**Project Documentation**

**Done By: Maroof Tahir and Faisal Rafiq**

**Overview**

The project involves four main phases: data collection, data cleaning, data storage and visualization of that data (stats). In the data collection phase, data is scraped from ESPNcricinfo.com using APIs to gather the latest statistics and rankings. The data cleaning phase involves processing the scraped data to ensure it is in a usable format, addressing issues such as missing values, inconsistencies, and errors. Finally, in the data storage phase, the cleaned data is written into CSV files for analysis and reporting. This project involves extracting player and team data from the ESPN Cricinfo website. The goal is to scrape the data, convert it into a CSV format, and visualize various metrics such as the number of players, top players, and most winnings.

**Technologies Used**

* Python
* Matplotlib
* Pandas
* Pyplot
* BeautifulSoup4
* Requests

**Project Structure**

The project directory contains the following files and directories:

* **10 Players Bio/**: A folder containing biographies of 10 selected players.
* **ICC\_ODI\_Rankings.csv**: CSV file containing ICC ODI rankings data.
* **ICC\_Project.py**: Python script for extracting and processing ICC rankings data.
* **ICC\_T20I\_Rankings.csv**: CSV file containing ICC T20I rankings data.
* **ICC\_Test\_Rankings.csv**: CSV file containing ICC Test rankings data.
* **ICC\_WODI\_Rankings.csv**: CSV file containing ICC Women's ODI rankings data.
* **ICC\_WT20I\_Rankings.csv**: CSV file containing ICC Women's T20I rankings data.
* **index.txt**: Text file with indexing information.
* **Individual Players/**: A folder containing detailed information about individual players.
* **Individual**.py: Used to extracting all player data in a team.
* **Players Ranking/**: A folder containing various player ranking files.
* **Players**.py: Used to extract top 10 players from every team.
* **team.py**: Python script for extracting and processing team data.

**Steps to Extract and Process Data**

1. **Extract Data from Website**: The data is extracted from the ESPN Cricinfo website using web scraping techniques.
2. **Scrape Data**: The ‘ICC\_Project.py’ and ‘team.py’ scripts are used to scrape player and team data, respectively.
3. **Convert to CSV**: The scraped data is processed and converted into CSV files for easier analysis and visualization.
4. **Visualizations**: We have visualized the data of team ranking across the season, individual player stats and top players of each team.

**Visualized Data**

The data is visualized to show various metrics, including:

* Number of players
* Top players
* Most winnings by teams in season

**File Details**

**1. Ten­ Players Bio/**

This directory contains detailed biographies of 10 selected players. Each file in this directory provides in-depth information about a player, including their career statistics and achievements.

**2. ICC\_ODI\_Rankings.csv**

This CSV file contains the ICC ODI rankings data. The columns include player names, team names, and their respective rankings.

**3. ICC\_Project.py**

This Python script is responsible for extracting and processing ICC rankings data. It includes functions to scrape data from the website and convert it into a structured format.

**4. ICC\_T20I\_Rankings.csv**

This CSV file contains the ICC T20I rankings data. Similar to the ODI rankings file, it includes player names, team names, and their respective rankings.

**5. ICC\_Test\_Rankings.csv**

This CSV file contains the ICC Test rankings data. It includes player names, team names, and their respective rankings.

**6. ICC\_WODI\_Rankings.csv**

This CSV file contains the ICC Women's ODI rankings data. It includes player names, team names, and their respective rankings.

**7. ICC\_WT20I\_Rankings.csv**

This CSV file contains the ICC Women's T20I rankings data. It includes player names, team names, and their respective rankings.

**8. index.txt**

This text file contains indexing information that helps in organizing and accessing the data files.

**9. Individual Players/**

This directory contains detailed information about individual players. Each file provides comprehensive data on a specific player, including their statistics and career highlights.

**10. Individual.py**

This Python script is responsible for extracting and processing individual player data of each team. It includes functions to scrape data from the website and convert it into a structured format.

**11. Players Ranking/**

This directory contains various player ranking files. Each file provides rankings for different categories and formats.

**12. players.py**

This Python script is responsible for extracting and processing Top 10 players from every team. It includes functions to scrape data from the website and convert it into a structured format.

**13. team.py**

This Python script is responsible for extracting and processing team data. It includes functions to scrape data from the website and convert it into a structured format.

**Usage**

1. Run the ‘ICC\_Project.py’ script to scrape and process the ICC rankings data.
2. Run the ‘team.py’ script to scrape and process the team data.
3. Use the generated CSV files to visualize the data and analyze various metrics.

**Visualization Examples**

* Number of players: A bar chart showing the number of players from each team.
* Top players: A table listing the top players based on their rankings and performance.
* Most winnings: A graph showing the teams with the most winnings in different formats.