

**Python Web Scraping**

**Project Title: Web Scraping MS Dhoni's ODI Statistics Using Python**

****

**Name: Caesar Edwin**

**Email ID:** [**edwincaesar029@gmail.com**](mailto:edwincaesar029@gmail.com)

**Guided By : Mr. Sameer Warsolkar Sir**

**Institution: IT VEDANT**

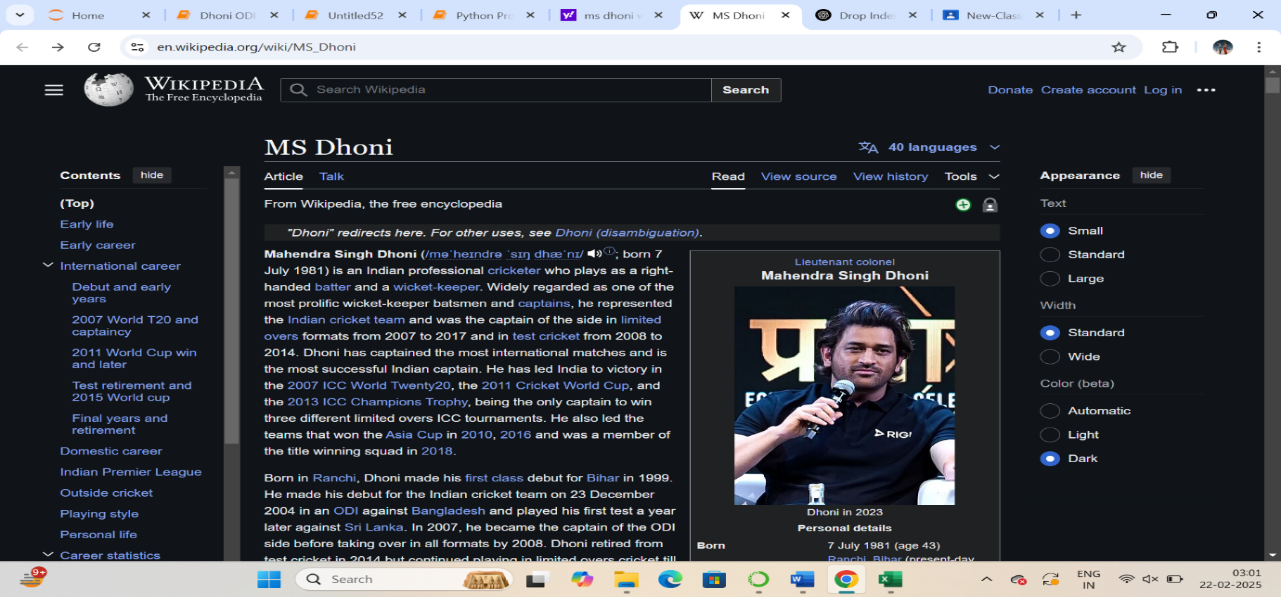
**Project Descriptions:**

**This project focuses on extracting and analysing MS Dhoni's One Day International (ODI) statistics using Python web scraping techniques. The goal is to collect data such as matches played, runs scored, batting average, strike rate, centuries, half-centuries, dismissals as a wicketkeeper, and other key performance metrics from reliable cricket statistic websites like Wikipedia.**

**The scraped data will be processed, cleaned, and stored in a structured format CSV for further analysis. Using Python libraries like BeautifulSoup and Requests, the project will automate data extraction, ensuring up-to-date insights into Dhoni's legendary ODI career.**

**To start with the web scraping process**

Select the target website: Choose website that provides necessary data e.g. MS. Dhoni’s ODI Statistics

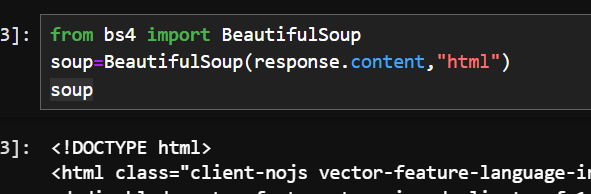
****

**To begin with the web scraping process, we need to import the essential libraries that will handle the data extraction, parsing, and storage.**

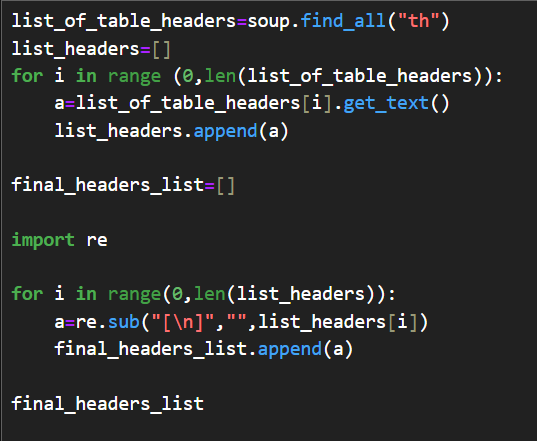
**Here are the libraries we will be using:**

* **Python for web scraping and processing**
* **BeautifulSoup and requests for web scraping**
* **Pandas for data manipulation**

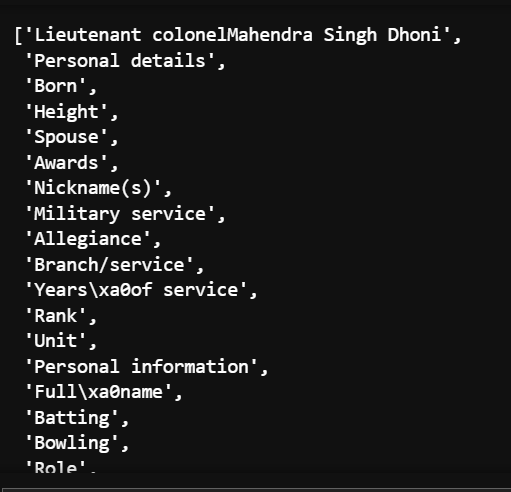
****

****

**Extracting and cleaning table data (<th> elements) from an HTML page using BeautifulSoup, ensuring newline characters (\n) are removed."**

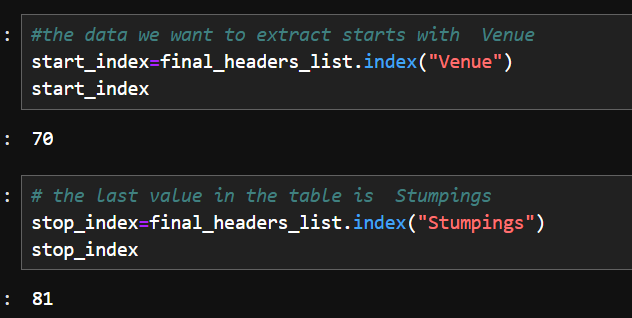
****

**OUTPUT**

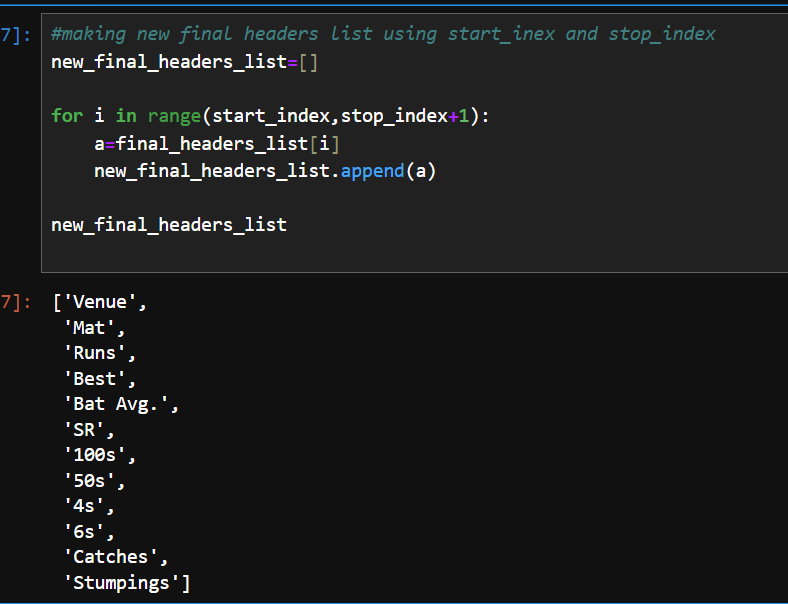
****

**To extract specific table headers from final\_headers\_list using indexes.**

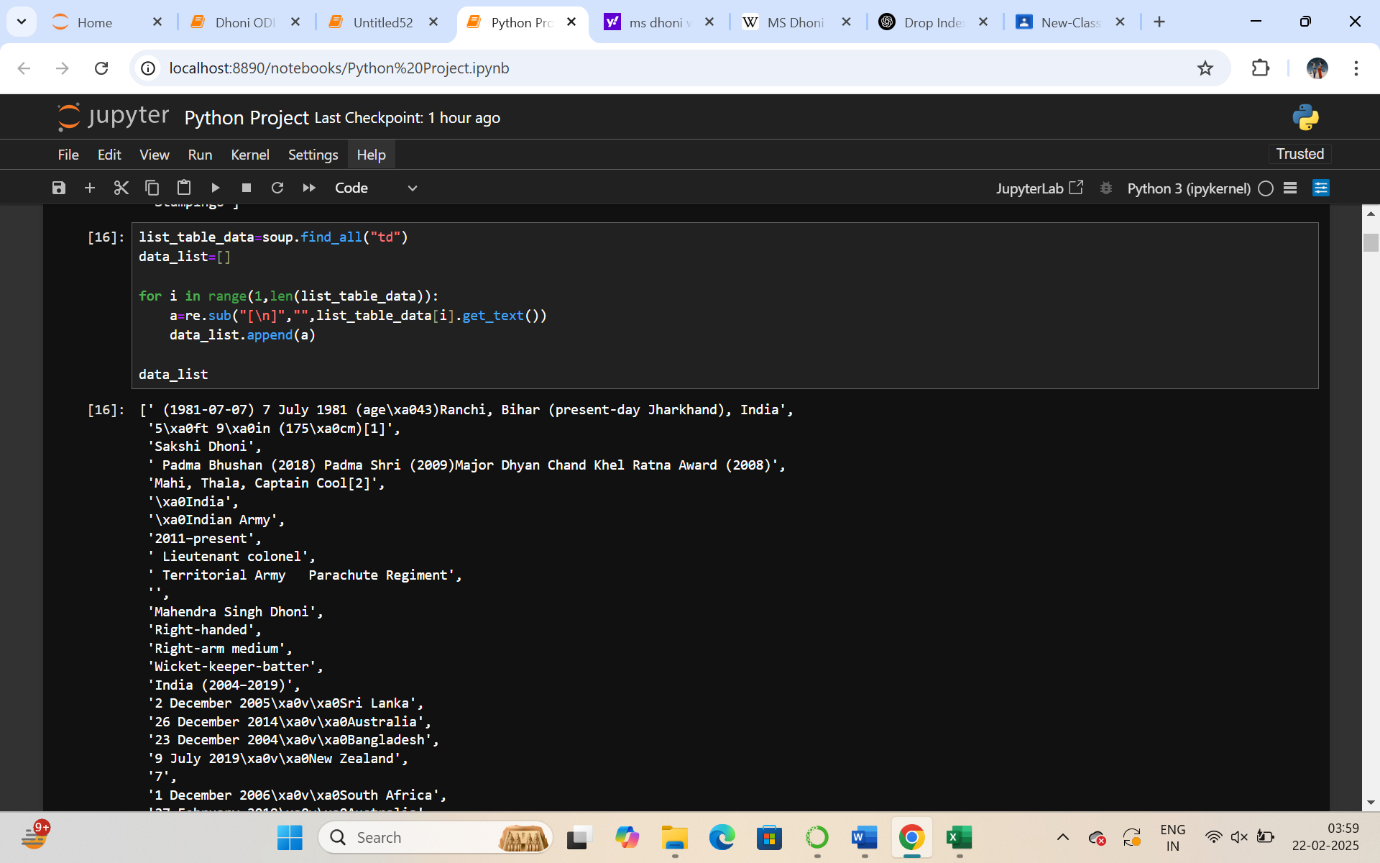
**Step 1 : Using index to find the index of required table headers**

****

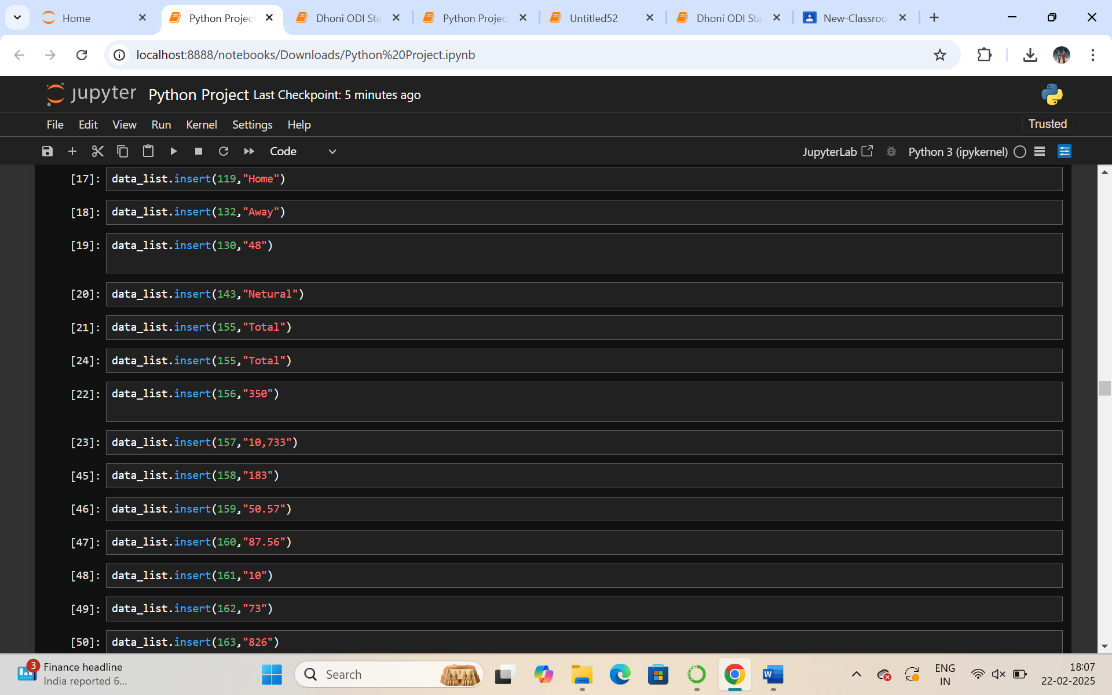
**Step 2: Creating new\_final\_headers\_list with the required values from final\_headers\_list**

****

**Extracting and cleaning table data (<td> elements) from an HTML page using BeautifulSoup, ensuring newline characters (\n) are removed."**

****

**Adding additional Details in data\_list**

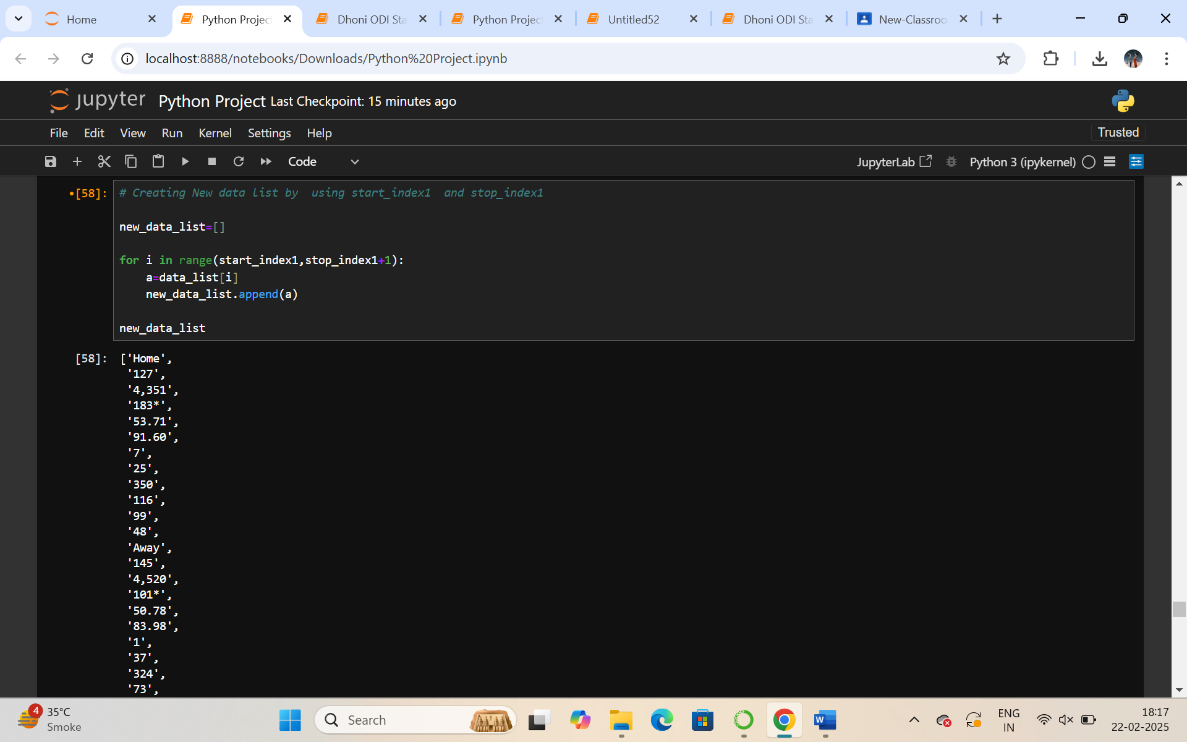
****

**To extract specific table data from data\_list using indexes**

**Step 1 : Using index to find the index of required table data**

****

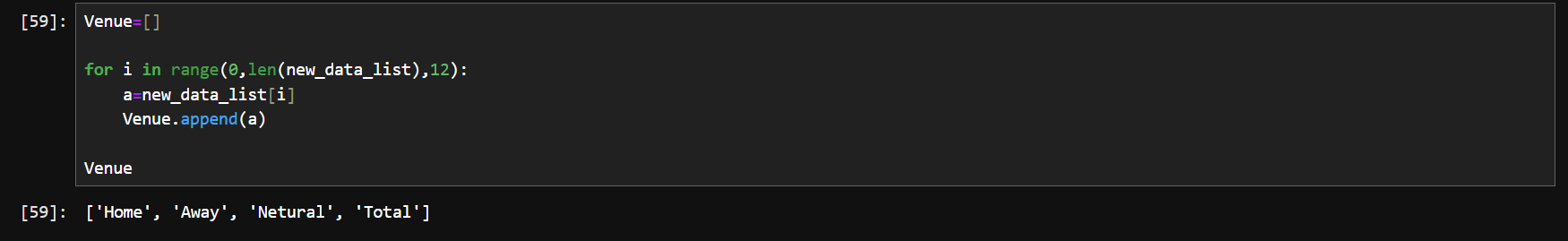
**Step 2: Creating new\_data\_list with the required values from data\_list**

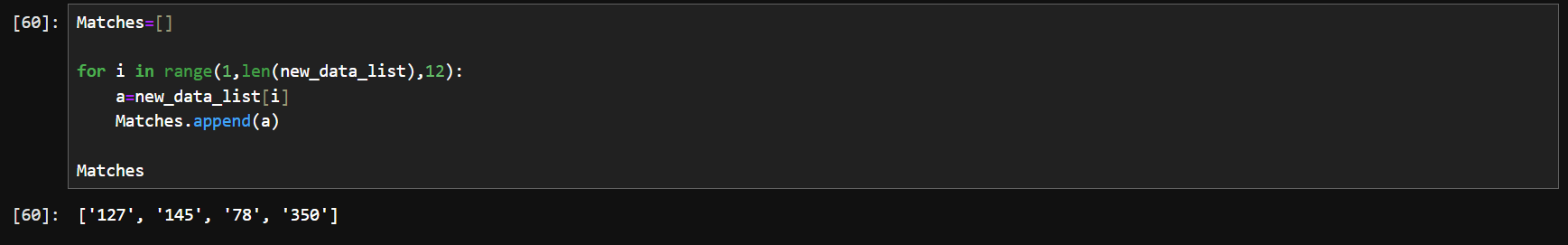
****

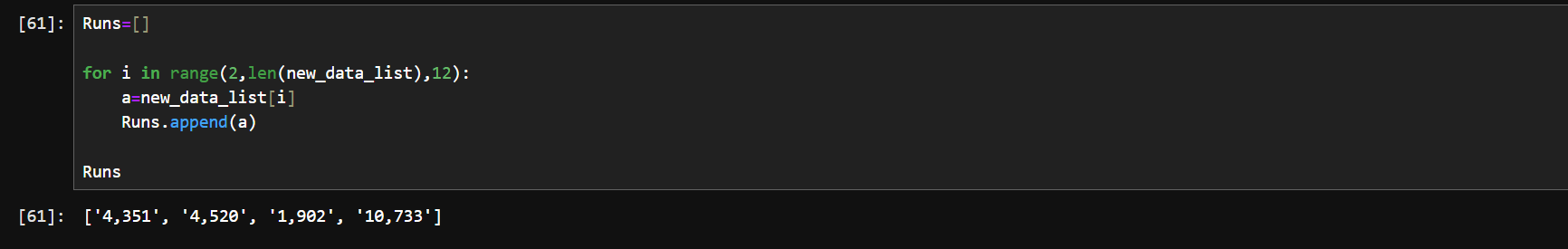
**Implementing a loop function to extract table content from the given list efficiently.**

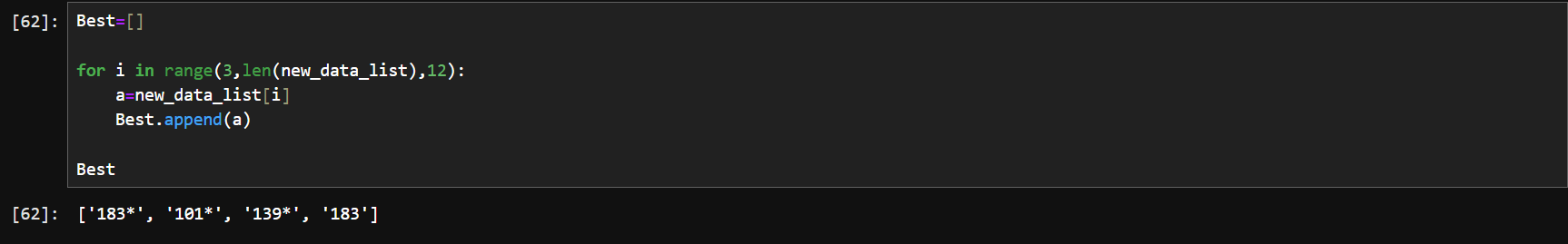
**Content to be extracted are**

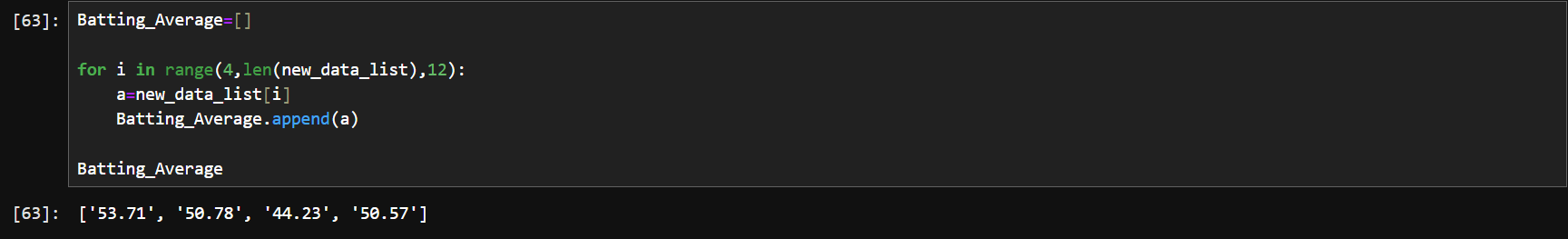
* **Venue**
* **Matches**
* **Runs**
* **Best Score**
* **Batting Average**
* **Strike Rate**
* **Hundreds**
* **Fifties**
* **Fours**
* **Sixes**
* **Catches**
* **Stumpings**

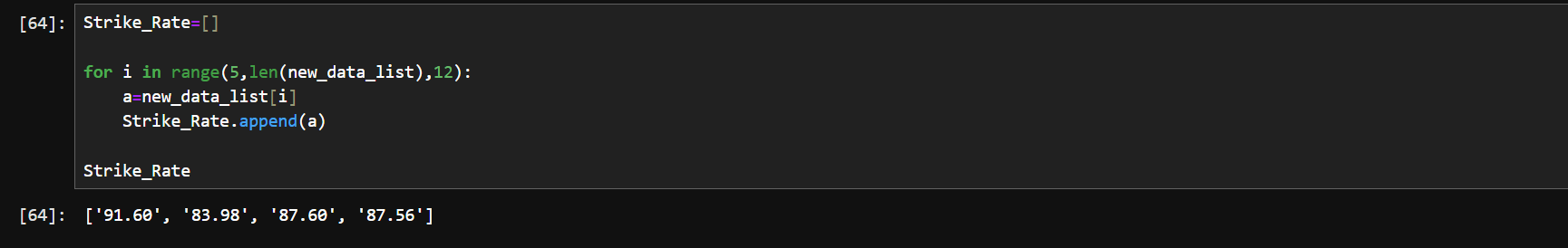
****

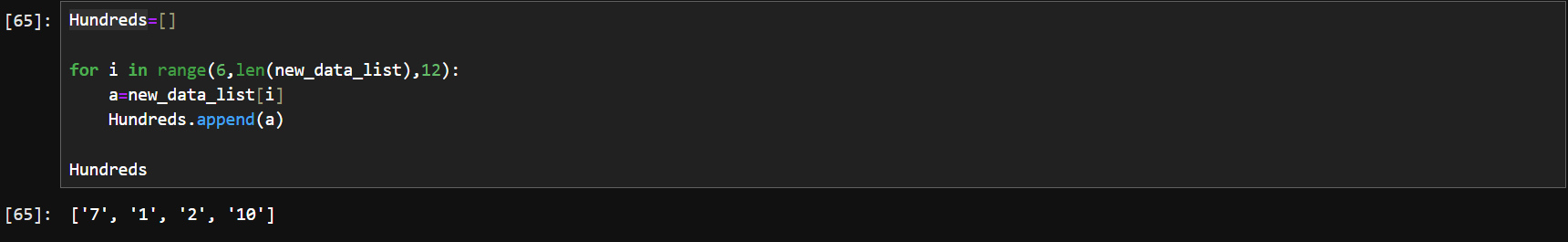
****

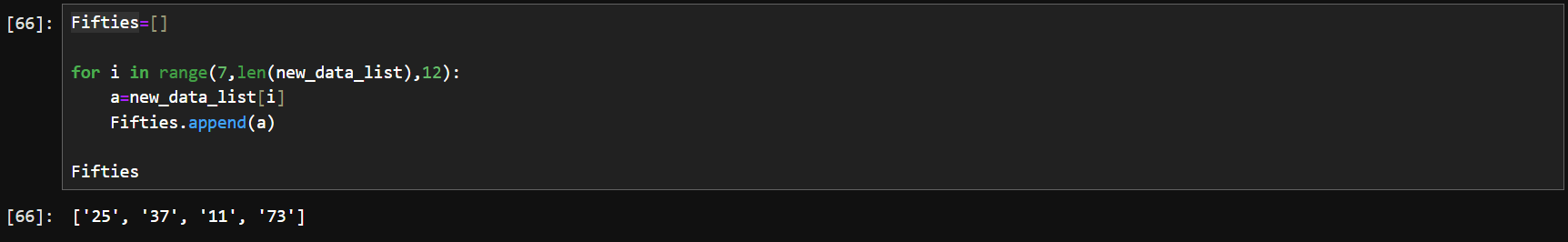
****

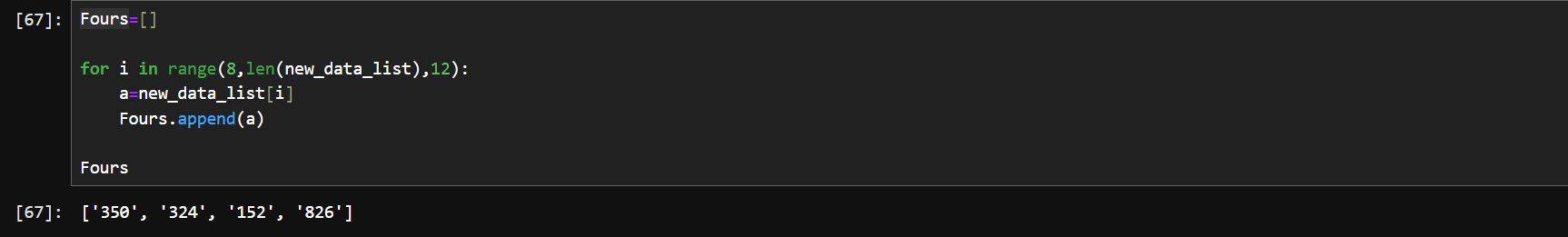
****

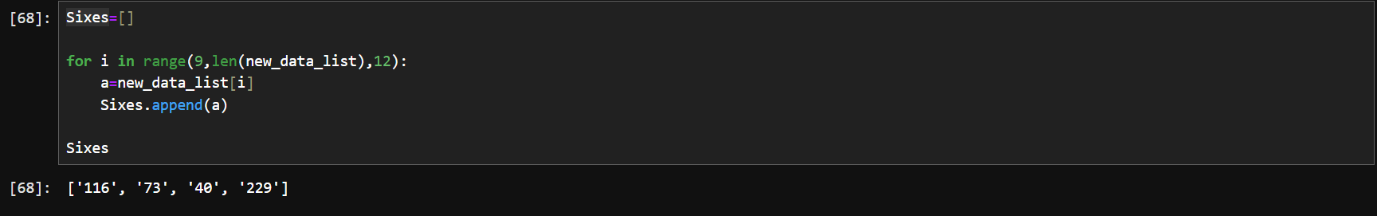
****

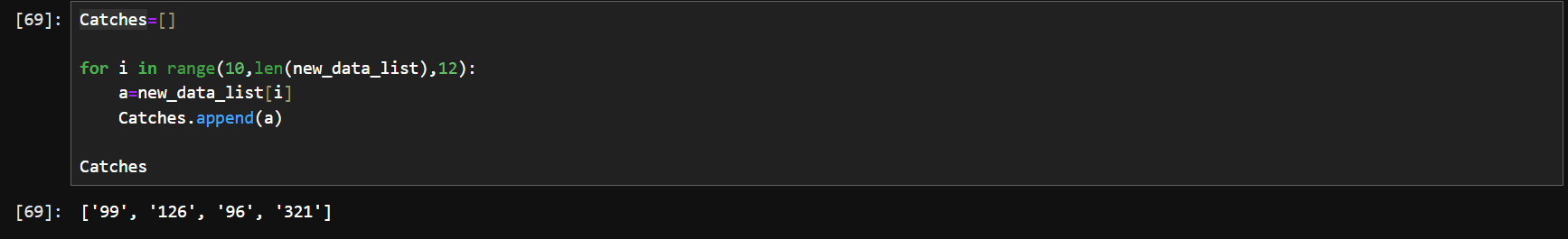
****

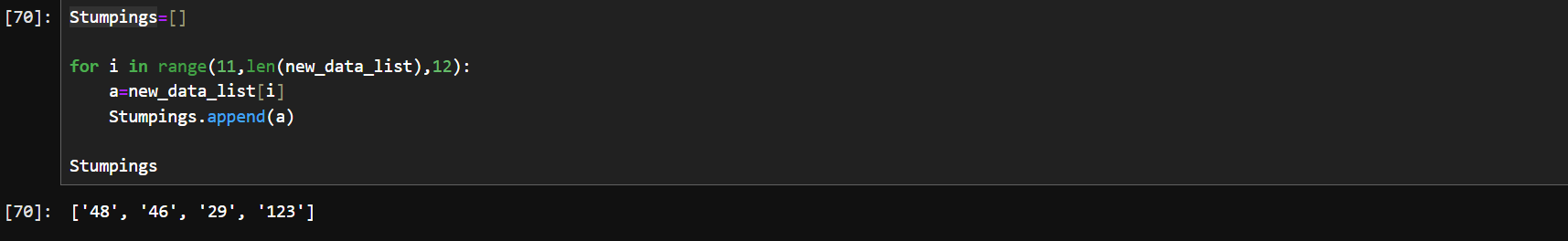
****

****

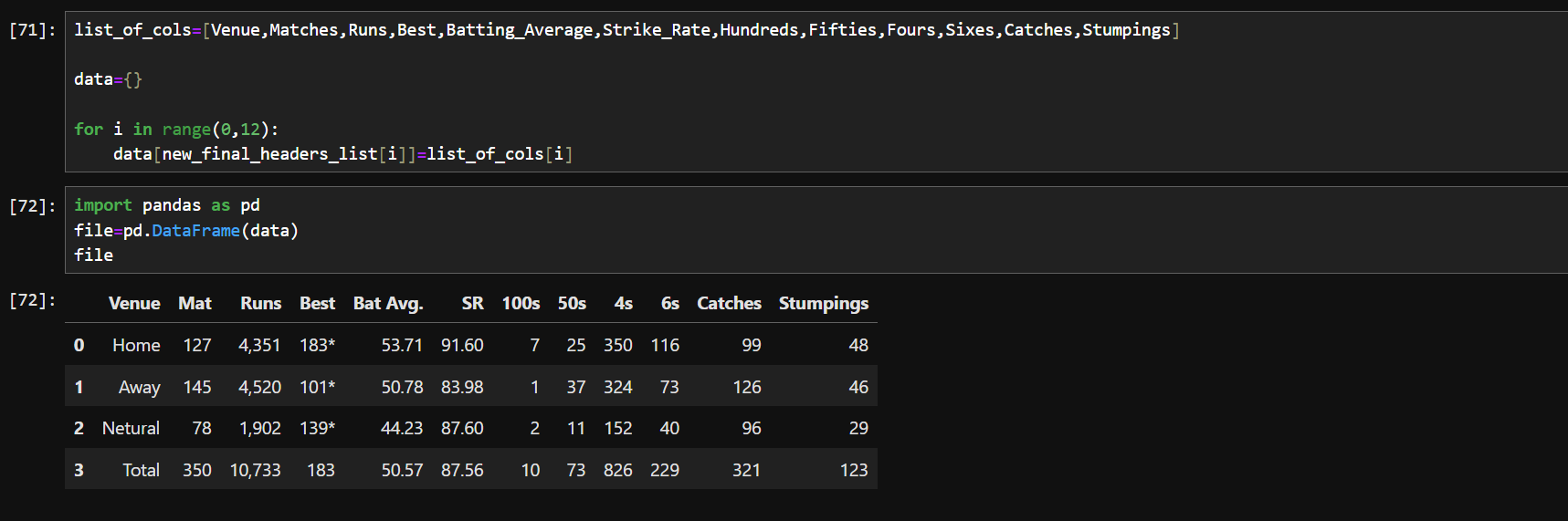
****

****

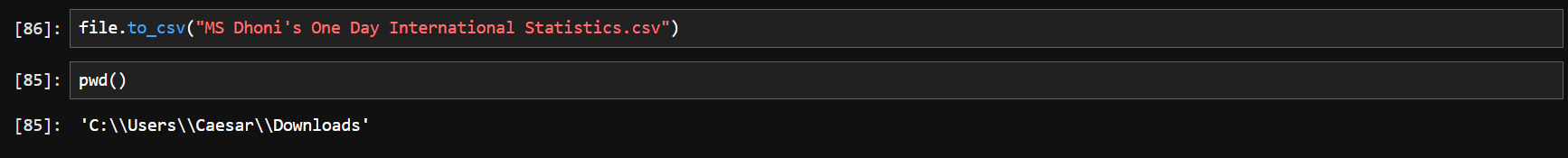
****

****

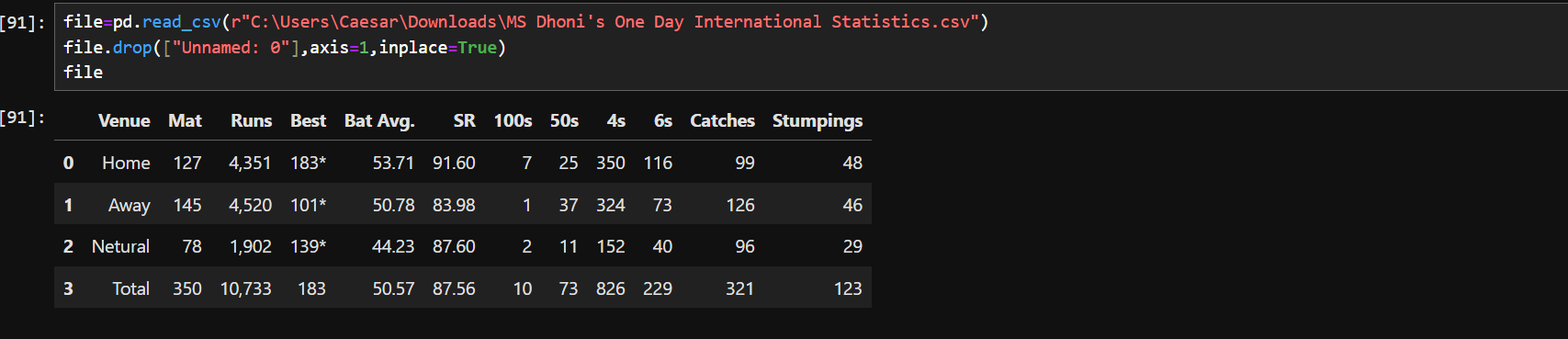
**Transforming the extracted content into a structured table format using DataFrame and Pandas libraries.**

****

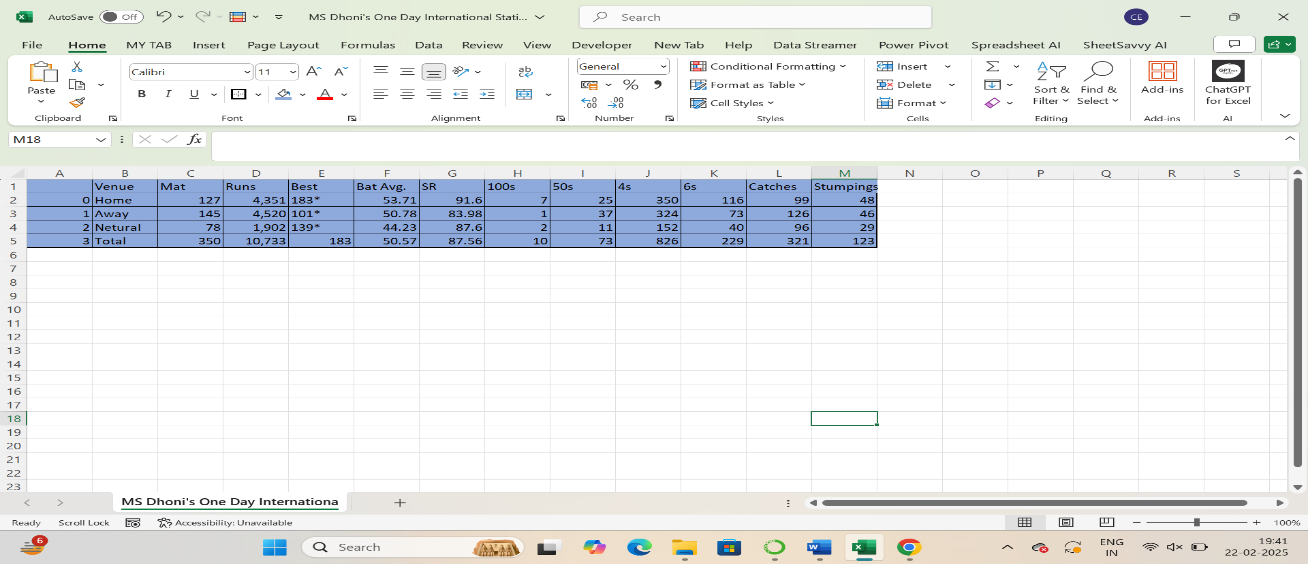
**Exporting the data into csv file**

****

**Opening the CSV file in read mode for reading**

****

**Open the CSV file in the Excel application to view its contents.**

****

**Conclusion:**

**This project successfully scraped and analysed MS Dhoni's ODI stats using Python. It provided insights into his performance, including runs, averages, and dismissals. The results highlight Dhoni’s impact as a batsman and wicketkeeper.**