Crisil ETL Assignment

This is an ETL assignment for Data Engineers

The Premier League, also known as the English Premier League or the EPL is the top level of the English football league system. Contested by 20 clubs, it operates on a system of promotion and relegation with the English Football League (EFL).

The EPL data has information on seasons between 09-10 and 18-19. They hire you as the Data Engineer to process and stage the raw data for all seasons.

For this assignment

- You need to create an ETL job that prepares the data for the EPL Research team. Please consider the following:
    - The Job must be an ETL code developed in Python
    - Deployment of the code is required. Please prefer using open-source tools
- Show us the overall architecture of your pipeline. Include all the steps of the workflow
- We want a deployment for ETL, please remember that the best deployments are the easiest to execute

ETL Job

The job should ingest the files and generate the following outputs:

- The position table for all the seasons
- The best scoring team by season

The report needs to be exported to a file, you can choose the extension of the file but remember this is an ETL Job. Please save your code in the src folder

Data Dictionary

Abbreviation | 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

HBP | Home Team Bookings Points (10 | yellow, 25 | red)

ABP | Away Team Bookings Points (10 | yellow, 25 | red)

Deployment

You need to describe how to deploy your job and run it in any environment, so please provide very clear and detailed instructions. Please save the deployment in the "deploy" folder

Datasets

The files are located under the "data" folder

Points to note

- If you need to provide any additional information or instructions, please create a new README.md file
- We would like to evaluate your coding skills, so remember to use the best practices of Software engineering
- This is an ETL Job, show us all you know about good practices of doing ETLs
- Please share the link to your GitHub public account or share the results via zip file

Keep in mind that this is a timed exercise, so please send in your completed assignment before the deadline! Good Luck!