Web Scraping

Table of Contents

Please note *that running these cells will create a "NBA.txt" file which is required to run subsequent cells, which will create a "clean.csv".*

In this lesson, you will learn about another method data scientists use to collect large amounts of data. Along with APIs, web scraping is a technique commonly used to scrape data from websites, forums, social media platforms, and other online sources. Web scraping is done by writing scripts or using existing tools to extract relevant information.

One useful function for web scraping with MATLAB is **webread()**. The **webread()** function is used to read the data directly into MATLAB as a character array. The function **webread()** requires a URL as an input.

Read the URL

In the example, below you will extract NBA player statistics from the following website: NBA Stats

```
%clear the workspace and command window
clc;clear;

% Retrieve the HTML content from the website using its URL
NBA = webread(['https://sports.yahoo.com/nba/stats/individual/' ...
    '?
guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_'
...

'sig=AQAAACcxyP0zC_9xC_p4e21LXs0wPttsnh6LxuXVlCf1z8wm6XQpZteIS2bcNpDoLUonhbzw
2m0f3n84C3CE3gjICj' ...
    '-10Dk4VjkE4yr9QO7PvfdK_GXYMeh0-
ltK9H1O4bpHkJyO1uwGKfWNYY4pe_sM4C9718DyR4q8wKLl9SWtyMpN'])
```

```
NBA =
    '<!DOCTYPE html><html class="ys-design desktop" id="atomic" lang="en-US"><head><script>window.performa
     window.addEventListener('pageshow', function (e) {if (e.persisted && window.rapidInstance) { window.r
     </script><script id="wafer-db-config" type="application/json">{"name":"sports-site","version":1}</scr
     if (!window.YAHOO || !window.YAHOO.i13n || !window.YAHOO.i13n.Rapid) { return; }
     var rapidConfig = {"keys":{"ver":"y20","site":"sports","navtype":"server","pt":"utility","pct":"stats
     window.rapidInstance = new window.YAHOO.i13n.Rapid(rapidConfig);
```

```
})();</script><link rel="manifest" href="/manifest.json"/></head><body><div id="app"><div class="H(10
/*! Copyright 2017 Yahoo Holdings, Inc. All rights reserved. */
⋮
```

Extract Text from HTML

If you look at the variable **NBA** in its current state you will only see HTML content. At this point, you have to look through HTML content to locate the table and the specific elements containing the player's statistics. Let's extract the text from the variable **NBA** using the function **extractHTMLText().**

```
% Extract text from the variable NBA and store it in a variable called text
text = extractHTMLText(NBA)
```

```
text =
    'Qualified Leaders

     Qualified Status  Qualified Leaders   All Players

     NBA

     NBA   Eastern   Western

     All Positions

     All Positions   Point Guard   Shooting Guard   Guard   Guard-Forward   Small Forward   Power Forward

     All Splits/Situations

     Situations  All Splits/Situations   Home   Away   Day   Night   Pre-All Star   Post-All Star   In Win

     Player  Team  G  Min  FGM  FGA  FG%  3PM  3PA  3P%  FTM  FTA  FT%  OR  DR  Reb  Ast  TO  Stl  Blk  PF

     Luka Doncic

     DAL  70  37:29  11.5  23.6  48.7  4.1  10.6  38.2  6.8  8.7  78.6  0.8  8.4  9.2  9.8  4.0  1.4  0.5

     Giannis Antetokounmpo

     MIL  73  35:10  11.5  18.8  61.1  0.5  1.7  27.4  7.0  10.7  65.7  2.7  8.8  11.5  6.5  3.4  1.2  1.1

     Shai Gilgeous-Alexander

     OKC  75  34:02  10.6  19.8  53.5  1.3  3.6  35.3  7.6  8.7  87.4  0.9  4.7  5.5  6.2  2.2  2.0  0.9

     Jalen Brunson

     NY  77  35:24  10.3  21.4  47.9  2.7  6.8  40.1  5.5  6.5  84.7  0.6  3.1  3.6  6.7  2.4  0.9  0.2  1

     Kevin Durant

     PHO  75  37:13  10.0  19.1  52.3  2.2  5.4  41.3  4.8  5.6  85.6  0.5  6.1  6.6  5.0  3.3  0.9  1.2

     Devin Booker

     PHO  68  35:59  9.4  19.2  49.2  2.2  6.1  36.4  6.0  6.7  88.6  0.8  3.7  4.5  6.9  2.6  0.9  0.4  3

     Jayson Tatum
```

BOS 74 35:45 9.1 19.3 47.1 3.1 8.2 37.6 5.6 6.7 83.3 0.9 7.2 8.1 4.9 2.5 1.0 0.6 2

De'Aaron Fox

SAC 74 35:56 9.7 20.9 46.5 2.9 7.8 36.9 4.2 5.7 73.8 0.9 3.7 4.6 5.6 2.6 2.0 0.4 2

Stephen Curry

GS 74 32:43 8.8 19.5 45.0 4.8 11.8 40.8 4.0 4.4 92.3 0.5 4.0 4.5 5.1 2.8 0.7 0.4 1

Nikola Jokic

DEN 79 34:39 10.4 17.9 58.3 1.1 2.9 35.9 4.5 5.5 81.7 2.8 9.5 12.4 9.0 3.0 1.4 0.9

Anthony Edwards

MIN 79 35:04 9.1 19.7 46.1 2.4 6.7 35.7 5.4 6.4 83.6 0.7 4.8 5.4 5.1 3.1 1.3 0.5 1

Tyrese Maxey

PHI 70 37:31 9.1 20.3 45.0 3.0 8.1 37.3 4.7 5.4 86.8 0.5 3.2 3.7 6.2 1.7 1.0 0.5 2

Kyrie Irving

DAL 58 35:00 9.7 19.5 49.7 3.0 7.3 41.1 3.3 3.6 90.5 0.8 4.2 5.0 5.2 1.8 1.3 0.5 1

Damian Lillard

MIL 73 35:20 7.4 17.5 42.4 3.0 8.5 35.4 6.5 7.0 92.0 0.5 3.9 4.4 7.0 2.6 1.0 0.2 1

DeMar DeRozan

CHI 79 37:50 8.2 17.2 48.0 0.9 2.8 33.3 6.6 7.7 85.3 0.5 3.8 4.3 5.3 1.7 1.1 0.6 2

Kawhi Leonard

LAC 68 34:16 9.0 17.1 52.5 2.1 4.9 41.7 3.7 4.2 88.5 1.2 4.9 6.1 3.6 1.8 1.6 0.9 1

Jaylen Brown

BOS 70 33:28 9.0 17.9 49.9 2.1 5.9 35.4 3.0 4.3 70.3 1.2 4.3 5.5 3.6 2.4 1.2 0.5 2

Zion Williamson

NO 70 31:32 8.9 15.6 57.0 0.1 0.3 33.3 5.0 7.1 70.2 1.7 4.1 5.8 5.0 2.8 1.1 0.7 2.

Cade Cunningham

DET 62 33:27 8.5 18.8 44.9 1.9 5.4 35.5 3.8 4.4 86.9 0.5 3.8 4.3 7.5 3.4 0.9 0.4 2

Paul George

LAC 74 33:49 7.9 16.7 47.1 3.3 7.9 41.3 3.6 3.9 90.7 0.5 4.7 5.2 3.5 2.1 1.5 0.5 2

Paolo Banchero

ORL 80 34:59 8.0 17.6 45.5 1.5 4.4 33.9 5.1 7.0 72.5 1.0 5.9 6.9 5.4 3.1 0.9 0.6 1

Dejounte Murray

ATL 78 35:41 8.6 18.8 45.9 2.6 7.1 36.3 2.7 3.4 79.4 0.8 4.5 5.3 6.4 2.6 1.4 0.3 1

Jaren Jackson Jr.

MEM 66 32:11 7.8 17.6 44.4 1.8 5.5 32.0 5.1 6.3 80.8 1.3 4.2 5.5 2.3 2.4 1.2 1.6 3

Cam Thomas

BKN 66 31:26 8.0 18.0 44.2 2.2 6.0 36.4 4.3 5.1 85.6 0.4 2.8 3.2 2.9 1.9 0.7 0.2 2

Kyle Kuzma

WAS 70 32:35 8.7 18.8 46.3 2.2 6.4 33.6 2.7 3.4 77.5 0.9 5.7 6.6 4.2 2.7 0.5 0.7 2

Karl-Anthony Towns

MIN 62 32:41 7.7 15.3 50.4 2.2 5.3 41.6 4.1 4.7 87.3 1.5 6.8 8.3 3.0 2.9 0.7 0.7 3

Pascal Siakam

IND 80 33:13 8.5 15.9 53.6 1.1 3.1 34.6 3.6 5.0 73.2 1.7 5.3 7.1 4.3 1.8 0.8 0.3 2

Victor Wembanyama

SA 71 29:40 7.8 16.7 46.5 1.8 5.5 32.5 4.1 5.2 79.6 2.3 8.4 10.6 3.9 3.7 1.2 3.6 2

Jamal Murray

DEN 59 31:33 8.0 16.7 48.1 2.5 5.8 42.5 2.7 3.1 85.3 0.7 3.4 4.1 6.5 2.1 1.0 0.7 1

Alperen Sengun

HOU 63 32:29 8.4 15.6 53.7 0.5 1.8 29.7 3.9 5.6 69.3 2.9 6.4 9.3 5.0 2.6 1.2 0.7 3

Miles Bridges

CHA 69 37:24 8.1 17.5 46.2 2.3 6.5 34.9 2.5 3.1 82.5 1.0 6.3 7.3 3.3 2.0 0.9 0.5 1

Jimmy Butler

MIA 60 34:02 6.6 13.2 49.9 1.0 2.4 41.4 6.6 7.7 85.8 1.8 3.6 5.3 5.0 1.7 1.3 0.3 1

Brandon Ingram

NO 64 32:52 7.8 15.9 49.2 1.3 3.8 35.5 3.8 4.8 80.1 0.7 4.4 5.1 5.7 2.5 0.8 0.6 2.

RJ Barrett

TOR 58 31:41 7.5 15.2 49.5 1.6 4.3 36.0 3.6 5.0 71.5 0.9 4.5 5.4 3.3 2.2 0.5 0.4 2

CJ McCollum

NO 66 32:43 7.3 16.0 45.9 3.6 8.4 42.9 1.7 2.1 82.7 0.6 3.7 4.3 4.6 1.7 0.9 0.6 1.

Scottie Barnes

TOR 60 34:54 7.5 15.7 47.5 1.7 4.9 34.1 3.3 4.2 78.1 2.4 5.9 8.2 6.1 2.8 1.3 1.5 2

Terry Rozier

MIA 61 33:27 7.2 16.4 44.3 2.4 6.7 36.3 2.8 3.2 86.9 0.6 3.5 4.0 5.6 1.7 1.0 0.3 1

Franz Wagner

ORL 72 32:28 7.3 15.2 48.2 1.3 4.6 28.1 3.8 4.4 85.0 1.0 4.3 5.3 3.7 1.9 1.1 0.4 2

Mikal Bridges

```
BKN  82  34:48  6.9  15.8  43.6  2.7  7.2  37.2  3.1  3.9  81.4  0.8  3.7  4.5  3.6  2.0  1.0  0.4  1
Jalen Green
HOU  82  31:43  6.9  16.2  42.3  2.5  7.4  33.2  3.5  4.3  80.4  0.5  4.7  5.2  3.5  2.3  0.8  0.3  1
Devin Vassell
SA  68  33:04  7.3  15.5  47.2  2.4  6.6  37.2  2.4  3.0  80.1  0.4  3.4  3.8  4.1  1.6  1.1  0.3  1.
Domantas Sabonis
SAC  82  35:42  7.7  13.0  59.4  0.4  1.1  37.9  3.6  5.1  70.4  3.6  10.1  13.7  8.2  3.3  0.9  0.6
Bam Adebayo
MIA  71  34:02  7.5  14.3  52.1  0.2  0.6  35.7  4.1  5.5  75.5  2.2  8.1  10.4  3.9  2.3  1.1  0.9
Coby White
CHI  79  36:28  6.8  15.3  44.7  2.6  7.0  37.6  2.8  3.3  83.8  0.6  4.0  4.5  5.1  2.1  0.7  0.2  2
Jalen Williams
OKC  71  31:19  7.5  14.0  54.0  1.5  3.4  42.7  2.5  3.1  81.4  0.5  3.5  4.0  4.5  1.7  1.1  0.6  2
Collin Sexton
UTA  78  26:36  6.5  13.3  48.7  1.6  4.2  39.4  4.1  4.7  85.9  0.9  1.7  2.6  4.9  2.1  0.8  0.2  2
Nikola Vucevic
CHI  76  34:21  7.7  15.9  48.4  1.2  4.1  29.4  1.4  1.7  82.2  2.8  7.8  10.5  3.3  1.6  0.7  0.8
Klay Thompson
GS  77  29:40  6.4  14.7  43.2  3.5  9.0  38.7  1.6  1.8  92.7  0.5  2.8  3.3  2.3  1.5  0.6  0.5  1.
Fred VanVleet
HOU  73  36:46  5.8  13.9  41.6  3.1  8.0  38.7  2.7  3.1  86.0  0.5  3.4  3.8  8.1  1.7  1.4  0.8  2
Jordan Poole
WAS  78  30:05  6.3  15.2  41.3  2.4  7.2  32.6  2.5  2.8  87.7  0.4  2.3  2.7  4.4  2.4  1.1  0.3  3
1
2
3
4'
```

Select Relevant Text

Notice that there is additional information that is not relevant. To extract only the players statistics use **extractAfter().**

1. Use **extractAfter()** to get information found after the the phrase "3+ Days Rest" and store the new text in a variable called **new.**


```
new = extractAfter(text, "3+ Days Rest")
```

```
new =
    '

    Player  Team  G  Min  FGM  FGA  FG%  3PM  3PA  3P%  FTM  FTA  FT%  OR  DR  Reb  Ast  TO  Stl  Blk  PF

    Luka Doncic

    DAL  70  37:29  11.5  23.6  48.7  4.1  10.6  38.2  6.8  8.7  78.6  0.8  8.4  9.2  9.8  4.0  1.4  0.5

    Giannis Antetokounmpo

    MIL  73  35:10  11.5  18.8  61.1  0.5  1.7  27.4  7.0  10.7  65.7  2.7  8.8  11.5  6.5  3.4  1.2  1.1

    Shai Gilgeous-Alexander

    OKC  75  34:02  10.6  19.8  53.5  1.3  3.6  35.3  7.6  8.7  87.4  0.9  4.7  5.5  6.2  2.2  2.0  0.9

    Jalen Brunson

    NY  77  35:24  10.3  21.4  47.9  2.7  6.8  40.1  5.5  6.5  84.7  0.6  3.1  3.6  6.7  2.4  0.9  0.2  1

    Kevin Durant

    PHO  75  37:13  10.0  19.1  52.3  2.2  5.4  41.3  4.8  5.6  85.6  0.5  6.1  6.6  5.0  3.3  0.9  1.2

    Devin Booker

    PHO  68  35:59  9.4  19.2  49.2  2.2  6.1  36.4  6.0  6.7  88.6  0.8  3.7  4.5  6.9  2.6  0.9  0.4  3

    Jayson Tatum

    BOS  74  35:45  9.1  19.3  47.1  3.1  8.2  37.6  5.6  6.7  83.3  0.9  7.2  8.1  4.9  2.5  1.0  0.6  2

    De'Aaron Fox

    SAC  74  35:56  9.7  20.9  46.5  2.9  7.8  36.9  4.2  5.7  73.8  0.9  3.7  4.6  5.6  2.6  2.0  0.4  2

    Stephen Curry

    GS  74  32:43  8.8  19.5  45.0  4.8  11.8  40.8  4.0  4.4  92.3  0.5  4.0  4.5  5.1  2.8  0.7  0.4  1

    Nikola Jokic

    DEN  79  34:39  10.4  17.9  58.3  1.1  2.9  35.9  4.5  5.5  81.7  2.8  9.5  12.4  9.0  3.0  1.4  0.9

    Anthony Edwards

    MIN  79  35:04  9.1  19.7  46.1  2.4  6.7  35.7  5.4  6.4  83.6  0.7  4.8  5.4  5.1  3.1  1.3  0.5  1

    Tyrese Maxey
```

PHI 70 37:31 9.1 20.3 45.0 3.0 8.1 37.3 4.7 5.4 86.8 0.5 3.2 3.7 6.2 1.7 1.0 0.5 2

Kyrie Irving

DAL 58 35:00 9.7 19.5 49.7 3.0 7.3 41.1 3.3 3.6 90.5 0.8 4.2 5.0 5.2 1.8 1.3 0.5 1

Damian Lillard

MIL 73 35:20 7.4 17.5 42.4 3.0 8.5 35.4 6.5 7.0 92.0 0.5 3.9 4.4 7.0 2.6 1.0 0.2 1

DeMar DeRozan

CHI 79 37:50 8.2 17.2 48.0 0.9 2.8 33.3 6.6 7.7 85.3 0.5 3.8 4.3 5.3 1.7 1.1 0.6 2

Kawhi Leonard

LAC 68 34:16 9.0 17.1 52.5 2.1 4.9 41.7 3.7 4.2 88.5 1.2 4.9 6.1 3.6 1.8 1.6 0.9 1

Jaylen Brown

BOS 70 33:28 9.0 17.9 49.9 2.1 5.9 35.4 3.0 4.3 70.3 1.2 4.3 5.5 3.6 2.4 1.2 0.5 2

Zion Williamson

NO 70 31:32 8.9 15.6 57.0 0.1 0.3 33.3 5.0 7.1 70.2 1.7 4.1 5.8 5.0 2.8 1.1 0.7 2.

Cade Cunningham

DET 62 33:27 8.5 18.8 44.9 1.9 5.4 35.5 3.8 4.4 86.9 0.5 3.8 4.3 7.5 3.4 0.9 0.4 2

Paul George

LAC 74 33:49 7.9 16.7 47.1 3.3 7.9 41.3 3.6 3.9 90.7 0.5 4.7 5.2 3.5 2.1 1.5 0.5 2

Paolo Banchero

ORL 80 34:59 8.0 17.6 45.5 1.5 4.4 33.9 5.1 7.0 72.5 1.0 5.9 6.9 5.4 3.1 0.9 0.6 1

Dejounte Murray

ATL 78 35:41 8.6 18.8 45.9 2.6 7.1 36.3 2.7 3.4 79.4 0.8 4.5 5.3 6.4 2.6 1.4 0.3 1

Jaren Jackson Jr.

MEM 66 32:11 7.8 17.6 44.4 1.8 5.5 32.0 5.1 6.3 80.8 1.3 4.2 5.5 2.3 2.4 1.2 1.6 3

Cam Thomas

BKN 66 31:26 8.0 18.0 44.2 2.2 6.0 36.4 4.3 5.1 85.6 0.4 2.8 3.2 2.9 1.9 0.7 0.2 2

Kyle Kuzma

WAS 70 32:35 8.7 18.8 46.3 2.2 6.4 33.6 2.7 3.4 77.5 0.9 5.7 6.6 4.2 2.7 0.5 0.7 2

Karl-Anthony Towns

MIN 62 32:41 7.7 15.3 50.4 2.2 5.3 41.6 4.1 4.7 87.3 1.5 6.8 8.3 3.0 2.9 0.7 0.7 3

Pascal Siakam

IND 80 33:13 8.5 15.9 53.6 1.1 3.1 34.6 3.6 5.0 73.2 1.7 5.3 7.1 4.3 1.8 0.8 0.3 2

Victor Wembanyama

SA 71 29:40 7.8 16.7 46.5 1.8 5.5 32.5 4.1 5.2 79.6 2.3 8.4 10.6 3.9 3.7 1.2 3.6 2

Jamal Murray

DEN 59 31:33 8.0 16.7 48.1 2.5 5.8 42.5 2.7 3.1 85.3 0.7 3.4 4.1 6.5 2.1 1.0 0.7 1

Alperen Sengun

HOU 63 32:29 8.4 15.6 53.7 0.5 1.8 29.7 3.9 5.6 69.3 2.9 6.4 9.3 5.0 2.6 1.2 0.7 3

Miles Bridges

CHA 69 37:24 8.1 17.5 46.2 2.3 6.5 34.9 2.5 3.1 82.5 1.0 6.3 7.3 3.3 2.0 0.9 0.5 1

Jimmy Butler

MIA 60 34:02 6.6 13.2 49.9 1.0 2.4 41.4 6.6 7.7 85.8 1.8 3.6 5.3 5.0 1.7 1.3 0.3 1

Brandon Ingram

NO 64 32:52 7.8 15.9 49.2 1.3 3.8 35.5 3.8 4.8 80.1 0.7 4.4 5.1 5.7 2.5 0.8 0.6 2.

RJ Barrett

TOR 58 31:41 7.5 15.2 49.5 1.6 4.3 36.0 3.6 5.0 71.5 0.9 4.5 5.4 3.3 2.2 0.5 0.4 2

CJ McCollum

NO 66 32:43 7.3 16.0 45.9 3.6 8.4 42.9 1.7 2.1 82.7 0.6 3.7 4.3 4.6 1.7 0.9 0.6 1.

Scottie Barnes

TOR 60 34:54 7.5 15.7 47.5 1.7 4.9 34.1 3.3 4.2 78.1 2.4 5.9 8.2 6.1 2.8 1.3 1.5 2

Terry Rozier

MIA 61 33:27 7.2 16.4 44.3 2.4 6.7 36.3 2.8 3.2 86.9 0.6 3.5 4.0 5.6 1.7 1.0 0.3 1

Franz Wagner

ORL 72 32:28 7.3 15.2 48.2 1.3 4.6 28.1 3.8 4.4 85.0 1.0 4.3 5.3 3.7 1.9 1.1 0.4 2

Mikal Bridges

BKN 82 34:48 6.9 15.8 43.6 2.7 7.2 37.2 3.1 3.9 81.4 0.8 3.7 4.5 3.6 2.0 1.0 0.4 1

Jalen Green

HOU 82 31:43 6.9 16.2 42.3 2.5 7.4 33.2 3.5 4.3 80.4 0.5 4.7 5.2 3.5 2.3 0.8 0.3 1

Devin Vassell

SA 68 33:04 7.3 15.5 47.2 2.4 6.6 37.2 2.4 3.0 80.1 0.4 3.4 3.8 4.1 1.6 1.1 0.3 1.

Domantas Sabonis

SAC 82 35:42 7.7 13.0 59.4 0.4 1.1 37.9 3.6 5.1 70.4 3.6 10.1 13.7 8.2 3.3 0.9 0.6

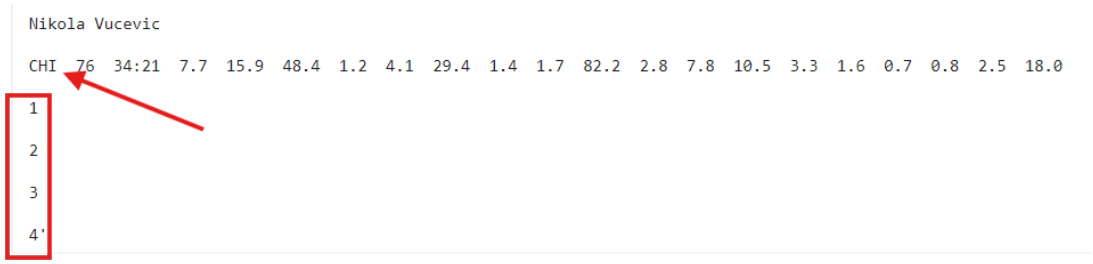
Bam Adebayo

MIA 71 34:02 7.5 14.3 52.1 0.2 0.6 35.7 4.1 5.5 75.5 2.2 8.1 10.4 3.9 2.3 1.1 0.9

Coby White

```
CHI  79  36:28  6.8  15.3  44.7  2.6  7.0  37.6  2.8  3.3  83.8  0.6  4.0  4.5  5.1  2.1  0.7  0.2  2

Jalen Williams

OKC  71  31:19  7.5  14.0  54.0  1.5  3.4  42.7  2.5  3.1  81.4  0.5  3.5  4.0  4.5  1.7  1.1  0.6  2

Collin Sexton

UTA  78  26:36  6.5  13.3  48.7  1.6  4.2  39.4  4.1  4.7  85.9  0.9  1.7  2.6  4.9  2.1  0.8  0.2  2

Nikola Vucevic

CHI  76  34:21  7.7  15.9  48.4  1.2  4.1  29.4  1.4  1.7  82.2  2.8  7.8  10.5  3.3  1.6  0.7  0.8

Klay Thompson

GS  77  29:40  6.4  14.7  43.2  3.5  9.0  38.7  1.6  1.8  92.7  0.5  2.8  3.3  2.3  1.5  0.6  0.5  1.

Fred VanVleet

HOU  73  36:46  5.8  13.9  41.6  3.1  8.0  38.7  2.7  3.1  86.0  0.5  3.4  3.8  8.1  1.7  1.4  0.8  2

Jordan Poole

WAS  78  30:05  6.3  15.2  41.3  2.4  7.2  32.6  2.5  2.8  87.7  0.4  2.3  2.7  4.4  2.4  1.1  0.3  3

1

2

3

4'
```

There is still some extra information at the end of the extracted text, but it will be easier to remove once you have created a table of the data. Additionally, the statistics corresponding to each player are not in the same row as their name.



Save Text in a .txt File

To save the extracted text into a .txt file use the code segment below.

```
% Specify the file name (e.g., 'extracted_text.txt')
fileName = 'NBA.txt';

% Open the file for writing
```

```
fileID = fopen(fileName, 'w');

% Write the extracted text to the file
fprintf(fileID,'%s\n', new);

% Close the file
fclose(fileID);
```

Clean Data in .txt File

Read the newly created .txt file using **readcell()** to create a cell array**.** In this case, you will need to use readcell('NBA.txt', 'Delimiter', 'space').

Write the function in the code block below and store the output in a variable called **Data.**

```
% Read the .txt file and store in the appropriate variable
Data = readcell('NBA.txt', 'Delimiter', 'space')
```

Data = 105x1 cell

	1
1	'Player Team G Min FGM FGA FG% 3PM 3PA 3P% FTM FTA FT% OR DR Reb Ast TO Stl Blk PF Pts'
2	'Luka Doncic'
3	'DAL 70 37:29 11.5 23.6 48.7 4.1 10.6 38.2 6.8 8.7 78.6 0.8 8.4 9.2 9.8 4.0 1.4 0.5 2.1 33.9'
4	'Giannis Antetokounmpo'
5	'MIL 73 35:10 11.5 18.8 61.1 0.5 1.7 27.4 7.0 10.7 65.7 2.7 8.8 11.5 6.5

	1
	3.4 1.2 1.1 2.9 30.4'
6	'Shai Gilgeous- Alexander'
7	'OKC 75 34:02 10.6 19.8 53.5 1.3 3.6 35.3 7.6 8.7 87.4 0.9 4.7 5.5 6.2 2.2 2.0 0.9 2.5 30.1'
8	'Jalen Brunson'
9	'NY 77 35:24 10.3 21.4 47.9 2.7 6.8 40.1 5.5 6.5 84.7 0.6 3.1 3.6 6.7 2.4 0.9 0.2 1.9 28.7'
10	'Kevin Durant'
11	'PHO 75 37:13 10.0 19.1 52.3 2.2 5.4 41.3 4.8 5.6 85.6 0.5 6.1 6.6 5.0 3.3 0.9 1.2 1.8 27.1'
12	'Devin Booker'
13	'PHO 68 35:59 9.4 19.2 49.2 2.2 6.1 36.4 6.0 6.7 88.6 0.8 3.7 4.5 6.9 2.6 0.9 0.4 3.0 27.1'
14	'Jayson Tatum'

⋮

The variable **Data** still has the same errors as before; therefore, you need to extract the information in it and reorganize it. You only need to extract information for the first 20 players.

Begin by extracting all the players names and save them in a one column cell array with 20 rows. In the code block below, fill in the missing inputs in the **cell()** function to make the cell array.

```
% Create a column array with 50 rows to extract the players names
allNames = cell(20,1);

% Extract names from every even numbered row
for i = 2:2:40
    allNames(i) = Data(i);
end

% Removes all empty cells in the cell array allNames
allNames = allNames(~cellfun('isempty', allNames));
```

Now repeat the same procedure as above to have each of the players stats in a cell array. Fill in the missing inputs in the code block below.

```
% Create a column array with 20 rows to extract the stats
allstats = cell(20,1);

% Extract stats from all odd numbered rows starting at 3.
for i = 3:2:41
    allstats(i) = Data(i);
end

% Fill in the missing inputs to remove the empty cells from the array
allstats
allstats = allstats(~cellfun('isempty', allstats))
```

```
allstats = 20x1 cell
'DAL  70  37:29  11.5  23.6  48.7  4.1  10.6  38.2  6.8  8.7  78.6  0.8  8.4  9. ···
'MIL  73  35:10  11.5  18.8  61.1  0.5  1.7  27.4  7.0  10.7  65.7  2.7  8.8  11
'OKC  75  34:02  10.6  19.8  53.5  1.3  3.6  35.3  7.6  8.7  87.4  0.9  4.7  5.5
'NY  77  35:24  10.3  21.4  47.9  2.7  6.8  40.1  5.5  6.5  84.7  0.6  3.1  3.6
'PHO  75  37:13  10.0  19.1  52.3  2.2  5.4  41.3  4.8  5.6  85.6  0.5  6.1  6.6
'PHO  68  35:59  9.4  19.2  49.2  2.2  6.1  36.4  6.0  6.7  88.6  0.8  3.7  4.5
'BOS  74  35:45  9.1  19.3  47.1  3.1  8.2  37.6  5.6  6.7  83.3  0.9  7.2  8.1
'SAC  74  35:56  9.7  20.9  46.5  2.9  7.8  36.9  4.2  5.7  73.8  0.9  3.7  4.6
'GS  74  32:43  8.8  19.5  45.0  4.8  11.8  40.8  4.0  4.4  92.3  0.5  4.0  4.5
'DEN  79  34:39  10.4  17.9  58.3  1.1  2.9  35.9  4.5  5.5  81.7  2.8  9.5  12.
    ⋮
```

Store Cleaned Data

Save the new cell array as a .csv.

1. Convert the cell array to table
2. Use **writetable()** to make the a .csv file named 'clean.csv'

```
newTable = cell2table([allNames allstats]);
writetable(newTable, 'clean.csv');
Final = readtable('clean.csv')
```

```
Final = 20x22 table
```

...

	Var1	Var2	Var3	Var4	Var5	Var6	Var7
1	'Luka'	'Doncic,DAL'	70	'37:29'	11.5000	23.6000	48.7000
2	'Giannis'	'Antetokounmpo,MIL'	73	'35:10'	11.5000	18.8000	61.1000
3	'Shai'	'Gilgeous-Alexander,OKC'	75	'34:02'	10.6000	19.8000	53.5000
4	'Jalen'	'Brunson,NY'	77	'35:24'	10.3000	21.4000	47.9000
5	'Kevin'	'Durant,PHO'	75	'37:13'	10	19.1000	52.3000
6	'Devin'	'Booker,PHO'	68	'35:59'	9.4000	19.2000	49.2000
7	'Jayson'	'Tatum,BOS'	74	'35:45'	9.1000	19.3000	47.1000
8	'De'Aaron'	'Fox,SAC'	74	'35:56'	9.7000	20.9000	46.5000
9	'Stephen'	'Curry,GS'	74	'32:43'	8.8000	19.5000	45
10	'Nikola'	'Jokic,DEN'	79	'34:39'	10.4000	17.9000	58.3000
11	'Anthony'	'Edwards,MIN'	79	'35:04'	9.1000	19.7000	46.1000
12	'Tyrese'	'Maxey,PHI'	70	'37:31'	9.1000	20.3000	45
13	'Kyrie'	'Irving,DAL'	58	'35:00'	9.7000	19.5000	49.7000
14	'Damian'	'Lillard,MIL'	73	'35:20'	7.4000	17.5000	42.4000

⋮

Now add names to each of the a variables above.

```
% Specify the new variable names
newNames = {'FirstName', 'LastName,Team', 'G', 'Min', 'FGM', 'FGA',
'FG%','3PM','3PA','3P%', 'FTM', 'FTA', 'FT%', 'OR', 'DR','Reb', 'Ast', 'TO',
'Stl', 'Blk', 'PF', 'Pts'};

% Rename the variables
```

```
Final = renamevars(Final, 1:width(Final), newNames)
```

```
Final = 20x22 table
```

...

	FirstName	LastName,Team	G	Min	FGM	FGA	FG%
1	'Luka'	'Doncic,DAL'	70	'37:29'	11.5000	23.6000	48.7000
2	'Giannis'	'Antetokounmpo,MIL'	73	'35:10'	11.5000	18.8000	61.1000
3	'Shai'	'Gilgeous-Alexander,OKC'	75	'34:02'	10.6000	19.8000	53.5000
4	'Jalen'	'Brunson,NY'	77	'35:24'	10.3000	21.4000	47.9000
5	'Kevin'	'Durant,PHO'	75	'37:13'	10	19.1000	52.3000
6	'Devin'	'Booker,PHO'	68	'35:59'	9.4000	19.2000	49.2000
7	'Jayson'	'Tatum,BOS'	74	'35:45'	9.1000	19.3000	47.1000
8	'De'Aaron'	'Fox,SAC'	74	'35:56'	9.7000	20.9000	46.5000
9	'Stephen'	'Curry,GS'	74	'32:43'	8.8000	19.5000	45
10	'Nikola'	'Jokic,DEN'	79	'34:39'	10.4000	17.9000	58.3000
11	'Anthony'	'Edwards,MIN'	79	'35:04'	9.1000	19.7000	46.1000
12	'Tyrese'	'Maxey,PHI'	70	'37:31'	9.1000	20.3000	45
13	'Kyrie'	'Irving,DAL'	58	'35:00'	9.7000	19.5000	49.7000
14	'Damian'	'Lillard,MIL'	73	'35:20'	7.4000	17.5000	42.4000

⋮

Summary

In this lesson, you were introduced to web scraping and the functions needed to scrape data from a website. While working through this lesson, you might have noticed that it mostly involved cleaning the data scraped from the website. Therefore, it is recommended to select websites with easy-to-read tables to reduce the amount of work required to get your data ready for analysis.