

FInal.R

shubham.shrestha

2024-01-17

```
esports <- read.csv("/srv/R/CSC285_public/Shubham(Final)/ESport_Earnings.csv", header = TRUE, sep = ',',
```

```
#Summarzing the esport data to find summary of diiferent fields to find odd and unwanted data  
summary(esports)
```

```
##      IdNo      TotalMoney      GameName      Genre  
## Min.   :151.0  Min.    :      0  Length:504  Length:504  
## 1st Qu.:276.8  1st Qu.:   1375  Class :character  Class :character  
## Median :454.5  Median :   36009  Mode  :character  Mode  :character  
## Mean   :441.0  Mean    :  1744743  
## 3rd Qu.:580.2  3rd Qu.:   205752  
## Max.   :706.0  Max.    :227419905  
##      PlayerNo      TournamentNo      Top_Country      Top_Country_Earnings  
## Min.    :  0.0  Min.    :  0.00  Length:504  Min.    :      0  
## 1st Qu.:  4.0  1st Qu.:  1.00  Class :character  1st Qu.:   391  
## Median : 20.0  Median :  5.00  Mode  :character  Median :   17117  
## Mean   : 186.7  Mean    : 80.75  Mean    : 601082  
## 3rd Qu.: 86.0  3rd Qu.: 25.25  3rd Qu.: 103604  
## Max.   :13032.0  Max.    :5909.00  Max.    :65060611  
##      Releaseyear  
## Min.    : 11  
## 1st Qu.:2006  
## Median :2013  
## Mean    :2007  
## 3rd Qu.:2017  
## Max.    :2020
```

```
#Removing duplicated data if their is any  
esports <- esports[!duplicated(esports), ]
```

```
#Since we are looking for total earning for different game we will remove all the data that has total e  
#since it doesnt match up with out objectives  
#Removing the data that has total money earned 0  
esports <- esports[esports$TotalMoney != 0, ]
```

```
#addition we can find there are mutiple data with top country with none value as it is not useful for o  
esports <- esports[esports$Top_Country != "None", ]
```

```
#The variable PlayerNo and Top_Country_Earning also has data with 0 so we removing those as well  
esports <- esports[esports$Top_Country_Earnings != 0, ]
```

```
#Releaseyear we can find that one of the data has 11 which doesnt make any sense  
#since it is a outlier and the data is of time period of 1989 to 2020 we can presume the data to be 201
```

```
esports$Releaseyear[esports$Releaseyear == 11] <- 2011
```

```
head(esports)
```

```
##      IdNo TotalMoney      GameName      Genre PlayerNo TournamentNo
## 1   613     310.00 Acceleration of SUGURI 2 Fighting Game        6          2
## 2   510  190075.61      Age of Empires      Strategy     246          95
## 3   179 1492489.51      Age of Empires II      Strategy    956         576
## 4   177   68113.85      Age of Empires III      Strategy   106          76
## 5   540   2668.00      Age of Empires Online      Strategy    16           7
## 6   178   52360.00      Age of Mythology      Strategy    22           6
##
##      Top_Country Top_Country_Earnings Releaseyear
## 1      United States          310.00          2018
## 2      Viet Nam      123408.43          1997
## 3      China      174252.44          1999
## 4      United States      18523.52          2005
## 5      Germany      1264.00          2011
## 6 Taiwan, Republic of China      22000.00          2002
```

```
summary(esports)
```

```
##      IdNo      TotalMoney      GameName      Genre
## Min.   :151.0   Min.    :    10   Length:444   Length:444
## 1st Qu.:274.8   1st Qu.:   4457   Class :character   Class :character
## Median :455.5   Median :   50000   Mode  :character   Mode  :character
## Mean   :442.4   Mean    :  1978937
## 3rd Qu.:587.2   3rd Qu.:   283092
## Max.   :706.0   Max.    :227419905
##      PlayerNo      TournamentNo      Top_Country      Top_Country_Earnings
## Min.    :    1.0   Min.    :    1.00   Length:444   Min.    :    10
## 1st Qu.:    7.0   1st Qu.:    2.00   Class :character   1st Qu.:   1819
## Median :   31.5   Median :    7.00   Mode  :character   Median :   24528
## Mean    :   211.8   Mean    :   91.49           Mean :   682310
## 3rd Qu.:  101.8   3rd Qu.:   35.00           3rd Qu.:  153137
## Max.    :13032.0   Max.    :5909.00           Max.    :65060611
##      Releaseyear
## Min.    :1989
## 1st Qu.:2006
## Median :2013
## Mean    :2011
## 3rd Qu.:2017
## Max.    :2020
```