

Ilham Wahyu Adli

Link: [ilhamwahyu8/IYKRA_EDA \(github.com\)](https://github.com/ilhamwahyu8/IYKRA_EDA)

Soal 1

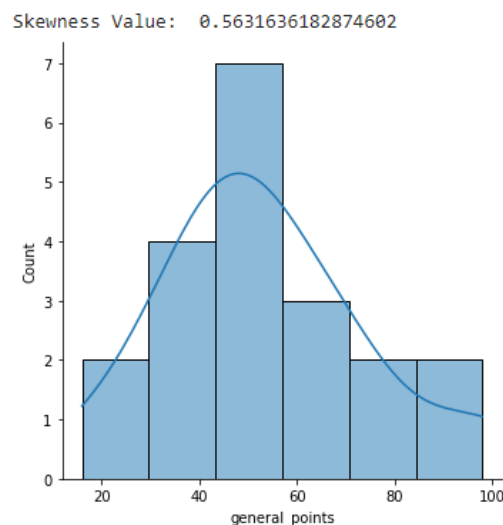
Terdapat data type dengan tipe objek ternyata data tersebut seharusnya merupakan integer, sehingga perlu adanya pergantian type data dan replace symbol ‘,’.

```
Column Name: attack_passes
[ '26,581', '23,638', '25,078', '21,295', '20,805', '19,202', '16,618', '16,494', '17,219', '16,358', '15,462', '15,243', '13,490', '15,943', '13,011', '14,297', '14,109', '10,226', '17,699', '15,902' ]
Column Name: attack_passes_long
[ '1,814', '2,250', '1,774', '2,267', '2,062', '2,242', '2,552', '2,310', '2,335', '2,315', '2,437', '2,259', '2,680', '1,932', '2,898', '2,394', '2,623', '2,513', '2,417', '2,603' ]
Column Name: attack_passes_back
[ '4,240', '3,416', '3,874', '3,191', '3,102', '2,634', '2,515', '2,362', '2,593', '2,273', '2,314', '2,233', '1,999', '2,249', '2,046', '2,040', '2,072', '1,418', '2,750', '2,492' ]
Column Name: defence_clearances
[ '543', '639', '593', '770', '762', '785', '824', '825', '945', '855', '956', '940', '1,120', '1,019', '1,080', '1,068', '921', '978', '971', '658' ]
```

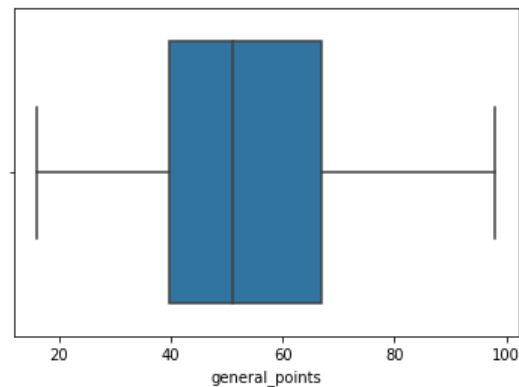
Tipe data lainnya setelah dilakukan explorasi tidak ada. Tidak ada data duplikat juga pada dataset yang disediakan. Mungkin terdapat beberapa outliers pada data ini tetapi outliers tersebut merupakan performa dari team tersebut dan bukan termasuk sebuah anomali. Sehingga dataset dapat terbilang bersih.

Soal 2

Melihat suatu distribusi dapat dilakukan dengan cara melakukan plot histogram dan densitynya. Seaborn telah mengimplementasikan kedua fitur tersebut dengan nama fungsi distplot(). Sehingga hasilnya adalah:



Kemudian selanjutnya dilakukan pengecekan outliers mengenai data tersebut, dan hasilnya tidak ada outliers pada data tersebut:

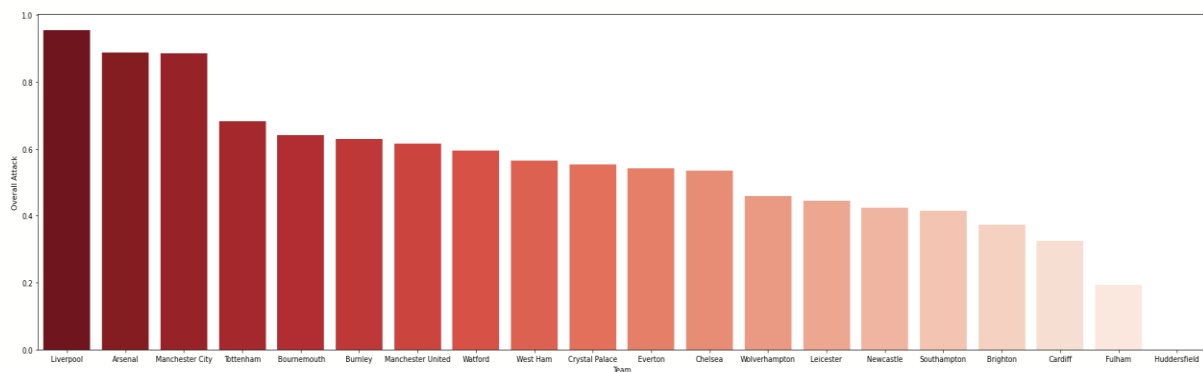


Sehingga tidak ada team yang bersifat anomaly pada data ini.

Soal 3

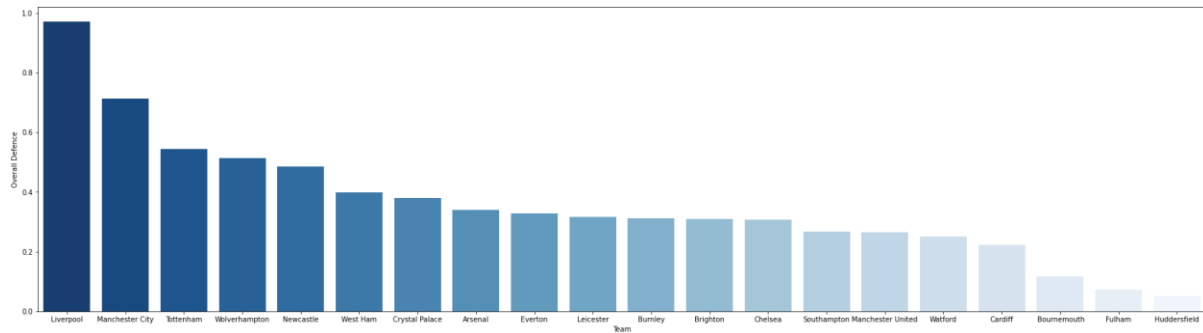
Untuk mencari team dengan penyerangan terbaik saya menghitung dulu efisiensi penyerangan dari team tersebut dengan menggunakan atribut: `['attack_shots', 'attack_shots_on_target', 'attack_posession', 'attack_pass_accuracy']`

Efisiensi atribut tersebut akan ditentukan oleh atribut lain yaitu `'attack_scored'` efisiensi ini dihitung dengan cara membagi `'attack_scored'` dengan atribut yang telah ditentukan untuk menghitung overall penyerangan. Perhitungan overall dihitung menggunakan mentotalkan seluruh atribut tersebut yang sebelumnya telah melalui tahap normalisasi sehingga memiliki skala yang sama. Sehingga hasilnya dapat dilihat pada gambar dibawah ini dengan Liverpool yang merupakan team dengan penyerangan terbaik:



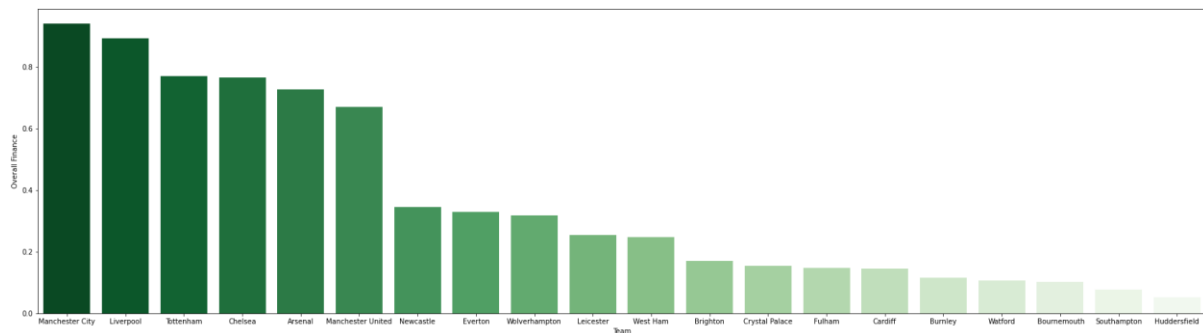
Soal 4

Cara pengejaan soal 4 hampir sama seperti soal 3 dengan atribut yang digunakan adalah `['defence_saves', 'defence_blocks', 'defence_interceptions', 'defence_tackles', 'defence_clearances']` yang akan dihitung efisiensinya menggunakan atribut `'defence_goals_conceded'`.



Soal 5

Cara pengerjaan soal 5 adalah dikarenakan atribut yang berkaitan dengan finansial pada televisi terdapat 2 yaitu ['finance_live_games_televised', 'finance_tv_revenue'] sehingga saya melakukan pembagian terhadap 2 atribut tersebut untuk menghitung seberapa efektif pendapatan yang didapatkan dari tayangan live televisi. Kemudian beberapa atribut pendukung yang dipilih antara lain ['finance_team_market', 'finance_market_average'] sehingga hasil yang didapatkan adalah seperti dibawah ini:



Soal 6

Saya disini mencoba mencari tahu apakah overall attack dan overall defence dari sebuah team mempengaruhi peringkat dari sebuah team atau tidak. Ternyata mempengaruhinya -0,79 dan -0,63 yang termasuk cukup berpengaruh.

```
dfCreativity.corr()

sns.heatmap(dfCreativity.corr());
```

