Cricket World Cup Tournament Toppers
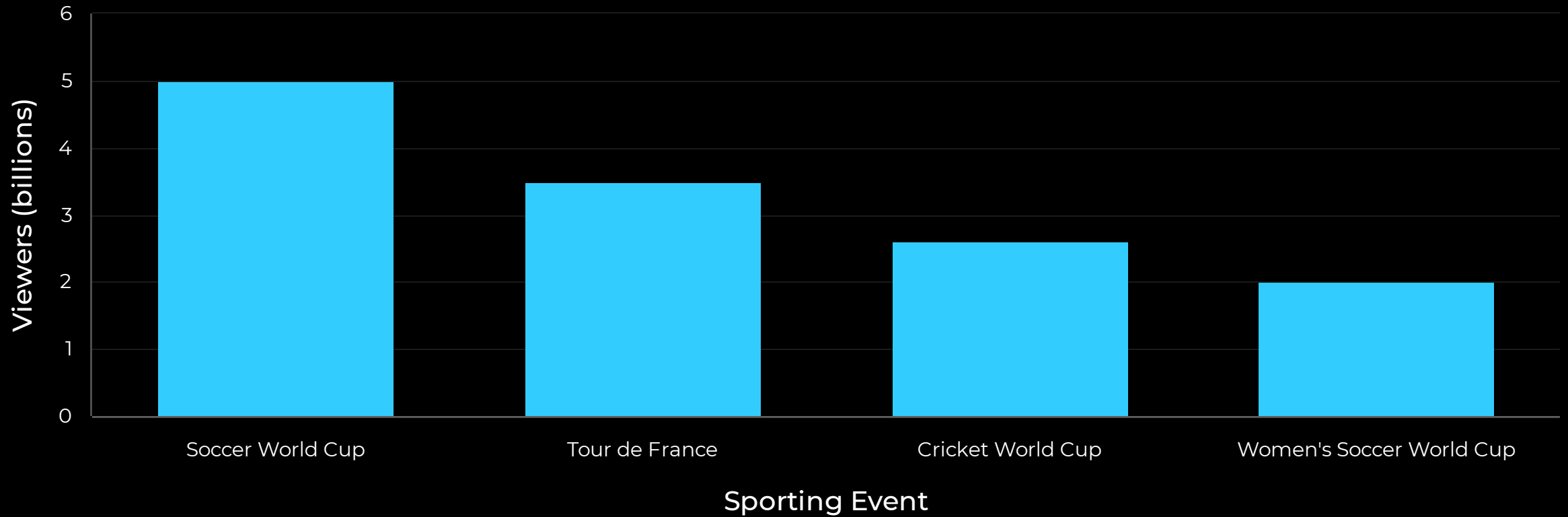
Scraper to extract cricket match data and player stats from ICC-Cricket
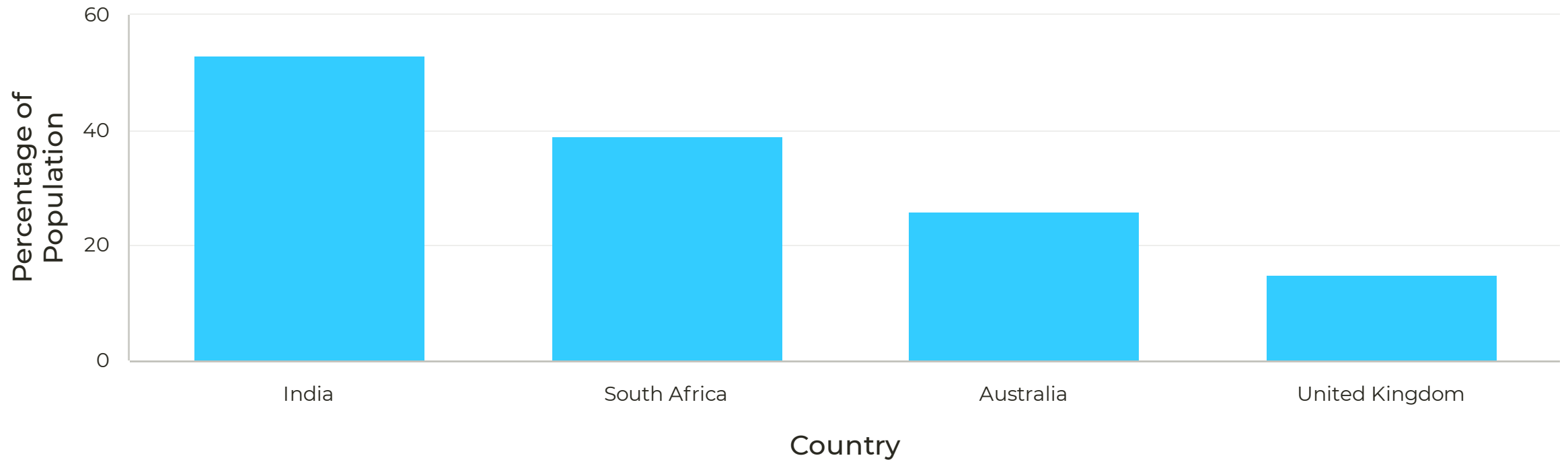
Cricket Popularity by Country

The International Cricket Council (ICC) has expressed intentions to enter cricket into the Olympics Games with plans to submit a bid to make an appearance at the 2028 Games in Los Angeles.

Cricket is most popular in India, followed by South Africa, Australia, and the UK.

The World of Cricket

Test cricket

The original and longest format of cricket, played over 5 days.

One Day International

Shorter format played over 1 day, introduced in 1971. (50.0 overs)

Twenty20

Fast paced format played over about 3 hours, introduced in 2003. (20.0 overs)

Cricket has evolved over the years into several popular formats.

Cricket Terminology

- **The game is played on a large grass field with a rectangular pitch in the center**

  The field has a boundary edge and rope marking the perimeter.

- **Each team has 11 players**

  One team bats and the other fields.

- **The bowler throws the ball to the batter**

  The batter tries to hit the ball with a wooden bat.

- **Runs are scored when batters run between wickets**

  The batter hits the ball and runs back and forth to score runs.

- **The bowler delivers 6 balls at a time**

  This is called an over. An innings consists of multiple overs.

- **Each team gets two innings per game.**

  One team's turn to bat and the other to bowl

Greatest Cricket Batsmen

Sachin Tendulkar (India)

Known as the God of Cricket; scored 100 centuries in international cricket

Brian Lara (West Indies)

One of the greatest left-handed batsmen; scored 400 runs in a single test innings

Ricky Ponting (Australia)

Known for his perfect pull shot; second highest run scorer for Australia in tests

These players are considered among the best batsmen in cricket history due to their remarkable achievements and consistency.

Data Sources

ESPNcricinfo

sports news website exclusively for the game of cricket

ICC-Cricket

official website for the International Cricket Council, the global governing body for cricket

HowSTAT!

most comprehensive collection of cricket records, statistics and graphs relating to every facet of international cricket

Key Technologies

Python

Python was used to build the web scraper and flask application

BeautifulSoup

Selenium library was used to parse and extract data from HTML pages

SQLite

SQLite database was used to efficiently store and query the scraped data

These key technologies were fundamental building blocks of the cricket scraper project allowing efficient web scraping, data storage and analysis.

Overview

Scrape stats daily

Automating a task using cronjob to scrape cricket stats from public websites daily

Parse stats by player and team

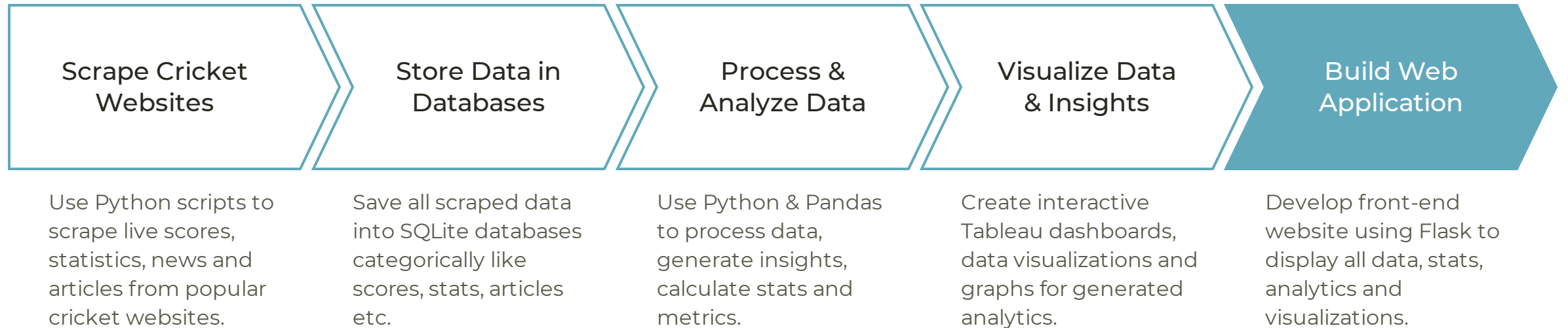Extract stats for each player and team into structured data

Display in Flask Route

Retrieve stats from SQLite database and calculate streaks based on consecutive appearances of a player

Daily automated cricket stats scraping pipeline storing structured data in a database for Flask App

Daily Process

Scrape Cricket Websites	Store Data in Databases	Process & Analyze Data	Visualize Data & Insights	Build Web Application
Use Python scripts to scrape live scores, statistics, news and articles from popular cricket websites.	Save all scraped data into SQLite databases categorically like scores, stats, articles etc.	Use Python & Pandas to process data, generate insights, calculate stats and metrics.	Create interactive Tableau dashboards, data visualizations and graphs for generated analytics.	Develop front-end website using Flask to display all data, stats, analytics and visualizations.

Top Players Tracked

Virat Kohli, Joe Root, Kane Williamson and Steve Smith are some of the top ranked batsmen in Test cricket as of 2022.

Although, Virat Kohli was the top run scorer and Mohammed Shami was the top wicket taker of the tournament,

- **Dilshan Madhushanka** held the longest streak for MOST WICKETS at **5 days**

- **Quinton de Kock** held the longest streak for HIGHEST SCORE at **13 days**

- **Quinton de Kock** held the longest streak for MOST RUNS at **16 days**

/25 days of data were scraped