|  |  |
| --- | --- |
|  | **Data Fellowship Program Batch 6**  by IYKRA | Analytics for Everyone |

**Practice Case - Exploratory data analysis**

|  |  |
| --- | --- |
| **Module name** | Exploratory data analysis |
| **Key objectives** | To assess student’s understanding and creativity by making insightful data exploration before and/or after analysis |
| **Tools needed** | Python |
| **Dataset** | (epl\_1819.csv)  <https://github.com/Syukrondzeko/Fellowship> |
| **Effort (time/duration)** | 1 x 4 hours |
| **Students workflow** | * Read the questions * Make data explorations * Write the answer into pdf for the explanations and give a link to github for source code (your link must be embedded on that pdf) |
| **Output** | PDF, uploaded on Olaas |

**Scenario**

Looking back at the season that was 2018-2019 and looking to delve into sight deeper insights. Using the data to see how clubs are similar stylistically, in the way they pass, attack and score goals.

This data set is wide ranging in the sense it encompasses stats seen on a regular league table but goes beyond looking at how teams pass and keep possession, how they defend, tackle as well as looking at market values of a team and how much money each team was allotted from the TV rights deal.

This data was gathered from

1) BBC Sports Football,

2) Premierleague.com

3) Transfermarkt.co.uk

Some goals of this project:

1. Is this data clean?
2. How is the point distribution of the epl team? and which team is an anomaly?
3. Which team has the best attack?
4. Which team has the best defence?
5. Which team is good in the financial aspect?
6. Explore your creativity!

**Aspects Rated Description (Total Score: 100)**

|  |  |
| --- | --- |
| **Completeness** | The points will be deducted by 20 for one problem that is not done at all ( Question 1 – 5). But if all is complete then there is no deduction. |
| **Correctness for number 1 – 2** | Each question has a maximum 15 point |
| **Rationality of selecting variables and answering the question number 3 - 5** | Each question has a maximum 15 point |
| **Creativity** | This aspect assesses how students choose the right and interesting visualization to explore data. Maximum point is 15 |
| **How deep the insight** | This is an additional point where students are expected to be able to find hidden insight in the data. Maximum point is 10 |