

Actividad Módulo 14: Interacción Con Datos

Para cada archivo:

- Leer el archivo y guardarlo en una lista
- Obtener las primeras 10 filas
- Obtener las últimas 10 filas
- Seleccionar 2-4 columnas del dataset y generar un archivo nuevo

fifa_eda.csv

```
In [ ]: # 1 - Guardar archivo en una lista
import csv
import numpy as np

with open('d:/Documentos/DataAnalysis/EBAC/Python/Modulo14/fifa_eda.csv', encoding=
    csvreader = csv.reader(f)

    header = next(csvreader)

    players = []
    for row in csvreader:
        players.append(row)

print(type(players), '\n')

# 2 - Primeros 10 elementos
print(players[:10], '\n')

# 3 - Últimos 10 elementos
print(players[-10:-1])

# 4 - Generar un archivo nuevo
data = np.array(players)

data = data[:,(1,2,3,6)]

print(data)
with open('new_players.csv', "w", encoding='UTF8') as f:
    writer = csv.writer(f)
    writer.writerow(['Name', 'Age', 'Nationality', 'Club'])
    for row in data:
        writer.writerow(row)
```

```
<class 'list'>
```

```
[[['158023', 'L. Messi', '31', 'Argentina', '94', '94', 'FC Barcelona', '110500.0',
'565.0', 'Left', '5.0', '4.0', 'RF', '2004', '2021-01-01', '5.583333333333333', '15
9.0', '226500.0'], ['20801', 'Cristiano Ronaldo', '33', 'Portugal', '94', '94', 'Ju
ventus', '77000.0', '405.0', 'Right', '5.0', '5.0', 'ST', '2018', '2022-01-01', '6.
166666666666667', '183.0', '127100.0'], ['190871', 'Neymar Jr', '26', 'Brazil', '92
', '93', 'Paris Saint-Germain', '118500.0', '290.0', 'Right', '5.0', '5.0', 'LW', '
2017', '2022-01-01', '5.75', '150.0', '228100.0'], ['193080', 'De Gea', '27', 'Spai
n', '91', '93', 'Manchester United', '72000.0', '260.0', 'Right', '4.0', '1.0', 'GK
', '2011', '2020-01-01', '6.333333333333333', '168.0', '138600.0'], ['192985', 'K.
De Bruyne', '27', 'Belgium', '91', '92', 'Manchester City', '102000.0', '355.0', 'R
ight', '4.0', '4.0', 'RCM', '2015', '2023-01-01', '5.916666666666667', '154.0', '19
6400.0'], ['183277', 'E. Hazard', '27', 'Belgium', '91', '91', 'Chelsea', '93000.0
', '340.0', 'Right', '4.0', '4.0', 'LF', '2012', '2020-01-01', '5.666666666666667',
'163.0', '172100.0'], ['177003', 'L. Modrić', '32', 'Croatia', '91', '91', 'Real Ma
drid', '67000.0', '420.0', 'Right', '4.0', '4.0', 'RCM', '2012', '2020-01-01', '5.6
666666666666667', '146.0', '137400.0'], ['176580', 'L. Suárez', '31', 'Uruguay', '91
', '91', 'FC Barcelona', '80000.0', '455.0', 'Right', '5.0', '3.0', 'RS', '2014', '
2021-01-01', '6.0', '190.0', '164000.0'], ['155862', 'Sergio Ramos', '32', 'Spain',
'91', '91', 'Real Madrid', '51000.0', '380.0', 'Right', '4.0', '3.0', 'RCB', '2005
', '2020-01-01', '6.0', '181.0', '104600.0'], ['200389', 'J. Oblak', '25', 'Sloveni
a', '90', '93', 'Atlético Madrid', '68000.0', '94.0', 'Right', '3.0', '1.0', 'GK',
'2014', '2021-01-01', '6.166666666666667', '192.0', '144500.0']]
```

```
[[['246269', 'G. Nugent', '16', 'England', '46', '66', 'Tranmere Rovers', '60.0', '
1.0', 'Right', '1.0', '2.0', 'CM', '2018', '2019-01-01', '5.833333333333333', '176.
0', '165.0'], ['246268', 'D. Walker-Rice', '17', 'England', '47', '66', 'Tranmere R
overs', '60.0', '1.0', 'Right', '1.0', '2.0', 'RW', '2018', '2019-01-01', '5.833333
333333333', '154.0', '143.0'], ['241638', 'B. Worman', '16', 'England', '47', '67',
'Cambridge United', '60.0', '1.0', 'Right', '1.0', '2.0', 'ST', '2017', '2021-01-01
', '5.666666666666667', '148.0', '165.0'], ['243165', 'N. Christoffersson', '19', '
Sweden', '47', '63', 'Trelleborgs FF', '60.0', '1.0', 'Right', '1.0', '2.0', 'ST',
'2018', '2020-01-01', '6.25', '170.0', '113.0'], ['238813', 'J. Lundstram', '19', '
England', '47', '65', 'Crewe Alexandra', '60.0', '1.0', 'Right', '1.0', '2.0', 'CM
', '2017', '2019-01-01', '5.75', '134.0', '143.0'], ['243413', 'D. Walsh', '18', 'R
epublic of Ireland', '47', '68', 'Waterford FC', '60.0', '1.0', 'Left', '1.0', '2.0
', 'RB', '2018', '2018-01-01', '6.083333333333333', '168.0', '153.0'], ['231381', '
J. Young', '18', 'Scotland', '47', '62', 'Swindon Town', '60.0', '1.0', 'Left', '1.
0', '2.0', 'ST', '2015', '2019-01-01', '5.75', '157.0', '143.0'], ['244677', 'M. Ba
ldisimo', '18', 'Canada', '47', '69', 'Vancouver Whitecaps FC', '70.0', '1.0', 'Rig
ht', '1.0', '2.0', 'CM', '2018', '2021-01-01', '5.5', '150.0', '175.0'], ['242844',
'J. Livesey', '18', 'England', '47', '70', 'Burton Albion', '60.0', '1.0', 'Right',
'1.0', '1.0', 'GK', '2018', '2021-01-01', '5.916666666666667', '154.0', '165.0']]
```

```
[['L. Messi' '31' 'Argentina' 'FC Barcelona']
['Cristiano Ronaldo' '33' 'Portugal' 'Juventus']
['Neymar Jr' '26' 'Brazil' 'Paris Saint-Germain']
...
['B. Worman' '16' 'England' 'Cambridge United']
['D. Walker-Rice' '17' 'England' 'Tranmere Rovers']
['G. Nugent' '16' 'England' 'Tranmere Rovers']]
```

Top Selling Books

```
In [ ]: import openpyxl

xlsx_file = 'Amazon_top100_bestselling_books_2009to2021.xlsx'
wb_obj = openpyxl.load_workbook(xlsx_file)
sheet = wb_obj.active

data = []

# 1 - Guardar en una Lista
for row in sheet.iter_rows(values_only=True):
    data.append(row)

print(type(data), '\n')

# 2 - Primeros 10 elementos
print(data[:10], '\n')

# 3 - Últimos 10 elementos
print(data[-10:-1], '\n')

# 4 - Guardar en un nuevo archivo
data = np.array(data)

data = data[:,(1,2,3,6)]

wb = openpyxl.Workbook()
sheet = wb.active

# Escribir datos en la hoja de excel
for i in range(len(data)):
    for j in range(len(data[i])):
        sheet.cell(row=i+1, column=j+1).value = data[i,j]

# Guardar archivo de Excel
wb.save('new_books.xlsx')

# with open('NewBooks.csv', "w", encoding='UTF8') as f:
#     writer = csv.writer(f)
#     for row in data:
#         writer.writerow(row)
```

```
<class 'list'>
```

```
[(None, 'price', 'ranks', 'title', 'no_of_reviews', 'ratings', 'author', 'cover_type', 'year', 'genre'), (0, 12.49, 1, 'The Lost Symbol', 16118, 4.4, 'Dan Brown', 'Hardcover', 2009, 'Fiction'), (1, 13.4, 2, 'The Shack: Where Tragedy Confronts Eternity', 23392, 4.7, 'William P. Young', 'Paperback', 2009, 'Fiction'), (2, 9.93, 3, 'Liberty and Tyranny: A Conservative Manifesto', 5036, 4.8, 'Mark R. Levin', 'Hardcover', 2009, 'Non Fiction'), (3, 14.3, 4, 'Breaking Dawn (The Twilight Saga, Book 4)', 16912, 4.7, 'Stephenie Meyer', 'Hardcover', 2009, 'Fiction'), (4, 9.99, 5, 'Going Rogue: An American Life', 1572, 4.6, 'Sarah Palin', 'Hardcover', 2009, 'Non Fiction'), (5, 18.29, 6, 'StrengthsFinder 2.0', 7082, 4.1, 'Gallup', 'Hardcover', 2009, 'Non Fiction'), (6, 12.66, 7, 'The Help', 18068, 4.8, 'Kathryn Stockett', 'Hardcover', 2009, 'Fiction'), (7, 17.56, 8, 'New Moon (The Twilight Saga)', 12329, 4.7, 'Stephenie Meyer', 'Paperback', 2009, 'Fiction'), (8, 58.9, 9, 'The Twilight Saga Collection', 6100, 4.7, 'Stephenie Meyer', 'Hardcover', 2009, 'Fiction')]
```

```
[(1290, 13.29, 100, 'First Little Readers Parent Pack: Guided Reading Level A: 25 Irresistible Books That Are Just the Right Level for Beginning Readers', 27332, 4.7, 'Deborah Schecter', 'Paperback', 2021, 'Non Fiction'), (1289, 7.59, 99, 'The Truth About COVID-19: Exposing The Great Reset, Lockdowns, Vaccine Passports, and the New Normal', 7246, 4.8, 'Doctor Joseph Mercola', 'Hardcover', 2021, 'Non Fiction'), (1288, 8.95, 98, 'Dragons Love Tacos', 15753, 4.8, 'Adam Rubin', 'Hardcover', 2021, 'Fiction'), (1287, 7.49, 97, 'Think and Grow Rich: The Landmark Bestseller Now Revised and Updated for the 21st Century (Think and Grow Rich Series)', 79872, 4.7, 'Napoleon Hill', 'Paperback', 2021, 'Non Fiction'), (1286, 16.69, 96, 'Will', 21401, 4.8, 'Will Smith', 'Hardcover', 2021, 'Non Fiction'), (1285, 13.4, 95, 'The Wish', 35292, 4.7, 'Nicholas Sparks', 'Hardcover', 2021, 'Fiction'), (1284, 11.89, 94, 'What Should Danny Do? (The Power to Choose Series) (The Power to Choose, 1)', 22169, 4.8, 'Adir Levy', 'Hardcover', 2021, 'Fiction'), (1283, 16.53, 93, 'A Promised Land', 125017, 4.9, 'Barack Obama', 'Hardcover', 2021, 'Non Fiction'), (1282, 12.32, 92, 'The Vanishing Half: A Novel', 96412, 4.4, 'Brit Bennett', 'Hardcover', 2021, 'Fiction')]
```

Subtítulos de Game of Thrones

```
In [ ]: import json

with open('season1.json') as f:
    subtitles = json.load(f)

header = np.array(['episode', 'subtitle'])
subtitle_list = []

# Convertir de la estructura del json a un formato tabular - Mejor para una posteri
for episode_name, episode in subtitles.items():
    for line_number, line in episode.items():
        subtitle_list.append([episode_name.replace('.srt', ''), line_number, line])

# 1 - Guardar el archivo en una lista
print(type(subtitle_list), '\n')

# 2 - Imprimir las primeras 10 líneas
print(subtitle_list[:10], '\n')

# 3 - Imprimir las últimas 10 líneas
print(subtitle_list[:-10:-1], '\n')

# 4 - Seleccionar 2 - 4 columnas y guardarlas en un nuevo archivo
np_subtitles = np.array(subtitle_list)

# Borrar número de línea y dejar únicamente capítulo y el subtítulo
np_subtitles = np_subtitles[:,(0,2)]

# --- CHANGED CODE --- #
# Dado que cargué los datos a un formato tabular, pude omitir parte del código de L
results = {}

for subtitle in np_subtitles:
    if subtitle[0] not in results:
        results[subtitle[0]] = []
    results[subtitle[0]].append(subtitle[1])

with open('new_subtitles.json', 'w') as f:
    f.write(json.dumps(results, indent=4))
```

```
<class 'list'>
```

```
[['Game Of Thrones S01E01 Winter Is Coming', '1', 'Easy, boy.'], ['Game Of Thrones S01E01 Winter Is Coming', '10', 'Our orders were to track the wildlings.'], ['Game Of Thrones S01E01 Winter Is Coming', '100', '- Right. Give it here. - No!'], ['Game Of Thrones S01E01 Winter Is Coming', '101', 'Put away your blade.'], ['Game Of Thrones S01E01 Winter Is Coming', '102', '- I take orders from your father, not you. - Please, Father!'], ['Game Of Thrones S01E01 Winter Is Coming', '103', "I'm sorry, Bran."], ['Game Of Thrones S01E01 Winter Is Coming', '104', 'Lord Stark?'], ['Game Of Thrones S01E01 Winter Is Coming', '105', 'There are five pups.'], ['Game Of Thrones S01E01 Winter Is Coming', '106', 'One for each of the Stark children.'], ['Game Of Thrones S01E01 Winter Is Coming', '107', 'The direwolf is the sigil of your house.']]
```

```
[['Game Of Thrones S01E10 Fire And Blood', '99', "This one's your father."], ['Game Of Thrones S01E10 Fire And Blood', '98', 'No! Please, no!'], ['Game Of Thrones S01E10 Fire And Blood', '97', "Mother says that shouldn't be long."], ['Game Of Thrones S01E10 Fire And Blood', '96', "...and as soon as you've had your blood, I'll put a son in you."], ['Game Of Thrones S01E10 Fire And Blood', '95', "Do as you're bid, child."], ['Game Of Thrones S01E10 Fire And Blood', '94', 'Walk with me. I want to show you something.'], ['Game Of Thrones S01E10 Fire And Blood', '93', 'Stop! No!'], ['Game Of Thrones S01E10 Fire And Blood', '92', '- Thank you, my lord. - "Your Grace." I\'m King now.'], ['Game Of Thrones S01E10 Fire And Blood', '91', 'You look quite nice.']]
```