

**Data Wrangling Assignment**

**Name:** Abhishek Dubey

**Student ID:** D20123718

**Full time:** MSc Data Science 2020-21

**Date:** 18/11/2020

**Table of Content**

* Part 1 2
  + Importing Data Set
  + Merging
  + Cleaning
* Part 2 3
  + Variable Understanding
  + Exploration of Data Set
* Part 3 4
  + Defining Questions in Data Set
  + Visualization of Data Set
  + Answering all Questions with help of visualization
* Appendix 6
  + Tables
  + Plots

**PART 1**

**Final Dataset**

<https://github.com/Abhidubey96/Data-Visualization-Project/blob/main/champions_data.csv>

I have used python 3.8.1, pip 20.2.1, pandas, matplotlib and many more for analysis.

Python file: <https://github.com/Abhidubey96/Data-Visualization-Project/blob/main/Data%20Assignment.ipynb>

By merging all three-sub data set

Because of insufficient data and wrong variables in original single data file I have to collaborate it with different source of data of different types to sounds it meaningful and to perform better analysis. Basically, data gathered from all 3-public repository.

Data file from public repository.

Champions League final match Team data from 1956 – 2017 (extracting csv file)

<https://data.world/sportsvizsunday/sports-viz-sundays-2018>

Champions League final match Stadium data from 1956- 2017 (extracting data by using html reader)

<https://www.stadiumguide.com/figures-and-statistics/lists/champions-league-final-venues/>

European Stadium Capacity Data (extracting table from webpage)

<https://www.stadiumguide.com/figures-and-statistics/lists/europes-largest-football-stadiums/>

As data file having empty cells and irrelevant columns (variables) so it’s better to do cleaning process.

All cleaning done in python; you can refer above python script file.

**Topic:**

Best Team in Champions League History?

**Description:**

The UEFA Champions League is an annual club football competition organized by the Union of European Football Associations (UEFA) and contested by top-division European clubs, deciding the competition winners through a group and knockout format. It is one of the most prestigious football tournaments in the world and the most prestigious club competition in European football, played by the national league champions ( and, for some nations, one or more runners-up) of their national associations.  
For Data Visualization I am analyzing final matches played between teams from 1956 till 2017 also will going to find out so many hidden answers with using graph plots and tableau Dashboard.

**Description of intended audience:**

* Football Teams, especially managers of team to identify the ranking and motivation for winning next title.
* Common football fan to identify the best team in champions league history. Also ranking of teams based on goal scored and titles
* Champions league federation for statistics and history evaluation

**PART 2**

**Data Exploration:**

Key variables and their explanation for understanding:

Winner: Team Wins Title

Finalist: Runner up team

Winner score: winning team scored goals

Finalist score: Runner up team scored goals

Stadium: Venue where game was played

Attendees: people who watched the match live on stadium

Capacity: Actual capacity of stadium

People attendant match in excess: (Attendees – Capacity)

Note: Because of some stadium are not in use and demolished, I assumed capacity of stadium equals to attendees

Note: People attendant match in excess shows you the popularity of team and fans visited the stadium to watch live.

**Reference: Appendix, Table 1**

Data Type of each variable

Winner object

Winner Country object

Winner Score int64

Finalist Score int64

Finalist object

Finalist Country object

Attendance int64

Year int64

Stadium object

Capacity int32

Creating total goals column by adding winner score and finalist score

**Reference: Appendix, Table 2**

62 rows × 11 columns

Total Goals scored by both teams in final

Let’s identify the data by describe function

**Reference: Appendix, Table 3**

**Now we will analyze each variable by plotting in graphs:**

Below graph shows team having total number of goals throughout the champions league history.

From analysis we can say Real Madrid score the most 35 goals.

**Reference: Appendix, Graph 1**

Below graph shows team match having total number of scored goals per match

Can from analysis we can say Real Madrid vs Eintracht Frankfurt scored the most 10 goals in single match.

**Reference: Appendix, Graph 2**

The analysis of capacity of stadium vs attendance of public while live match.

Below graph we have calculated the exceed number by

People attendent match = Capacity of Stadium – Attendees of live match

And we changed the graph with team per year to analyse for which team stadium was over crowded.

We sorted the graph per year and analyze that Real Madrid having more popular games than anyother team.

**Reference: Appendix, Graph 3**

Below Graph shows most number of champions league title holder team.

Real Madrid tops the position with 12 hits.

**Reference: Appendix, Graph 4**

**Part 3**

We will analyze and extract answers for below question from insights.

**Most Dominant team in champions league history as per titles?**

**Most Dominant team in champions league history as per total goal scored?**

**Most popular Team with fan base and most anticipated match in history?**

When stadium was most occupied by people (outnumbered) and for which team?

**Insight 1:**

To answer this question: **Most Dominant team in champions league history as per titles?**

**First, we gathered and modify data as per titles. We draw graph between winner and sum of winning year.**

**Champions League play every year so each year one team wins like world cup. So, count the total number of wins with winner team name and plot bar graph.**

**As you can see in below graph Real Madrid wins the most number of champions league title.**

**Reference: Appendix, Graph 4**

**Insight 2:**

To answer this question: **Most Dominant team in champions league history as per total goal scored?**

**First, we gathered and modify data as per goals. We draw graph between winner team vs winning score column.**

**Winning score column’s gives goals scored for each game. So, I add them by using sum function and draw graph between winning team vs total goals per team.**

**As you can see in below graph Real Madrid scored the most number goals in champions league final.**

**Reference: Appendix, Graph 1**

**Insight 3:**

To answer this question:

**Most popular Team with fan base and most anticipated match in history?**

When stadium was most occupied by people (outnumbered) and for which team?

First, we gathered the data, initial file was having the attendance of the public watched the match live. So, we import another file from different data source about stadium per capacity. By merging these two files in python we have final data file having attendees of match in stadium and actual capacity of stadium.

\*\* As some of the stadiums as not in use and demolished we cannot say the capacity so I assume the attendees == capacity of stadium

We calculated new field excess in attendees = attendees – capacity

Excess in attendees field show which match having most popularity.

Below graph shows attendees in blue bar graph and capacity in orange circle per year with team.

Clearly shows Real Wins the popularity Ranking

**Reference: Appendix, Graph 3**

**Appendix**

**Table 1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Winner** | **Winner Country** | **Winner Score** | **Finalist Score** | **Finalist** | **Finalist Country** | **Attendance** | **Year** | **Stadium** | **Capacity** | **Match Goals** | **Match Played** | **people attendent match in excess** | **Team with year** |
| Real Madrid | Spain | 4 | 3 | Stade de Reims | France | 38239 | 1956 | Stade Velodrome Paris | 38239 | 7 | Real Madrid vs Stade de Reims | 0 | Real Madrid : 1956 |
| Real Madrid | Spain | 2 | 0 | Fiorentina | Italy | 124000 | 1957 | Estadio Santiago Bernabeu | 81044 | 2 | Real Madrid vs Fiorentina | 42956 | Real Madrid : 1957 |
| Real Madrid | Spain | 3 | 2 | Milan | Italy | 67000 | 1958 | Heysel Stadium | 67000 | 5 | Real Madrid vs Milan | 0 | Real Madrid : 1958 |
| Real Madrid | Spain | 2 | 0 | Stade de Reims | France | 72000 | 1959 | Neckarstadion | 72000 | 2 | Real Madrid vs Stade de Reims | 0 | Real Madrid : 1959 |
| Real Madrid | Spain | 7 | 3 | Eintracht Frankfurt | West Germany | 127621 | 1960 | Hampden Park | 52000 | 10 | Real Madrid vs Eintracht Frankfurt | 75621 | Real Madrid : 1960 |
| Benfica | Portugal | 3 | 2 | Barcelona | Spain | 26732 | 1961 | Wankdorfstadion | 26732 | 5 | Benfica vs Barcelona | 0 | Benfica : 1961 |

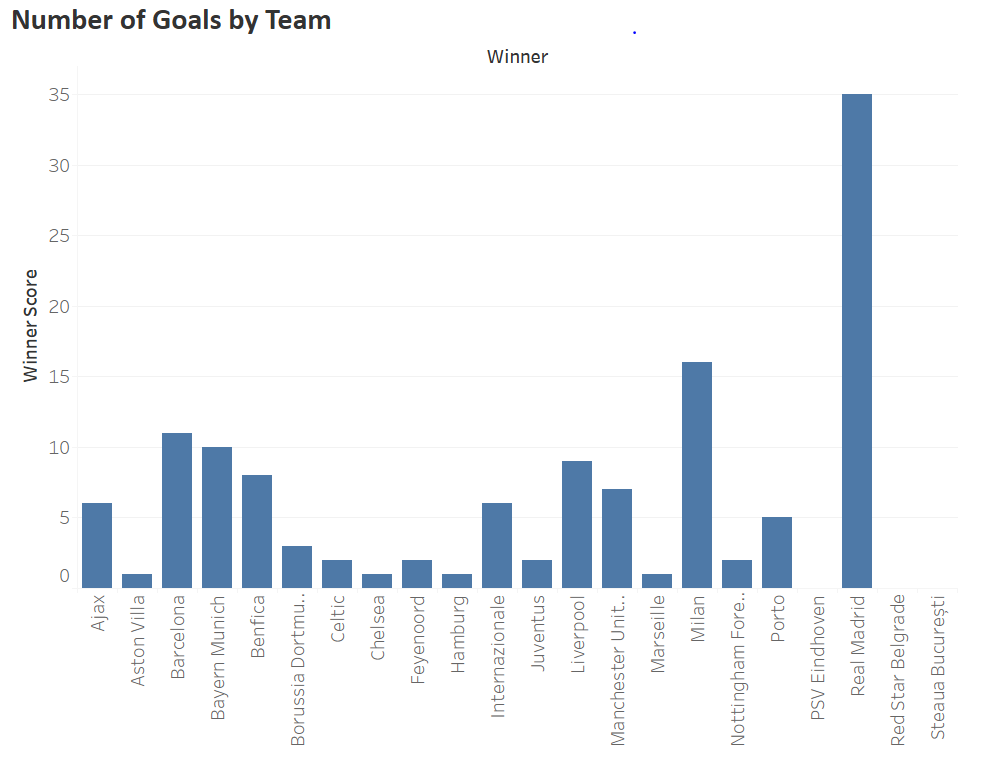
**Table 2**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Winner** | **Winner Country** | **Winner Score** | **Finalist Score** | **Finalist** | **Finalist Country** | **Attendance** | **Year** | **Stadium** | **Capacity** | **Match Goals** |
| **0** | Real Madrid | Spain | 4 | 3 | Stade de Reims | France | 38239 | 1956 | Stade Velodrome Paris | 38239 | 7 |
| **1** | Real Madrid | Spain | 2 | 0 | Fiorentina | Italy | 124000 | 1957 | Estadio Santiago Bernabeu | 81044 | 2 |
| **2** | Real Madrid | Spain | 3 | 2 | Milan | Italy | 67000 | 1958 | Heysel Stadium | 67000 | 5 |
| **3** | Real Madrid | Spain | 2 | 0 | Stade de Reims | France | 72000 | 1959 | Neckarstadion | 72000 | 2 |
| **4** | Real Madrid | Spain | 7 | 3 | Eintracht Frankfurt | West Germany | 127621 | 1960 | Hampden Park | 52000 | 10 |

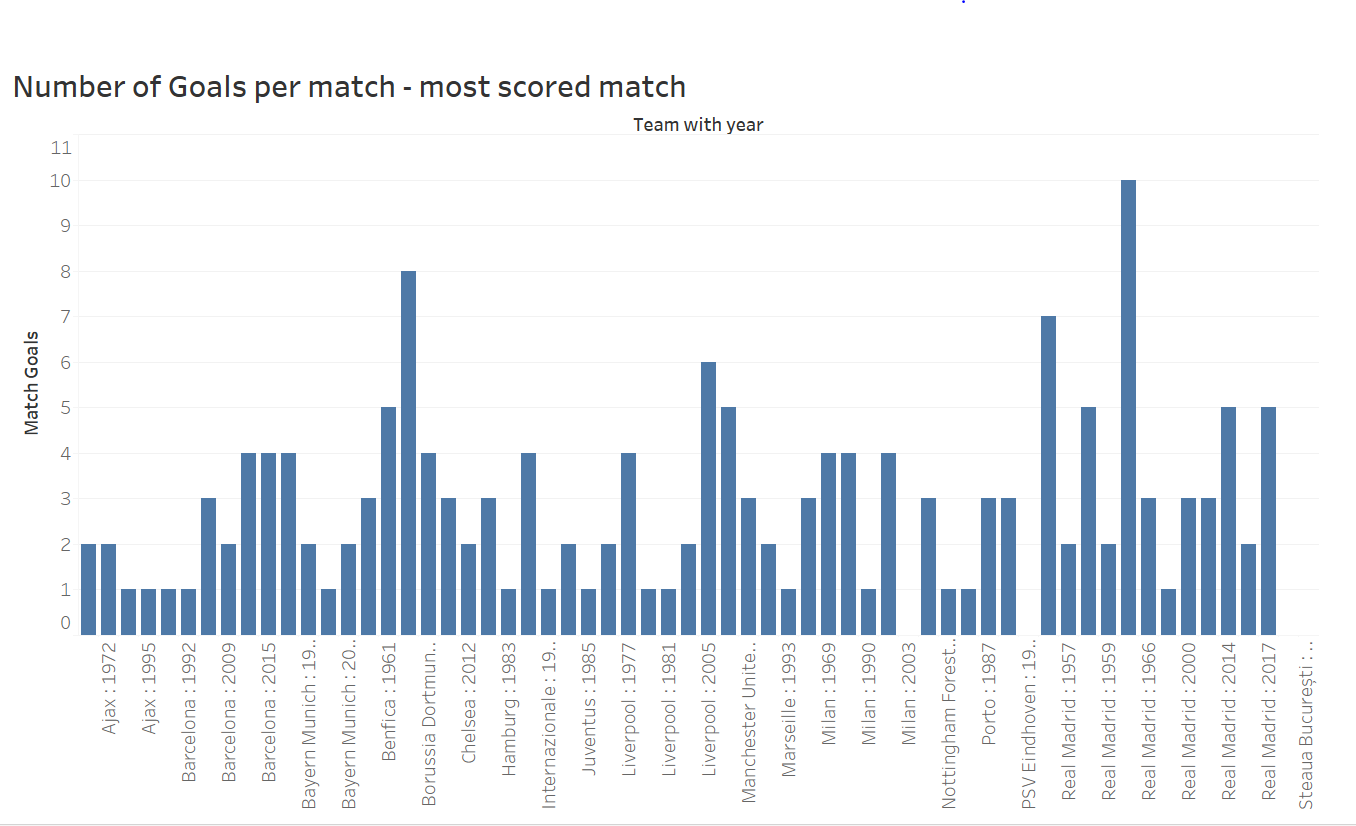
**Table 3**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Winner Score** | **Finalist Score** | **Attendance** | **Year** | **Capacity** |
| **count** | 62.000000 | 62.000000 | 62.000000 | 62.000000 | 62.000000 |
| **mean** | 2.064516 | 0.661290 | 66394.645161 | 1986.500000 | 69303.225806 |
| **std** | 1.341286 | 0.828658 | 18594.730692 | 18.041619 | 15571.942732 |
| **min** | 0.000000 | 0.000000 | 26732.000000 | 1956.000000 | 26732.000000 |
| **25%** | 1.000000 | 0.000000 | 53606.250000 | 1971.250000 | 57514.500000 |
| **50%** | 2.000000 | 0.500000 | 65121.000000 | 1986.500000 | 70887.500000 |
| **75%** | 3.000000 | 1.000000 | 72035.250000 | 2001.750000 | 80018.000000 |
| **max** | 7.000000 | 3.000000 | 127621.000000 | 2017.000000 | 99354.000000 |

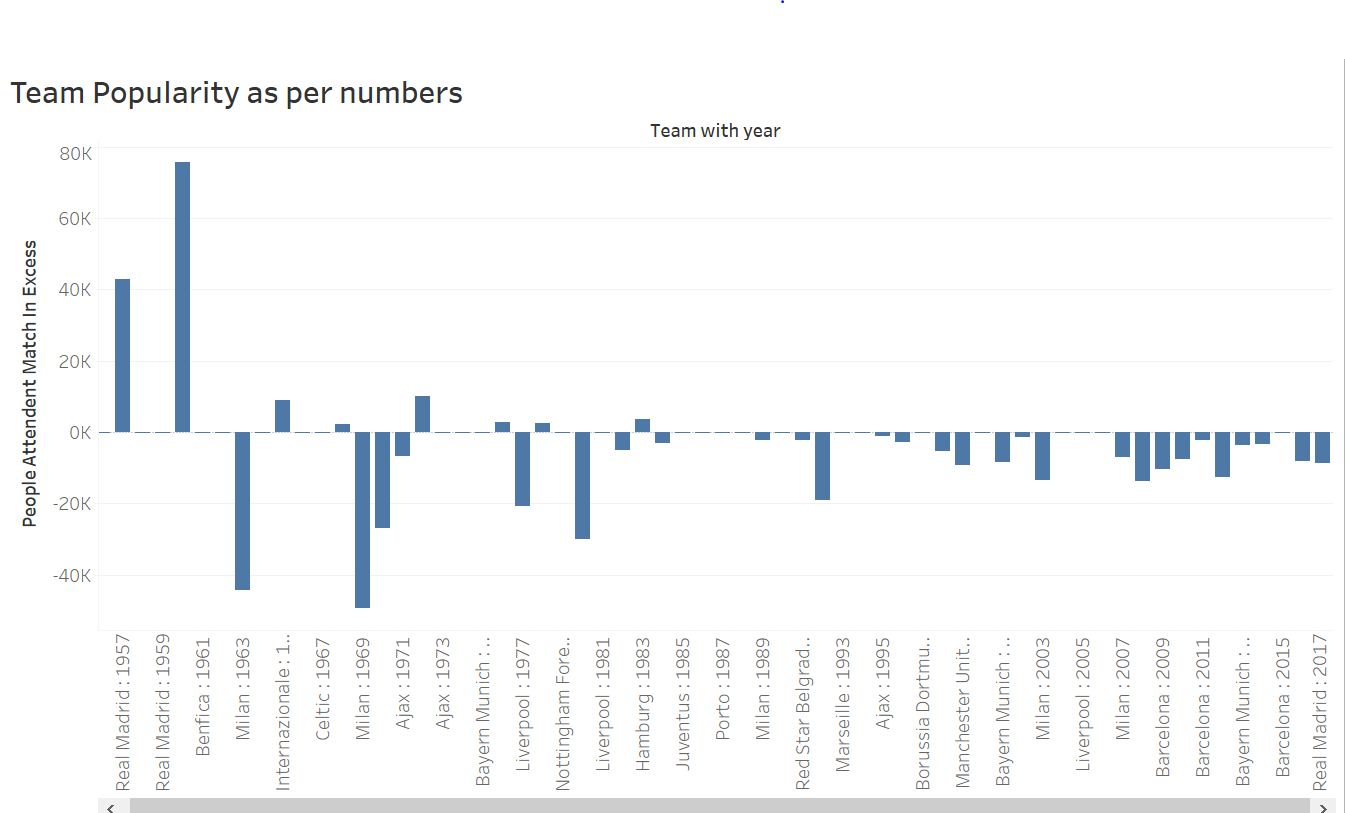
**Graph 1**



**Graph 2**



**Graph 3**



**Graph 4**

