Assignment 1

This assignment is based on the English Premier League 2018-19 dataset. The dataset is found on the GitHub repo. The following table is the legend for the features.

Field Name	Order	Type (Format)	Description
Date	1	date (%Y-%m-%d)	Match Date (dd/mm/yy)
HomeTeam	2	string (default)	Home Team
AwayTeam	3	string (default)	Away Team
FTHG	4	integer (default)	Full Time Home Team Goals
FTAG	5	integer (default)	Full Time Away Team Goals
FTR	6	string (default)	Full Time Result (H=Home Win, D=Draw, A=Away Win)
HTHG	7	integer (default)	Half Time Home Team Goals
HTAG	8	integer (default)	Half Time Away Team Goals
HTR	9	string (default)	Half Time Result (H=Home Win, D=Draw, A=Away Win)
Referee	10	string (default)	Match Referee
HS	11	integer (default)	Home Team Shots
AS	12	integer (default)	Away Team Shots
HST	13	integer (default)	Home Team Shots on Target
AST	14	integer (default)	Away Team Shots on Target
HF	15	integer (default)	Home Team Fouls Committed
AF	16	integer (default)	Away Team Fouls Committed
HC	17	integer (default)	Home Team Corners
AC	18	integer (default)	Away Team Corners
HY	19	integer (default)	Home Team Yellow Cards
AY	20	integer (default)	Away Team Yellow Cards
HR	21	integer (default)	Home Team Red Cards
AR	22	integer (default)	Away Team Red Cards

Do the following in a Colab notebook as before.

1. Import the dataset and rename the column names to more suitable, understandable feature names.

2. Print a list of all the competing teams listed alphabetically
3. Find the maximum number of shots taken during a match
4. Find the average shots on target to total shots taken ratio across the competition
5. What is the number of cards given out by each referee across the competition?
6. Which team has accumulated the most cards across the competition?
7. Which referee results in the home team winning the match most?
8. What was the percentage of matches in which the teams winning at half time went on to win the match at full time?

The above questions are just to get you riled up and confident that you're getting a hang of things. The following two will make you question yourself if that's true.

1. THE CHALLENGE. Make and display your own version of the PL table. It should look something like this (without the logos of course).

Position	Club	Played	Won	Drawn	Lost	GF	GA	GD	Points
1	Manchester City	38	32	2	4	95	23	+72	**98**
2	Liverpool	38	30	7	1	89	22	+67	**97**
3	Chelsea	38	21	9	8	63	39	+24	**72**
4	Tottenham Hotspur	38	23	2	13	67	39	+28	**71**
5	Arsenal	38	21	7	10	73	51	+22	**70**
6	Manchester United	38	19	9	10	65	54	+11	**66**
7	Wolverhampton Wanderers	38	16	9	13	47	46	+1	**57**
8	Everton	38	15	9	14	54	46	+8	**54**
9	Leicester City	38	15	7	16	51	48	+3	**52**
10 ▲	West Ham United	38	15	7	16	52	55	-3	**52**
11 ▼	Watford	38	14	8	16	52	59	-7	**50**
12	Crystal Palace	38	14	7	17	51	53	-2	**49**
13 ▲	Newcastle United	38	12	9	17	42	48	-6	**45**
14 ▼	Bournemouth	38	13	6	19	56	70	-14	**45**
15	Burnley	38	11	7	20	45	68	-23	**40**

2. Use your newfound EDA and/or VDA skills to find any/many amazing facts. This is an open question to help you go crazy with the dataset. It might need you to sit with the dataset for some time.

Make your submission in a similar way to your first one, download the notebook using the ‘download .ipynb’ option in Colab and submit it in a folder of your name in ‘Assignment 1’ directory.

We’ll discuss the best results and facts found out by you guys sometime in the middle of the coming week.

Also, ANY DOUBTS, PM one of us and above all, enjoy yourself!