

# **DATA SCIENCE**

## **PROJECT REPORT**

**TOPIC: The Spanish league (LIGA BBVA) Analysis from Past 47 year's dataset.**

**GROUP MEMBER:**

- **k163762 (Basit Ali Khatri)**

**SECTION: GR2**

**DATASET Description:**

This dataset contains information about a football matches with the follow features:

**1111, 1988-89, 1, 4, FC Barcelona, Real Madrid, 5, 1, 29/11/1988, 247014000**

- id (1111): Unique identifier of each match
- season (1988-89): Season in which the match was played
- division (1): División in which the match was played (1st '1', 2nd '2')
- round (4): round in which the match was played
- localTeam (FC Barcelona): Home Team
- visitorTeam (Real Madrid): Visitor Team
- localGoals (5): Goals scored by the home team
- visitorGoals (1): Goals scored by the Visitor team
- date (29/11/1988): Date in which the match was played

- timestamp (247014000): Timestamp in which the match was played

## PROJECT DESCRIPTION:

Data Set with the football matches of the Spanish league of the 1st and 2nd division from the 1970-71 to 2016-17 season, has been created in order to analyze the performance of different teams in different divisions season-wise. Moreover to analyze the winners of Spanish league each year and the dominant teams of Spanish League.

## Data Cleaning:

In the below picture it is mentioned that our original dataset contains no null or missing values for any columns. In total there are 10 Columns and each column have 37147 Values.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37147 entries, 0 to 37146
Data columns (total 10 columns):
id                37147 non-null int64
season            37147 non-null object
division          37147 non-null int64
round             37147 non-null int64
localTeam         37147 non-null object
visitorTeam       37147 non-null object
localGoals        37147 non-null int64
visitorGoals      37147 non-null int64
date              37147 non-null object
timestamp         37147 non-null int64
dtypes: int64(6), object(4)
memory usage: 2.8+ MB
Program Ended Gracefully.
```

## Data Formatting:

In order to format data and make data more clean and organized to manipulate it I have also format some of the columns of my dataset which description is given below.

---

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37147 entries, 0 to 37146
Data columns (total 10 columns):
id                37147 non-null int64
Season            37147 non-null object
division          37147 non-null int64
MatchDay          37147 non-null int64
home_team         37147 non-null object
away_team         37147 non-null object
Home_Goals        37147 non-null int64
Away_Goals        37147 non-null int64
date              37147 non-null object
timestamp         37147 non-null int64
dtypes: int64(6), object(4)
memory usage: 2.8+ MB
Program Ended Gracefully.
```

## Exploratory Data Analysis:

I have not done the proper EDA because of my nature of data. As we know that EDA is mainly used to detect outliers and unusual values in our dataset but because of the nature of my dataset there can be unusual values which can't be omitted.

## Visualization of Data And Findings:

Now we will start visualizing our data.

- **Standings of past 47 years:**

Firstly the Standings of each Laliga (Division=1) is generated in the separate .csv file in the directory (Project\_Output\Spanish\_League\_1\_Analysis\Standings).

One instance of Standings file looks like this.

Season	16	non-null	object
Football_Club	16	non-null	object
Home_Wins	16	non-null	int64
Away_Wins	16	non-null	int64
Match_Drawn_x	16	non-null	int64
Match_Drawn_y	16	non-null	int64
Home_Losses	16	non-null	int64
Away_Losses	16	non-null	int64
Total_Matches_Win	16	non-null	int64
Total_Matches_Lost	16	non-null	int64
Total_Matches_Drawn	16	non-null	int64
total_matches	16	non-null	int64
Points	16	non-null	int64
home_goals	16	non-null	int64
away_goals	16	non-null	int64
Goals_Against_x	16	non-null	int64
Goals_Against_y	16	non-null	int64
Goals_for	16	non-null	int64
Goals_Against	16	non-null	int64
Goal_Difference	16	non-null	int64

- **Head-to-Head Statistics:**

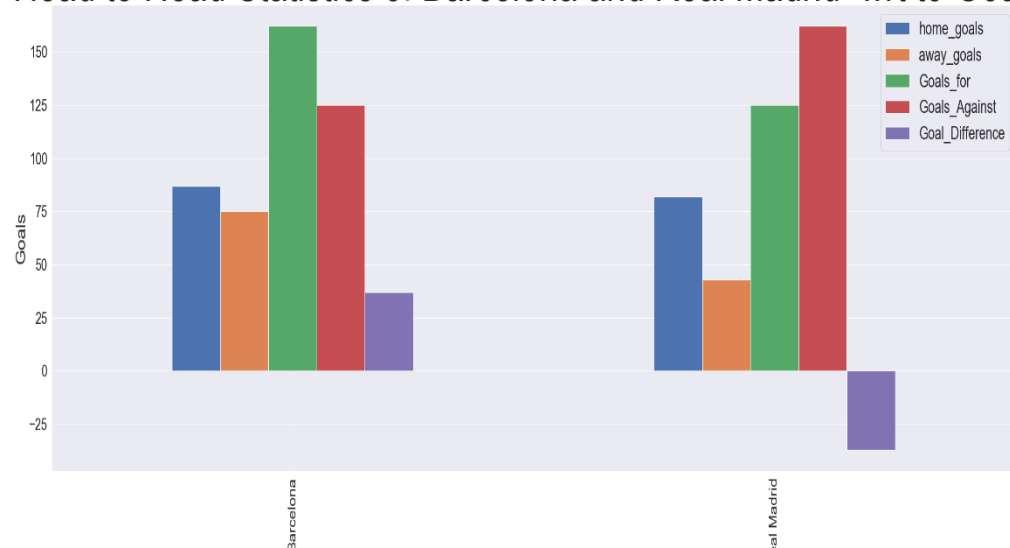
In football head-to-head statistics are very important in order to compare between the two teams. Therefore in order to compare teams I have also show how teams performance with respect to goals and results differ from the other teams.

The head-to-head statistics can be found under  
(Project\_Output\Spanish\_League\_1\_Analysis\Figures\Head\_to\_Head\_Statistics)

For instance :

### With Respect To Goals :

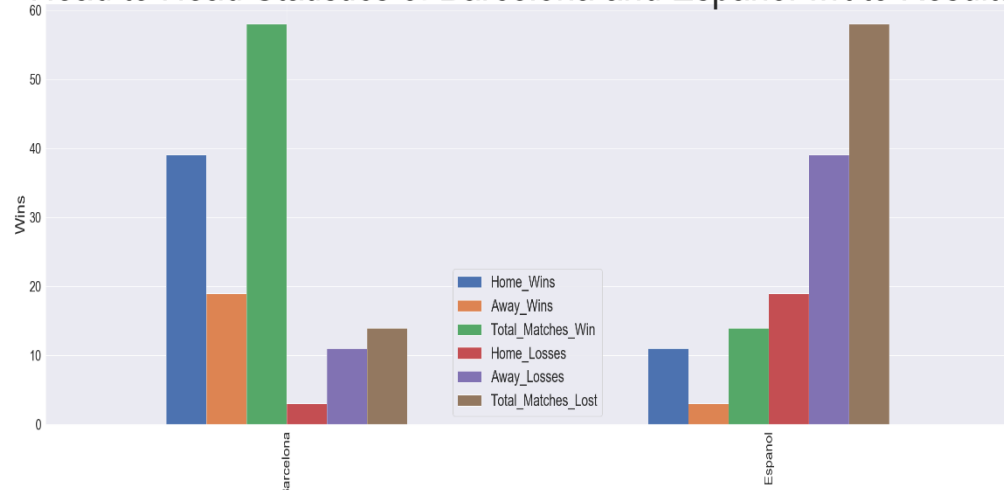
Head to Head Statistics of Barcelona and Real Madrid wrt to Goals.



The result shows that Barcelona have scored more goals against Real Madrid because of negative goal difference between them and Barcelona have wins in away and home goals also over Real Madrid.

### With Respect to Results:

Head to Head Statistics of Barcelona and Espanol wrt to Results.

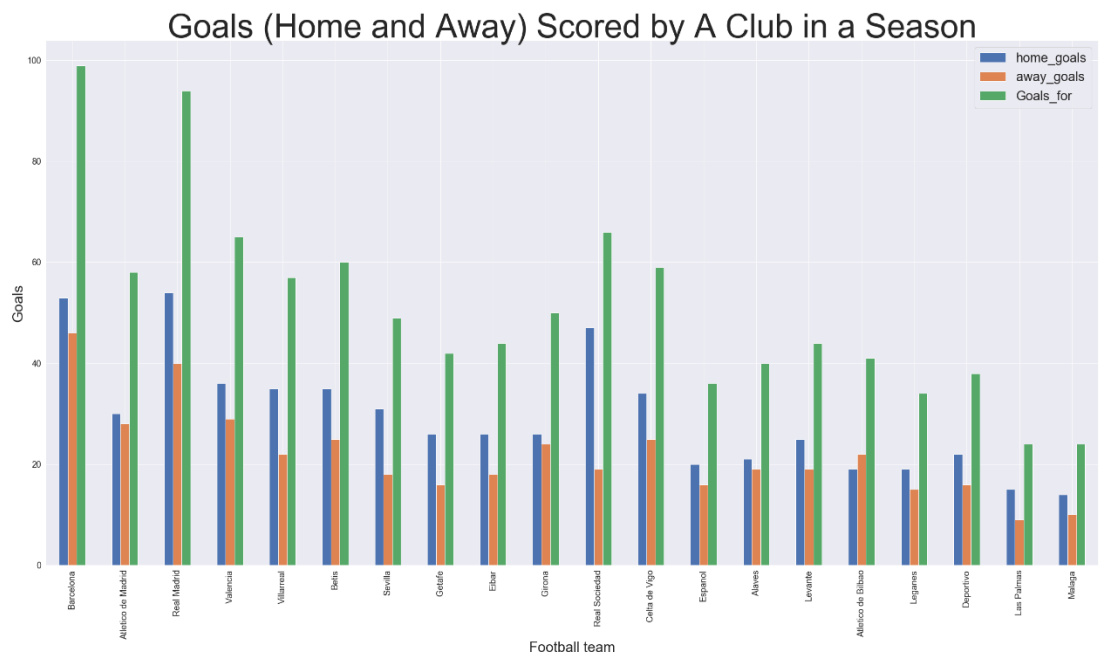


This clearly shows that how two teams have performed against each other in past years and which side is dominant ,For example in the above picture we can clearly see that Barcelona have won more matches (away+home+total) then Española.

- **Season-Wise Total Goals (Home and away):**

Now we will find out How many goals were scored in a particular season (Home and away). The output of this can be found under :  
(Project\_Output\Spanish\_League\_1\_Analysis\Figures\Season\_Wise\_Total\_Goals)

For instance : Season (2017-2018)



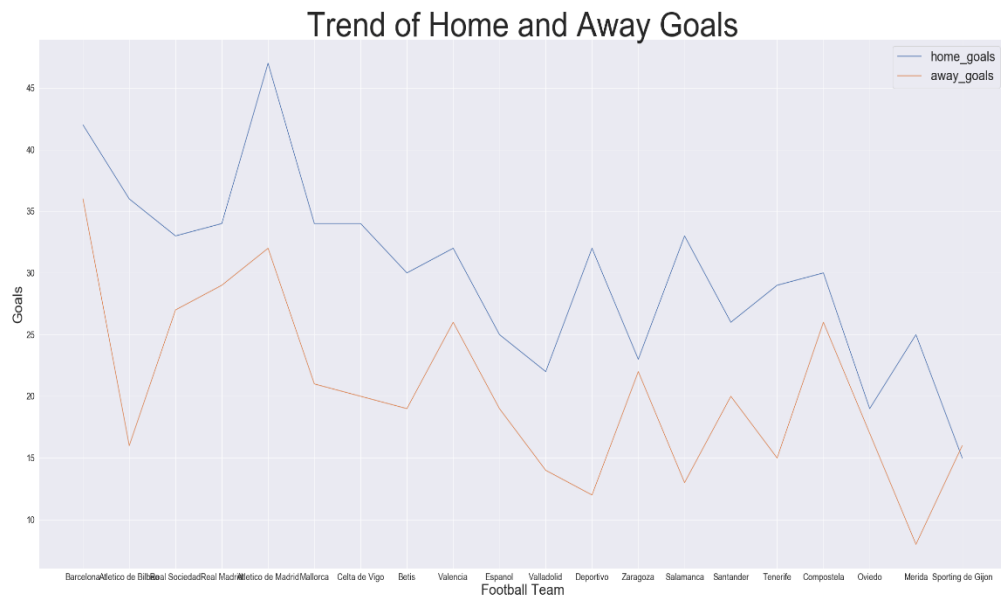
- **Trend of Home and Away goals:**

The following image shows how home and away goals vary by team to team in a particular season.

Directory:

(Project\_Output\Spanish\_League\_1\_Analysis\Figures\Season\_Wise\_Trend\_of\_Home\_Away\_Goals)

For instance: Season (1997-1998)



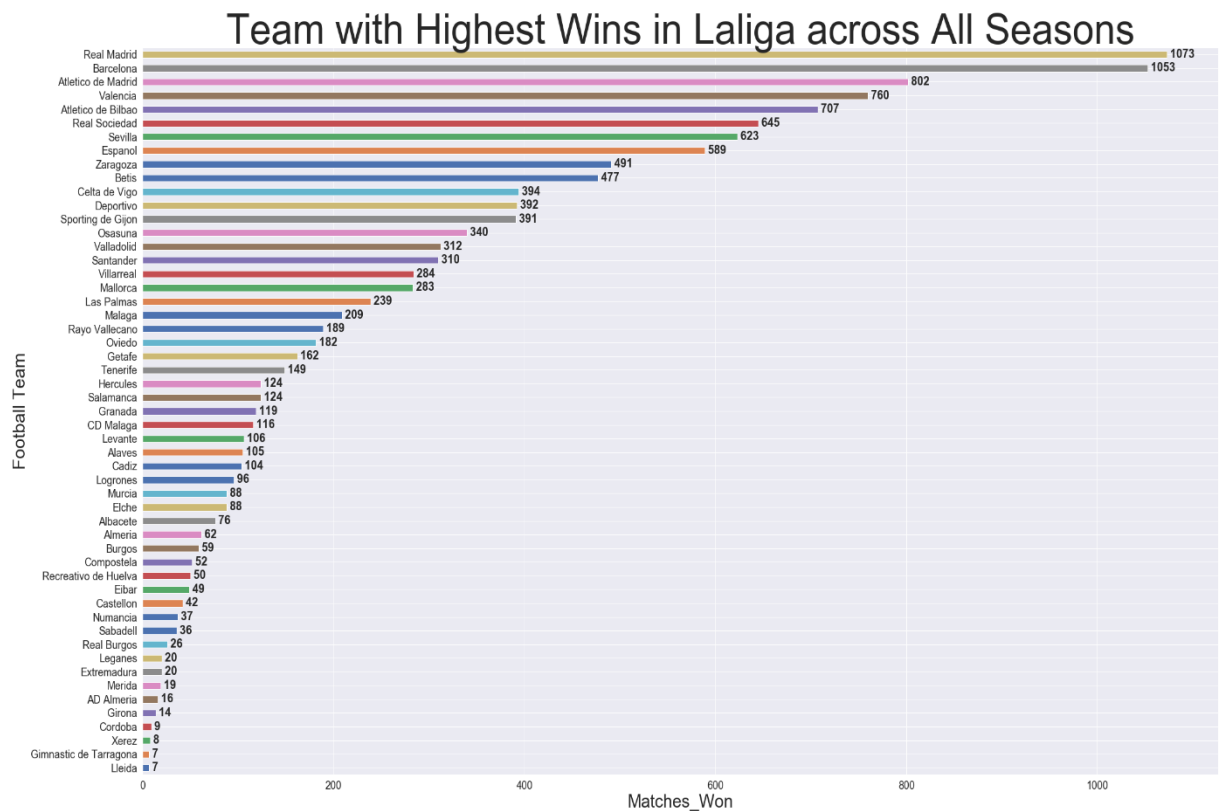
As we can see that no team has scored more away goals then home goals. So we can conclude that Team often score in their home and relatively less in their away game.

- **All Seasons Combined Stats for All the teams :**

Directory :

(C:\Users\basit\Project\_Output\Spanish\_League\_1\_Analysis\Figures\All\_Seasons\_Combined\_Stats)

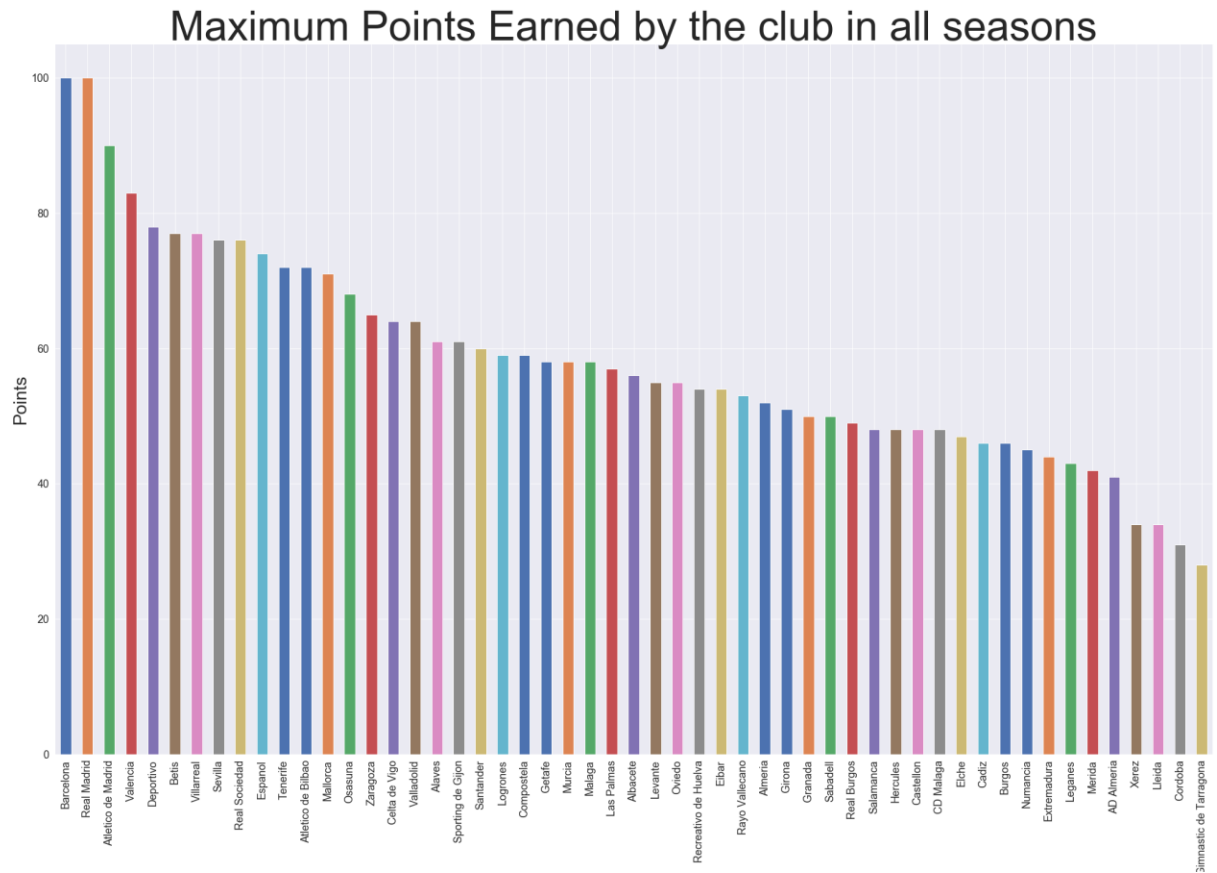
**Teams with number of Laliga Wins :**



From this picture we can clearly see that Real Madrid and Barcelona have registered the greatest number of wins across all seasons.



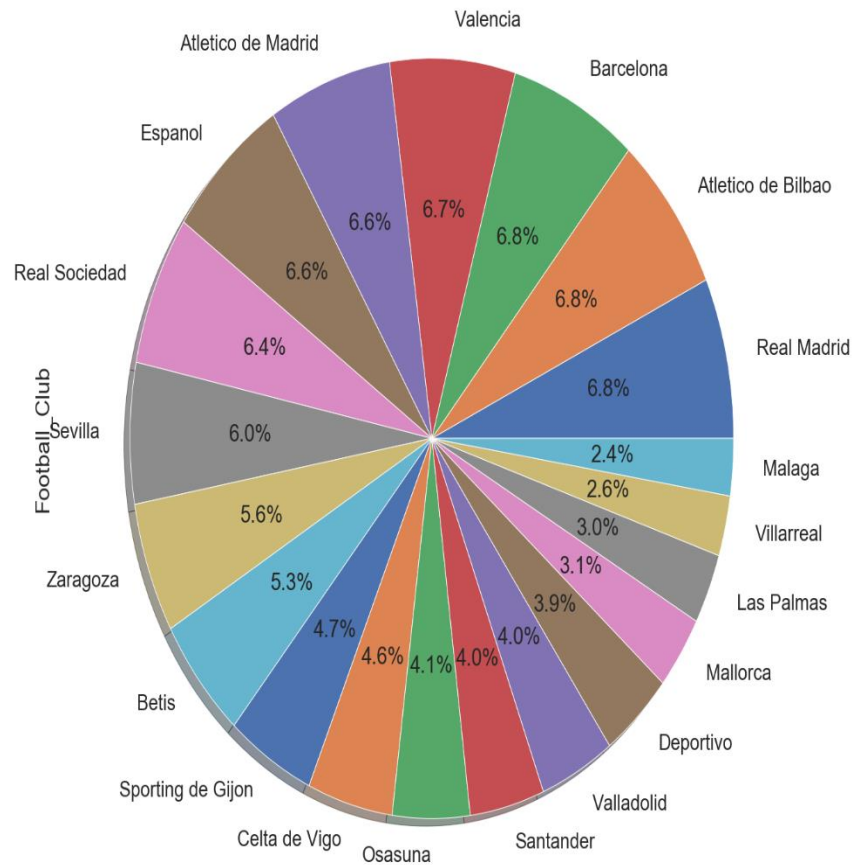
**Team with Maximum Points across all seasons:**



From this picture we can clearly see that Real Madrid and Barcelona have registered the maximum of 100 points across all seasons.

## Top 20 teams with maximum percentage of Laliga Appearance:

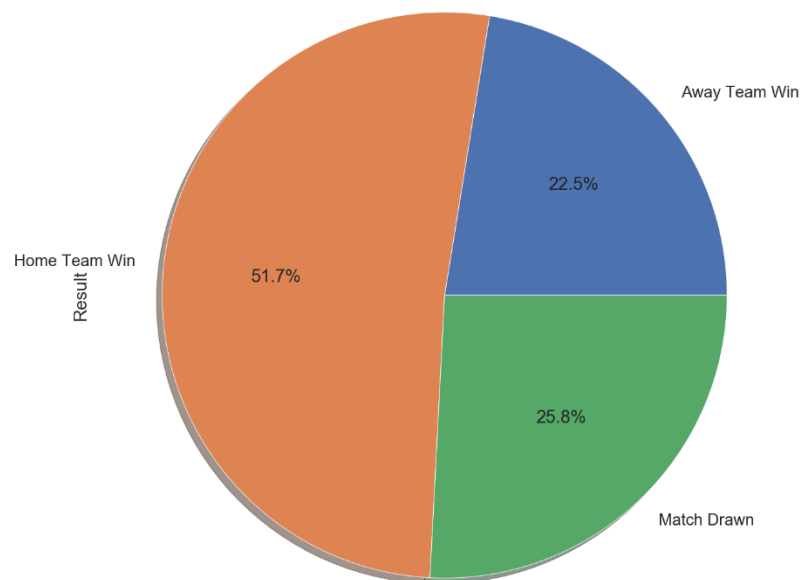
### Top 20 Team with Numbers of Laliga BBVA they Played



From this picture we can clearly see that Real Madrid and Barcelona have 6.8% of Laliga appearance overall which is the maximum of all.

### Matches Outcome across All seasons:

Percentage Showing Which team Won in a Laliga BBVA match



From this picture we can see that Across all seasons of laliga Home team have won around 51.7% and away team has just won 22.5 % which is relatively very less infect less than half of home wins.

## **CONCLUSION:**

In order to conclude our results we can state that In laliga Past 47 years The home team have always being dominant regardless of the quality of the team and have scored average more goals then the Away team and moreover Real madrid , Barcelona and Atletico Madrid have been the three giants of this league over the years .

And in terms of winning the laliga Real Madrid have won most of them in these years whereas Barcelona was second to it.