**Learning outcomes**

* Select appropriate approaches to perform statistical analysis on a set of data (A3, PLO9)
* Demonstrate the ability to solve specific problem with R scripts (C3, PLO2)

**Coursework Title**

**2019-2020 England Premier League season**

**Coursework Description**

You will be advice to study the England Premier League championship to gain the basic understanding of the football matches before proceed to analyse the given dataset. The given dataset is provided by the sports-statistics.com on the England Premier League for the season 2019 to 2020. Explore and examine the 208 records provided using any suitable statistical analysis techniques to obtain the desired results. Statistical analysis techniques include data manipulation / exploitation process using appropriate R concepts and techniques.

The dataset comprised of 208 matches that were extracted from http://www.football-data.co.uk/ website. These records are based on achievement of 20 football clubs. Following are the core variables in the given dataset.

|  |  |
| --- | --- |
| **Column** | **Description** |
| Div | League Division |
| Date | Match Date (dd/mm/yy) |
| HomeTeam | Home Team |
| AwayTeam | Away Team |
| FTHG | Full Time Home Team Goals |
| FTAG | Full Time Away Team Goals |
| FTR | Full Time Result (H |
| Home Win, D | Draw, A |
| Away Win) |  |
| HTHG | Half Time Home Team Goals |
| HTAG | Half Time Away Team Goals |
| HTR | Half Time Result (H |
| Home Win, D | Draw, A |
| Away Win) |  |
| Attendance | Crowd Attendance |
| Referee | Match Referee |
| HS | Home Team Shots |
| AS | Away Team Shots |
| HST | Home Team Shots on Target |
| AST | Away Team Shots on Target |
| HHW | Home Team Hit Woodwork |
| AHW | Away Team Hit Woodwork |
| HC | Home Team Corners |
| AC | Away Team Corners |
| HF | Home Team Fouls Committed |
| AF | Away Team Fouls Committed |
| HO | Home Team Offsides |
| AO | Away Team Offsides |
| HY | Home Team Yellow Cards |
| AY | Away Team Yellow Cards |
| HR | Home Team Red Cards |
| AR | Away Team Red Cards |

**Requirements**

* You are required to study, analyse and interpret the dataset provided by producing different reports out of the 208 records.
* Develop your solution with RStudio.
* Document any assumptions made during data exploration.
* The statistical analysis process through data manipulation or exploitation with R techniques have to be documented with an explanation.
* Report your finding from the analysis carried out.

*(Recommends visualizing the reports produced with graph/chart)*

**Deliverable**

1. A report which includes the following

* Introduction (assumption and what do you want to achieve)
* Solution
  + Explain the applied techniques / Concepts for data manipulation
  + Implemented R scripts
  + Screen capture of the outcome of the program execution such as
  1. Results
  2. Reports
  3. Visualization (if any)
* References

1. An R script file of your solution.

**Assessment Criteria**

The assignment comprised of 2 components; Approach (50%) and R Scripts (50%). The details breakdown of each component are as follows:

1. **Approach**

* Understanding of the scenario (5%)
* Analysis Approach – Exploration and Manipulation
  + Importing data from an external resource into the working environment and cleaning / pre-processing (5%)
  + Choosing the Data Exploration and manipulation techniques – filtering, sampling, categorization, etc (25%)
  + Discussion on the finding (15%)

1. **R Scripts**

* R Scripts provided are relevant and based on the applied techniques (25%)
* Apply and implements additional logics, functions to extend the analysis (25%)