Creating and Accessing Pandas DataFrames

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10/15/24
Section: CPE 21S4	**Date Submitted**: 10/15/24
Name: Dominic Joseph P. Virtucio	**Instructor**: Ma'am Sayo

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:

Instructions	Documentation
1. Loading your dataset:	

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

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True)
```

```
[17] import pandas as pd
```

```
[18] #Loading your dataset:
     path="/content/drive/MyDrive/DataSet/Football-Team.csv"
     df=pd.read_csv(path)
     df.head(5)
```

	team	goals_scored	goals_conceded	wins	draws	losses	points	goal_difference	rank
0	Manchester City	179	66	55	12	9	177	113	1
1	Liverpool	154	83	44	19	13	151	71	2
2	Arsenal	146	68	46	12	18	150	78	3
3	Manchester United	130	102	39	17	20	134	28	4
4	Chelsea	135	99	37	19	20	130	36	5

2. Creating a dataframe from your CSV file:

```
#Creating a dataframe from your CSV file:
print(df)
```

```
                       team  goals_scored  goals_conceded  wins  draws  \
0             Manchester City           179              66    55     12
1                   Liverpool           154              83    44     19
2                     Arsenal           146              68    46     12
3           Manchester United           130             102    39     17
4                     Chelsea           135              99    37     19
5           Tottenham Hotspur           142             106    38     14
6                 Aston Villa           131             107    36     15
7             West Ham United           122             121    33     18
8                     Everton            87              99    30     17
9            Newcastle United           131             124    30     15
10             Crystal Palace            98             124    25     18
11   Wolverhampton Wanderers             86             117    25     16
12  Brighton and Hove Albion             95             108    21     26
13                    Fulham            82             114    18     21
14             Leicester City            68              50    20      6
15                    Burnley            74             133    15     18
16               Leeds United            62              54    18      5
17                Bournemouth            54              67    13      9
18                Southampton            47              68    12      7
19                  Brentford            56              65    10      9
20           Sheffield United            55             167    10      9
21          Nottingham Forest            49              67     9      9
22                  Luton Town            52              85     6      8
23      West Bromwich Albion              35              76     5     11

    losses  points  goal_difference  rank
0        9     177             113     1
1       13     151              71     2
2       18     150              78     3
3       20     134              28     4
4       20     130              36     5
5       24     128              36     6
6       25     123              24     7
7       25     117               1     8
8       29     107             -12     9
9       31     105               7    10
10      33      93             -26    11
11      35      91             -31    12
12      29      89             -13    13
13      37      75             -32    14
14      12      66              18    15
15      43      63             -59    16
16      15      59               8    17
17      16      48             -13    18
18      19      43             -21    19
19      19      39              -9    20
20      57      39            -112    21
21      20      36             -18    22
22      24      26             -33    23
23      22      26             -41    24
```

3. Creating a dataframe from a dictionary of lists:

```
[21] #Creating a dictionary with lists:
     data = {
         'team': ['Manchester City', 'Liverpool', 'Arsenal', 'Manchester United', 'Chelsea'],
         'goals_scored': [179, 154, 146, 130, 135],
         'goals_conceded': [66, 83, 68, 102, 99],
         'wins': [55, 44, 46, 39, 37],
         'draws': [12, 19, 12, 17, 19],
         'losses': [9, 13, 18, 20, 20],
         'points': [177, 151, 150, 134, 130],
         'goal_difference': [113, 71, 78, 28, 36],
         'rank': [1, 2, 3, 4, 5]Z
     }

     df = pd.DataFrame(data)
     print(df)
```

```
                team  goals_scored  goals_conceded  wins  draws  losses  \
0    Manchester City           179              66    55     12       9
1          Liverpool           154              83    44     19      13
2            Arsenal           146              68    46     12      18
3  Manchester United           130             102    39     17      20
4            Chelsea           135              99    37     19      20

   points  goal_difference  rank
0     177              113     1
1     151               71     2
2     150               78     3
3     134               28     4
4     130               36     5
```

4.Creating a dataframe from a list of dictionaries:

```
[23] #Creating a dataframe from a list of dictionaries:
     data = [
         {'team': 'Manchester City', 'goals_scored': 179, 'goals_conceded': 66, 'wins': 55, 'draws': 12, 'losses': 9, 'points': 177, 'goal_difference': 113, 'rank': 1},
         {'team': 'Liverpool', 'goals_scored': 154, 'goals_conceded': 83, 'wins': 44, 'draws': 19, 'losses': 13, 'points': 151, 'goal_difference': 71, 'rank': 2},
         {'team': 'Arsenal', 'goals_scored': 146, 'goals_conceded': 68, 'wins': 46, 'draws': 12, 'losses': 18, 'points': 150, 'goal_difference': 78, 'rank': 3},
         {'team': 'Manchester United', 'goals_scored': 130, 'goals_conceded': 102, 'wins': 39, 'draws': 17, 'losses': 20, 'points': 134, 'goal_difference': 28, 'rank': 4},
         {'team': 'Chelsea', 'goals_scored': 135, 'goals_conceded': 99, 'wins': 37, 'draws': 19, 'losses': 20, 'points': 130, 'goal_difference': 36, 'rank': 5},
     ]

     df = pd.DataFrame(data)
     print(df)
```

```
                team  goals_scored  goals_conceded  wins  draws  losses  \
0    Manchester City           179              66    55     12       9
1          Liverpool           154              83    44     19      13
2            Arsenal           146              68    46     12      18
3  Manchester United           130             102    39     17      20
4            Chelsea           135              99    37     19      20

   points  goal_difference  rank
0     177              113     1
1     151               71     2
2     150               78     3
3     134               28     4
4     130               36     5
```

5. Selecting dataframe columns:

Single Columns:

```
[25] #Selecting dataframe columns:
     single_column = df['team']
     print(single_column)
```

```
0      Manchester City
1            Liverpool
2              Arsenal
3    Manchester United
4              Chelsea
Name: team, dtype: object
```

Multiple Columns:

```
multiple_columns = df[['team', 'wins']]

print(multiple_columns)
```

```
                team  wins
0    Manchester City    55
1          Liverpool    44
2            Arsenal    46
3  Manchester United    39
4            Chelsea    37
```

6. Selecting data frame rows

Python Indexing

SIngle Rows:

```
[34] #Creating Single Rows using python indexing
     single_row_python = df[2:3]
     print(single_row_python)
```

```
       team  goals_scored  goals_conceded  wins  draws  losses  points  \
2  Arsenal           146              68    46     12      18     150

   goal_difference  rank
2               78     3
```

Multiple Rows:

```
#Creating Single Rows using python indexing
multiple_rows = df[1:4]
print(multiple_rows)
```

```
                team  goals_scored  goals_conceded  wins  draws  losses  \
1          Liverpool           154              83    44     19      13
2            Arsenal           146              68    46     12      18
3  Manchester United           130             102    39     17      20

   points  goal_difference  rank
1     151               71     2
2     150               78     3
3     134               28     4
```

Pandas Indexing

SIngle Rows:

```
single_row_pandas = df.iloc[[0]]
print(single_row_pandas)
```

```
             team  goals_scored  goals_conceded  wins  draws  losses  points  \
0  Manchester City           179              66    55     12       9     177

   goal_difference  rank
0              113     1
```

Multiple Rows:

```
multiple_rows_pandas = df.iloc[[0, 1, 2]]
print(multiple_rows_pandas)
```

```
              team  goals_scored  goals_conceded  wins  draws  losses  points  \
0  Manchester City           179              66    55     12       9     177
1        Liverpool           154              83    44     19      13     151
2          Arsenal           146              68    46     12      18     150

   goal_difference  rank
0              113     1
1               71     2
2               78     3
```