- Analyzed Dataset

For this first project, I analyzed and investigated the European Football League dataset.

- Questions Posed

1. Which League Scored the Most Goals in the time period between 2008 and 2016?
2. What teams won the most games?
3. Does a Correlation Exist between Players' Average Overall Rating and their average Potential Rating?

- Investigation Descriptions

Which League Scored the Most Goals in the time period between 2008 and 2016?

To answer this question, I evaluated the Total Number of goals scored in Each Match(matches_dataset) and then group the tally of total goals scored by their specific European Football Leagues.

What teams won the most games?

I first created a column to hold the names of winning teams and DRAW if teams had drawn, then filtered out the draw games and remained with those that had named winners. I then grouped the dataset by seasons and counted the number of games won in each season. Juventus had the most wins of 33 in the 2013/2014 Season.

Does a Correlation Exist between Players' Average Overall Rating and their average Potential Rating?

I plotted a scatter plot to visualise the relationship between the two features in question. I discovered that the scatter plot implies a Positive Correlation that as Player

overall increases, so does their potential. Therefore players with greater overall rating often have higher potential ratings.

Data Wrangling Performed;

1. Check for Missing Values across the Teams, Players, and Matches dataset.
2. Check for Duplicates
3. Fix Datatype Issues

Wrangling Process:

I merged the necessary tables to create a more inclusive table to allow for deeper analysis.

First, check for missing values across the Teams, Players and Matches datasets was performed. The Players and Matches data had __no missing values__ and __no duplicate entries__. Howerver, The Teams' Dataset contained __1 duplicate row__ and __969 Missing values__.

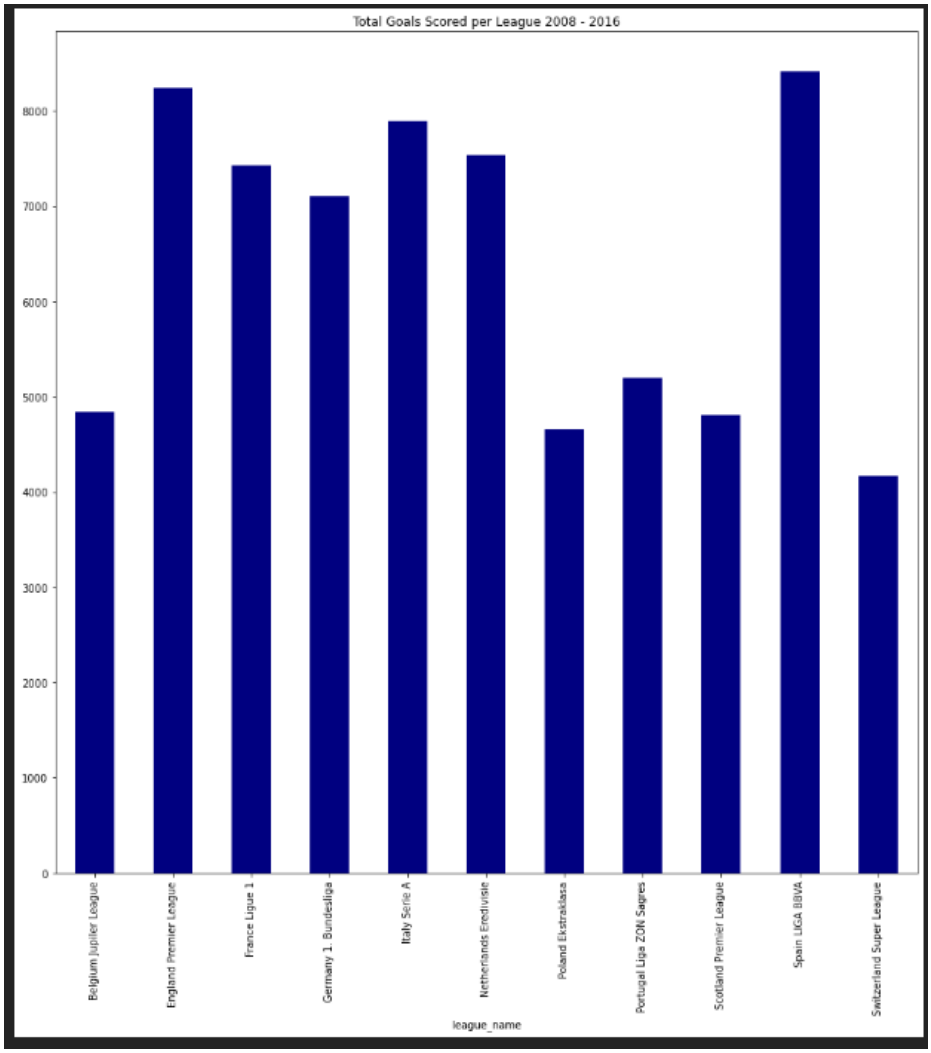
The duplicate row was dropped and the missing values from the __buildUpPlayDribbling__ feature were filled in with the mean/average values of the specified column.

Lastly, there were instances of inaccurate datatypes in place for the Match_Data's __game_day__ column/feature as well as the Team_Data's __date__ which were tagged as `objects` instead of a `datetime` features. Their datatypes were corrected accordingly.

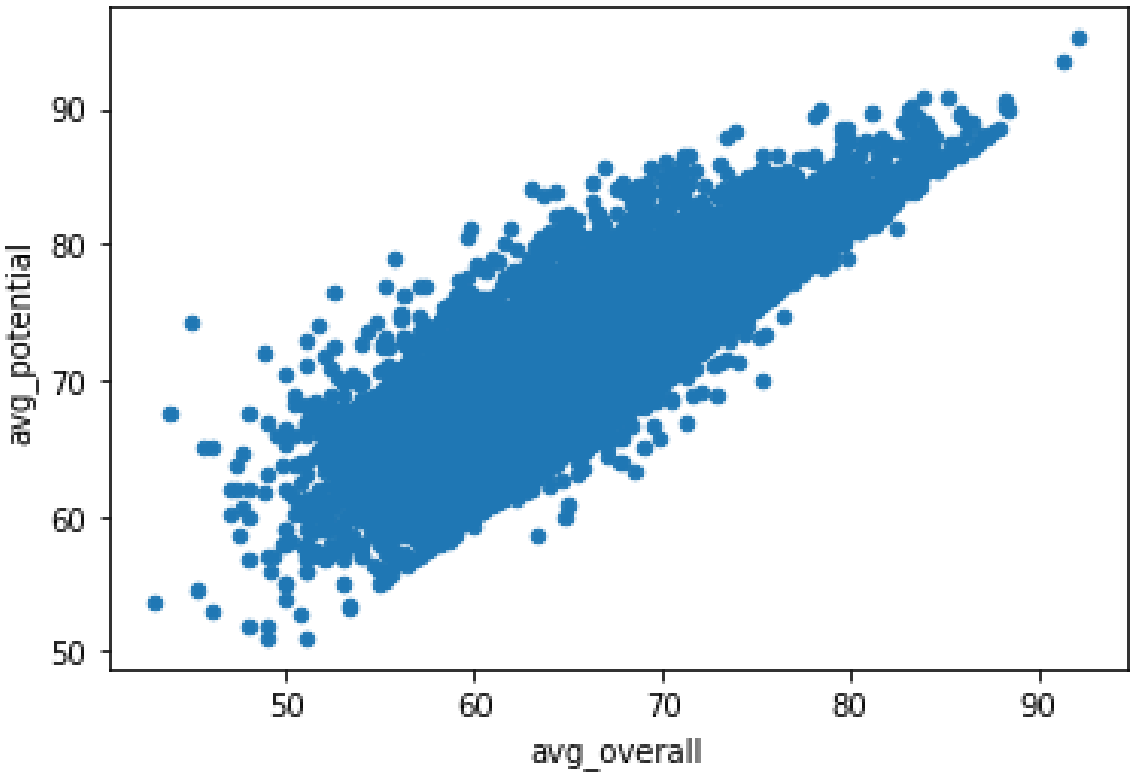
Analytics

1. From The Analysis, Notice that the Spanish League's __Spain LIGA BBVA__ scored the most goals between 2008 - 2016 with __8412 Goals__, followed closely by the __English Premier League__ with __8240 Goals__. The Italian League's __Italy Serie A__ came in third with __7895 Goals__ scored.

Based on the number of Goals scored per league it would be reasonable to conclude that the Spanish, English and Italian leagues are the most competitive leagues in Europe in terms of Total Number of Goals Scored.



2. Notice that the scatter plot implies a Positive Correlation that as Player overall increases, so does their potential. Therefore players with greater overall rating often have higher potential ratings.



3. From the Players' Dataset, notice the __average weight was 168.38 Pounds__, __average height was 181.87 centimeters__. The Collective __average overall rating__ for each player was __66.82 Overall__ and the collective __average player potential rating__ was __72.09__ Potential Overall Rating.

   As for the Matches Dataset, From all analyzed matches, the __home team__ scored an average of __1.5445__ goals while the away team scored an average of __1.1609__ goals.

   This shows that Home Teams often outscore Away Teams in European Football Matches. Therefore we could argue that the Home Advantage does influence a team's ability to score more goals in a match.

	player_api_id	avg_weight	avg_height	avg_overall	avg_potential
count	11060.000000	11060.000000	11060.000000	11060.000000	11060.000000
mean	156582.427215	168.380289	181.867445	66.821384	72.090373
std	160713.700624	14.990217	6.369201	6.237697	5.800325
min	2625.000000	117.000000	157.480000	43.000000	51.000000
25%	35555.500000	159.000000	177.800000	62.820000	68.000000
50%	96619.500000	168.000000	182.880000	66.720000	72.000000
75%	212470.500000	179.000000	185.420000	70.952500	76.000000
max	750584.000000	243.000000	208.280000	92.190000	95.230000

	home_team_goal	away_team_goal	match_id
count	25979.000000	25979.000000	2.597900e+04
mean	1.544594	1.160938	1.195429e+06
std	1.297158	1.142110	4.946279e+05
min	0.000000	0.000000	4.831290e+05
25%	1.000000	0.000000	7.684365e+05
50%	1.000000	1.000000	1.147511e+06
75%	2.000000	2.000000	1.709852e+06
max	10.000000	9.000000	2.216672e+06