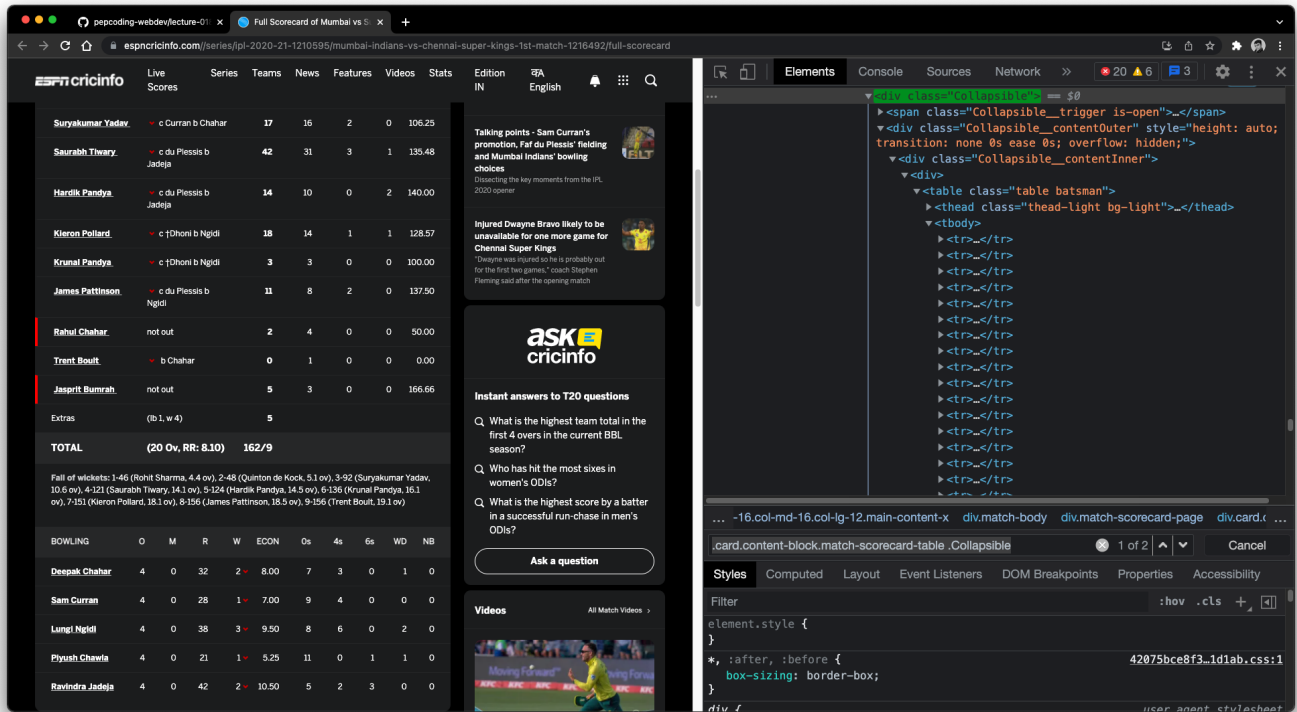


Lecture 20 | Web Scraping

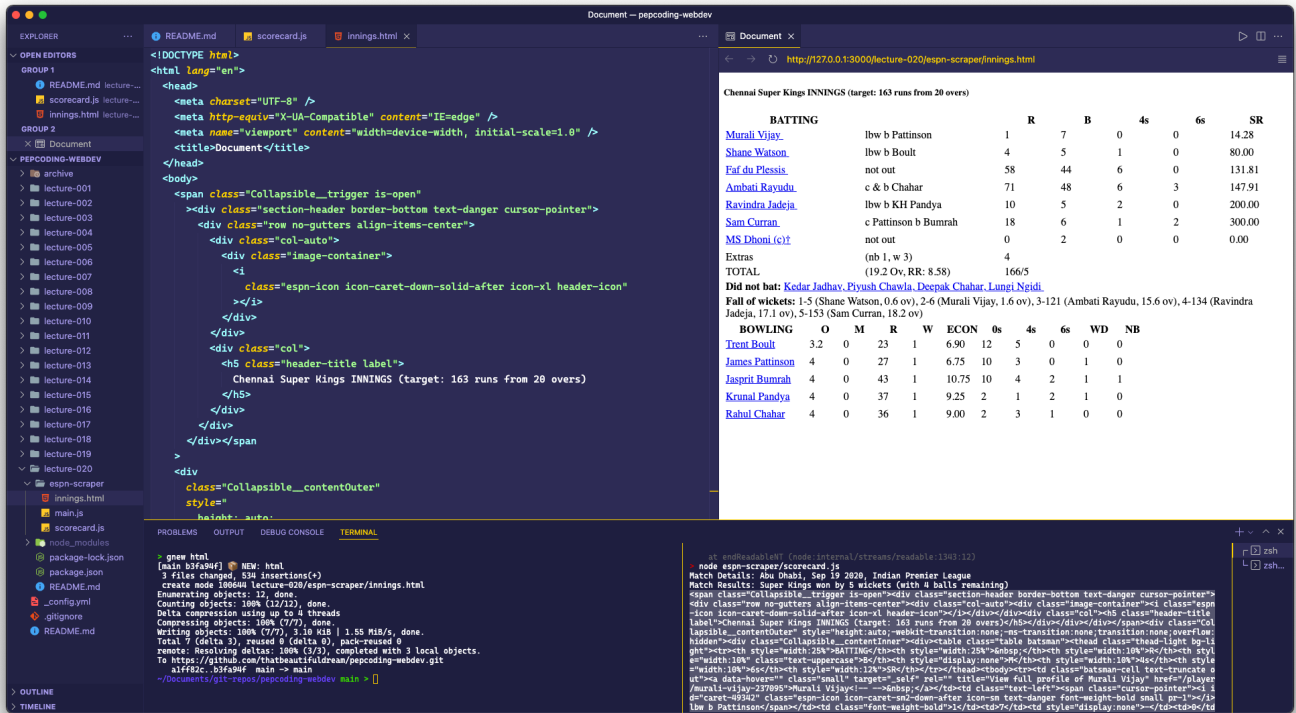
- Scrape Team performance details from [this](#)
- Access the data using class selectors



- grabbed the table contents from Mumbai and Chennai stats table as HTML and store in a html file

```
// grabbing Mumbai and Chennai stats table
let innings = $(".card.content-block.match-scorecard-table
.Collapsible");
let htmlString = "";
for (let i = 0; i < innings.length; i++) {
  htmlString += $(innings[i]).html();
}
console.log(htmlString);
}
```

- Added a new file to store the HTML data `innings.html`



```
// idea for the task was to use split() on the string and then use a for
loop to iterate through the array
```

```
const url =
  "https://www.espn-cricinfo.com/series/ipl-2020-21-1210595/mumbai-indians-
vs-chennai-super-kings-1st-match-1216492/full-scorecard";
```

```
const cheerio = require("cheerio");
const request = require("request");
```

```
request(url, cb);
```

```
function cb(error, response, html) {
  if (error) {
    console.error(error);
  } else {
    extractMatchDetails(html);
  }
}
```

```
// note : to select the desired class use parent class and then child
class (desendent)
```

```
function extractMatchDetails(html) {
  const $ = cheerio.load(html);
  let descString = $(".header-info .description").text();
  let descStringArr = descString.split(",");
  // console.log(descStringArr);
  let venue = descStringArr[1].trim(); // using trim() to remove the white
space is a good practice
  let date = descStringArr[2].trim();
```

```

let matchType = descStringArr[3].trim();
let results = $(".event .status-text span").text();
console.log(`Match Details: ${venue}, ${date}, ${matchType}`);
console.log(`Match Results: ${results}`);

// grabbing Mumbai and Chennai stats table
let innings = $(".card.content-block.match-scorecard-table
.Collapsible");
let htmlString = "";
for (let i = 0; i < innings.length; i++) {
  htmlString += $(innings[i]).html();
  let teamName = $(innings[i]).find("h5").text();
  teamName = teamName.split("INNINGS")[0].trim();
  console.log(`${teamName}`);
}
// console.log(htmlString);
}

```

Output:

```

> node scorecard.js
Match Details: Abu Dhabi, Sep 19 2020, Indian Premier League
Match Results: Super Kings won by 5 wickets (with 4 balls remaining)
Mumbai Indians
Chennai Super Kings

```

- Both teams are in a **h5** container

The screenshot shows a web browser displaying the Mumbai Indians Innings table and the Chennai Super Kings Innings table. The DOM tree on the right highlights the **h5** container for the Mumbai Indians Innings table.

Mumbai Indians INNINGS (20 overs maximum)

	R	B	4s	6s	SR
c Curran b Chawla	12	10	2	0	120.00
c Watson b Curran	33	20	5	0	165.00
c Curran b Chahar	17	16	2	0	106.25
c du Plessis b Jadeja	42	31	3	1	135.48
c du Plessis b Jadeja	14	10	0	2	140.00
c Dhoni b Ngidi	18	14	1	1	128.57
c Dhoni b Ngidi	3	3	0	0	100.00
c du Plessis b Ngidi	11	8	2	0	137.50
not out	2	4	0	0	50.00
b Chahar	0	1	0	0	0.00
not out	5	3	0	0	166.66
Extras (lb 1, w 4)	5				
TOTAL (20 Ov, RR: 8.10)	162/9				

Fall of wickets: 1-46 (Rohit Sharma, 4.4 ov), 2-48 (Quinton de Kock, 5.1 ov), 3-92 (Suryakumar Yadav, 10.6 ov), 4-121 (Saurabh Tiwary, 14.1 ov), 5-124 (Hardik Pandya, 14.5 ov), 6-136 (Krunal Pandya, 16.1 ov), 7-151 (Kieron Pollard, 18.1 ov), 8-156 (James Pattinson, 18.5 ov), 9-156 (Trent Boult, 19.1 ov)

BOWLING

	O	M	R	W	ECON	0s	4s	6s	WD	NB
Deepak Chahar	4	0	32	2	8.00	7	3	0	1	0
Sam Curran	4	0	28	1	7.00	9	4	0	0	0
Lungi Ngidi	4	0	38	3	9.50	8	6	0	2	0
Piyush Chawla	4	0	21	1	5.25	11	0	1	1	0
Ravindra Jadeja	4	0	42	2	10.50	5	2	3	0	0

Chennai Super Kings INNINGS (target: 163 runs from 20 overs)

BATTING

	R	B	4s	6s	SR
lbw b Pattinson	1	7	0	0	14.28
lbw b Boult	4	5	1	0	80.00
not out	58	44	6	0	131.81
c & b Chahar	71	48	6	3	147.91
lbw b KH Pandya	10	5	2	0	200.00
c Pattinson b Bumrah	18	6	1	2	300.00
not out	0	2	0	0	0.00
Extras (nb 1, w 3)	4				
TOTAL (19.2 Ov, RR: 8.58)	166/5				

Did not bat: Kedar Jadhav, Piyush Chawla, Deepak Chahar, Lungi Ngidi

Fall of wickets: 1-5 (Shane Watson, 0.6 ov), 2-6 (Murali Vijay, 1.6 ov), 3-121 (Ambati Rayudu, 15.6 ov), 4-134 (Ravindra Jadeja, 17.1 ov), 5-153 (Sam Curran, 18.2 ov)

BOWLING

	O	M	R	W	ECON	0s	4s	6s	WD	NB
Deepak Chahar	4	0	32	2	8.00	7	3	0	1	0
Sam Curran	4	0	28	1	7.00	9	4	0	0	0
Lungi Ngidi	4	0	38	3	9.50	8	6	0	2	0
Piyush Chawla	4	0	21	1	5.25	11	0	1	1	0
Ravindra Jadeja	4	0	42	2	10.50	5	2	3	0	0

- Based on the teams deciding the opponent, we can use the **h5** container to extract the team name

```
// grabbing Mumbai and Chennai stats table
let innings = $(".card.content-block.match-scorecard-table
.Collapsible");
let htmlString = "";
for (let i = 0; i < innings.length; i++) {
  htmlString += $(innings[i]).html();
  let teamName = $(innings[i]).find("h5").text();
  teamName = teamName.split("INNINGS")[0].trim();
  let opponentIndex = i == 0 ? 1 : 0;
  let opponentName = $(innings[opponentIndex]).find("h5").text();
  opponentName = opponentName.split("INNINGS")[0].trim();

  console.log(`${teamName} v/s ${opponentName}`);
}
// console.log(htmlString);
}
```

Output:

```
> node scorecard.js
Match Details: Abu Dhabi, Sep 19 2020, Indian Premier League
Match Results: Super Kings won by 5 wickets (with 4 balls remaining)
Mumbai Indians v/s Chennai Super Kings
Chennai Super Kings v/s Mumbai Indians
```

```
let cInning = $(innings[i]);

let allRows = cInning.find(".table.batsman tbody tr");

for (let j = 0; j < allRows.length; j++) {
  let allCols = $(allRows[j]).find("td");
  let isWorthy = $(allCols[0]).hasClass("batsman-cell");

  if (isWorthy == true) {
    let playerName = $(allCols[0]).text().trim();

    let runs = $(allCols[2]).text().trim();
    let balls = $(allCols[3]).text().trim();
    let fours = $(allCols[5]).text().trim();
    let sixes = $(allCols[6]).text().trim();
    let STR = $(allCols[7]).text().trim();

    console.log(
      `${playerName} | ${runs} | ${balls} | ${fours} | ${sixes} |
      ${STR}`
    );
  }
}
```

```

    }

    console.log(
        "-----"
    );
}
}

```

Output:

```

Rohit Sharma (c) | 12 |10 | 2 | 0 | 120.00
Quinton de Kock † | 33 |20 | 5 | 0 | 165.00
Suryakumar Yadav | 17 |16 | 2 | 0 | 106.25
Saurabh Tiwary | 42 |31 | 3 | 1 | 135.48
Hardik Pandya | 14 |10 | 0 | 2 | 140.00
Kieron Pollard | 18 |14 | 1 | 1 | 128.57
Krunal Pandya | 3 |3 | 0 | 0 | 100.00
James Pattinson | 11 |8 | 2 | 0 | 137.50
Rahul Chahar | 2 |4 | 0 | 0 | 50.00
Trent Boult | 0 |1 | 0 | 0 | 0.00
Jasprit Bumrah | 5 |3 | 0 | 0 | 166.66

```

```

-----
Murali Vijay | 1 |7 | 0 | 0 | 14.28
Shane Watson | 4 |5 | 1 | 0 | 80.00
Faf du Plessis | 58 |44 | 6 | 0 | 131.81
Ambati Rayudu | 71 |48 | 6 | 3 | 147.91
Ravindra Jadeja | 10 |5 | 2 | 0 | 200.00
Sam Curran | 18 |6 | 1 | 2 | 300.00
MS Dhoni (c)† | 0 |2 | 0 | 0 | 0.00
-----

```

Modularity

```

// exporting the function to be used in other files
module.exports = {
    ps: processScoreCard,
};

```

allMatch.js

```

const cheerio = require("cheerio");
const request = require("request");

const scorecardObj = require("../scorecard");

function getAllMatchLink(uri) {

```

```
request(uri, function (error, response, html) {
  if (error) {
    console.log(error);
  } else {
    extractAllLink(html);
  }
});
}

function extractAllLink(html) {
  let $ = cheerio.load(html);

  let scoreCardArr = $('a[data-hover="Scorecard"]');

  for (let i = 0; i < scoreCardArr.length; i++) {
    let link = $(scoreCardArr[i]).attr("href");
    let fullLink = "https://www.espnricinfo.com/" + link;
    //console.log(fullLink);

    scorecardObj.ps(fullLink);
  }
}

module.exports = {
  getAllMatch: getAllMatchLink,
};
```