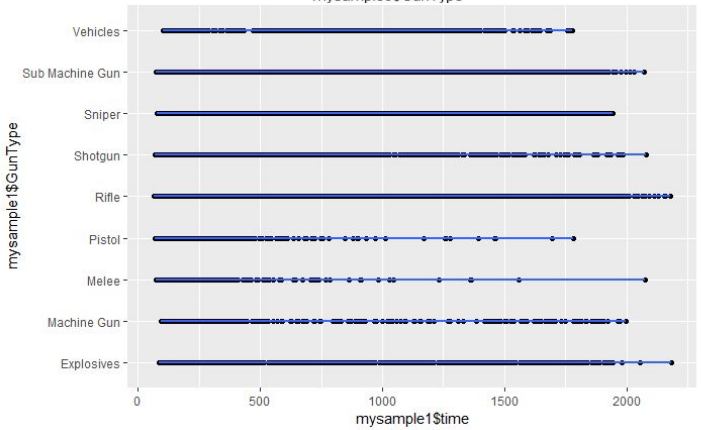
- Clean data set has 39,705,577 rows
- Computer couldn’t handle so many rows
- So we sampled 3 different data sets of 30000 rows each
- Each data set has random values
- All the computations will be performed on the three data sets and compared for validation

Correlation Matrix

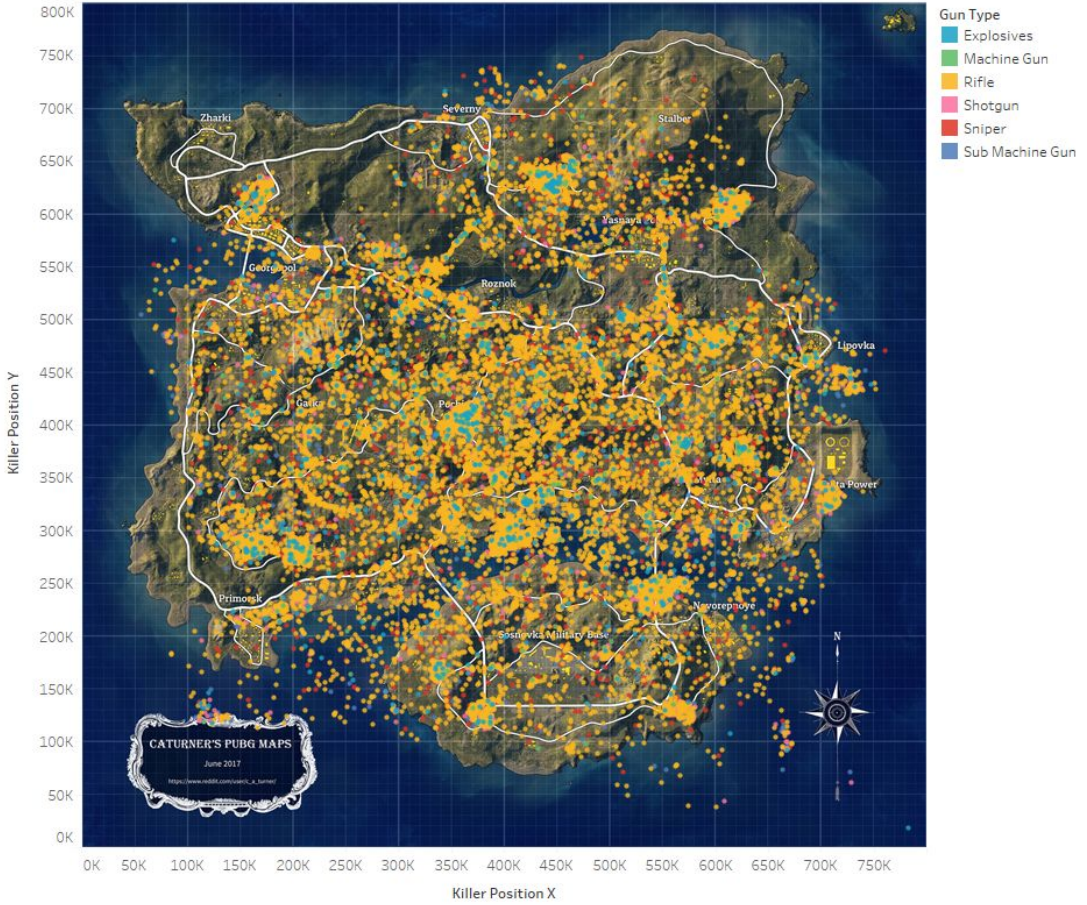
- Used one hot encoding on dummy variables and ran correlation analysis
- Explosives, Rifle and Sniper = **positive correlation** with time
- Rifles, Snipers = **positive correlation** with killer elevation
- SMG **positive correlation** when both, **victim and killer are in normal elevations**

Distributions





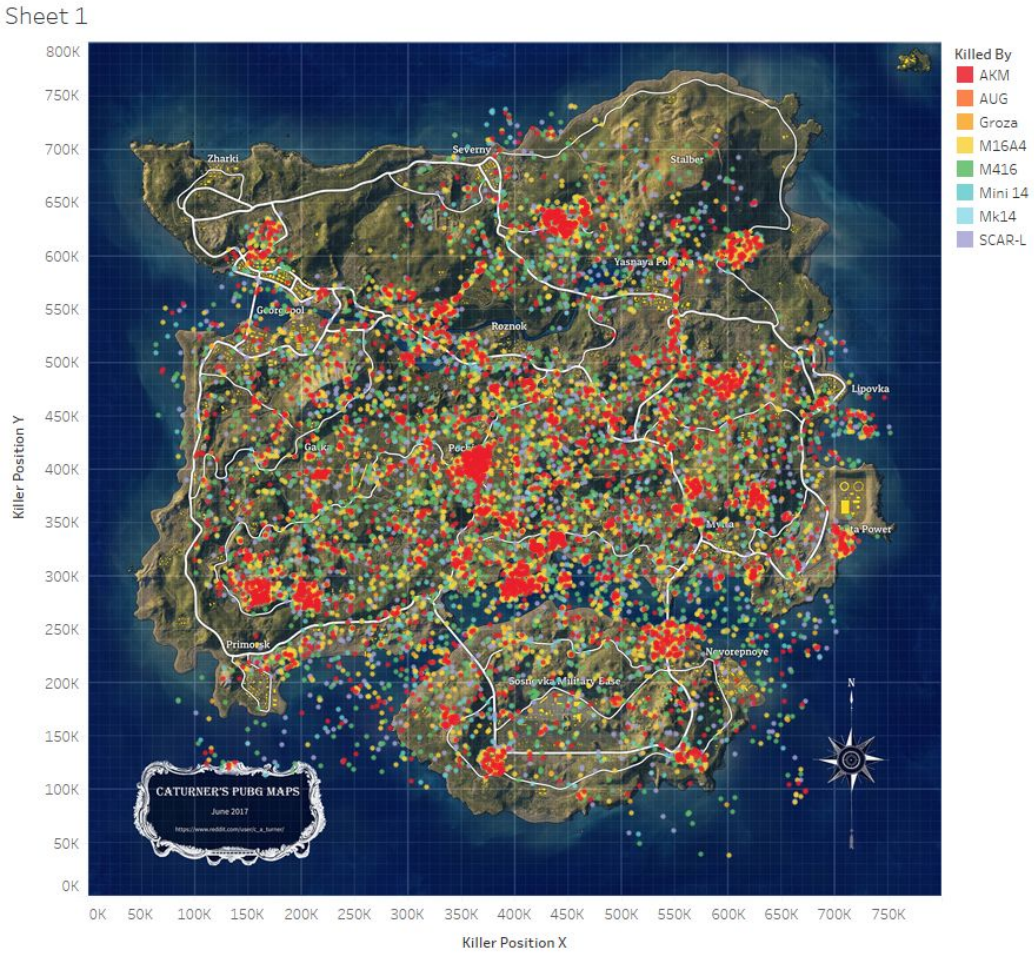
Sheet 1



Killer Position X vs. Killer Position Y. Color shows details about Gun Type. The view is filtered on Gun Type, which excludes Melee, Pistol and Vehicles.

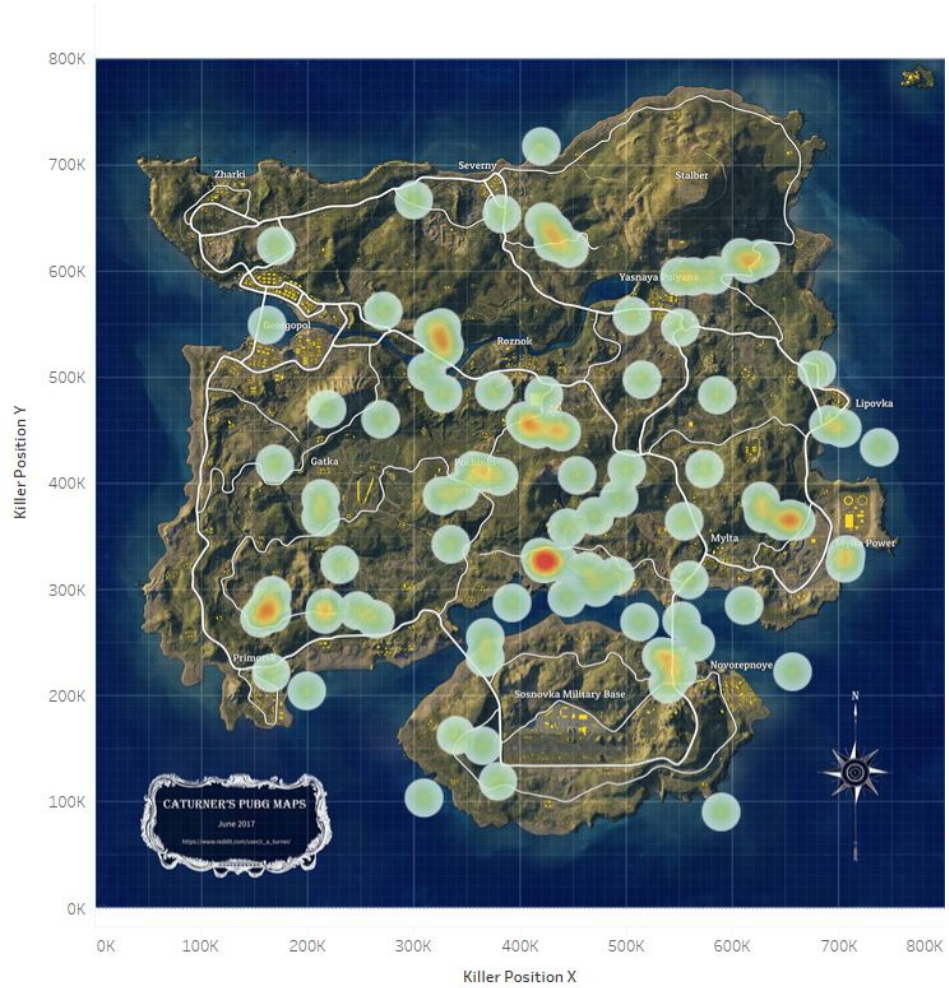
RIFLES

- AKM is used around urban structures.
- AUG and Groza are rarely used.



Killer Position X vs. Killer Position Y. Color shows details about Killed By. The data is filtered on Gun Type, which keeps Rifle, Sniper and Sub Machine Gun. The view is filtered on Killed By, which excludes 10 members.

Sheet 1



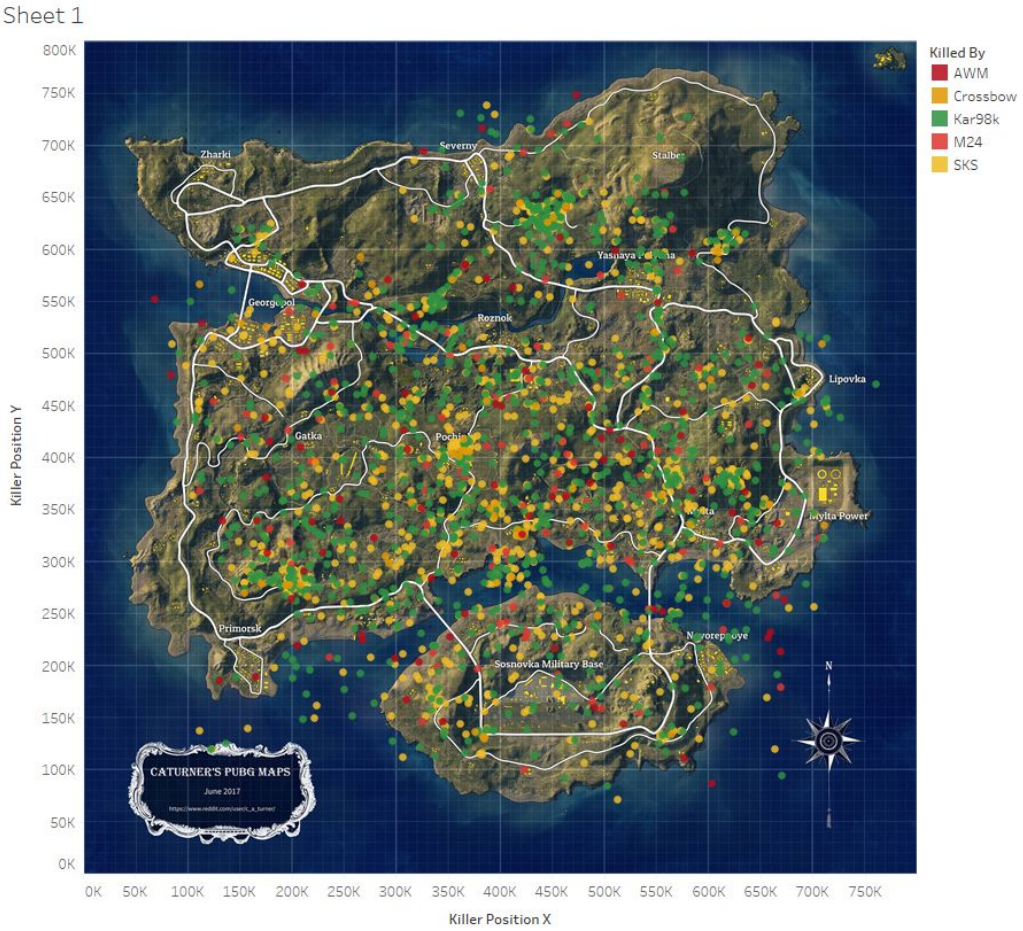
Killer Position X vs. Killer Position Y.

Rifles

- Used near Urban Structures
- Near hills and around the outskirts

SNIPERS

- Most used - Kar98k & SKS
- Crossbows are rarely used.
- AWM is a very rare sniper.



Killer Position X vs. Killer Position Y. Color shows details about Killed By. The data is filtered on Gun Type, which keeps Rifle, Sniper and Sub Machine Gun. The view is filtered on Killed By, which excludes 13 members.

Snipers

- Used at high altitude places
- More scattered than the other type of guns

SUBMACHINE GUNS

- Most used gun - Micro UZI
- Used mostly near structures.

Killer Position X vs. Killer Position Y. Color shows details about Killed By. The view is filtered on Killed By, which keeps Micro UZI, Tommy Gun, UMP9, Vector and VSS.

Sheet 1

Killer Position X vs. Killer Position Y.

Submachine Gun

- Used Near Structures

Data Splitting

- The three data sets were split:
  - Training and Testing
  - 0.75 Training and 0.25 Testing
- Columns used for training and testing
  - We are going to predict the gun type for a scenario
  - Killer position (x,y)
  - Victim position (x,y)
  - Time in the game, since the battlefield reduces in size as time progresses
  - Killer Elevation
  - Victim Elevation

kNN

- Performed kNN classification with k fold cross validation
- Moderate accuracy
- Bad Kappa
- Bottom line, the model is pretty inaccurate and even more unreliable

```
Overall Statistics

               Accuracy : 0.4811
                 95% CI : (0.4697, 0.4924)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0912
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.4849
                 95% CI : (0.4736, 0.4963)
    No Information Rate : 0.5564
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0913
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.4907
                 95% CI : (0.4793, 0.502)
    No Information Rate : 0.5586
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1066
 Mcnemar's Test P-Value : NA
```

Rparts

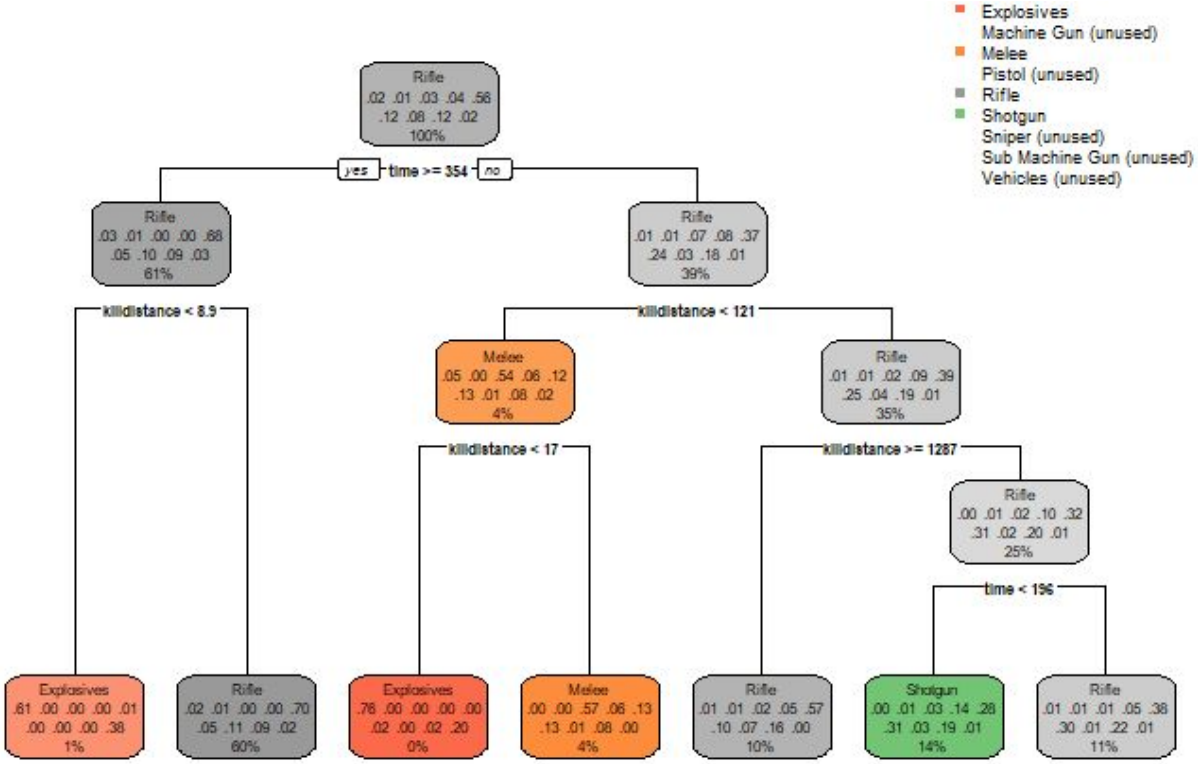
- An accuracy of approx 0.60
- With repeated cross validation
- Moderate Kappa value of approx 0.54
- Overall this model is an pretty good model

```
Overall Statistics

               Accuracy : 0.5902
                 95% CI : (0.5789, 0.6013)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 4.323e-08

                  Kappa : 0.2286
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.5844
                 95% CI : (0.5732, 0.5956)
    No Information Rate : 0.5564
    P-Value [Acc > NIR] : 5.273e-07

                  Kappa : 0.2323
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.5875
                 95% CI : (0.5763, 0.5987)
    No Information Rate : 0.5586
    P-Value [Acc > NIR] : 2.228e-07

                  Kappa : 0.2017
 Mcnemar's Test P-Value : NA
```

Rparts tree


Explosives
Machine Gun (unused)
Melee
Pistol (unused)
Rifle
Shotgun
Sniper (unused)
Sub Machine Gun (unused)
Vehicles (unused)
Rifle
.02 .01 .03 .04 .56
.12 .08 .12 .02
100%
yes
time >= 354
no
Rifle
.03 .01 .00 .00 .68
.05 .10 .09 .03
61%
Rifle
.01 .01 .07 .08 .37
.24 .03 .18 .01
39%
killdistance < 8.9
killdistance < 121
Melee
.05 .00 .54 .06 .12
.13 .01 .08 .02
4%
Rifle
.01 .01 .02 .09 .39
.25 .04 .19 .01
35%
killdistance < 17
killdistance >= 1287
Rifle
.00 .01 .02 .10 .32
.31 .02 .20 .01
25%
time < 196
Explosives
.61 .00 .00 .00 .01
.00 .00 .00 .38
1%
Rifle
.02 .01 .00 .00 .70
.05 .11 .09 .02
60%
Explosives
.76 .00 .00 .00 .00
.02 .00 .02 .20
0%
Melee
.00 .00 .57 .06 .13
.13 .01 .08 .00
4%
Rifle
.01 .01 .02 .05 .57
.10 .07 .16 .00
10%
Shotgun
.00 .01 .03 .14 .28
.31 .03 .19 .01
14%
Rifle
.01 .01 .01 .05 .38
.30 .01 .22 .01
11%

Rule Learner model

- Mediocre Accuracy
- Low Kappa
- Reliability score bit better than kNN

```
Overall Statistics

               Accuracy : 0.5814
                 95% CI : (0.5701, 0.5926)
    No Information Rate : 0.5586
    P-Value [Acc > NIR] : 3.545e-05

                  Kappa : 0.1035
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.584
                 95% CI : (0.5728, 0.5952)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 9.453e-06

                  Kappa : 0.1214
 Mcnemar's Test P-Value : NA
```

```
Overall Statistics

               Accuracy : 0.5815
                 95% CI : (0.5702, 0.5927)
    No Information Rate : 0.5564
    P-Value [Acc > NIR] : 6.289e-06

                  Kappa : 0.1063
 Mcnemar's Test P-Value : NA
```

Rule Learner Model

```
JRIP rules:
===========

(killdistance <= 1.552417) and (time >= 537) and (killer_position_x <= 334228.4) => GunType=Vehicles (51.0/23.0)
(killdistance <= 0) and (time >= 704) and (killer_position_y >= 367525.4) and (time <= 847) => GunType=Vehicles (19.0/6.0)
(killdistance <= 1.389244) and (time >= 914) and (killer_position_y <= 210491.7) => GunType=Vehicles (7.0/1.0)
(killdistance <= 0) => GunType=Explosives (254.0/73.0)
(killdistance <= 98.595588) and (time <= 174) and (killer_position_x <= 354326.3) => GunType=Melee (115.0/21.0)
(killdistance <= 131.158263) and (time <= 128) => GunType=Melee (314.0/108.0)
(killdistance <= 93.407708) and (time <= 284) and (killdistance >= 78.751317) => GunType=Melee (190.0/69.0)
(killdistance <= 115.702031) and (killer_position_x >= 488458.7) and (killer_position_x <= 535550.8) => GunType=Melee (17.0/6.0)
(killdistance <= 128.232016) and (time <= 160) and (killer_position_x >= 432195.8) => GunType=Melee (37.0/18.0)
(killdistance >= 9867.813132) and (time >= 1678) and (killdistance <= 15483.686419) => GunType=Sniper (166.0/81.0)
(killdistance >= 10981.405343) and (killer_position_y >= 322453.1) and (killer_position_y <= 369829.1) and (time >= 1505) => GunType=Sniper (55.0/23.0)
(killdistance <= 1043.960708) and (killer_position_y >= 414874.8) and (victim_position_y <= 627478.2) and (victim_position_y >= 613295.3) and (killdistance >= 242.88699) and
(time <= 138) and (victim_position_x <= 451097.7) => GunType=Shotgun (59.0/18.0)
(time <= 268) and (killdistance <= 685.914936) and (killdistance >= 603.52299) and (killer_position_x >= 420193.6) and (victim_position_y >= 325850.6) => GunType=Shotgun
(111.0/51.0)
 => GunType=Rifle (21107.0/8738.0)

Number of Rules : 14
```

SVM

- Used the radial kernel with repeated cross validation
- Other kernels never converged
- Mediocre accuracy
- 0 Kappa score
- The reliable at all

```
Overall Statistics

               Accuracy : 0.5593
                 95% CI : (0.548, 0.5706)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 0.5141

                  Kappa : -2e-04
 Mcnemar's Test P-Value : NA


Overall Statistics

               Accuracy : 0.5564
                 95% CI : (0.5451, 0.5677)
    No Information Rate : 0.5564
    P-Value [Acc > NIR] : 0.5048

                  Kappa : 0
 Mcnemar's Test P-Value : NA


Overall Statistics

               Accuracy : 0.5586
                 95% CI : (0.5472, 0.5698)
    No Information Rate : 0.5586
    P-Value [Acc > NIR] : 0.5048

                  Kappa : 0
 Mcnemar's Test P-Value : NA
```

Extras

- Using scaled data, we ran supervised learning using kNN, rparts and jrips rule learner respectively
- Using scaled data didn't produce any better results

```
Overall Statistics

               Accuracy : 0.4825
                 95% CI : (0.4712, 0.4939)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0941
 Mcnemar's Test P-Value : NA


Overall Statistics

               Accuracy : 0.584
                 95% CI : (0.5728, 0.5952)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 9.453e-06

                  Kappa : 0.1214
 Mcnemar's Test P-Value : NA


Overall Statistics

               Accuracy : 0.5902
                 95% CI : (0.5789, 0.6013)
    No Information Rate : 0.5595
    P-Value [Acc > NIR] : 4.323e-08

                  Kappa : 0.2286
 Mcnemar's Test P-Value : NA
```

Thank you