Creating and Accessing Pandas DataFrames

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10/15/24
Section: CPE21S4	**Date Submitted**:10/15/24
Name: Solis, Paul Vicnet M.	**Instructor**: Engr. Edcel B. Artificio

Intended Learning Outcomes (ILO):

By the end of this laboratory session, learners will be able to

- Construct and manipulate Pandas DataFrames from various data structures (such as lists, dictionaries, and NumPy arrays) while demonstrating an understanding of DataFrame attributes and methods. This includes loading the dataset, creating DataFrames with appropriate column labels and accessing data from rows and columns.

Instructions:

1. **Loading your dataset**: Refer back to your chosen dataset from the PRELIM period. Whether you downloaded it or stored it in your Google Drive, you are required to load it into the Google Colab. Watch this video to learn more about how to read CSV files in Google Colab.**(Take a screenshot to document successful execution.)**

2. **Creating a dataframe from your CSV file:** Once you have successfully loaded your dataset, you need to create a dataframe from your uploaded CSV file.**(Take a screenshot to document successful execution.)**

3. **Creating a dataframe from a dictionary of lists:** Manually create a dictionary where each value is composed of a list from your original dataset, then load it into a dataframe, before printing it. You are required to provide at least five (5) observations in your list. **(Take a screenshot to document successful execution.)**

4. **Creating a dataframe from a list of dictionaries:** Manually create a list of dictionaries from your original dataset, then pass it into a dataframe, before printing it. You are required to provide at least five (5) observations in your list. **(Take a screenshot to document successful execution.)**

5. **Selecting dataframe columns:** Execute a method that would allow you to select a single and multiple dataframe columns. **(Take a screenshot to document successful execution.)**

6. **Selecting dataframe rows:**Execute a method that would allow you to select a single and multiple dataframe rows using panda indexing and python indexing.

Output:

1.

```
[28] from google.colab import drive
     drive.mount('/content/drive')

Mounted at /content/drive

[29] path = '/content/drive/MyDrive/DATASET/labactivity_nba - all_seasons.csv'
     df = pd.read_csv(path)
     df.head(5)
```

	Unnamed: 0	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_round	...	dreb_pct	usg_pct	ts_pct	ast_pct	season	Unnamed: 22	Unnamed: 23
0	0	Randy Livingston	HOU	22	193.04	94.800728	Louisiana State	USA	1996	2	...	0.071	0.169	0.487	0.248	1996-97	NaN	NaN
1	1	Gaylon Nickerson	WAS	28	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.111	0.174	0.497	0.043	1996-97	NaN	NaN
2	2	George Lynch	VAN	26	203.20	103.418976	North Carolina	USA	1993	1	...	0.185	0.175	0.512	0.125	1996-97	NaN	NaN
3	3	George McCloud	LAL	30	203.20	102.058200	Florida State	USA	1989	1	...	0.111	0.206	0.527	0.125	1996-97	NaN	NaN
4	4	George Zidek	DEN	23	213.36	119.748288	UCLA	USA	1995	1	...	0.169	0.195	0.500	0.064	1996-97	NaN	NaN

5 rows × 27 columns

2.

```
data = pd.read_csv(path)
data
```

	Unnamed: 0	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_round	...	dreb_pct	usg_pct	ts_pct	ast_pct	season	Unnamed: 22	Unna
0	0	Randy Livingston	HOU	22	193.04	94.800728	Louisiana State	USA	1996	2	...	0.071	0.169	0.487	0.248	1996-97	NaN	
1	1	Gaylon Nickerson	WAS	28	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.111	0.174	0.497	0.043	1996-97	NaN	
2	2	George Lynch	VAN	26	203.20	103.418976	North Carolina	USA	1993	1	...	0.185	0.175	0.512	0.125	1996-97	NaN	
3	3	George McCloud	LAL	30	203.20	102.058200	Florida State	USA	1989	1	...	0.111	0.206	0.527	0.125	1996-97	NaN	
4	4	George Zidek	DEN	23	213.36	119.748288	UCLA	USA	1995	1	...	0.169	0.195	0.500	0.064	1996-97	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
12839	12839	Joel Embiid	PHI	29	213.36	127.005760	Kansas	Cameroon	2014	1	...	0.243	0.370	0.655	0.233	2022-23	NaN	
12840	12840	John Butler Jr.	POR	20	213.36	86.182480	Florida State	USA	Undrafted	Undrafted	...	0.065	0.102	0.411	0.066	2022-23	NaN	
12841	12841	John Collins	ATL	25	205.74	102.511792	Wake Forest	USA	2017	1	...	0.180	0.168	0.593	0.052	2022-23	NaN	
12842	12842	Jericho Sims	NYK	24	208.28	113.398000	Texas	USA	2021	2	...	0.175	0.074	0.780	0.044	2022-23	NaN	
12843	12843	JaMychal Green	GSW	33	205.74	102.965384	Alabama	USA	Undrafted	Undrafted	...	0.164	0.169	0.650	0.094	2022-23	NaN	

12844 rows × 27 columns

3.

```
dictionary = {
    'player_name': ['Randy Livingston', 'Gaylon Nickerson', 'George Lynch ', 'George McCloud','George Zidek'],
    'team_abbreviation': ['HOU', 'WAS', 'VAN', 'LAL', 'DEN'],
    'age': ['22', '28', '26', '30', '23'],
    'player_height': ['193.04', '190.50 ', '203.20', '203.20  ', '213.36  '],
    'player_weight': ['94.800728', '86.182480 ', '103.418976', '102.058200', '119.748288']

}
dictionary = pd.DataFrame(dictionary)
dictionary
```

	player_name	team_abbreviation	age	player_height	player_weight
0	Randy Livingston	HOU	22	193.04	94.800728
1	Gaylon Nickerson	WAS	28	190.50\t	86.182480\t
2	George Lynch\t	VAN	26	203.20	103.418976
3	George McCloud	LAL	30	203.20\t	102.058200
4	George Zidek	DEN	23	213.36\t	119.748288

4.

```
DataList = [
    {'player_name': 'Randy Livingston', 'team_abbreviation': 'HOU', 'age': '22', 'player_height': '193.04', 'player_weight': '94.800728'},
    {'player_name': 'Gaylon Nickerson', 'team_abbreviation': 'WAS', 'age': '28', 'player_height': '190.50', 'player_weight': '86.182480'},
    {'player_name': 'George Lynch', 'team_abbreviation': 'VAN', 'age': '26', 'player_height': '203.20', 'player_weight': '103.418976'},
    {'player_name': 'George McCloud', 'team_abbreviation': 'LAL', 'age': '30', 'player_height': '203.20', 'player_weight': '102.058200'},
    {'player_name': 'George Zidek', 'team_abbreviation': 'DEN', 'age': '23', 'player_height': '213.36', 'player_weight': '119.748288'}
]
DataList = pd.DataFrame(DataList)
DataList
```

	player_name	team_abbreviation	age	player_height	player_weight
0	Randy Livingston	HOU	22	193.04	94.800728
1	Gaylon Nickerson	WAS	28	190.50	86.182480
2	George Lynch	VAN	26	203.20	103.418976
3	George McCloud	LAL	30	203.20	102.058200
4	George Zidek	DEN	23	213.36	119.748288

5.

```
print("Single Column Selection: Player Name")
print(df['player_name'].head(5))
print("\n")
print("Multiple Column Selection: Player Name, Team Abbreviation")
print(df[['player_name', 'team_abbreviation']].head(5))
print("\n")
```

```
Single Column Selection: Player Name
0    Randy Livingston
1    Gaylon Nickerson
2        George Lynch
3      George McCloud
4        George Zidek
Name: player_name, dtype: object


Multiple Column Selection: Player Name, Team Abbreviation
       player_name team_abbreviation
0  Randy Livingston               HOU
1  Gaylon Nickerson               WAS
2      George Lynch               VAN
3    George McCloud               LAL
4      George Zidek               DEN
```

6.

```
print("Single Row Selection (First Obeservaion)")
print(df.iloc[0])
print("\n")
print("Multiple Row Selection (First 5 Obeservation)")
print(df.iloc[0:5])
print("\n")
```

```
Single Row Selection (First Obeservaion)
Unnamed: 0                    0
player_name      Randy Livingston
team_abbreviation             HOU
age                            22
player_height              193.04
player_weight           94.800728
college          Louisiana State
country                      USA
draft_year                  1996
draft_round                    2
draft_number                  42
gp                            64
pts                          3.9
reb                          1.5
ast                          2.4
net_rating                   0.3
oreb_pct                   0.042
dreb_pct                   0.071
usg_pct                    0.169
ts_pct                     0.487
ast_pct                    0.248
season                   1996-97
Unnamed: 22                  NaN
Unnamed: 23                  NaN
Unnamed: 24                  NaN
Unnamed: 25                  NaN
Unnamed: 26                  NaN
Name: 0, dtype: object


Multiple Row Selection (First 5 Obeservation)
   Unnamed: 0       player_name team_abbreviation  age  player_height  \
0           0  Randy Livingston               HOU   22         193.04
1           1  Gaylon Nickerson               WAS   28         190.50
```