Data Visualization Assignment 1 

**Name:** Abhishek Dubey

**Student ID:** D20123718

**Class code:** SPEC9995, 2020-21

**Full time:** MSc Data Science

**Dataset Link:**

* <https://data.world/sportsvizsunday/sports-viz-sundays-2018>
* <https://www.stadiumguide.com/figures-and-statistics/lists/champions-league-final-venues/>
* <https://www.stadiumguide.com/figures-and-statistics/lists/europes-largest-football-stadiums/>

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

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

**Topic:**

Best Team in Champions League History?

Link: <https://public.tableau.com/views/VisualizationAssignmentFinal_16042689791190/FinalDashboard?:language=en-GB&:display_count=y&publish=yes&:origin=viz_share_link>

\*\* Please watch it in full screen for better clarity

**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

As per the finding there are few questions, we can answer in order to determine the best team:

**Most scoring match played in champions league?**

This will help us to understand which team is most competitive.

**Which Team scores the greatest number of goals till date?**

This will help us to understand which team is dominant.

**When stadium is most occupied by people (outnumbered) and for which match?**

This will help us to understand the popularity of team among people.

**Most number of wins?**

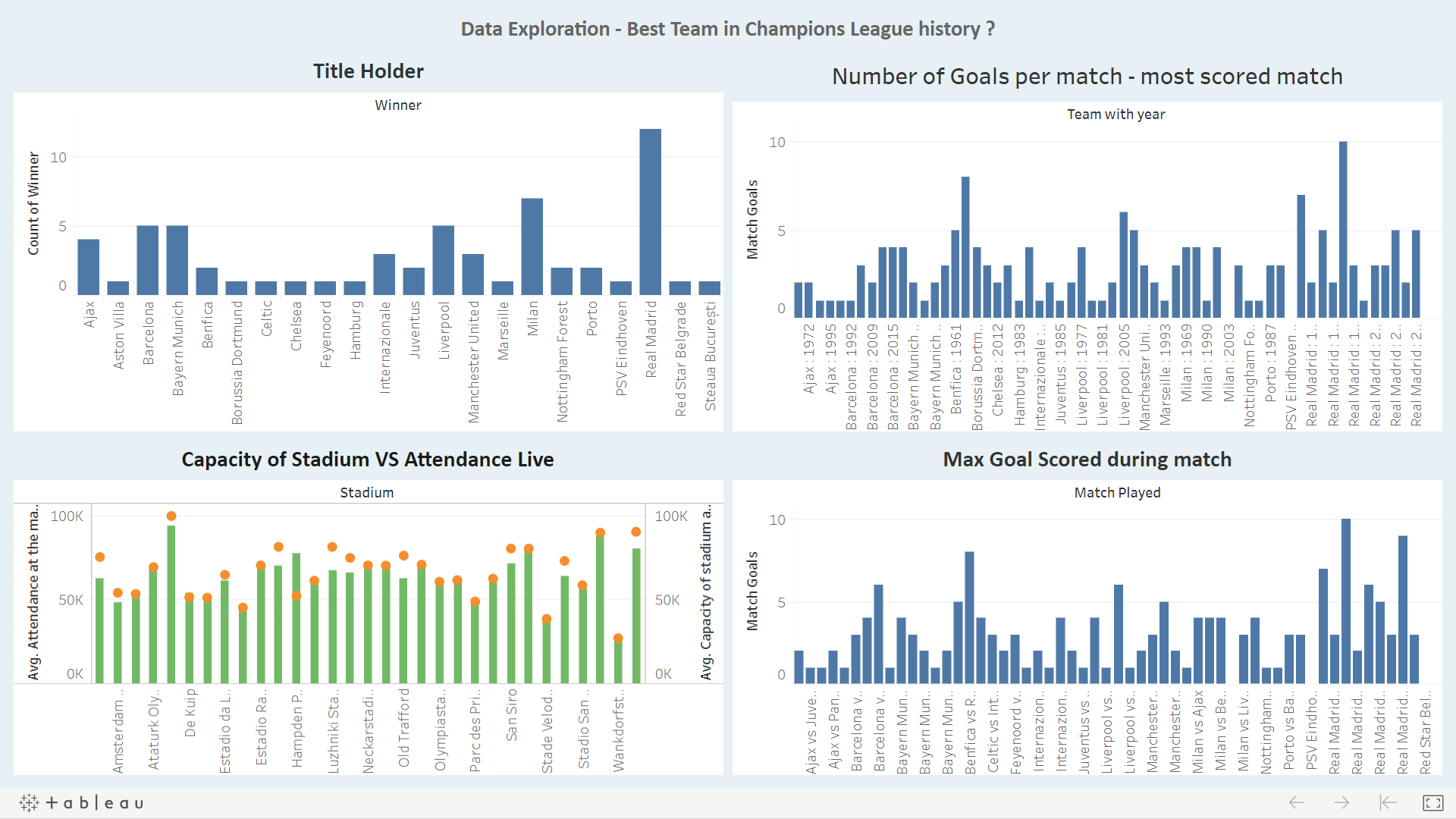
By analyzing above questions, we can definitely answer which team is most successful in terms of wins and popularity in champions league history.

**Data Exploration:**

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

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

Link for Tableau Dashboard: <https://public.tableau.com/views/VisualizationAssignmentFinal_16042689791190/DataExploration?:language=en-GB&:display_count=y&publish=yes&:origin=viz_share_link>



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

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

#### Data file from public repository.

#### Champions League final match Team data from 1956 - 2017

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

#### Champions League final match Stadium data from 1956- 2017

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

#### European Stadium Capacity Data

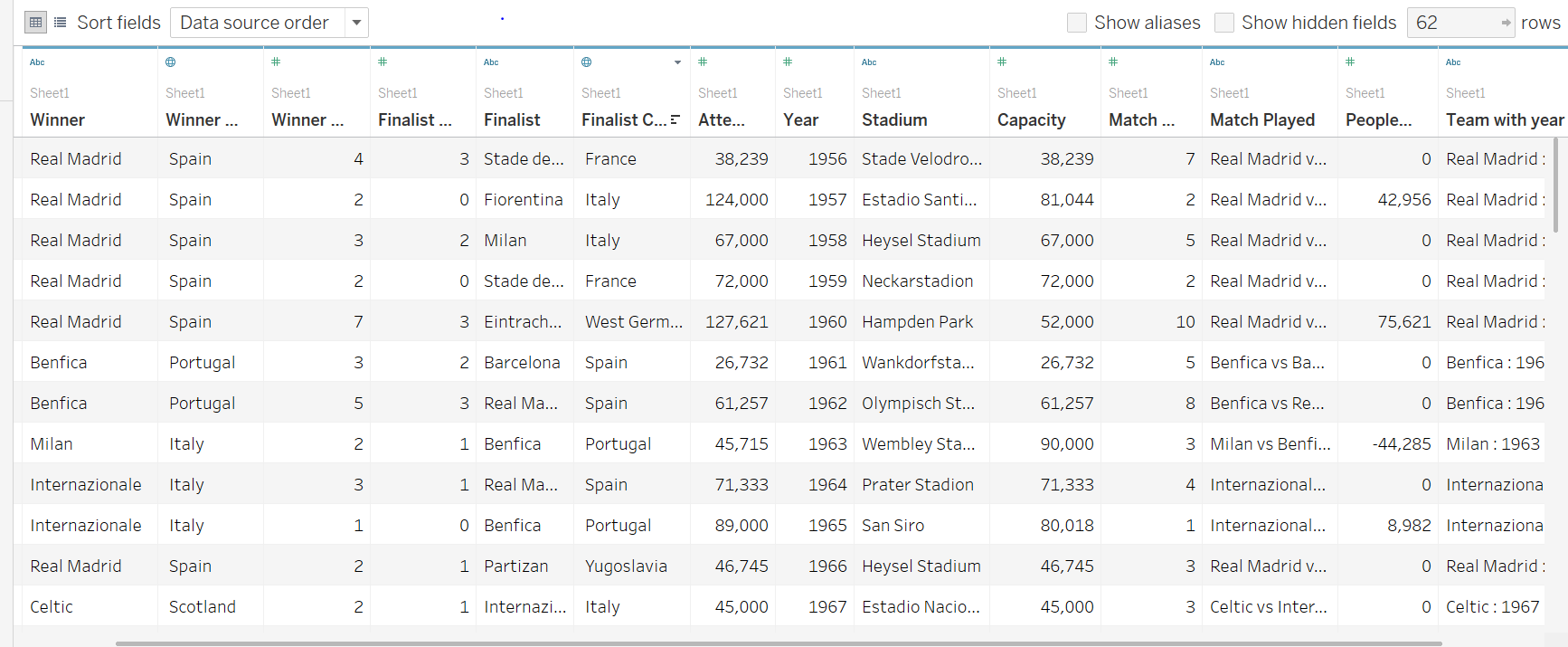
<https://www.stadiumguide.com/figures-and-statistics/lists/europes-largest-football-stadiums/>

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 file link to check from my GitHub account.

#### Exploration in Tableau:

#### Dataset:



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)

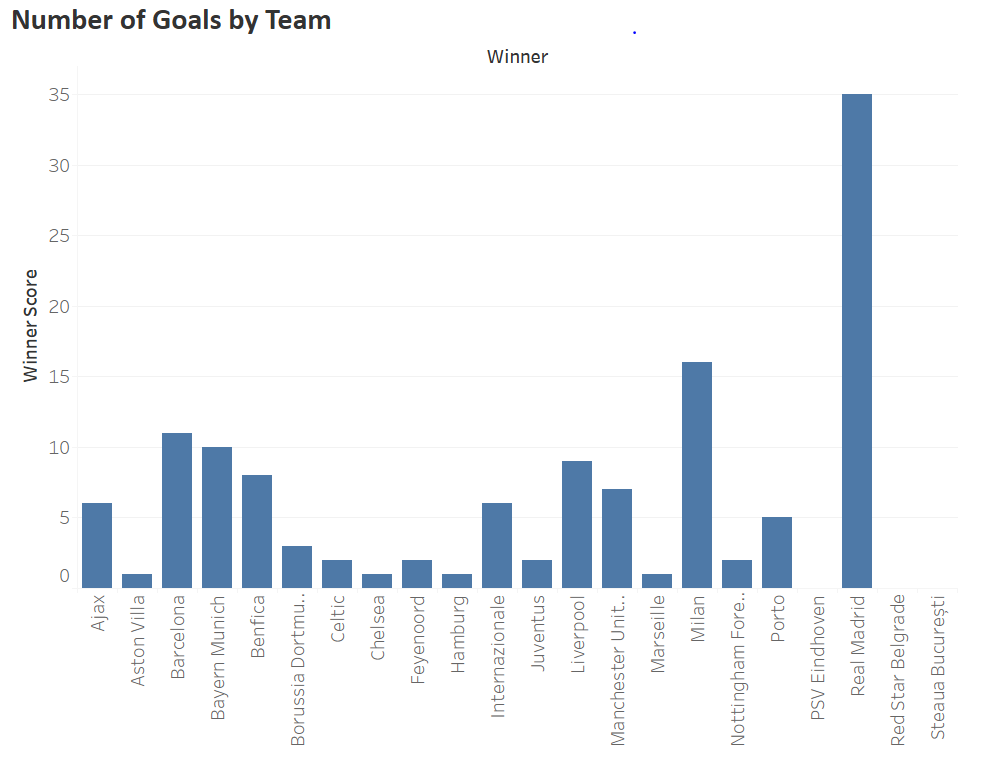
Note: Because of some stadium are not in use and demolished, I assumed capacity of stadium equals to attendees

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

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

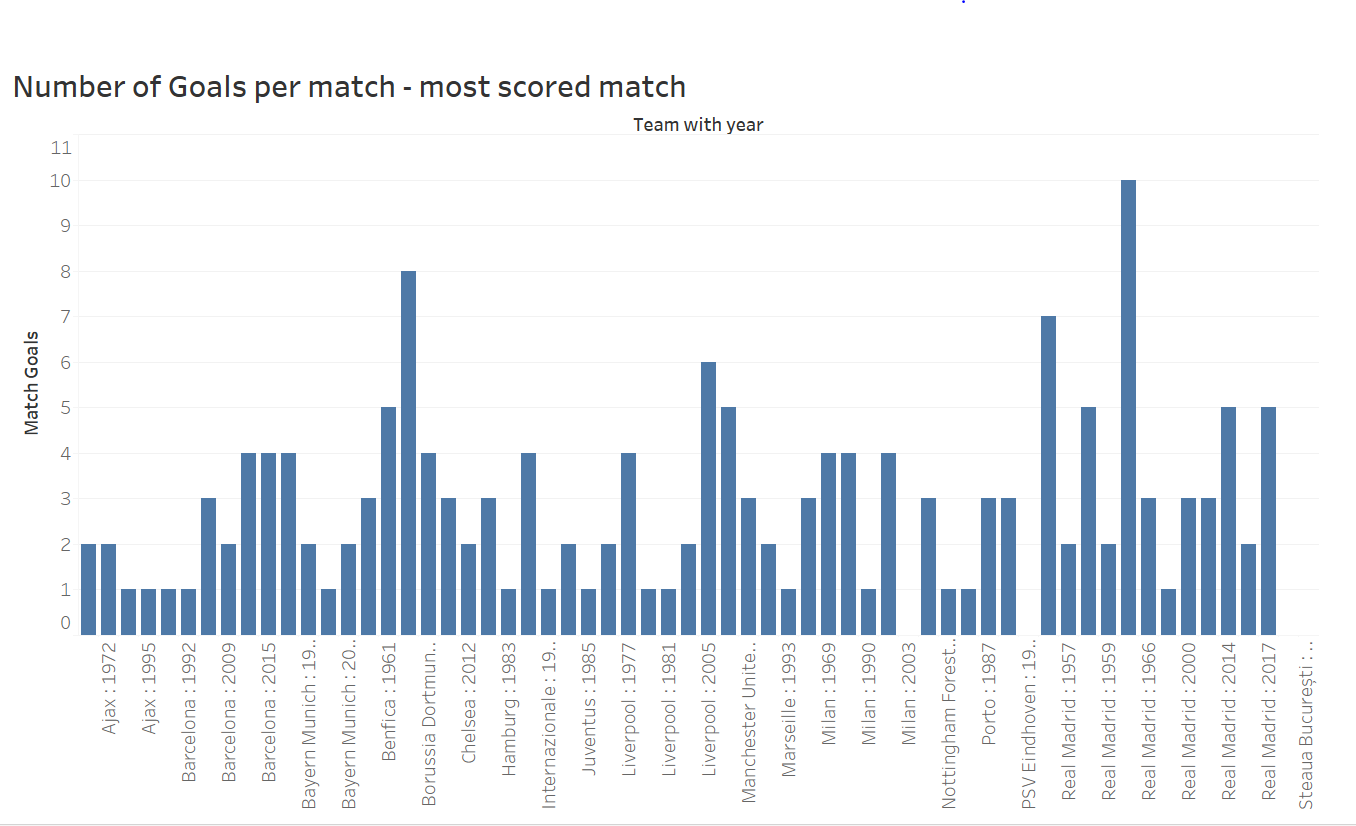
Below graph shows team having total number of goals through out the champions league history.

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



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

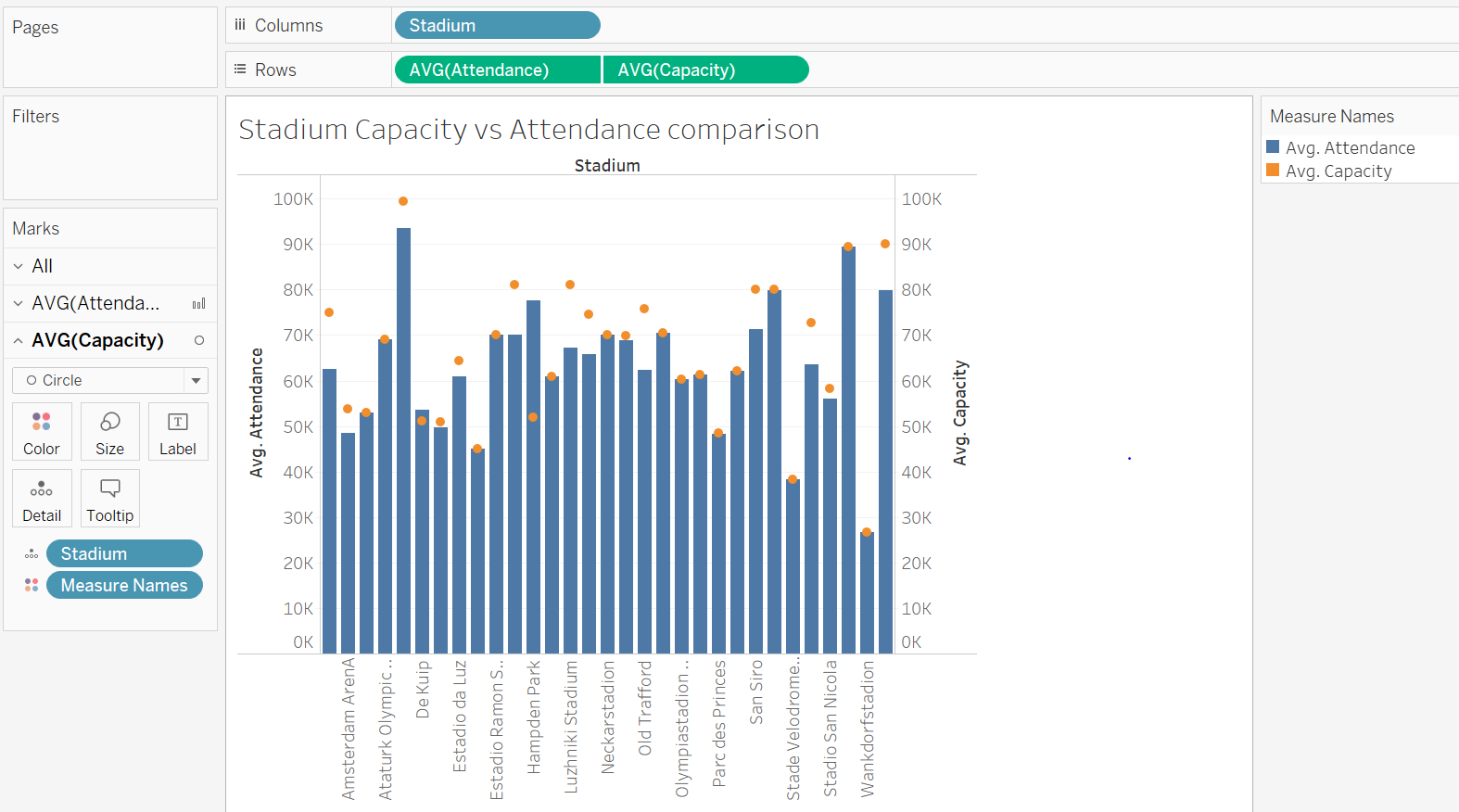
Can from analysis we can say Real Madrid vs Eintracht Frankfurt scored the most 10 goals in single match.



Below graphs the analysis of capacity of stadium vs attendance of public while live match.

As from analysis we can say Hamden park shows the largest difference in capacity vs attendance. As public exceed the capacity to watch match live.

Further we can analyze which team is most popular among people and because of that people are coming in excess.

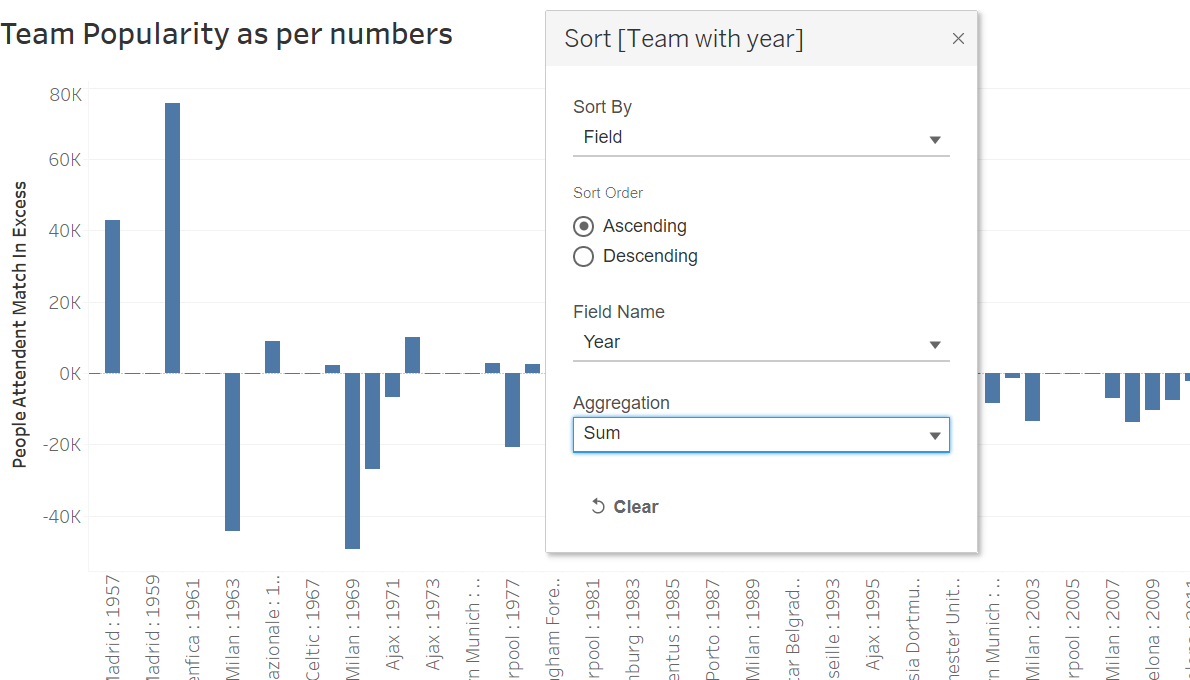


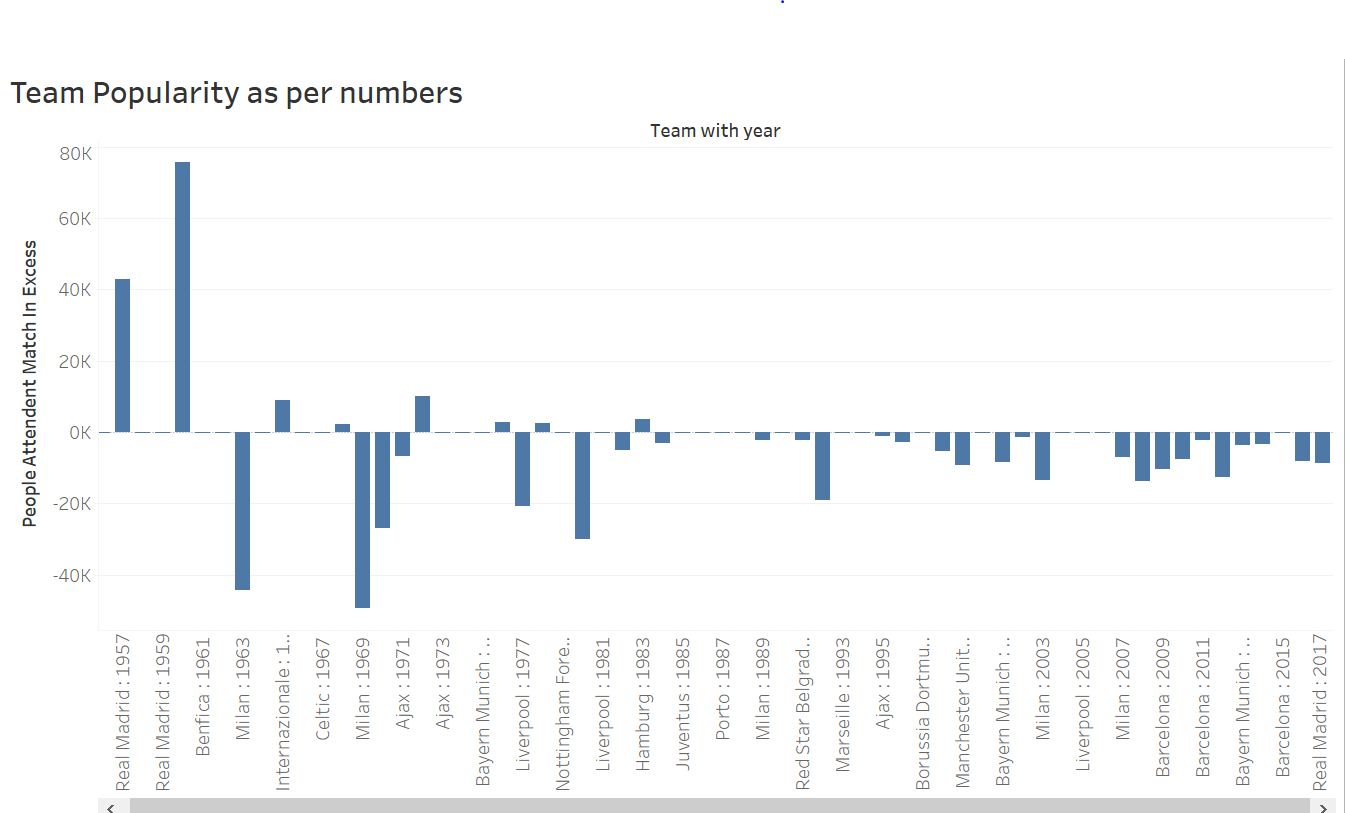
For the Below graph we have calculated the exceed number by

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.

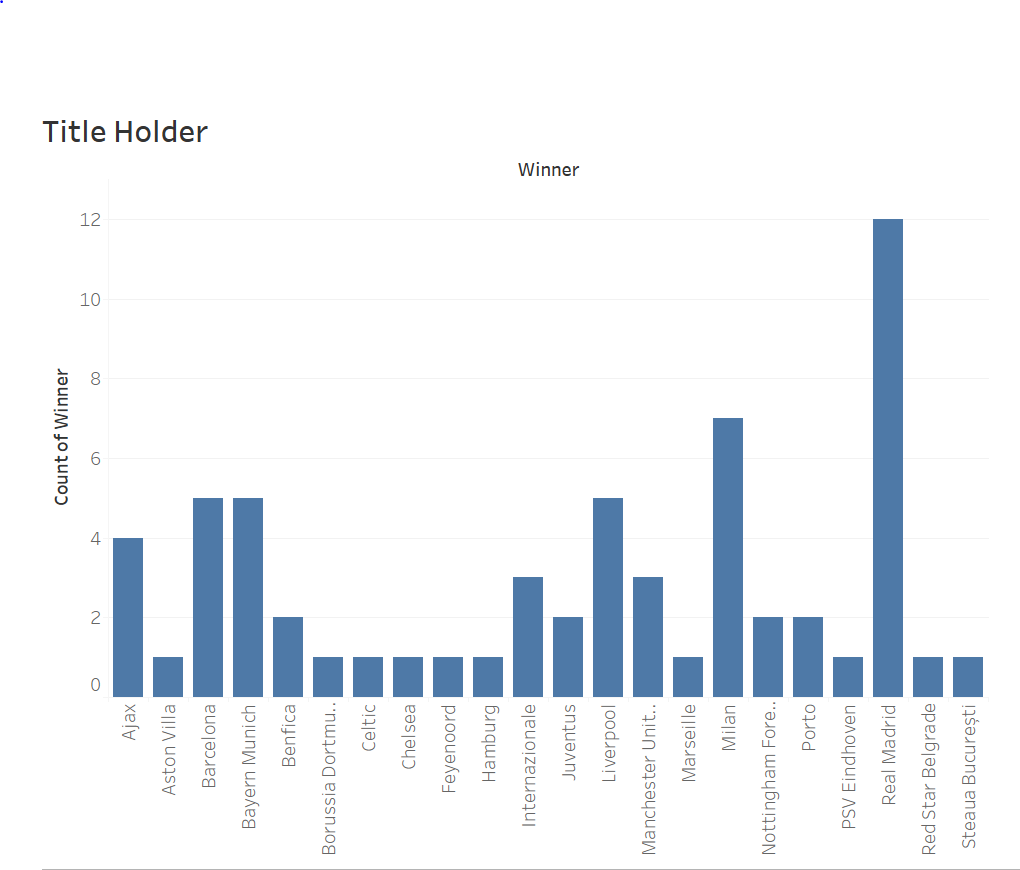
We sorted the graph per year and analyze that Real Madrid having more popular games than anyother team.





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

Real Madrid tops the position with 12 hits.



**Insights from Data:**

Link for Tableau Dashboard: <https://public.tableau.com/views/VisualizationAssignmentFinal_16042689791190/FinalDashboard?:language=en-GB&:display_count=y&publish=yes&:origin=viz_share_link>

\*\* Please watch it in full screen for better clarity

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?

**Topic: Best Team in Champions League history?**

Dashboard:



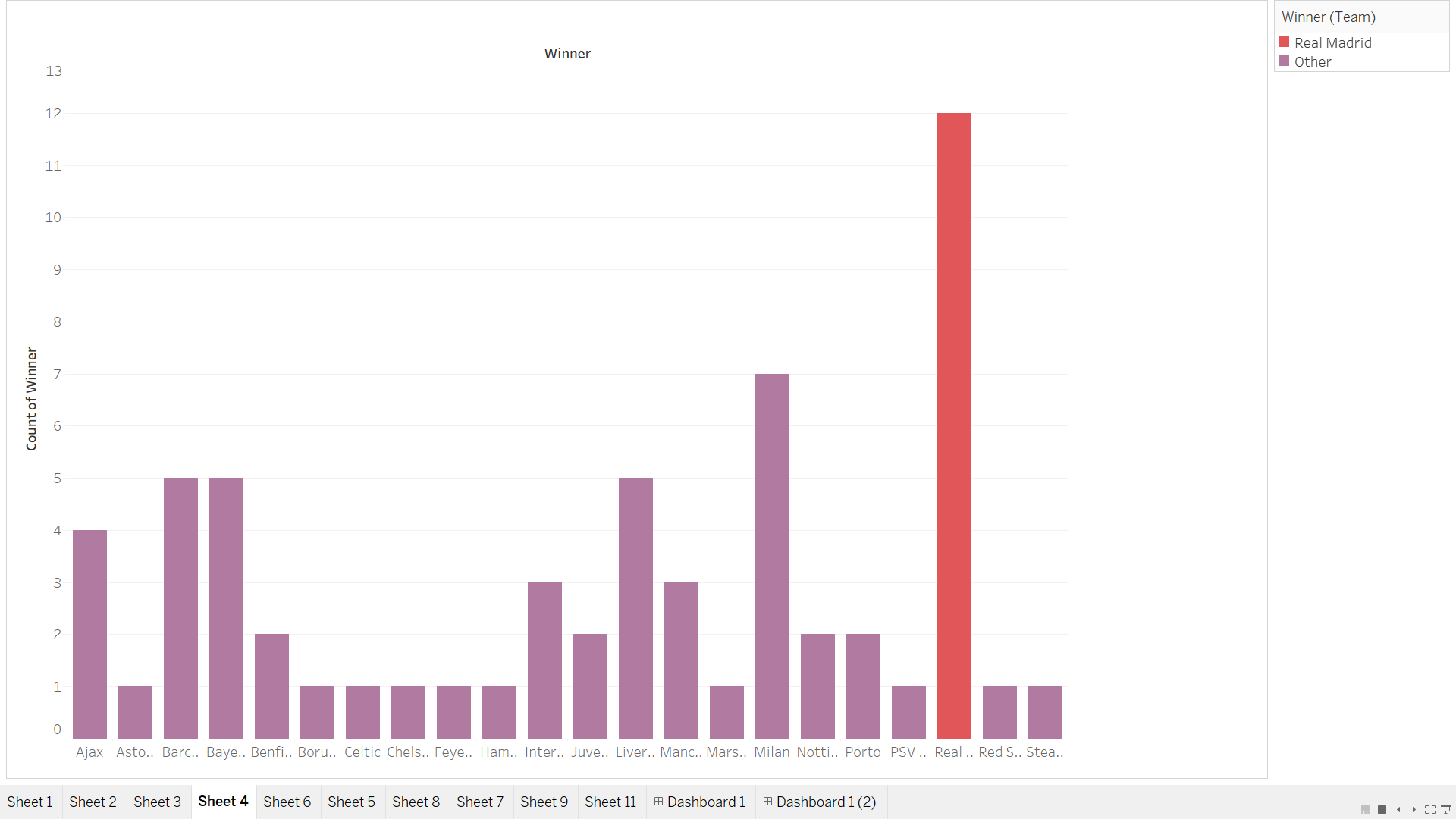
**Insight 1:**

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.**



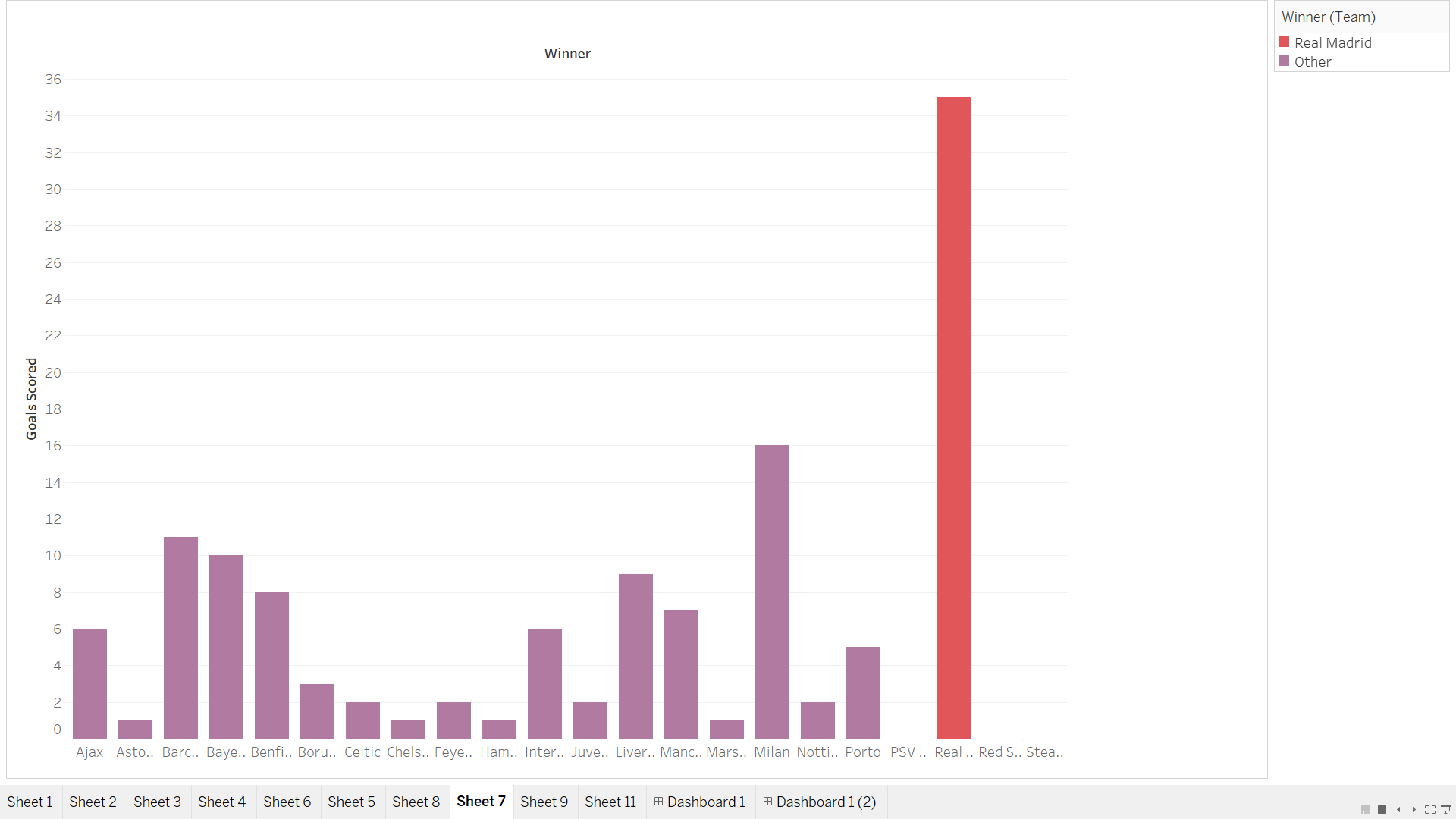
**Insight 2:**

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.**



**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?

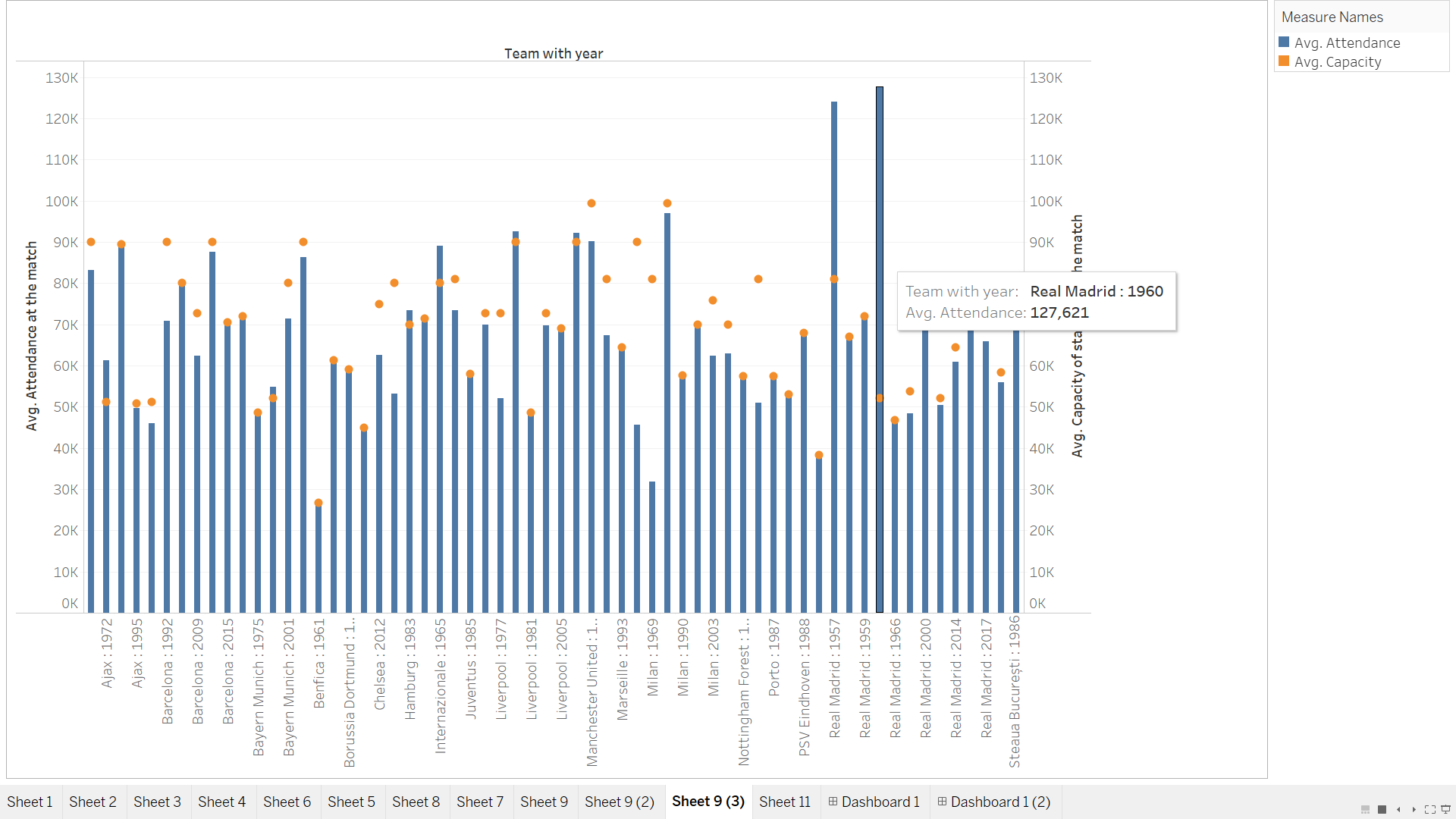
First, we gathered the data, initial file was having the attendance of the public watched the match live. So, we import another file from different data source about stadium per capacity. By merging these two files in python we have final data file having attendees of match in stadium and actual capacity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



I calculated the exceed amount for given particular team to analyze as a whole which team in total has more popularity.

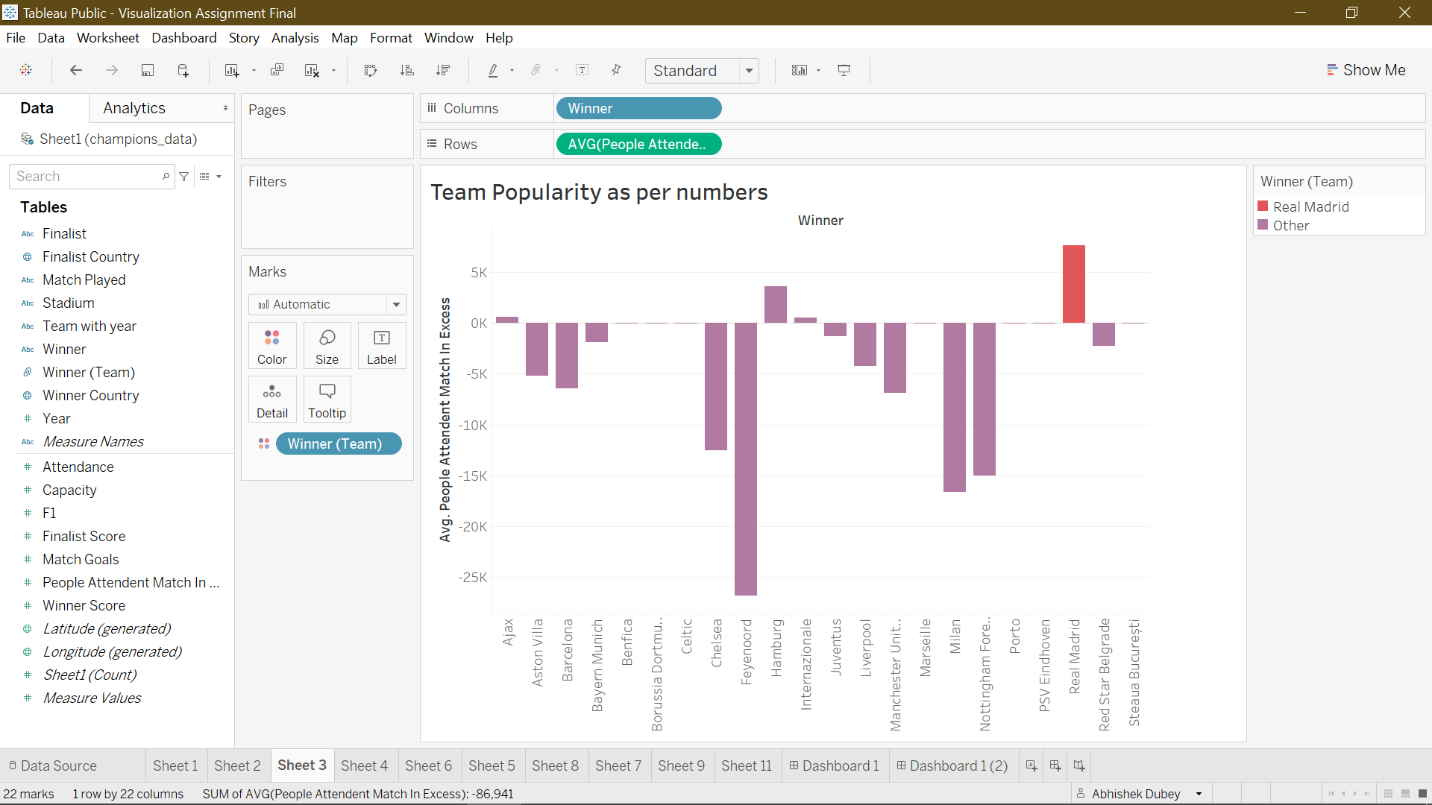
So, I draw the below graph in tableau showing teams with Average capacity and Average attendees’ number.

Highlighted Real Madrid because it is the only team through out champions league history having attendees exceeding the capacity of stadium.



Graph between Avg People attendant match in excess vs Team

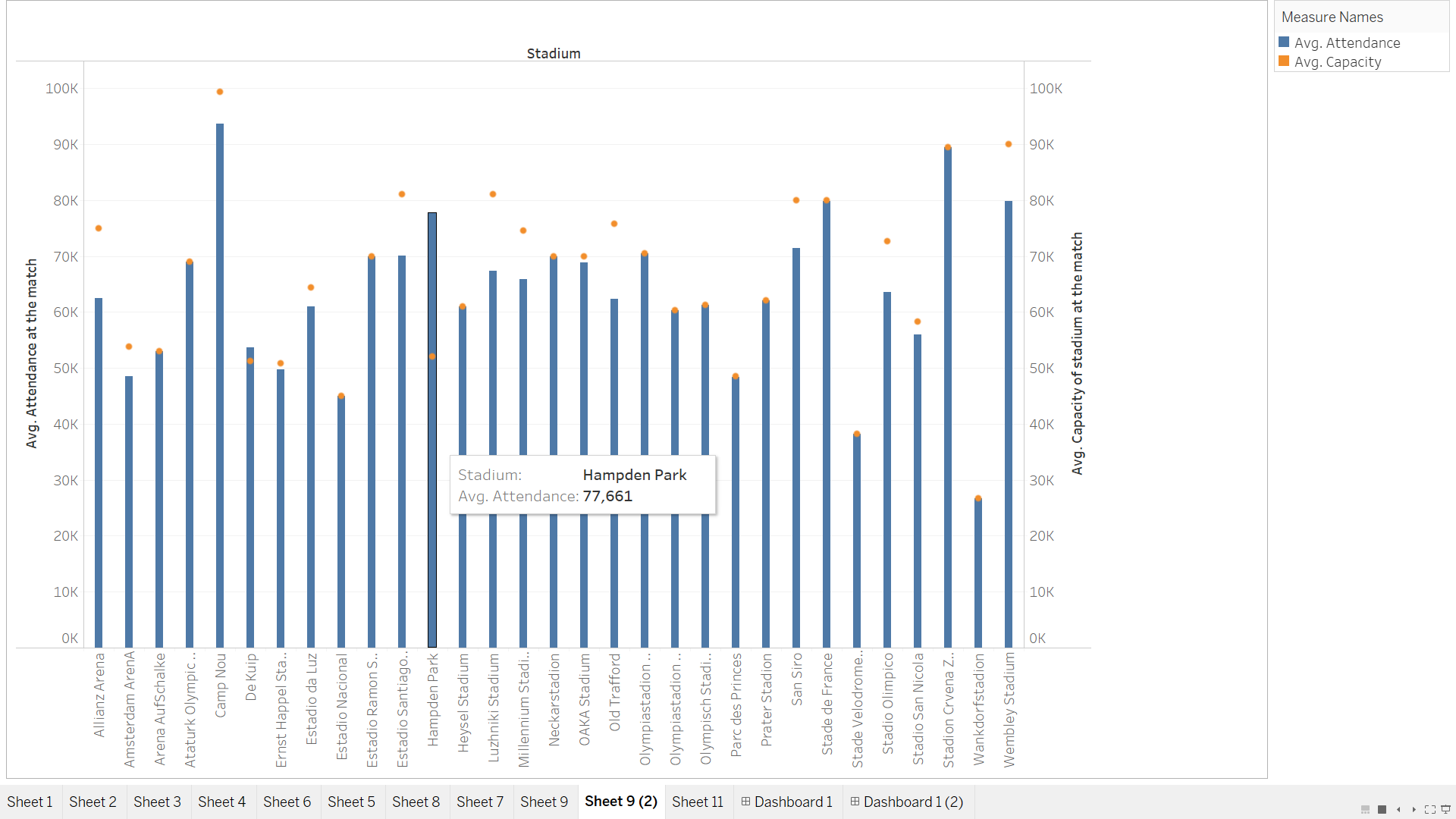
Clearly shows Real Wins the popularity Ranking



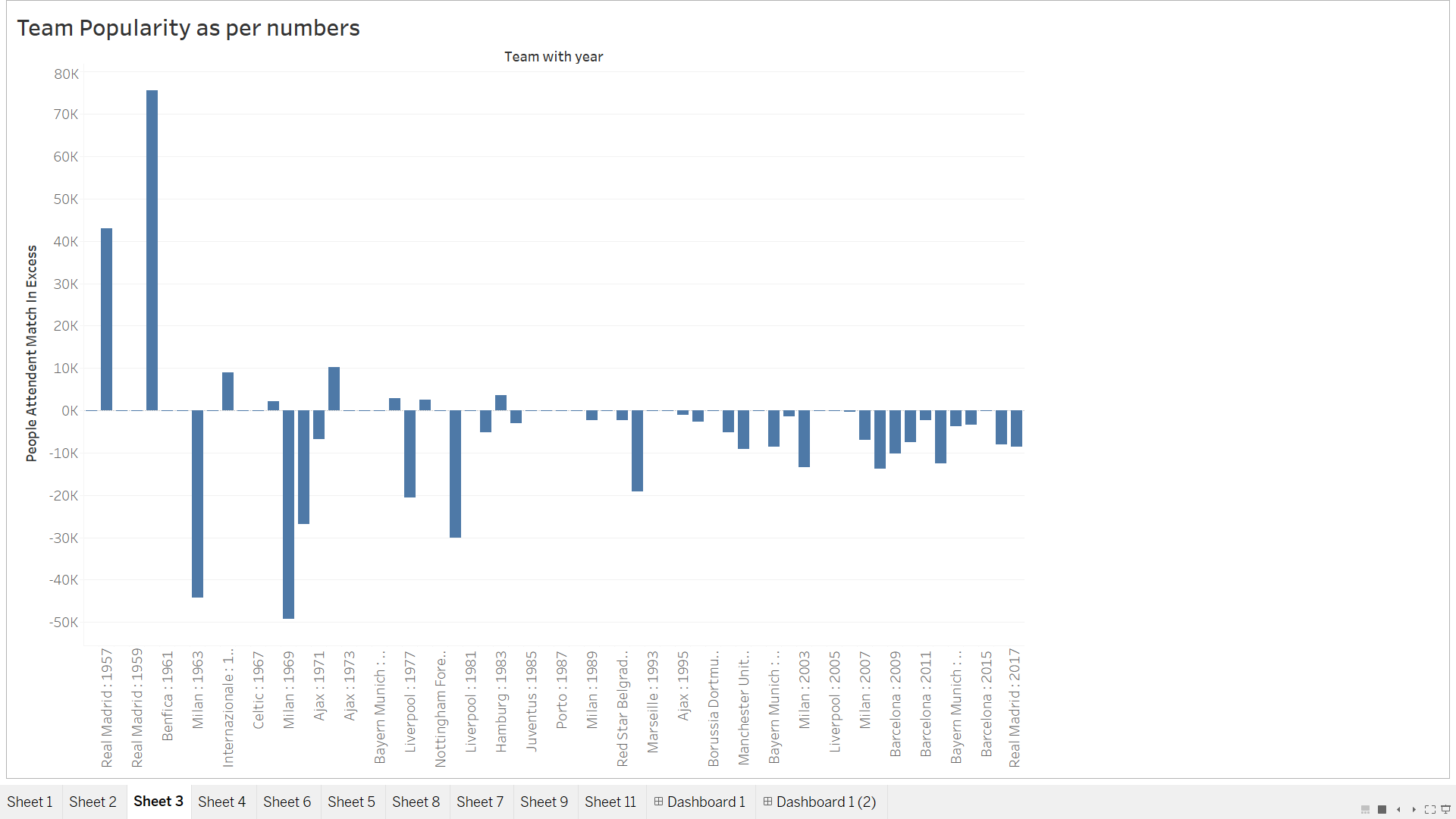
**Previous Iteration:**

Iteration for analyzing the popularity of team.

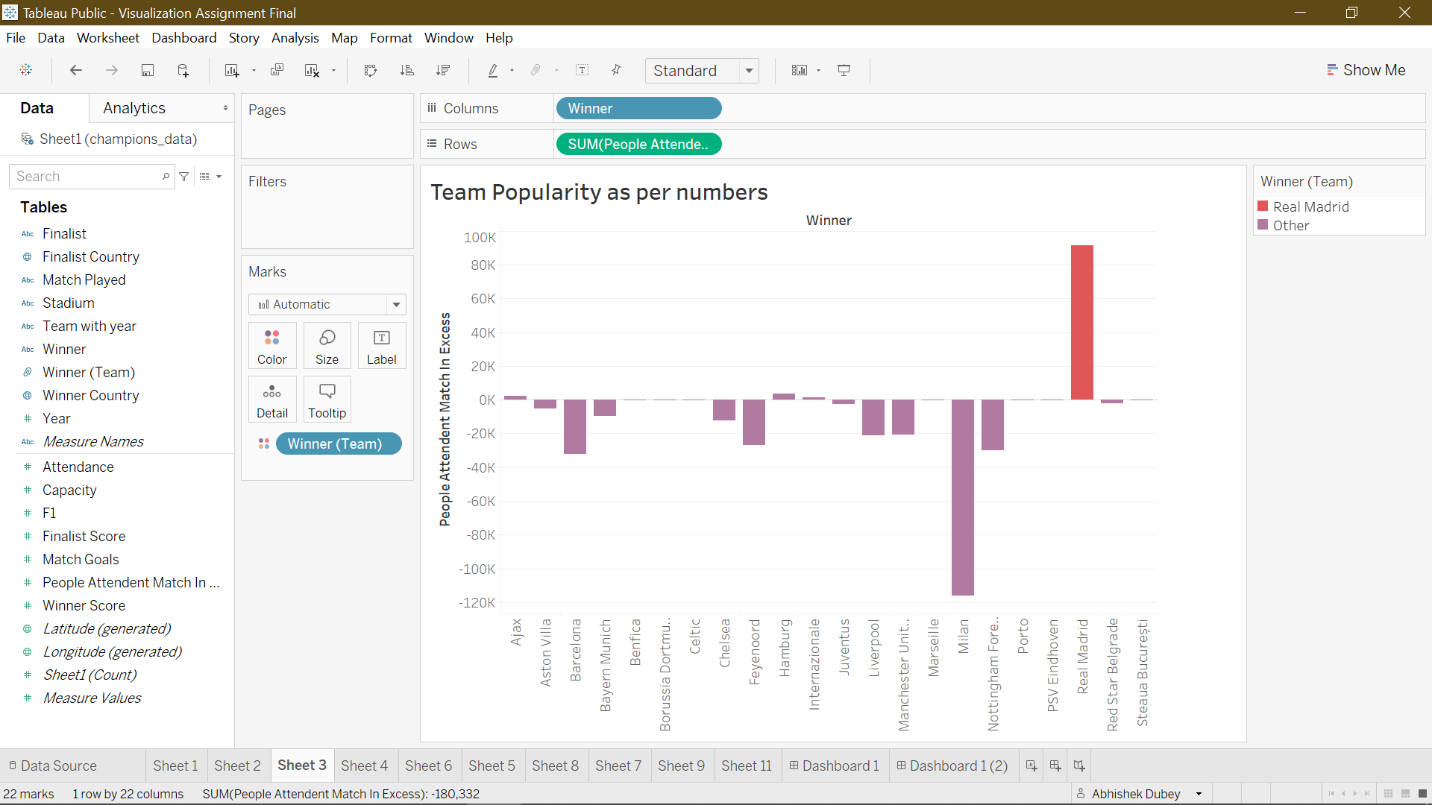
Iteration 1: Stadium vs average attended and average capacity



Iteration 2: We need to analyze particular team

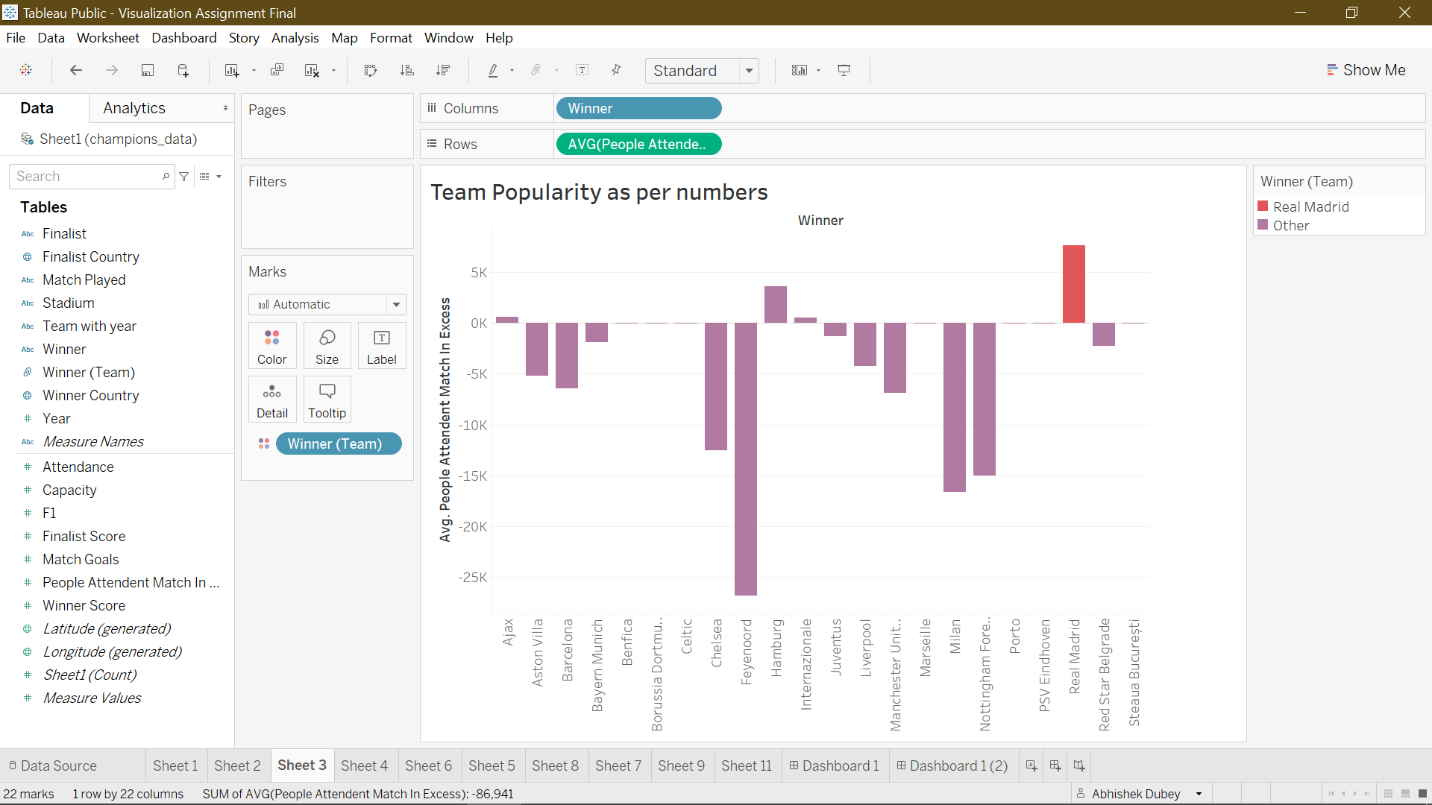


Iteration 3: We need total excess in number for particular team.



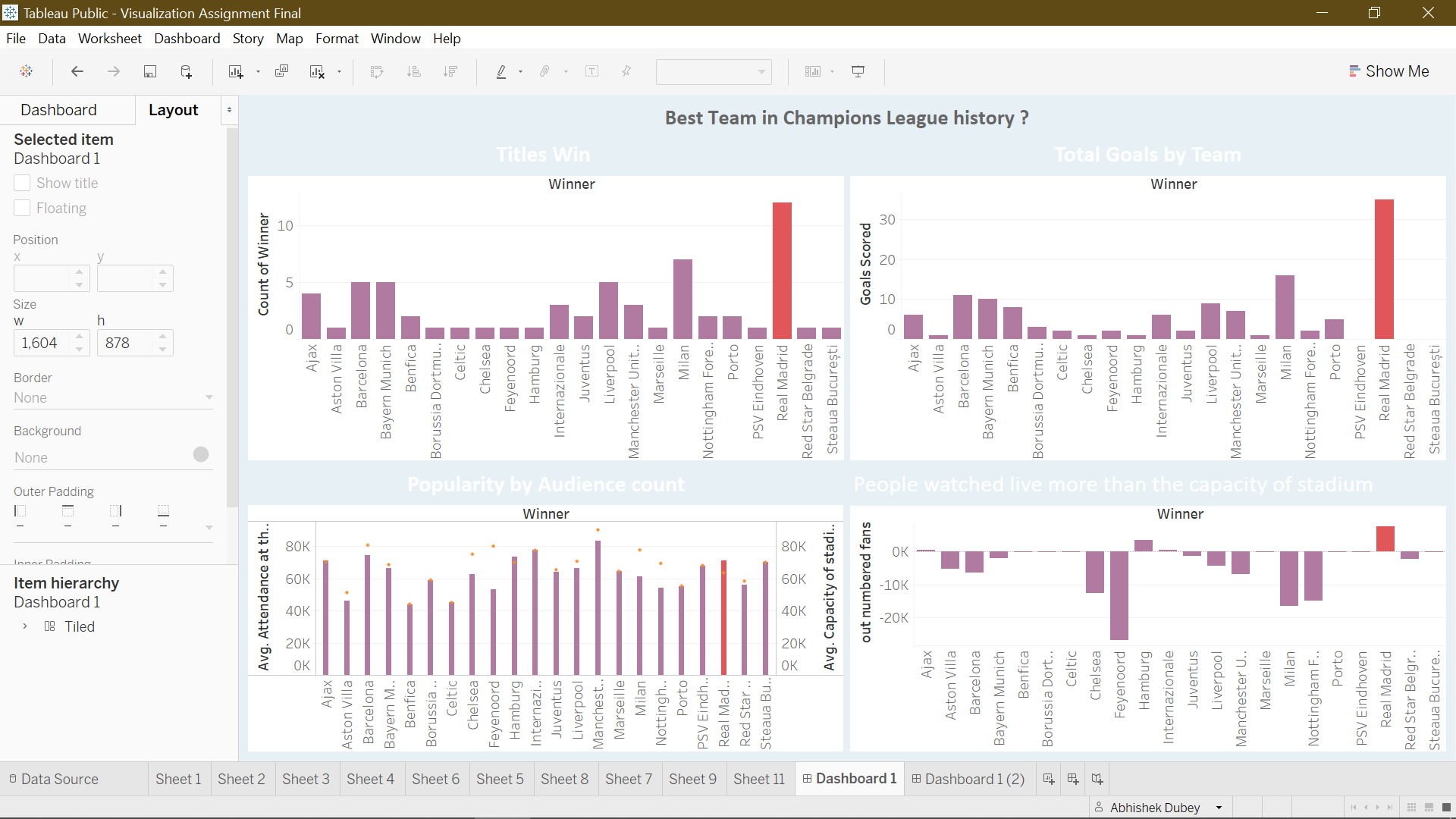
Iteration 4: We need Average for per team

And below is the final graph



**Iteration for Dashboard:**

Iteration 1: By Below graph we cannot analyze the average exceed number for particular team. Difficult to say something about popularity





Final Iteration:

Below graph shows every aspect for our data from winning titles, scoring goals and Public popularity.

So Overall we can say Real Madrid is the best Team in Champions League.

