README: FIFA 20 Players Data Analysis Notebook

1 Overview

This repository contains a Jupyter notebook (`players_20.ipynb`) that performs exploratory data analysis on the FIFA 20 players dataset (`players_20.csv`). The notebook focuses on analyzing player wages across clubs, culminating in a visualization of the top 10 clubs by total player wages. This README provides an overview of the notebook, prerequisites, dataset details, structure, and instructions for running the analysis.

2 Prerequisites

To run the notebook, ensure you have the following installed:

- **Python 3.6 or higher**: The notebook uses Python for data analysis.

- **Jupyter Notebook or JupyterLab**: To execute and interact with `players_20.ipynb`.

- **Python Libraries**:

  - `pandas`: For data manipulation and analysis.

  - `numpy`: For numerical computations.

  - `matplotlib`: For creating visualizations.

  - `seaborn`: For enhanced visualization styling.

  Install these libraries using:

  ```
  pip install pandas numpy matplotlib seaborn
  ```

- **Dataset**: The `players_20.csv` file, which contains FIFA 20 player data, must be in the same directory as the notebook.

3 Dataset Description

The dataset (`players_20.csv`) contains detailed information about soccer players from the FIFA 20 video game. It includes 18,278 rows and 104 columns, covering attributes such as:

- `sofifa_id`: Unique player identifier.

- `short_name`, `long_name`: Player names.

- `age`, `height_cm`, `weight_kg`: Physical attributes.

- `nationality`, `club`: Player affiliations.

- `wage_eur`: Weekly wage in euros.

- `overall`, `potential`: Player ratings.

- Various skill attributes (e.g., `mentality_penalties`, `goalkeeping_diving`).

The notebook specifically uses the `club` and `wage_eur` columns to compute and visualize total wages per club.

4   Notebook Structure

The notebook is organized into the following sections:

1. **ImportLibraries**: ImportsessentialPythonlibraries(`pandas`, `numpy`, `matplotlib`, `seaborn`) and suppresses warnings for cleaner output.

2. **Load Data**: Reads the `players_20.csv` file into a pandas DataFrame.

3. **Initial Exploration**:

    - Displays the first five rows of the DataFrame using `df.head()`.

    - Checks the dataset dimensions with `df.shape` (18,278 rows, 104 columns).

    - Summarizes numerical columns using `df.describe()`.

4. **Analysis and Visualization**:

    - Groups the data by `club` and sums the `wage_eur` column.

    - Selects the top 10 clubs by total wages.

    - Creates a bar plot visualizing the total wages for these clubs, with appropriate labels and formatting.

5   Running the Notebook

To run the notebook, follow these steps:

1. **Clone the Repository**:

   ```
   git clone <repository-url>
   cd <repository-directory>
   ```

2. **Ensure the Dataset is Available**: Place `players_20.csv` in the same directory as `players _20.ipynb`. The dataset can be sourced from platforms like Kaggle (e.g., FIFA 20 Complete Player Dataset).

3. **Set Up the Environment**:

- Create a virtual environment (optional but recommended):

```
python -m venv venv
 source venv/bin/activate % On Windows: venv\Scripts\activate
```

- Install dependencies:

```
pip install -r requirements.txt
```

(Create a `requirements.txt` with the libraries listed in Section 2 if not already present.)

4. **Launch Jupyter Notebook**:

```
jupyter notebook
```

Open `players_20.ipynb` in the Jupyter interface.

5. **Execute the Notebook**: Run all cells sequentially to load the data, perform the analysis, and generate the bar plot.

6 Notes

- The bar plot is displayed within the notebook and is not saved to a file. To save the plot, modifythevisualizationcodetoinclude`plt.savefig('top_10_clubs_wages.png')`.

- Ensure the `players_20.csv` file is correctly formatted and present, as the notebook assumes it exists in the working directory.

- The notebook is designed for Python 3. If using an older Python version, compatibility issues may arise with certain library functions.

7 Contact

For questions or issues, please open an issue on the GitHub repository or contact the repository maintainer.