ATP Tennis Rankings Web Scraper

ATP Singles Rankings

Project Overview

This project is a Python-based web scraper that collects and analyzes ATP tennis rankings data. It scrapes player rankings from the ATP Tour website, extracts player information such as rank, name, points, age, and nationality, and stores the data in a JSON file. Additionally, the data is analyzed and visualized using Matplotlib to show age distribution and nationality representation.

Project Structure

The project consists of three main Python scripts:

1. **`Main.py`** – Scrapes ATP tennis rankings data and saves it to `players_data.json`.
2. **`Age_Statistics.py`** – Analyzes and visualizes the age distribution of tennis players.
3. **`Countries_Statistics.py`** – Analyzes and visualizes the top 8 nationalities of ranked tennis players.

Dependencies

To run this project, you need the following Python libraries:

- Beautifulsoup4
- Matplotlib
- Requests

```
pip install requests beautifulsoup4 matplotlib
```

How It Works

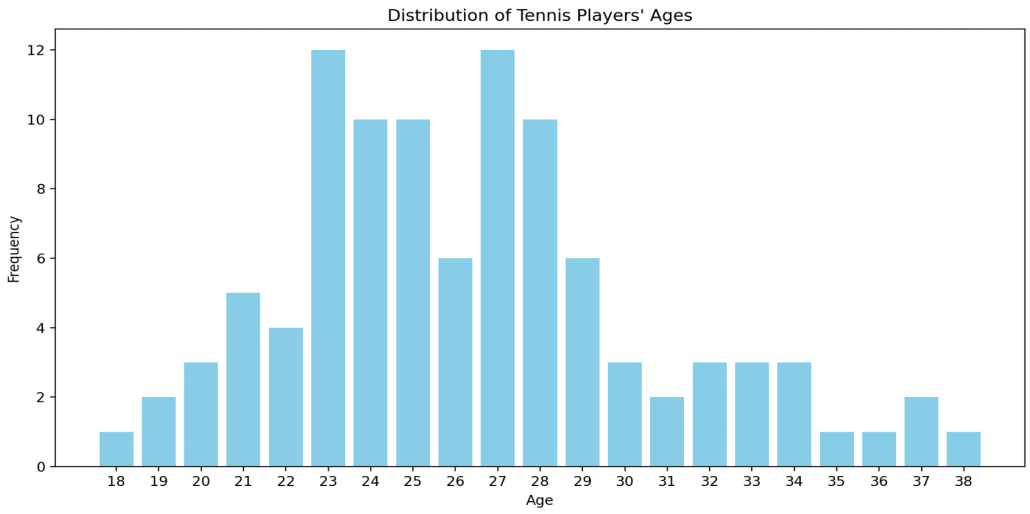
1. Web Scraping (`Main.py`)

- Fetches HTML content from the **ATP Tour rankings page** (`https://www.atptour.com/en/rankings/singles`).
- Uses **BeautifulSoup** to parse the page and extract ranking table data.
- Extracts:
  - **Rank**
  - **Name**
  - **Points**
  - **Age** (via ATP player API)
  - **Nationality** (via ATP player API)
- Saves the scraped data as **`players_data.json`**.

---

2. Age Distribution Analysis (`Age_Statistics.py`)

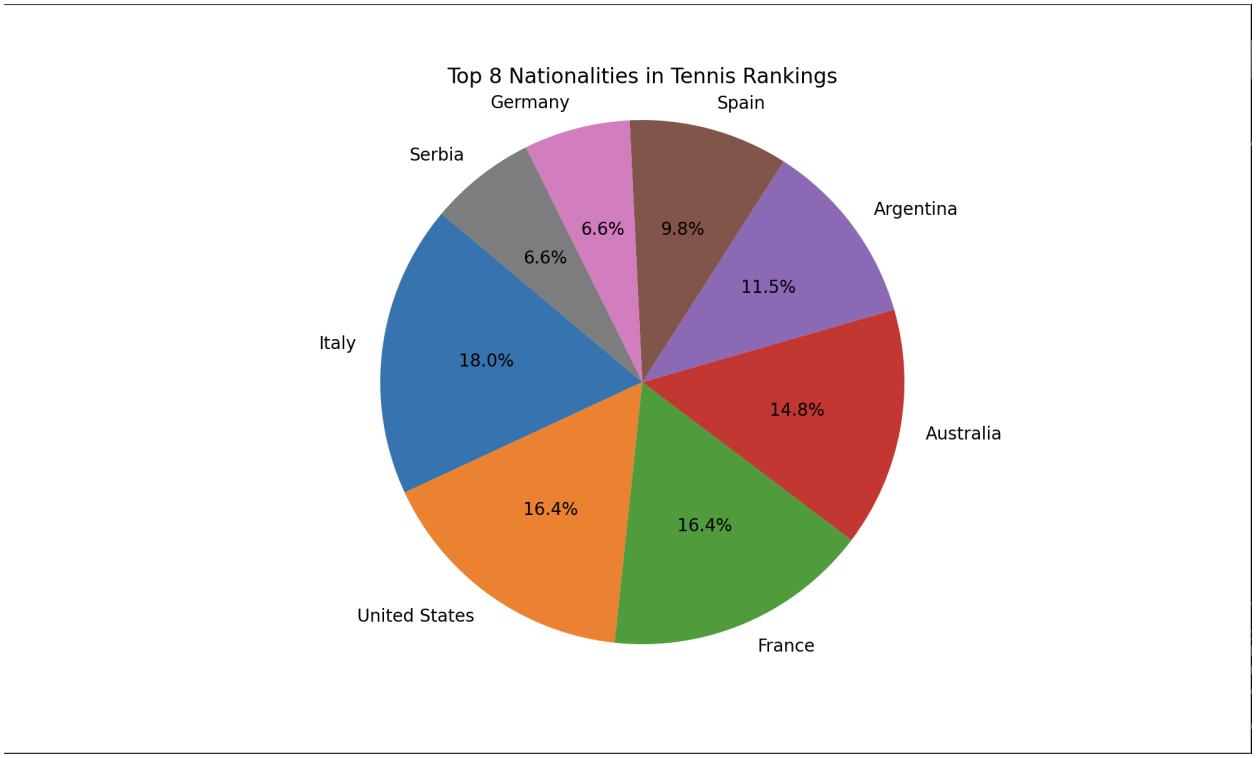
- Loads `players_data.json`.
- Extracts and counts the occurrence of each player's age.
- Creates a **bar chart** visualizing the distribution of ages in the ATP rankings.



---

3. Nationality Distribution Analysis (`Countries_Statistics.py`)

- Loads `players_data.json`.
- Extracts and counts the number of players from each nationality.
- Displays the **top 8 most common nationalities** in a **pie chart**.



How to Run

1. Run the Web Scraper

Execute `Main.py` to scrape the latest ATP rankings and generate `players_data.json`:

```
python Main.py
```

2. Run Age Distribution Analysis

To generate the bar chart for age distribution:

```
python Age_Statistics.py
```

3. Run Nationality Distribution Analysis

To generate the pie chart for top 8 nationalities:

```
python Countries_Statistics.py
```

Results

- The `players_data.json` file will store the scraped player information in JSON format.
- Running `Age_Statistics.py` will generate a **bar chart** showing the frequency of player ages.
- Running `Countries_Statistics.py` will generate a **pie chart** of the top 8 nationalities in the ATP rankings.

Potential Improvements

- Automate data updates using a scheduler.
- Expand analysis with additional statistics (e.g., average age of winning a first ATP tournament).