PYTHON DEVELOPER ASSIGNMENT

SUMMARY

A dataset is provided representing a set of events from a series of football matches. Each event represents the performance of a single player in a single match. The dataset is fully denormalised, so each event carries redundant information about the match, player and team. The task is to transform the data into a (somewhat) normalised form, and output the resulting datasets to files.

INPUT DATA SPECIFICATION

The input data is provided as a single CSV file. The file has a header row naming each column. The table below gives the definition of the columns:

Field name	Type	Description
Match Id	Int	The unique Id of the match during which the event occurred
Match Name	String	The name of the match during which the event occurred
Team Id	Int	The Id of the team that the player played for during the match
Team Name	String	The name of the team the player played for during the match
Is Home	Boolean	Whether the player's team was home or away during the match
Player Id	Int	The Id of the player represented by this event
Player Name	String	The name of the player represented by this event
Goals Scored	Int	The number of goals scored by the player during the match
Minutes Played	Int	How many minutes the player played for during the match

Notes:

- You can assume that each player only plays for one team across the entire dataset
- All matches are 90 minutes in duration.

OUTPUT DATA REQUIREMENTS

The program should produce several output datasets, which normalise the input data. The required output datasets are:

MATCH

Each record represents a single match.

Field Name	Type	Description
Match Id	Int	The unique Id of the match
Match Name	String	The name of the match
Home Team Id	Int	The team Id of the team playing at home for this match
Away Team Id	Int	The team Id of the team playing at away for this match
Home Goals	Int	How many goals the home team scored
Away Goals	Int	How many goals the away team scored

TEAM

Each record represents one of the teams playing in the various matches.

Field Name	Type	Description
Team Id	Int	The Id of the team
Team Name	String	The name of the team

PLAYER

Each record represents one of the players participating in the various matches.

Field Name	Type	Description
Player Id	Int	The Id of this player
Team Id	Int	The Id of the team the player plays for
Player Name	String	The name of this player

STATISTIC

Each record represents the performance of one player in one match

Field Name	Type	Description
Stat Id	Int	A unique Id for each record in the dataset
Player Id	Int	The Id of the player the statistic relates to
Match Id	Int	The match from which this player statistic originated
Goals Scored	Int	How many goals this player scored in this match
Minutes Played	Int	How many minutes the player played in this match
Fraction of total minutes played	Float	What proportion of the 90 minute match the player played. Should be between 0 and 1
Fraction of total goals scored	Float	What proportion of the total goals scored in the match were scored by this player. Should be between 0 and 1

OUTPUT FORMAT

Each output dataset should be saved to its own file in JSON Lines Format https://jsonlines.org/ (each row of the file is a valid JSON record representing one row in the data).

PROGRAM REQUIREMENTS

The assignment should be completed by developing a Python application to process the input data and produce the output data. It should run using the standard CPython interpreter.

We are looking for the following characteristics:

- Correctness. The application should produce output according to the specification above
- Readability. The application should be written to be easily understood and maintained by other developers.
- Testing. Unit testing in place to demonstrate the correctness of key parts of the solution.

Additional Notes:

- You can use additional libraries if appropriate but this is not mandatory
- You don't need to worry about issues of deployment such as CICD, container images as such. It is enough that the application runs on a local machine.