

Analysis of Manchester City Soccer Games

Our population data was sourced from the Manchester City games database spanning from 2018 to May 2023. This database, hosted on Google, stands as the exclusive repository for soccer game information with an extensive dataset, and it is highly reputable. You can access the database through the following link:

[“https://www.google.com/search?q=manchester+city+games&rlz=1C1UEAD_enUS931US931&oq=manchester+city+games&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTINCAEQLhiDARiXAxIABDIGCAIQIxnMgYIAXAjGCcyDQgEEAAYgwEYsQMYgAQyDQgFEAAYgwEYsQMYgAQyBggGEAAYAzIHCAcQABiABDINCAgQABiDARiXAxIABDIKCAkQABiXAxIABNIBCDU3MzBqMWO3qAIAA&sourceid=chrome&ie=UTF-8#cobssid=s&sie=m;/g/11v9_b85v0;2;/m/0c1q0;dt;fp;1;;”](https://www.google.com/search?q=manchester+city+games&rlz=1C1UEAD_enUS931US931&oq=manchester+city+games&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTINCAEQLhiDARiXAxIABDIGCAIQIxnMgYIAXAjGCcyDQgEEAAYgwEYsQMYgAQyDQgFEAAYgwEYsQMYgAQyBggGEAAYAzIHCAcQABiABDINCAgQABiDARiXAxIABDIKCAkQABiXAxIABNIBCDU3MzBqMWO3qAIAA&sourceid=chrome&ie=UTF-8#cobssid=s&sie=m;/g/11v9_b85v0;2;/m/0c1q0;dt;fp;1;;) We manually stored all the data of the games into our population dataset.

From the database, we collected a population of 331 games and then we applied `sample_n()` function in R to select a random of 200 rows and 15 columns of data. This will be our dataset for the research project. Here is the in-depth breakdown of all the columns of the data file.

- **Year:** The year the game was played on. (Categorical)
- **Goals Scored:** The number of goals scored by Manchester City. (Quantitative)
- **Goals Conceded:** The number of goals conceded by Manchester City. (Quantitative)
- **Win/Lose/Draw:** This column has three values:(Win, Lose or Draw). (Categorical)
- **Formation:** The formation in which Manchester City played the game in. (Categorical)
- **Shots:** The total shots of Manchester City in a game. (Quantitative)
- **Shots on Target:** The total shots on target of Manchester City in a game. (Quantitative)
- **Possession%:** The percentage of possession of the ball that Manchester City had in a game (Quantitative).
- **Passes:** The total passes of Manchester City in a game. (Quantitative)
- **Fouls:** The total fouls of Manchester City in a game. (Quantitative)
- **Yellow Cards:** The total number of yellow cards in a game for Manchester City. (Quantitative)
- **Red Cards:** The total number of red cards in a game for Manchester City. (Quantitative)
- **Offsides:** The total number of offsides of Manchester City (Quantitative).
- **Corners:** The total corners of Manchester City in a game. (Quantitative)
- **RedCardsFlag:** This column has two values. The value is Yes if Manchester City got a red card and no if they did not. (Categorical)

This is the code we used to sample out 200 random rows from the population data which consisted of 331 rows

```
# Load necessary package
install.packages("dplyr")
library(dplyr)
# Read the dataset
data <- read.csv("C:\\Users\\OSAKWEJ1\\Downloads\\STA305.csv")
# Randomly sample 200 rows
sampled_data <- data %>% sample_n(200)
```