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

**Practice Case - Statistics**

|  |  |
| --- | --- |
| ***Module name*** | Statistics |
| ***Key objectives*** | To assess student’s understanding about basic and fundamental statistics |
| ***Tools needed*** | Python |
| **Dataset** | Seasons\_Stats.csv  <https://github.com/Syukrondzeko/Fellowship> |
| **Effort (time/duration)** | 1 x 4 hours |
| **Students workflow** | * Read the questions * Make data analysis * 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**

The data-set contains aggregate individual statistics for 67 NBA seasons. from basic box-score attributes such as points, assists, rebounds etc., to more advanced money-ball like features such as Value Over Replacement.

**Pre-Processing**

In this task not all data will be used, only data in 2017. So it is necessary to do filtering at the beginning. Besides that there are some players who make team transfers in the NBA transfer market so that there is duplication of player data. Therefore you can use the df.drop\_duplicates() syntax to solve this to produce the same output as the trainer. Delete columns that have as many missing values as the entire row of data. Then you can do additional preprocessing if needed or you can immediately process the data.

Some goals of this project:

1. Who is the youngest and oldest player in the NBA in 2017 for each team (Tm) ?
2. Which player has the most minutes played (MP) in each position (Pos)?
3. Which team has the highest average total rebound percentage (TRB%), assist percentage (AST%), steal percentage (STL%), and block percentage (BLK%)?
4. Who is the best player in your opinion based on his record stats? note: you can refer to variables point (PTS), assists, rebounds, or anything else. A combination of several variables would be nice.
5. Which team has the best average stat record of their players? Note: you can refer to points, assists, rebounds, or anything else. A combination of several variables would be nice

**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 pre processing*** | Max points are 5. But if you fail in this stage, the rest will be wrong. |
| ***Creativity and Correctness on answering the question*** | * Each question has a maximum 15 point (for number 1 – 3) * Number 4 has a maximum 20 points * Number 5 has a maximum 25 points |
| ***Workflow*** | Here will be assessed how your story line is in presenting the results of the analysis. Max 5 points. |