Topic: Analysis of 164 Countries Based on Active Football Leagues

Prepared by Filip Ferens 420872 and Dominika Dunda 422722

1. Description of the topic and the web page

League321.com is a soccer website designed to serve as a valuable resource for individuals seeking comprehensive information on association football league tables worldwide. It aims to provide easily accessible and up-to-date season-by-season league standings from various countries. The website goes beyond league tables and also offers data on national cup results, club records, and player statistics for many countries.

Selected topic will allow us to receive generated links that lead us to pages in order to view general information about football in each of the selected countries. After going to the target page, we see a specific, selected country with its indication on the map, and information such as the football association of the country, year of foundation, addresses, affiliations, and the main league tables. We learn about the country's championship and cup, and about the league structure. We will also see and results and also links to other useful information such as national football teams.

2. Description of the scraper mechanics - what the program is technically doing

Beautiful Soup: The code fetches the HTML content of the webpage, extracts specific links based on filtering criteria, extracts the headers from those links in parallel, calculates the extraction time, and saves the filtered links and headers into separate CSV files.

Selenium: The code uses Selenium with a Firefox web driver to open the webpage, extract specific links using JavaScript execution, process each link concurrently to extract headers and save the filtered links and headers into separate CSV files.

Scrapy: This spider crawls the website starting from the specified URLs, extracts all links that match the defined criteria, saves the filtered links to a CSV file, and then follows each link to

extract the header information from the corresponding web pages using BeautifulSoup. The extracted headers are yielded as dictionaries.

3. Short technical description of the output

The outputs obtained from the codes are:

1. links_output.csv: This files contain a single column representing the filtered links extracted from the website http://www.league321.com/. Each row represents a link that satisfies the filtering criteria of containing '-football' and ending with '-football.html'.
2. output.csv: These CSV files contain a single column that stores the extracted headers from the web pages corresponding to the filtered links. Each row represents a header extracted from a web page. If the extraction fails for a particular link, the header value will be 'NA'.

Both output files are saved to the specified output paths. Additionally, the code measures the execution time for the web scraping process and prints it to the console.

The purpose of these output files is to provide information about the headers and links related to football leagues present on the league321.com website. The headers can be used to identify and categorize the football leagues by countries, while the links can be used to access further details or information related to each league.

SCRAPERS COMPARISON			
	Beautiful Soup	**Scrapy**	**Selenium**
Execution time (seconds)	133.45	33.82	383.63
No. of the links	164	164	164
No. of the headers	164	164	164
Issues encountered			Headers are duplicated, so final output without duplicates would be smaller

Source: Own elaboration.

After comparing this three approaches we may conclude that:

→ Scrapy has the shortest execution time at 33.82 seconds, followed by Beautiful Soup at 133.45 seconds, and Selenium with the longest execution time of 383.63 seconds. This indicates that Scrapy is the most efficient in terms of speed.

→ All three approaches extracted the same number of links and headers (164 in this case), suggesting that they are equally capable of retrieving the desired information.

→ Selenium faced the issue of duplicated headers, which would result in a smaller final output without duplicates. This implies that additional steps may be required to handle or eliminate duplicates when using Selenium.

3. Elementary data analysis

The findings from the analysis can be used to understand the global football landscape, identify best practices from successful leagues, and facilitate knowledge exchange among football associations. It can provide valuable insights for researchers, football enthusiasts, and stakeholders interested in the development and dynamics of football leagues worldwide.

This analysis aims to provide a snapshot of 164 countries based on the presence of active football leagues. By examining this information, we can gain insights into the global landscape of football and identify countries with prominent football cultures. The data was collected from reliable sources, and it provides a foundation for further investigation and in-depth analysis. Out of the 164 countries analyzed, 138 were found to have active football leagues. The remaining 26 countries did not have documented information regarding active football leagues.

Among the countries with active football leagues, certain regions emerged as hotspots for football:

Europe: Europe demonstrated a strong football culture, with a significant number of countries having active leagues. Major football powerhouses such as England, Spain, Germany, Italy, and France were present, along with countries like the Netherlands, Portugal, and Belgium, known for their footballing prowess.

South America: South American nations showcased a deep passion for football. Countries like Brazil and Argentina, known for producing world-class players, were at the forefront. Other countries like Uruguay, Colombia, and Chile also boasted active football leagues.

Africa: Football had a strong presence in various African nations. Countries like Nigeria, Egypt, South Africa, and Cameroon had vibrant football leagues and produced talented players.

Asia: Several Asian countries had active football leagues, contributing to the growth of the sport in the region. Nations like Japan, South Korea, China, and Saudi Arabia demonstrated a commitment to developing their football infrastructure and nurturing local talent.

Further investigation and analysis can delve into specific aspects such as league structures and historical footballing achievements of individual countries which provides a foundation for deeper research and understanding of the diverse football landscapes across the globe.

4. Description of which group participant wrote which part of the project

Filip Ferens:	Dominika Dunda
Selenium, Scrapy	Beautiful Soup