

Football Standings Scraper Report

Sameer

October 30, 2023

Introduction

This report discusses the Python script for scraping football standings data from the Sky Sports website for the Premier League, LaLiga, and Bundesliga. The script uses BeautifulSoup for web scraping and Peewee as an Object-Relational Mapping (ORM) tool to interact with a MySQL database.

Script Overview

The Python script connects to a MySQL database, creates a new database (if it doesn't exist), and defines a base model using Peewee. It then dynamically creates standings classes for each year and league, scrapes the data from the Sky Sports website, and inserts it into the corresponding database tables.

Script Execution

The script prompts the user for the database name, MySQL username, and password. It then asks the user to choose a football league (Premier League, LaLiga, or Bundesliga) and proceeds to scrape and insert the standings data for each year from 2000 to 2022.

Script Code

Below is a snippet of the main parts of the Python script:

```
1 #Defining the base class which becomes the table
2 class BaseModel(Model):
3     class Meta:
4         database = db
5 #Dynamically creating tables based on the year and league selected
6 def create_standings_class(db, class_name, attributes):
7     new_class = type(class_name, (BaseModel,), attributes)
8     return new_class
9
10 #Attributes which become the columns of the table
11 standings_attributes = {
12     'standing': IntegerField(),
13     'team': CharField(),
14     'matches_played': IntegerField(),
15     'matches_won': IntegerField(),
16     'matches_drawn': IntegerField(),
17     'matches_lost': IntegerField(),
18     'goals_for': IntegerField(),
19     'goals_against': IntegerField(),
20     'goal_difference': IntegerField(),
21     'points': IntegerField()
22 }
23
24 #Populating the table
25 Standing.create(
26     standing=int(data[i][0]),
27     team=data[i][1],
28     matches_played=int(data[i][2]),
29     matches_won=int(data[i][3]),
```

```
30     matches_drawn=int(data[i][4]),
31     matches_lost=int(data[i][5]),
32     goals_for=int(data[i][6]),
33     goals_against=int(data[i][7]),
34     goal_difference=int(data[i][8]),
35     points=int(data[i][9])
36 )
```

Listing 1: Football Standings Scraper Script

Conclusion

The script successfully scrapes and inserts football standings data into the MySQL database with the user given database name and table names following the format league-name-year. It can be extended to include additional leagues needed.