

# In-Depth Statistical Analysis and Survivability of PlayerUnknown's Battlegrounds

## Extended Abstract

Group: 02

Phi Trang

B.S. Computer Science, University  
of Colorado Boulder  
phi.trang@colorado.edu

Jacob Tran

B.S. Computer Science, University  
of Colorado Boulder  
jacob.tran@colorado.edu

Neil Nguyen

B.S. Computer Science, University  
of Colorado Boulder  
neil.nguyen@colorado.edu

## ABSTRACT

PlayerUnknown's Battlegrounds is a battle royale game-mode that will put up against 99 other players on a remote island for a winner-takes-all showdown. Players enter the island and try to locate weapons and supplies in a tactical battlefield that forces players into a shrinking playzone.

## CS CONCEPTS

- Data mining, aggregation, plotting, correlation, data analytics



**Figure 1: The official banner for PlayerUnknown's Battlegrounds**

## 1 Introduction and Problem Statement

A complex problem that players run into before every game is where to eject from the plane and which site to land. Our focus in this project will be to optimize player choices within the early, middle, and late game. Early game decisions such as areas to drop will greatly increase a player's ability to survive towards mid game. By analyzing the areas where players die the most, we can see areas to avoid, allowing players a higher probability of making it past the 40 players left mark. We will look at distance traveled to see if travelling more often will give a player a higher chance of surviving. This can affect player behavior during the middle and end of the game. If travelling decreases a players chance of surviving, we can draw the conclusion that "camping" can be an efficient method to win the game. We can also do analysis of weapons

to help players choose an arsenal that best fits their needs. We will aggregate weapon and death data to determine which weapons are most efficient at long range. We can also see which weapons have the most kills at the end of the game, so that players can determine what weapon they should have to win the battle royale.

These are all important questions and aspects to consider because they can explain to a player how to better their skills and increase their chance of surviving

## 2 Related Work

- “PUBG Analytics: Final Circle”

<https://www.kaggle.com/skihikingkevin/final-circle-heatmap>

This data mining project plotted a heat map of the frequency of where the final circle appears.

- “PUBG Data Analysis”

<https://www.kaggle.com/chegeerland/pubg-data-analysis>

This project explored where most people jump out in PUBG. Findings indicate that most players jump near military base, school, prison, and the northern cities.

- “PUBG 50 Million Death Map”

<https://public.tableau.com/profile/stewart.uyeda#! /vizhome/pubg50mildeathmap/pubg50mildeathmap>

This project explored where the deaths occurred during the game. They plotted a interactive Tableau heat map that steps through game time and illustrate the locations of deaths.

## 3 Proposed Work

First we will search for data containing null entries where there should be an entry and remove such instances. Then, we will split entries between the two different maps, because the two different maps have two different sets of weapons, and the location data is map specific. We can look at a lot different events, and so we can separate the data into the appropriate forms.

Our work will involve data manipulation in order to separate the large data set into relevant subsets. We will calculate the effectiveness of weapons using statistical analysis methods discussed in class. We will use players' location data to find the “hot” areas of the map and to find the distance between the players.

For drawing a conclusion on weapon effectiveness, we will look at entries from solo matches, in order to remove uncertainty in the data. This is because the game allows players to get back up after getting “knocked out” in games with teammates, and the cause of death will not reflect what knocked them down in the first place. Thus, we will be uncertain of a certain weapon’s actual effectiveness in the game. Therefore, we will only look for games in which the players do not have duplicate placements when data mining for weapon effectiveness.

We will use values such as killer\_position\_x, killer\_position\_y, and victim\_position\_x, victim\_position\_y in order to create heat maps of “hot” areas on each of the maps. We can also use these values as a method of gauging the effective range for each weapon type. We will publish heatmaps and player locations overlayed over the map of the play area.

Our work will use Jupyter Notebook in conjunction with the Pandas and Numpy

modules to reveal such things about the matches in Player Unknown's Battlegrounds.

After data cleaning and plotting, we will use the visuals and calculations to infer some key points from our data. For example, from our data, we can mine that snipers are more likely to kill for long distances. We can mine information such as the best weapon to use if a player wants to accrue the most kills, areas to avoid with high victim placements, areas to land to avoid killer positions, and much more.

## 4 Data Set

The original dataset:

<https://www.kaggle.com/skihikingkevin/pubg-mat-ch-deaths/data>

This csv / xlsx document contains the first 1 million entries from the Kaggle PUBG death dataset:

<https://github.com/jatr6994/csci4502/blob/master/PUBG%20Death%201million.xlsx>

The dataset contains information regarding:

**match\_id:** nominal string

**map:** ordinal string

**time:** interval integer

**killed\_by:** nominal string

**killer\_name:** nominal string

**killer\_placement:** ordinal int

**killer\_position\_x:** interval Int From 0 to 800000

**killer\_position\_y:** interval Int

**victim\_name:** nominal string

**victim\_placement:** ordinal int

**victim\_position\_x:** interval int

**victim\_position\_y:** interval int

## 5 Evaluation Methods

- Finding the best weapons for use early game
- Average range of kills with each gun
- Heat maps of kills vs deaths?
- Average heatmap of casualties for each map
- Common weapons for players in top 10

### K - Means Clustering

By using k-means to cluster the deaths on each map, we can see where the "regions" of each map is, as well as the average death location on each "region". We see that on Erangel, we have a Georgopol region, a Rozhok Northern region, an Eastern region, a Pochinki region, and a Military Island region. It is common for players to stay in their respective regions for the first half of the match as they are typically visiting nearby towns and villages to look for loot. The smaller a K-means cluster region is, the more data points are in that region. Therefore, the regions with a smaller area are hot regions and therefore you are more likely to find other players in such regions. The Military Island and the Pochinki regions on the map on Erangel are the smallest in area and therefore they are the areas players are dying in the most.

On Miramar, we see a Pecado region, and four other regions corresponding to NorthWest, NorthEast, SouthWest, and SouthEast. This is because Pecado is a very popular area for players as it is central in the map and therefore a lot of players are dying in that area. In the rest of the regions, players are less likely to find enemy players than in the Pecado region. So a player trying to traverse the map should probably avoid the pecado region as a whole unless they are looking to find other players on the way.

The final kill clusters show us where the final playzone should more or less be. We see that the regions are comparable to the death clusters because the final circle is always in a random area and the final kill may be anywhere. However, we see that the final kill is likely to be on land, so the final circle has a tendency to land further in land than bordering a body of water. A clever player could use this information to guess where their next playzone may be and therefore be able to be a couple steps ahead of their opponents.

## 6 Tools

We used:

- Anaconda/Python
  - High-level programming language focused on statistical analysis. We will be using this tool for most of our calculations.
- Jupyter Notebook
  - Open-source web application that allows us to create and share live code, equations and visualization of data. Most of our presentation will be done through a Jupyter Notebook. Our code, explanations, and plots can all be shown through a notebook.
- Tableau
  - Will be used to also visualize our data if necessary. Tableau is an easy to use drag and drop application used to create visualizations.
- Matlab Plot
  - Allow us to give visuals of our data. We will import Matlab Plot into our Jupyter Notebook and use it to plot important aspects of our data frame.

- Pandas

- Used to read in our dataset and store it into a dataframe. Allows for easy data cleaning and processing. We will use Pandas' functions to process and clean our dataset. Pandas provides us with functions that easily remove NaN values as well as outliers or values that just don't make sense.

- Numpy

- Fundamental python library used for scientific computing. It contains tools that are useful for algebra and arrays.

- Scipy

- Python-based software for mathematics, science and engineering.

- Other Python libraries as needed

## 7 Milestones to Do

- Analyze deaths according to position dropped. Analyze survival rates depending on position dropped.
- Statistical analysis on most efficient weapons, distance of bullets travelled, most popular guns.
- Consider using data from the "Aggregate" dataset which can be found on Kaggle to elaborate more on our findings.
- Calculate correlation between player distance travelled and player placement/rank.
- Calculate correlation between player walking distance and player placement/rank.
- Calculate the correlation between number of assists, dmg, and kills.

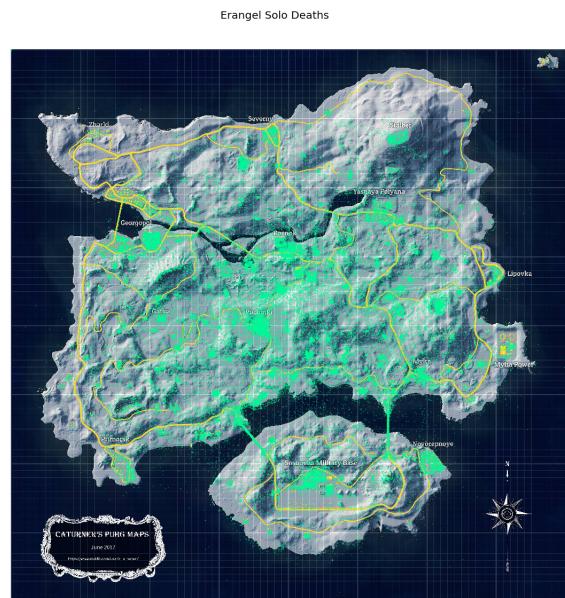
- Mine data from our plots and calculations to create an overall guide at the end of our Jupyter Notebook.
- More to come

## 8 Milestones Completed

- Clean and process data based on solo games
- Clean and process data based on maps
- Clean and process weapon data to remove null positions
- Cleaned outliers from victim and killer positions (i.e. a punch with a distance of 2000 units)
- Got data on most popular guns
- Got weapon data for each map
- Got weapon data for average distance killed for each weapon.
- Plotted and overlay of victim positions over the respective maps.
- Plotted the frequency of each weapon used on a victim on respective maps
- Plotted the average range of each weapon use on a victim on respective maps.
- Statistical analysis on most efficient weapons, distance of bullets travelled, most popular guns.
- Tips and tricks for players on where to go on each specific map and which weapons to use

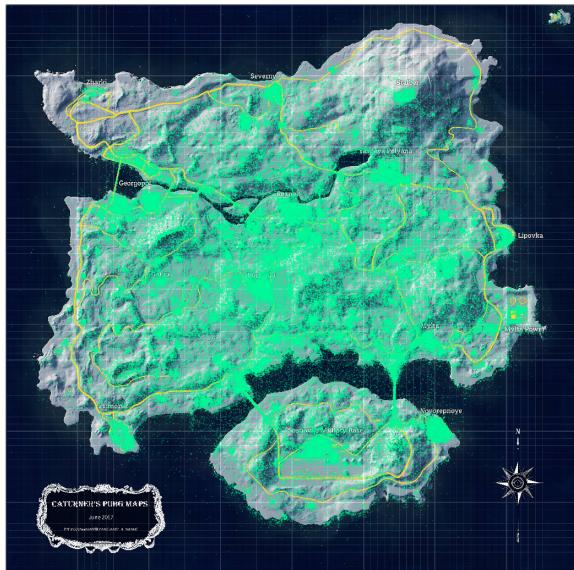
## 9 Results

- Heat map using a scatter plot to show where the most common deaths occur for the Erangel map for solo matches.



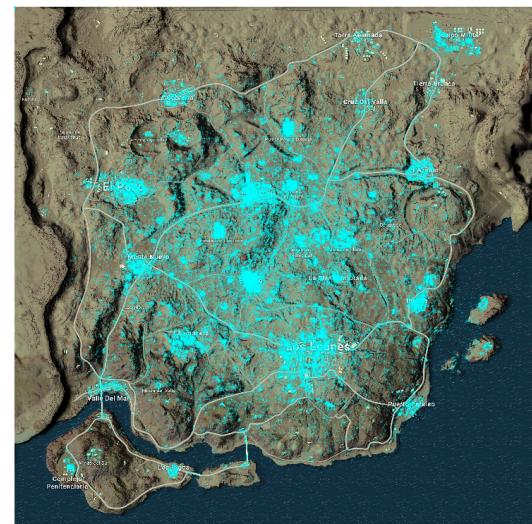
- Heat map using a scatter plot to show where the most common deaths occur for the Erangel map for all matches in dataframe

Erangel Deaths (Not just solo)



- Heat map using a scatter plot to show where the most common deaths occur for the Miramar map for all matches in dataframe

Miramar Deaths (Not just solo)

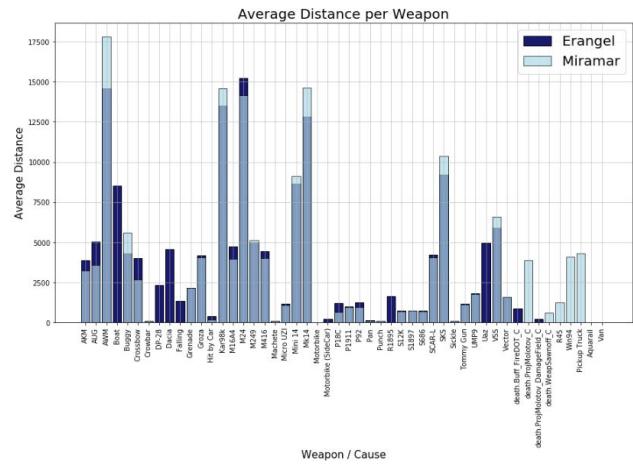
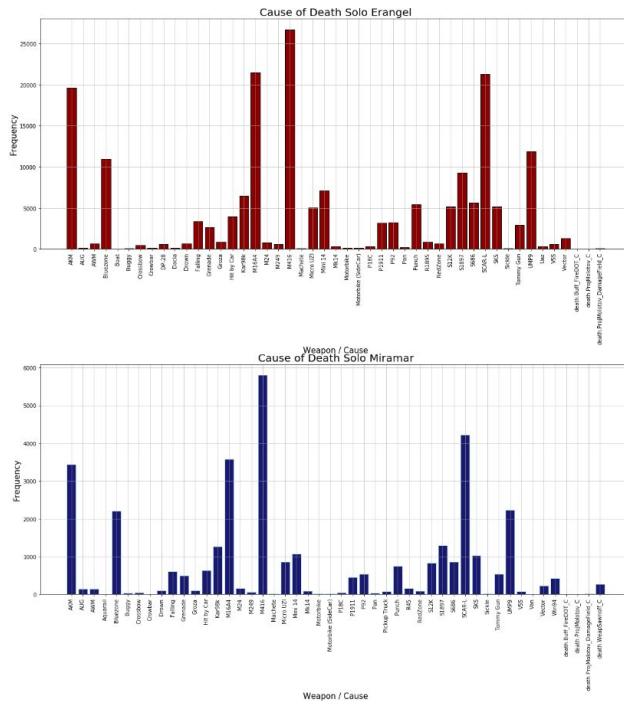


- Heat map using a scatter plot to show where the most common deaths occur for the Miramar map for solo matches.

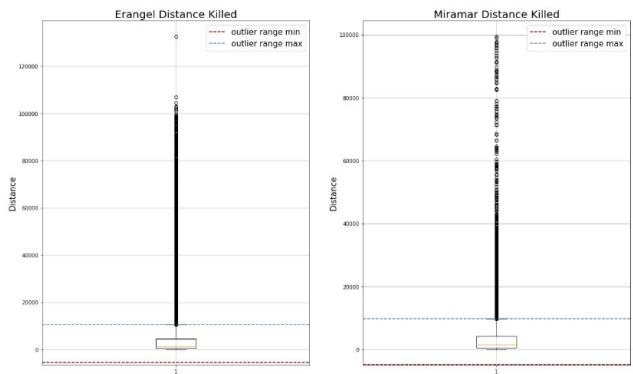
Miramar Solo Deaths



- We used the panda dataframe to plot the most popular weapons in each map. Popular weapons are based on how many kills a particular weapon accrues.

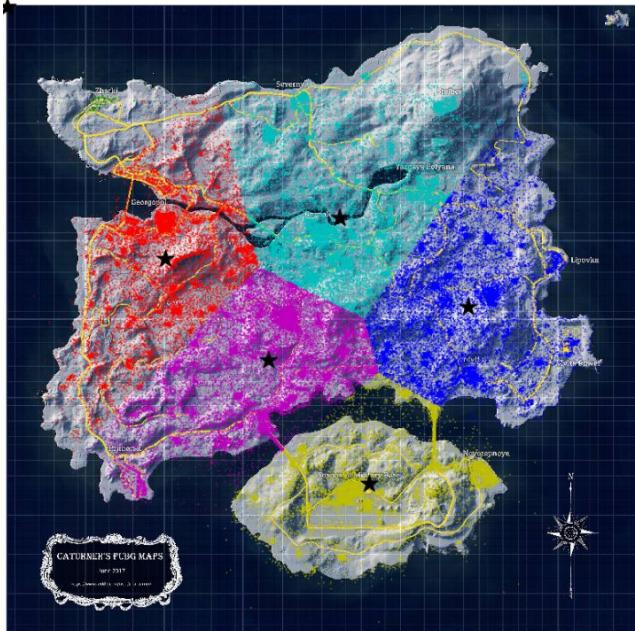


- We also calculated the average distance each weapon kills a victim. We did this by calculating distance based on the given killer and victim's x and y positions. We plotted both graphs on top of each other to see which weapons kill with which range. From an initial glance, it is obvious that weapons falling under the sniper class have the longest average range.



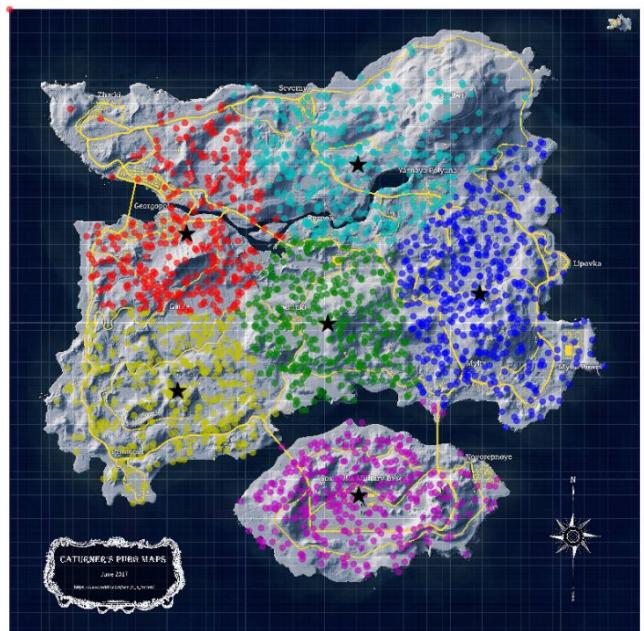
- We clustered the 5 most common death spots for Erangle by using kmeans.

Solo Erangel Death Cluster



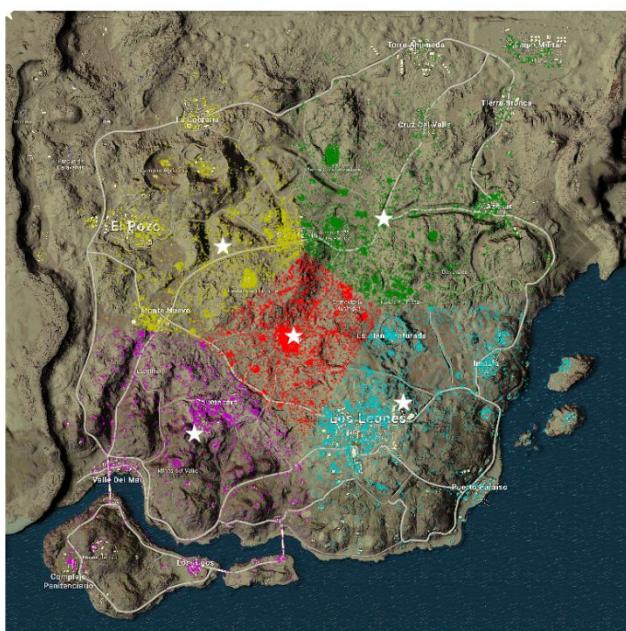
- We clustered the 5 most common final circle locations Erangel by using kmeans.

Solo Erangel Final Kill Cluster

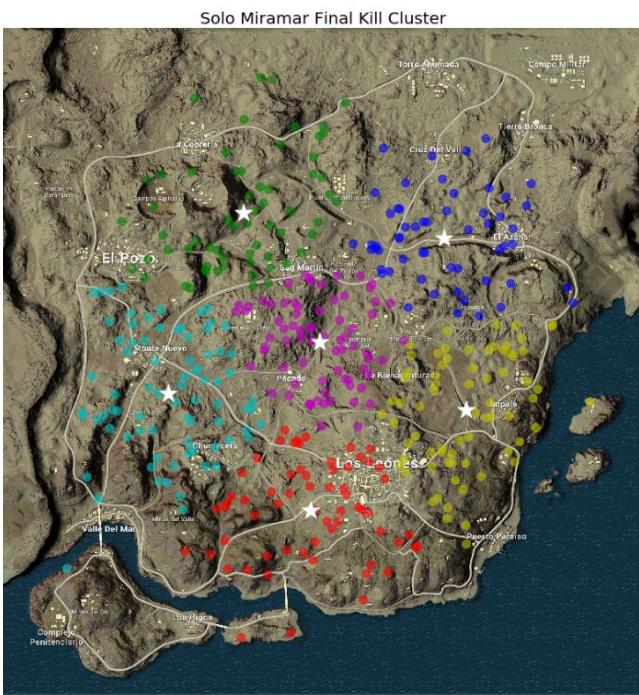


- We clustered the 5 most common death spots for Miramar by using kmeans.

Solo Miramar Death Cluster



- We clustered the 5 most common final circle locations Miramar by using kmeans.



- We included a “Tips and Tricks” as an interpretation of the results of our findings.
- We commented on weapon ranges, weapon use frequency, and interesting finds
- We included suggestions on where to avoid if a player is not looking to run into other players, and where a player should go if he is looking for find enemies.
- We have the final kills mapped so players can see where the final play zones are likely to be.
- Using the “Tips and Tricks” section of the notebook will provide players with a competitive edge over other players, as they will know the weapons to use for each map, the effective ranges for each weapon, and the hot zones as well as an idea of where the next play zone will be.
- We found interesting results such as the engagement distance is on average longer on Miramar than on Erangel. This

may be because of the high density grass on Erangel making it difficult to fight players at longer distances, while the desert-like conditions of Miramar allows players to see and fight at longer distances.

- We found that the AUG, while rare, has a longer average kill distance than even the M16A4, which is known amongst players as the best mid-range rifle for longer ranges.
- We found that the MK14 has a similar effective kill range to that of the Kar98K, although the two guns have a completely different feel, one being semi-automatic and the other being bolt-action.
- We can see where ridges and geographic points of interest are, as players are more likely to be using those spots to hide in an engagement. A clever player can use this to plan ahead and guess where another player may be hiding

## 10 Application and Key Results

### Tips and Tricks

#### Maps:

The heatmaps showing the locations of deaths on both maps show us where players are likely to engage their competitors. The heat maps show where the map is the hottest, so players looking to find some quick action should look to land or go to areas with opaque marks on the heatmaps. Players looking to play the long-con should avoid those hot spots and land in the small towns and villages that are less “hot”. By zooming in on the maps, we can see where in the cities a player is more likely to die, such as the shipment yard in South Geopol. A player looking to increase his chances of survival should avoid such an area, and a player looking to

score some kills should investigate those areas for players. A clever player should have the heatmaps memorized or readily available so as to increase their chances scoring kills, and decrease their chances of running into enemies.

## Weapons:

There is a lot of information to be obtained from the weapons data. We can see that on both maps, the M416 is king. However, when the M416 is not an option, it seems as the M16A4 is just as effective as the SCAR-L on Erangel, while the SCAR-L is a stronger killer on Miramar.

Another interesting find is that, in the shotguns category, the S1897 is more effective than the S686, and both of which are more effective than the S12K. This is despite the S12K being an automatic shotgun. This may be because the S12K is less common in the early-game, which is when shotguns are the most effective as players are still scrambling to find their weapons. However, a player with the choice of the double barrel shotgun and the pump-action shotgun should always choose the pump-action.

In terms of the long-range sniper rifles, we see that the Mini-14 scores kills more frequently than the Kar98K on Erangel, but not on Miramar. So on Erangel, it may be the wiser choice to use the Mini-14 than the Kar98K, while the opposite being true on Miramar.

This also coincides with the rarity/commonality of the weapons. We see that weapons only obtainable via air-dropped crates have very few kills.

## Weapons Distance:

Looking at the average killing distance of the weapons, we can see the effective range of each of the guns and the most likely engagement distance.

For the longer-range weapons, such as the AWM, Kar98K and the like, Miramar typically has a longer engagement distance. Therefore, we know that if a player is on the Miramar map, they are more likely to be engaging at a slightly longer distance than when playing on Erangel. A player with this knowledge should choose weapons that are more effective at further ranges on Miramar. This also makes long-range sights and scopes more valuable on Miramar than on Erangel.

An interesting fact to note from this comparison is the effective range of each of the guns. We see that the AWM is the best for engaging enemies at extreme distances, followed by the M24. One thing to note is that the Kar98K is has a similar average kill distance as the MK14. However, the Kar98K is a bolt-action rifle, and therefore has a firing rate much slower than the MK14. If a player is given a choice between the Kar98K and the MK14, it is perhaps a wise choice to choose the MK14 because it can push more bullets out faster while still being effective at the same range.

For the common assault rifles, we see that the M16A4 is the most effective at longer ranges, followed by the M416, and then the SCAR-L, and lastly the AKM. Therefore, the AKM is most effective battle-rifle for closer ranges and urban combat, while the M16A4 is well equipped to handle engagements at longer ranges. The AUG, though extremely rare, is even stronger at longer ranges than the M16A4, so a player with this rifle can typically trust it to be effective at longer distances than the other assault rifles.

Knowledge of the weapons' average distance for kills gives players a competitive edge. When the playing area is small, a player should probably switch to a closer range weapon that has a shorter average kill distance.

## **ACKNOWLEDGMENTS**

This work is overseen and supported by Liz Boese, Prof. University of Colorado Boulder.

## **REFERENCES**

NK, Mubaris. "K-Means Clustering in Python."  
Mubaris' Blog, 1 Oct. 2017,  
[mubaris.com/2017/10/01/kmeans-clustering-in-p  
ython/](http://mubaris.com/2017/10/01/kmeans-clustering-in-python/).