Data Analysis on PUBG Data with Excel,MySQL,R,Tableau

Players Unknown Battleground(PUBG) has taken the gaming world by storm , offering an Immersive battle royale experience . Beyond its gaming aspect , PUBG also provides a wealth Of data that can be mined and analyzed to gain insights into player behaviour , strategies , And performance.

Dataset Description:

Provide an overview of the PUBG dataset, including the types of data it contains (Player name , Player Kills , Player damage , Player Survival Time , etc.).

Data Preparation:

1. **Data Loading:**
    - Excel: Import the PUBG dataset into Excel.
    - MySQL: Load the dataset into a MySQL database.
    - R: Use R to read and store the dataset.
2. **Data Cleaning:**
    - Handle missing values, outliers, and any inconsistencies in the dataset.

Exploratory Data Analysis (EDA):

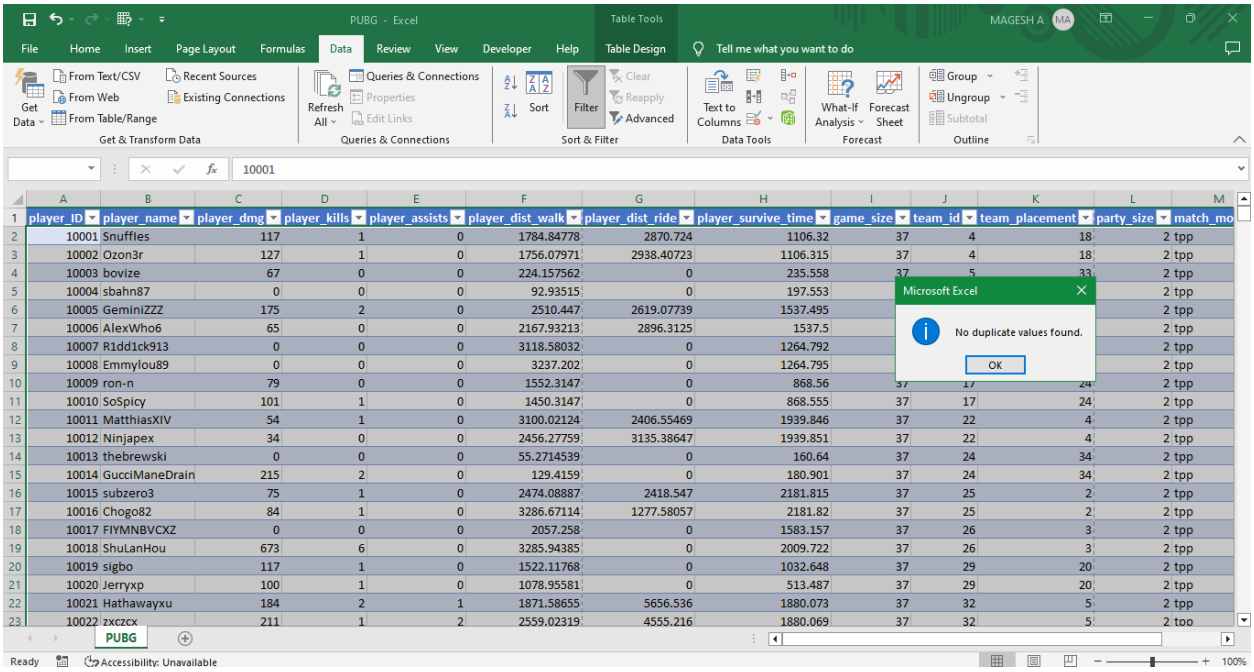
1. **Descriptive Statistics:**
    - Excel: Utilize Excel functions for basic statistics (mean, median, etc.).
    - MySQL: Run SQL queries to obtain summary statistics.
    - R: Use R functions for descriptive statistics.
2. **Visualization:**
    - Tableau: Import data and create interactive visualizations for deeper insights.

Data Cleaning

After uploading the CSV file to the Microsoft Excel and successfully done the Data cleaning process through Data Validation , Removing Duplicates , Removing null values

,etc.

Dataset -> [click here]



player_ID	player_name	player_dmg	player_kills	player_assists	player_dist_walk	player_dist_ride	player_survive_time	game_size	team_id	team_placement	party_size	match_mo
10001	Snuffles	117	1	0	1784.84778	2870.724	1106.32	37	4	18	2	tpp
10002	Ozon3r	127	1	0	1756.07971	2938.40723	1106.315	37	4	18	2	tpp
10003	bovize	67	0	0	224.157562	0	235.558	37	5	33	2	tpp
10004	sbahn87	0	0	0	92.93515	0	197.553	37	17	24	2	tpp
10005	GeminiZZZ	175	2	0	2510.447	2619.07739	1537.495	37	17	24	2	tpp
10006	AlexWho6	65	0	0	2167.93213	2896.3125	1537.5	37	17	24	2	tpp
10007	R1dd1ck913	0	0	0	3118.58032	0	1264.792	37	17	24	2	tpp
10008	Emmylou89	0	0	0	3237.202	0	1264.795	37	17	24	2	tpp
10009	ron-n	79	0	0	1552.3147	0	868.56	37	17	24	2	tpp
10010	SoSpicy	101	1	0	1450.3147	0	868.555	37	17	24	2	tpp
10011	MatthiasXIV	54	1	0	3100.02124	2406.55469	1939.846	37	22	4	2	tpp
10012	Ninjapex	34	0	0	2456.27759	3135.38647	1939.851	37	22	4	2	tpp
10013	thebrewski	0	0	0	55.2714539	0	160.64	37	24	34	2	tpp
10014	GucciManeDrain	215	2	0	129.4159	0	180.901	37	24	34	2	tpp
10015	subzero3	75	1	0	2474.08887	2418.547	2181.815	37	25	2	2	tpp
10016	Chogo82	84	1	0	3286.67114	1277.58057	2181.82	37	25	2	2	tpp
10017	FIYMNBVCXZ	0	0	0	2057.258	0	1583.157	37	26	3	2	tpp
10018	ShuLanHou	673	6	0	3285.94385	0	2009.722	37	26	3	2	tpp
10019	sigbo	117	1	0	1522.11768	0	1032.648	37	29	20	2	tpp
10020	Jerryxp	100	1	0	1078.95581	0	513.487	37	29	20	2	tpp
10021	Hathawayxu	184	2	1	1871.58655	5656.536	1880.073	37	32	5	2	tpp
10022	zxczcx	211	1	2	2559.02319	4555.216	1880.069	37	32	5	2	tpp

Analyzing using R

Installing Packages

```
install.packages("dplyr")
install.packages("tidyverse")
install.packages("tidyr")
install.packages("ggplot2")
install.packages("rmarkdown")
install.packages('tinytex')
library(dplyr)
library(tidyverse)
library(tidyr)
library(ggplot2)
library(rmarkdown)
library(tinytex)
```

Uploadng CSV File

```
PUBG <- read.csv("C:/Users/jansi pc/Desktop/pubg project/PUBG.csv")
```

Data Preprocessing

head()

```
head() is used to view first 6 rows of the dataset
```

```
head(PUBG)
```

```
##   player_ID player_name player_dmg player_kills player_assists player_dist_walk
## 1     10001    SnuffIes        117            1              0       1784.84778
## 2     10002      Ozon3r        127            1              0       1756.07971
## 3     10003      bovize         67            0              0        224.15756
## 4     10004     sbahn87          0            0              0         92.93515
## 5     10005   GeminiZZZ        175            2              0       2510.44700
## 6     10006    AlexWho6         65            0              0       2167.93213
##   player_dist_ride player_survive_time game_size team_id team_placement
## 1         2870.724            1106.320        37       4              18
## 2         2938.407            1106.315        37       4              18
## 3            0.000             235.558        37       5              33
## 4            0.000             197.553        37       5              33
## 5         2619.077            1537.495        37      14              11
## 6         2896.312            1537.500        37      14              11
##   party_size match_mode
## 1          2        tpp
## 2          2        tpp
## 3          2        tpp
## 4          2        tpp
## 5          2        tpp
## 6          2        tpp
```

```
##                                                                 match_id
## 1 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
## 2 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
## 3 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
## 4 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
## 5 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
## 6 2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO
##                       date
## 1 2017-11-26T20:59:40+0000
## 2 2017-11-26T20:59:40+0000
## 3 2017-11-26T20:59:40+0000
## 4 2017-11-26T20:59:40+0000
## 5 2017-11-26T20:59:40+0000
## 6 2017-11-26T20:59:40+0000
```

tail()

```
tail() is used to view bottom 6 rows of the dataset
```

```
tail(PUBG)
```

```
##      player_ID    player_name player_dmg player_kills player_assists
## 9951     19951 LuckyDolphin12         70            0              0
## 9952     19952        Succmul        341            0              2
## 9953     19953       2thikk4u         87            2              0
## 9954     19954   peckerpecker        137            2              0
## 9955     19955      Bawngfist        193            0              0
## 9956     19956  Coots_McGoots         87            1              0
##      player_dist_walk player_dist_ride player_survive_time game_size team_id
## 9951         270.7463            0.000             299.891        29       3
## 9952        3995.2910         3288.762            1333.002        29       6
## 9953        1629.5680         5947.790            1331.496        29       6
## 9954        3212.7978         4421.933            1161.466        29       6
## 9955         873.0563            0.000             361.651        29       6
## 9956         744.9182            0.000             497.617        29       7
##      team_placement party_size match_mode
## 9951             16          4        tpp
## 9952              7          4        tpp
## 9953              7          4        tpp
## 9954              7          4        tpp
## 9955              7          4        tpp
## 9956             14          4        tpp
```

```
##                                                                 match_id
## 9951 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
## 9952 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
## 9953 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
## 9954 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
## 9955 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
## 9956 2U4GBNA0Ymlm4JHKbbmQ9x9_rZbdELcOlVMjnXNbpV6MTm2TQUO4MffsjOqWPWcD
##                           date
## 9951 2017-11-20T01:00:32+0000
## 9952 2017-11-20T01:00:32+0000
## 9953 2017-11-20T01:00:32+0000
## 9954 2017-11-20T01:00:32+0000
## 9955 2017-11-20T01:00:32+0000
## 9956 2017-11-20T01:00:32+0000
```

column names

colnames() function is used to view the attributes of the dataset.

```
colnames(PUBG)
```

```
##  [1] "player_ID"           "player_name"         "player_dmg"
##  [4] "player_kills"        "player_assists"      "player_dist_walk"
##  [7] "player_dist_ride"    "player_survive_time" "game_size"
## [10] "team_id"             "team_placement"      "party_size"
## [13] "match_mode"          "match_id"            "date"
```

summary

```
summary() function is used to summarize the dataset
```

```
summary(PUBG)
```

```
##    player_ID     player_name          player_dmg      player_kills
##  Min.   :10001   Length:9956        Min.   :   0.0   Min.   : 0.0000
##  1st Qu.:12490   Class :character   1st Qu.:   0.0   1st Qu.: 0.0000
##  Median :14978   Mode  :character   Median :  76.0   Median : 0.0000
##  Mean   :14978                      Mean   : 122.7   Mean   : 0.8831
##  3rd Qu.:17467                      3rd Qu.: 172.0   3rd Qu.: 1.0000
##  Max.   :19956                      Max.   :3956.0   Max.   :42.0000
##  player_assists   player_dist_walk   player_dist_ride player_survive_time
##  Min.   :0.0000   Min.   :     0.0   Min.   :    0    Min.   :  68.02
##  1st Qu.:0.0000   1st Qu.:   198.8   1st Qu.:    0    1st Qu.: 241.22
##  Median :0.0000   Median :   815.3   Median :    0    Median : 641.34
##  Mean   :0.2216   Mean   :  1311.2   Mean   : 1146    Mean   : 777.78
##  3rd Qu.:0.0000   3rd Qu.:  2022.4   3rd Qu.: 1895    3rd Qu.:1279.00
##  Max.   :6.0000   Max.   :675124.5   Max.   :10595    Max.   :2184.73
##    game_size        team_id       team_placement   party_size
##  Min.   :13.00   Min.   :     1   Min.   : 1.0   Min.   :1.000
##  1st Qu.:27.00   1st Qu.:    11   1st Qu.:10.0   1st Qu.:2.000
##  Median :45.00   Median :    22   Median :20.0   Median :2.000
##  Mean   :45.58   Mean   : 16738   Mean   :23.8   Mean   :2.683
##  3rd Qu.:49.00   3rd Qu.:    41   3rd Qu.:32.0   3rd Qu.:4.000
##  Max.   :98.00   Max.   :100105   Max.   :98.0   Max.   :4.000
##   match_mode          match_id            date
##  Length:9956        Length:9956        Length:9956
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

str()

str () is an alternative function to display the summary of the output produced, especially when the data set is huge

```
str(PUBG)
```

```
## 'data.frame':    9956 obs. of  15 variables:
##  $ player_ID          : int  10001 10002 10003 10004 10005 10006 10007 10008 10009 10010 ...
##  $ player_name        : chr  "SnuffIes" "Ozon3r" "bovize" "sbahn87" ...
##  $ player_dmg         : int  117 127 67 0 175 65 0 0 79 101 ...
##  $ player_kills       : int  1 1 0 0 2 0 0 0 0 1 ...
##  $ player_assists     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ player_dist_walk   : num  1784.8 1756.1 224.2 92.9 2510.4 ...
##  $ player_dist_ride   : num  2871 2938 0 0 2619 ...
##  $ player_survive_time: num  1106 1106 236 198 1537 ...
##  $ game_size          : int  37 37 37 37 37 37 37 37 37 37 ...
##  $ team_id            : int  4 4 5 5 14 14 15 15 17 17 ...
##  $ team_placement     : int  18 18 33 33 11 11 17 17 24 24 ...
##  $ party_size         : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ match_mode         : chr  "tpp" "tpp" "tpp" "tpp" ...
##  $ match_id           : chr  "2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO" "2U4GBNA0YmnNZ
YkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO" "2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM
39pJO" "2U4GBNA0YmnNZYkzjkfgN4ev-hXSrak_BSey_YEG6kIuDG9fxFrrePqnqiM39pJO" ...
##  $ date               : chr  "2017-11-26T20:59:40+0000" "2017-11-26T20:59:40+0000" "2017-11-26T20:59:40+0000"
"2017-11-26T20:59:40+0000" ...
```

is.null

```
The R function is. null indicates whether a data object is of the data type NULL (i.e. a missing value). The func
tion returns TRUE in case of a NULL object and FALSE in case that the data object is not NULL
```

```
is.null(PUBG$player_ID)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_name)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_dmg)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_kills)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_assists)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_dist_walk)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_dist_ride)
```

```
## [1] FALSE
```

```
is.null(PUBG$player_survive_time)
```

```
## [1] FALSE
```

```
is.null(PUBG$game_size)
```

```
## [1] FALSE
```

```
is.null(PUBG$team_id)
```

```
## [1] FALSE
```

```
is.null(PUBG$team_placement)
```

```
## [1] FALSE
```

```
is.null(PUBG$party_size)
```

```
## [1] FALSE
```

```
is.null(PUBG$match_mode)
```

```
## [1] FALSE
```

```
is.null(PUBG$match_id)
```

```
## [1] FALSE
```

```
is.null(PUBG$date)
```

```
## [1] FALSE
```

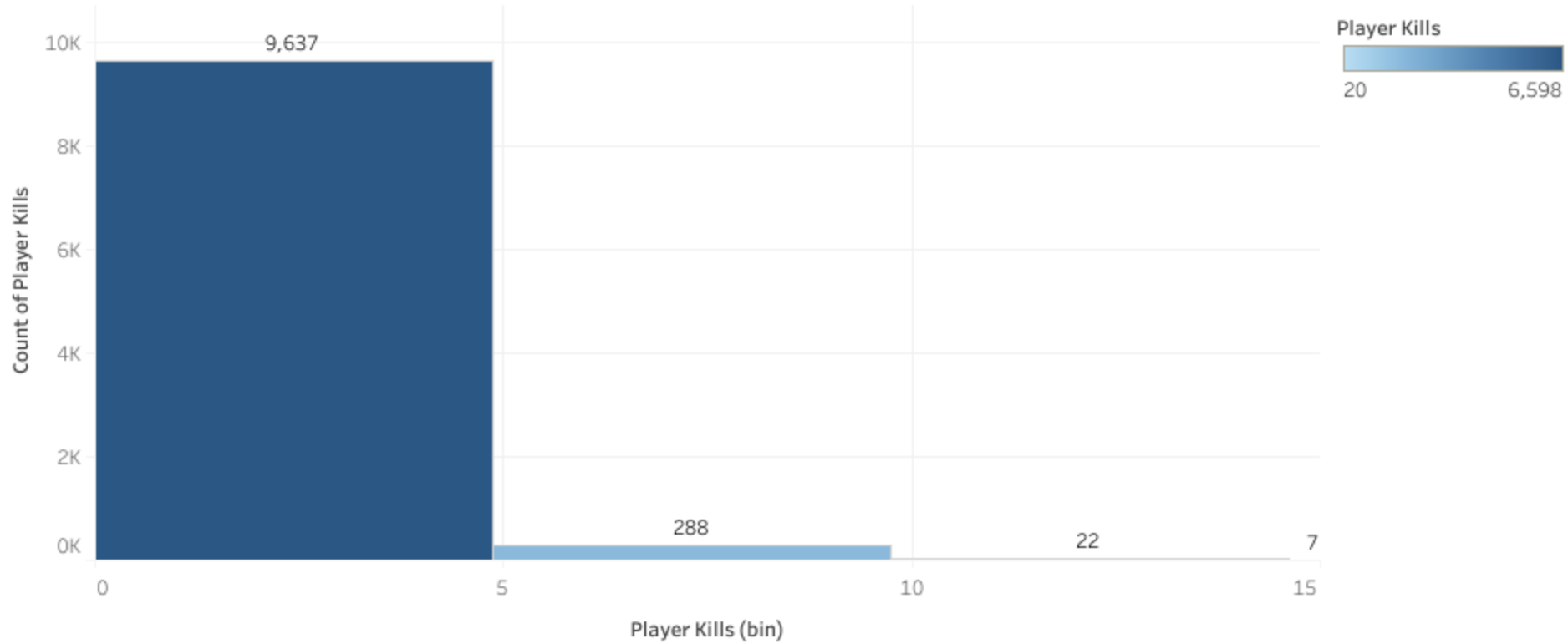
```
is.null(PUBG)
```

```
## [1] FALSE
```

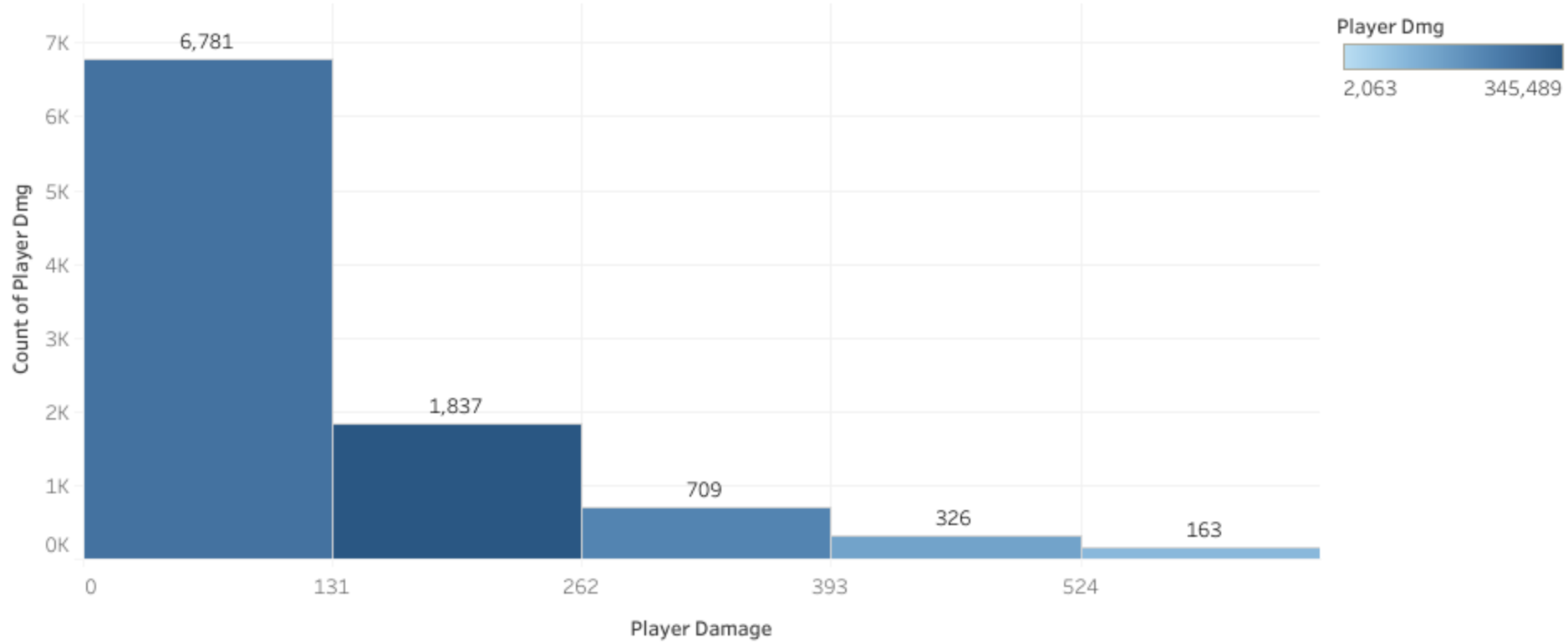
Visualization using Tableau

Tableau is a software company that offers collaborative data visualization software for organization working with business analytics. Organizations use Tableau to visualize data and reveal patterns for analysis in business intelligence , making the data more understandable.

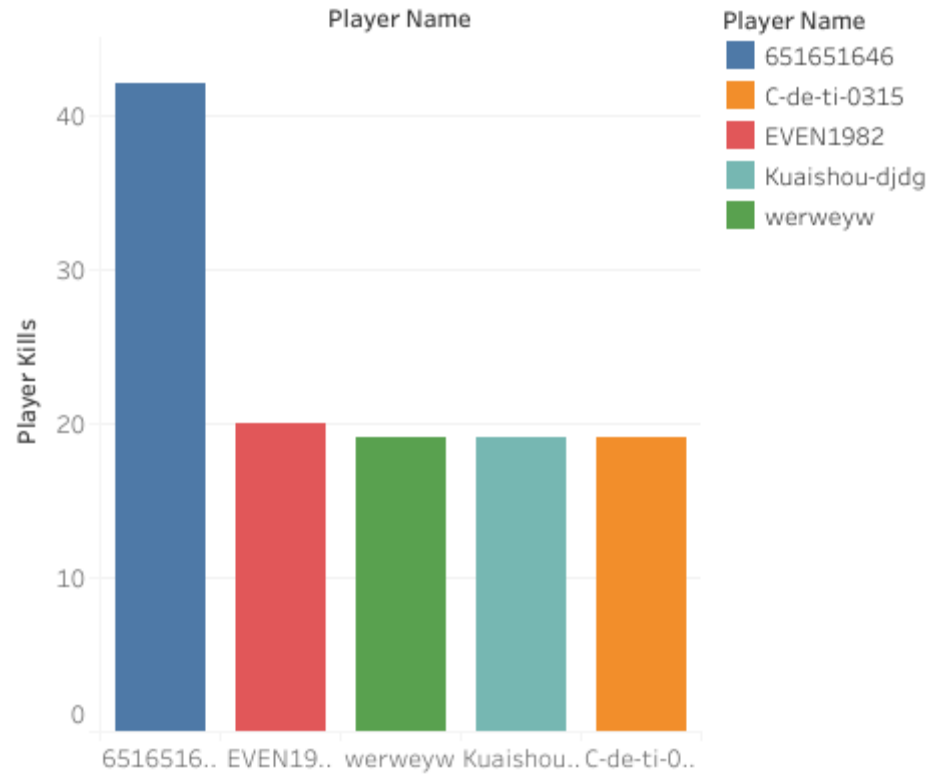
Histogram of Kills



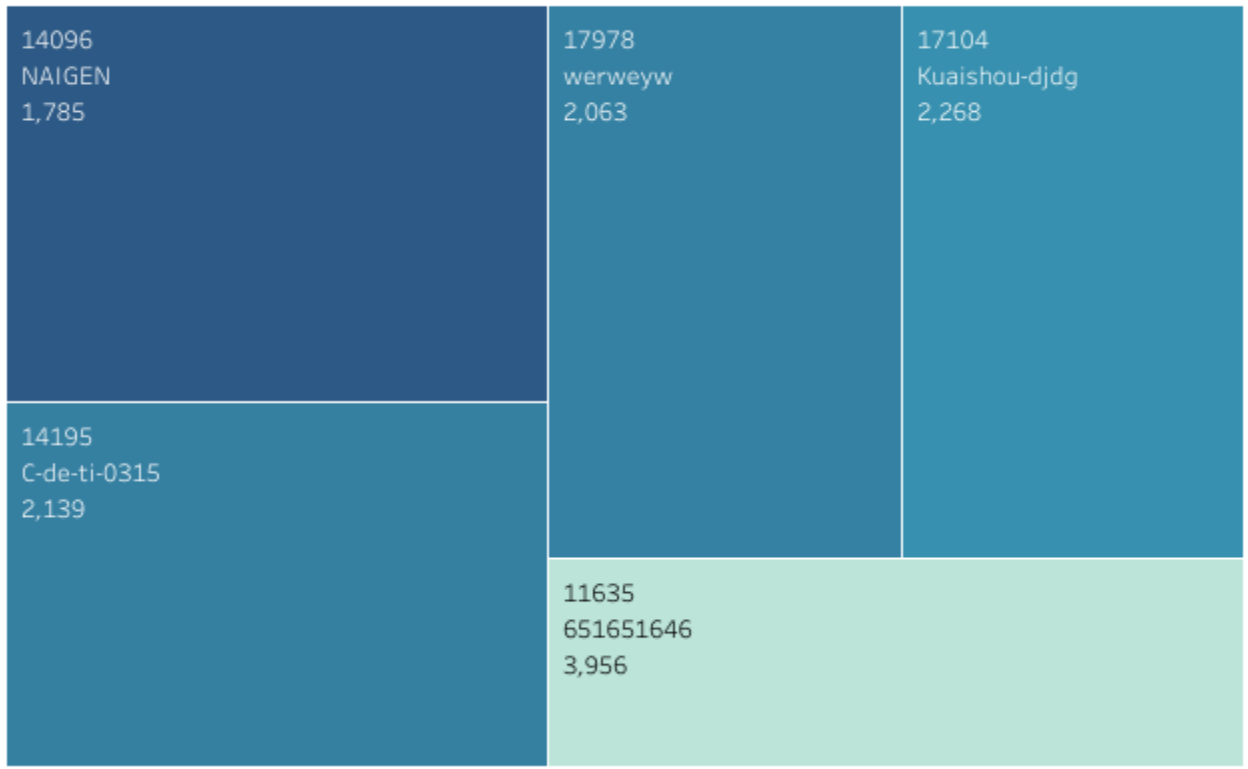
Histogram of Payer Damage



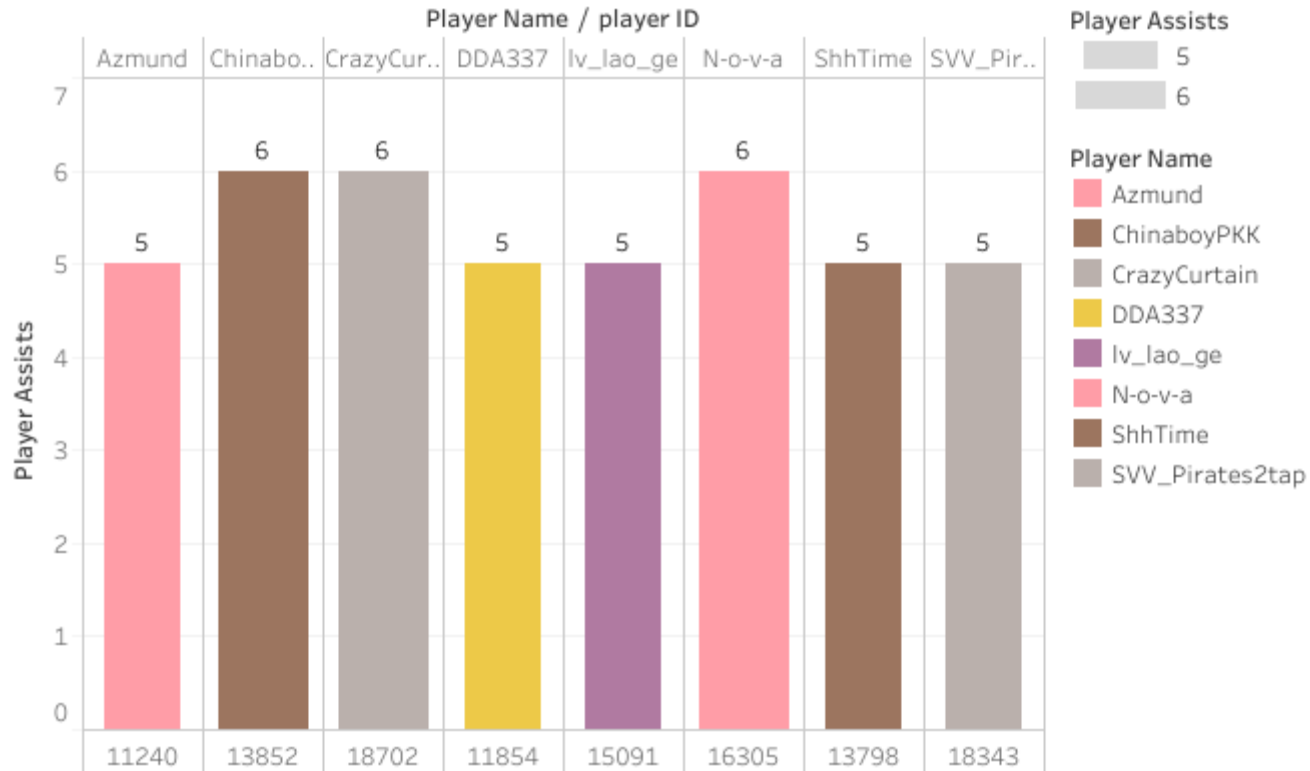
Top 5 Player By Kills



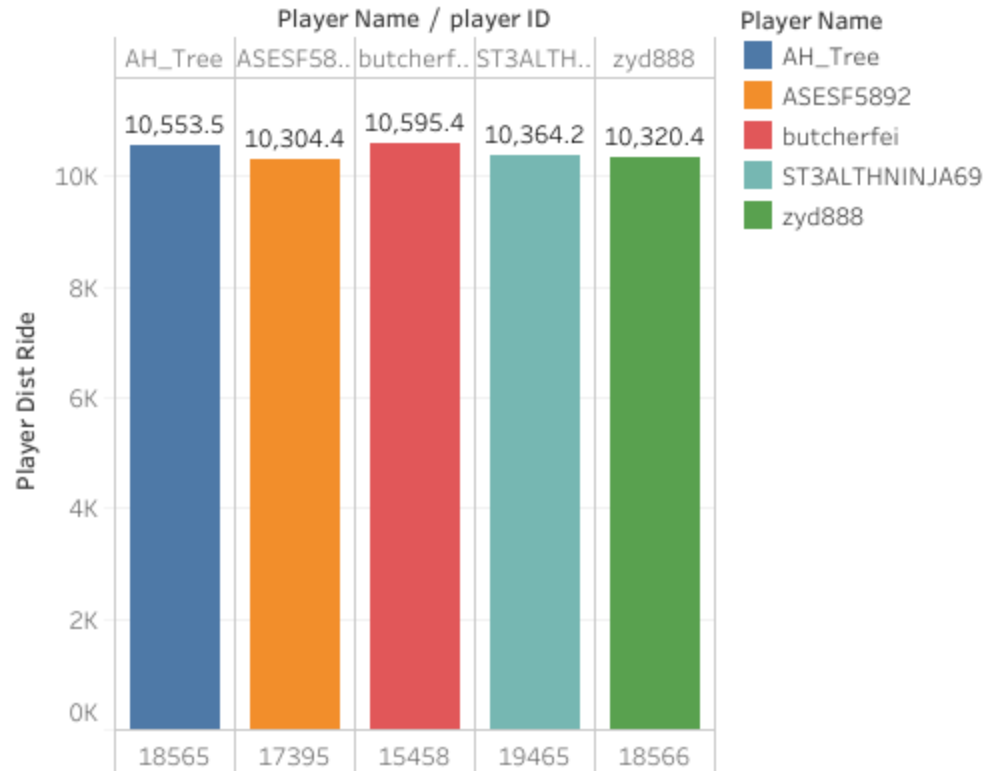
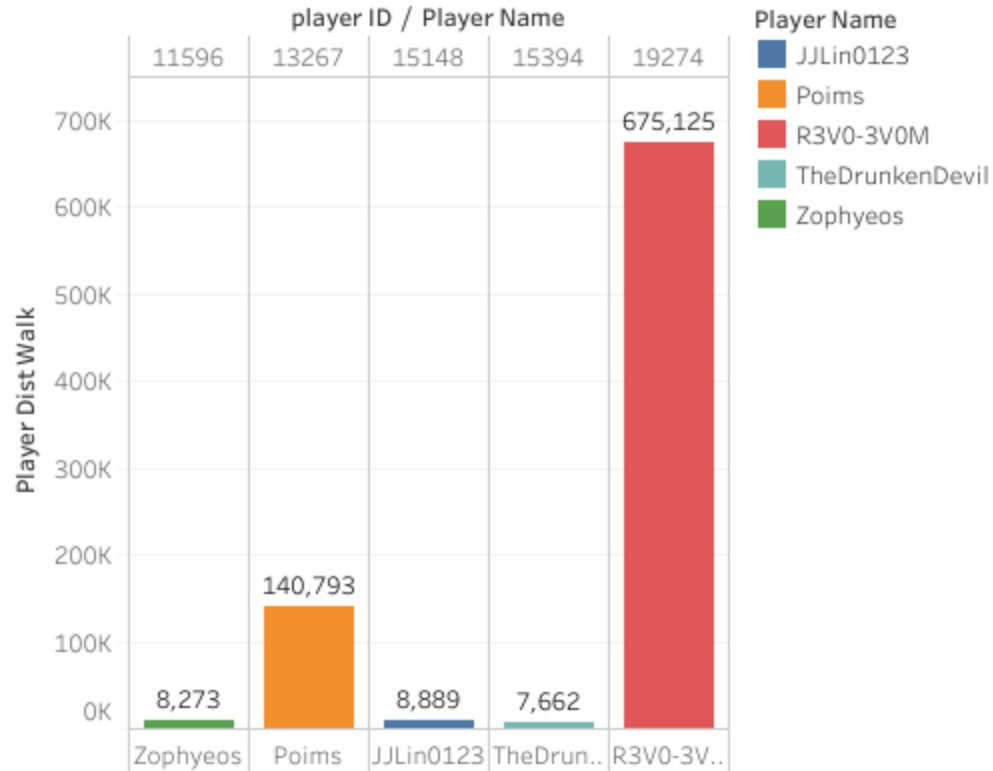
Top 5 Player By Damage



Top 8 Players By Assists



Top 5 Player Distance Ride



Top 5 Player By Distance Walk