

Bellevue University

Milestone 2 Screenshots

Timothy Jelinek

DSC350-T301

6/30/2024

JelinekMilestone2.py ×

JelinekMilestone2.py > ...

```
1 import csv
2 import pandas as pd
3
4 # define data
5 data = [
6     ["Player", "Hall of Fame Class", "College", "All-Pro
7     Selections", "Seasons"],
8     ["O.J. Simpson", 1985, "university of Southern California", 5,
9     11],
10    ["Roger Staubach", 1985, "Navy", "N/A", 11],
11    ["Paul Hornung", 1986, "University of Notre Dame", "N/A", 9],
12    ["Doak Walker", 1986, "Southern Methodist University", 5, 6],
13    ["Earl Campbell", 1991, "University of Texas", "N/A", 8],
14    ["Tony Dorsett", 1994, "University of Pittsburgh", "N/A", 12],
15    ["Marcus Allen", 2003, "University of Southern California", "N/
16    A", 16],
17    ["Barry Sanders", 2004, "Oklahoma State University", "N/A", 10],
18    ["Tim Brown", 2015, "University of Notre Dame", "N/A", 17],
19    ["Charles Woodson", 2021, "University of Michigan", "N/A", 18]
20 ]
21
22 # write data to csv file
23 with open("HalloFameFlatFile.csv", "w", newline="") as csvfile:
24     writer=csv.writer(csvfile)
25     writer.writerows(data)
26
27 print("CSV file created successfully HalloFameFlatFile.csv.")
28
29 df = pd.DataFrame(data[1:], columns=data[0])
30
31 # reformat headers
32 df.columns = [column.replace(" ", "_").lower() for column in df.
33 columns]
```

```

Columns]
30
31 # replace data to readable format
32 df.replace("N/A", pd.NA, inplace=True)
33
34 # find duplicates and bad data
35 threshold = 0
36 duplicates = df.duplicated()
37 bad_data = (df.isna().sum(axis=1) > threshold)
38
39 # fix casinng
40 df['college'] = df['college'].str.title()
41
42 # add player
43 new_row = {'Player': 'Ricky Williams', 'Hall of Fame Class': 2015,
44           'College': 'University of Texas', 'All-Pro Selections': "N/A",
45           'Seasons': 12}
44 df = pd.concat([df, pd.DataFrame([new_row])], ignore_index=True)
45
46 df.to_csv("HaloOfFameFlatFile.csv", sep="\t", index=False)

```

```

PS C:\Users\mstga\OneDrive\Desktop\Bellevue\DataWranglingForDataScience\Module4> &
C:/ProgramData/anaconda3/python.exe c:/Users/mstga/OneDrive/Desktop/Bellevue/DataWr
anglingForDataScience/Module4/JelinekMilestone2.py
CSV file created successfully HaloOfFameFlatFile.csv.

```