Final Project

Tushar Dubey

2023-05-12

SQL Connection

Connects to SQL database

```
library(RMySQL)
```

```
## Warning: package 'RMySQL' was built under R version 4.1.2
```

```
## Loading required package: DBI
```

```
## Warning: package 'DBI' was built under R version 4.1.2
```

```
mysqlconnection = dbConnect(RMySQL::MySQL(),
                            dbname='final_proj',
                            host='localhost',
                            port=3306,
                            user='root',
                            password='cmsc398e')
```

SQL

Problem: Valorant is a tactical first person shooter video game with a massive player base. There is a competitive scene in which professional players compete to win a championship for fame and glory. At the end of the year Valorant hosts "Champions" (akin to what the World Cup is to soccer) and 8 teams with 5 players each play to be the best. This data set shows performance statistics from each of those players and I will be analyzing each of their performances. I wonder if there is a correlation between a player's headshot percentage and their rating. There is a myth that in order to be a good valorant player you must hit a lot of headshots but I would like to see if this myth is true or false using data from the best players in the world. I will be testing to see if there is a correlation between a players headshot percentage and their associated rating. My hypothesis will be: a player's headshot percentage has no effect on their overall player rating. I will use a scatter plot and linear regression to determine whether or not my hypothesis is true or not.

SQL Components Used: I cleaned my data of null values. I used avg() as my aggregate function. I used rank() and format(). (note: sql comments are causing errors so I will explain each code block using text)

The player HS is our independent variable and the player rating is the dependent variable in our case.

```
SELECT player_name, player_ACS, player_rating, player_HS
FROM valorant
WHERE player_HS IS NOT NULL AND player_rating IS NOT NULL;
```

Table 1: Displaying records 1 - 10

player_name	player_ACS	player_rating	player_HS
yay	254.6	1.26	26%
kiNgg	260.0	1.26	23%
Laz	245.0	1.23	28%
pANcada	210.5	1.21	34%
Derke	247.9	1.20	34%
Shao	194.5	1.15	30%
Cryocells	232.2	1.15	24%
Sacy	194.0	1.14	25%
MaKo	207.8	1.13	29%
Less	219.9	1.13	25%

Remove null values and show all player headshot percentages

```
CREATE TABLE cleaned_data15 AS
SELECT
  player_name,
  COALESCE(player_HS, 0) AS player_HS
FROM
  valorant
WHERE
  player_HS IS NOT NULL;
```

```
SELECT * FROM cleaned_data15
```

Table 2: Displaying records 1 - 10

player_name	player_HS
yay	26%
kiNgg	23%
Laz	28%
pANcada	34%
Derke	34%
Shao	30%
Cryocells	24%
Sacy	25%
MaKo	29%
Less	25%

The following code will take each player and compare their headshot percentage to the average HS%. Then it will create a new column that will compare each players HS% to the average and compute the delta. This will give us an idea of how the player stands compared to the average headshot percentage. It does so by converting the percentages to values and then computing differences.

```
WITH cleaned_data AS (
  SELECT
    player_name,
    COALESCE(player_HS, 0) AS player_HS
  FROM
    valorant
```

```
  WHERE
    player_HS IS NOT NULL
),
average_hs AS (
  SELECT
    player_name,
    player_HS,
    FORMAT(AVG(player_HS) OVER (), 2) AS avg_HS_percentage,
    RANK() OVER (PARTITION BY NULL ORDER BY player_HS) AS hs_rank
  FROM
    cleaned_data
)
SELECT
  player_name,
  FORMAT(player_HS, 2) AS player_HS,
  avg_HS_percentage,
  FORMAT(player_HS - avg_HS_percentage, 2) AS HS_delta
FROM
  average_hs;
```

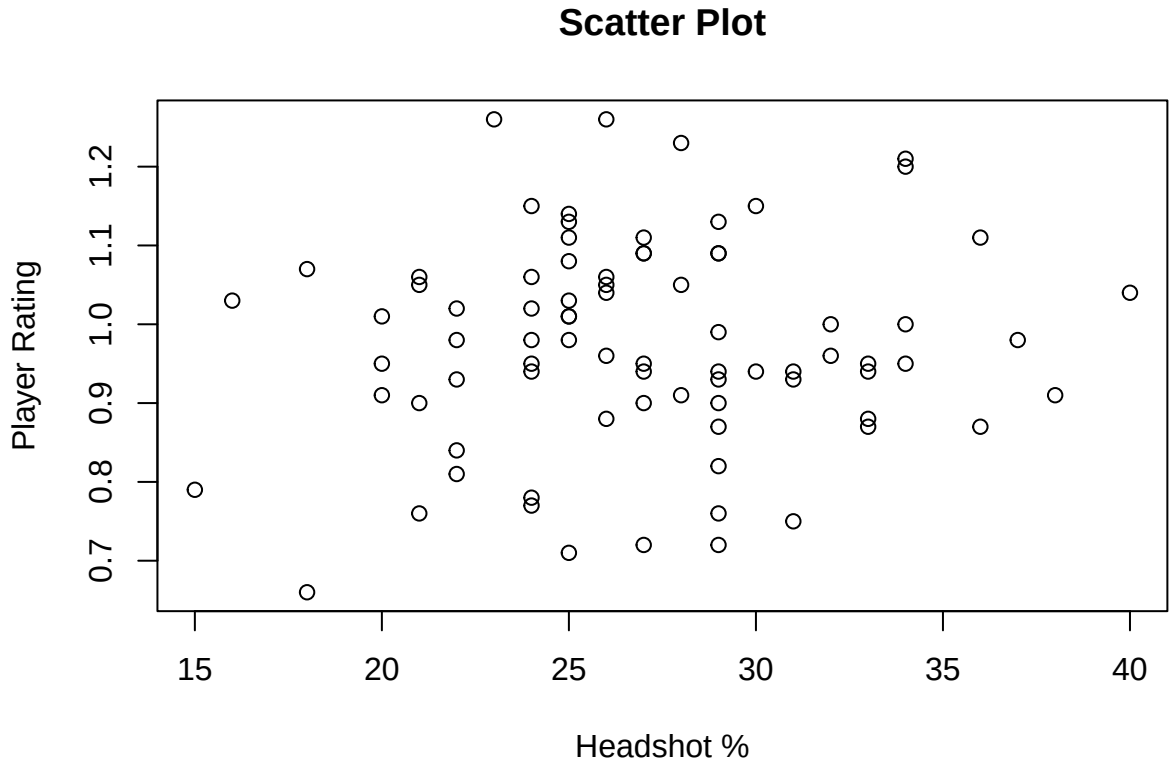
Table 3: Displaying records 1 - 10

player_name	player_HS	avg_HS_percentage	HS_delta
Famouz	15.00	26.93	-11.93
BcJ	16.00	26.93	-10.93
BerserX	18.00	26.93	-8.93
Will	18.00	26.93	-8.93
Jinggg	20.00	26.93	-6.93
AYRIN	20.00	26.93	-6.93
Boaster	20.00	26.93	-6.93
Sushiboys	21.00	26.93	-5.93
FNS	21.00	26.93	-5.93
saadhak	21.00	26.93	-5.93

```
# Read the CSV file and remove NAs
data <- read.csv("vlr_vct2022_data.csv")
data <- na.omit(data)

# Convert player_HS from a percentage to a number for plotting purposes
data$player_HS <- as.numeric(gsub("%", "", data$player_HS))

#This is the scatter plot with hs% on x axis and player rating on y
plot(data$player_HS, data$player_rating, xlab = "Headshot %", ylab = "Player Rating", main = "Scatter P
```




```r
# Perform linear regression
lm_model <- lm(player_rating ~ player_HS, data = data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = player_rating ~ player_HS, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.292788 -0.074749 -0.007668  0.092572  0.295012
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.908866   0.080145  11.340   <2e-16 ***
## player_HS   0.002440   0.002926   0.834    0.407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1318 on 79 degrees of freedom
## Multiple R-squared:  0.008724,   Adjusted R-squared:  -0.003824
## F-statistic: 0.6952 on 1 and 79 DF,  p-value: 0.4069
```

```r
set.seed(123)
train_indices <- sample(nrow(data), floor(0.7 * nrow(data)))
train_data <- data[train_indices, ]
test_data <- data[-train_indices, ]

# fitting linear regression model on train data
lm_model_train <- lm(player_rating ~ player_HS, data = train_data)
```

```r
test_data$player_HS <- as.numeric(gsub("%", "", test_data$player_HS))

predictions <- predict(lm_model_train, newdata = test_data)

# compute rmse and r^2 values
rmse <- sqrt(mean((predictions - test_data$player_rating)^2))
r_squared <- summary(lm_model_train)$r.squared

cat("Root Mean Squared Error (RMSE):", rmse)
```

```
## Root Mean Squared Error (RMSE): 0.1447397
```

```r
cat("\nR-squared (R^2):", r_squared)
```

```
##
## R-squared (R^2): 0.009525949
```

We can see here that the R^2 value is 0.00952. This means that there is a very weak correlation between HS% and player rating. We can say then that the players headshot percentage has no effect on the players rating.

My hypothesis was that a players headshot percentage has no effect on the players rating. A higher HS% does not lead to a higher player rating. I performed linear regression to see if there is a correlation between HS% and player rating. If the R^2 value is less than 0.7 then we can conclude that an increase in the players headshot percentage will have no effect on the players rating. I chose to use a scatter plot to compare my independent and dependent variables. We can tell visually that there does not seem to be any apparent correlation between the two variables. This is further supported by the value computed for R^2. The value is 0.00952 which is very low. We can conclude with this that the players headshot percentage has no effect on their rating. My hypothesis was accepted so we can conclude that the players' headshot percentages have no effect on their rating.