

| Creating and Accessing Pandas DataFrames  |  |
|---|--|
| <b>Course Code:</b> CPE 031   | <b>Program:</b> Computer Engineering       |
| <b>Course Title:</b> Visualization and Data Analysis  | <b>Date Performed:</b> 10/15/2024          |
| <b>Section:</b> CPE21S4   | <b>Date Submitted:</b> 10/15/2024          |
| <b>Name:</b> Reyes, Alexzander J.   | <b>Instructor:</b> Mrs. Maria Rizette Sayo |
| <b>Intended Learning Outcomes (ILO):</b><br><br>By the end of this laboratory session, learners will be able to <ul style="list-style-type: none"> <li>- 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.</li> </ul>   |  |
| <b>Instructions:</b><br><br><ol style="list-style-type: none"> <li>1. <b>Loading your dataset:</b> 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 <a href="#">Google Colab</a>. Watch this <a href="#">video</a> to learn more about how to read CSV files in Google Colab. <b>(Take a screenshot to document successful execution.)</b></li> <li>2. <b>Creating a dataframe from your CSV file:</b> Once you have successfully loaded your dataset, you need to create a dataframe from your uploaded CSV file. <b>(Take a screenshot to document successful execution.)</b></li> <li>3. <b>Creating a dataframe from a dictionary of lists:</b> 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. <b>(Take a screenshot to document successful execution.)</b></li> <li>4. <b>Creating a dataframe from a list of dictionaries:</b> 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. <b>(Take a screenshot to document successful execution.)</b></li> <li>5. <b>Selecting dataframe columns:</b> Execute a method that would allow you to select a single and multiple dataframe columns. <b>(Take a screenshot to document successful execution.)</b></li> <li>6. <b>Selecting dataframe rows:</b> Execute a method that would allow you to select a single and multiple dataframe rows using panda indexing and python indexing.</li> </ol> |  |

## Output:

```
[13] import pandas as pd

[18] 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).

[19] # Creating a dataframe from your CSV file
path = "/content/drive/MyDrive/DATA SET/NBA PLAYER SET - 2023_nba_player_stats.csv"
df = pd.read_csv(path)
df.head(5)
```

|   | Player Name             | POS | Team | Age | GP | W  | L  | Min    | PTS  |
|---|-------------------------|-----|------|-----|----|----|----|--------|------|
| 0 | Jayson Tatum            | SF  | BOS  | 25  | 74 | 52 | 22 | 2732.2 | 2225 |
| 1 | Joel Embiid             | C   | PHI  | 29  | 66 | 43 | 23 | 2284.1 | 2183 |
| 2 | Luka Doncic             | PG  | DAL  | 24  | 66 | 33 | 33 | 2390.5 | 2138 |
| 3 | Shai Gilgeous-Alexander | PG  | OKC  | 24  | 68 | 33 | 35 | 2416.0 | 2135 |
| 4 | Giannis Antetokounmpo   | PF  | MIL  | 28  | 63 | 47 | 16 | 2023.6 | 1959 |

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

## Creating a dataframe from a dictionary of lists:

```
# Create a dictionary of lists based on original data
data_dict = {
    'Player Name': ['Jayson Tatum', 'Joel Embiid', 'Luka Doncic',
                   'Shai Gilgeous-Alexander', 'Giannis Antetokounmpo'],
    'POS': ['SF', 'C', 'PG', 'PG', 'PF'],
    'Team': ['BOS', 'PHI', 'DAL', 'OKC', 'MIL'],
    'Age': [25, 29, 24, 24, 28],
    'PTS': [2225, 2183, 2138, 2135, 1959]
}

# Load the dictionary into a DataFrame
df_dict = pd.DataFrame(data_dict)

# Display the DataFrame to verify
print("\n==== DataFrame from Dictionary of Lists =====")
print(df_dict.to_string(index=False))
```

```
==== DataFrame from Dictionary of Lists =====
Player Name POS Team Age PTS
Jayson Tatum SF BOS 25 2225
Joel Embiid C PHI 29 2183
Luka Doncic PG DAL 24 2138
Shai Gilgeous-Alexander PG OKC 24 2135
Giannis Antetokounmpo PF MIL 28 1959
```

## Creating a dataframe from a list of dictionaries

```
[40] # Create a list of dictionaries with at least five observations
data_list = [
    {'Player Name': 'Jayson Tatum', 'POS': 'SF', 'Team': 'BOS', 'Age': 25, 'PTS': 2225},
    {'Player Name': 'Joel Embiid', 'POS': 'C', 'Team': 'PHI', 'Age': 29, 'PTS': 2183},
    {'Player Name': 'Luka Doncic', 'POS': 'PG', 'Team': 'DAL', 'Age': 24, 'PTS': 2138},
    {'Player Name': 'Shai Gilgeous-Alexander', 'POS': 'PG', 'Team': 'OKC', 'Age': 24, 'PTS': 2135},
    {'Player Name': 'Giannis Antetokounmpo', 'POS': 'PF', 'Team': 'MIL', 'Age': 28, 'PTS': 1959}
]

# Load the list into a DataFrame
df_list = pd.DataFrame(data_list)

# Display the DataFrame to verify
print("==== DataFrame from List of Dictionaries =====")
print(df_list)
print("\n")
```

```
==== DataFrame from List of Dictionaries =====
   Player Name POS Team  Age  PTS
0   Jayson Tatum  SF  BOS   25 2225
1   Joel Embiid  C  PHI   29 2183
2     Luka Doncic  PG  DAL   24 2138
3 Shai Gilgeous-Alexander  PG  OKC   24 2135
4  Giannis Antetokounmpo  PF  MIL   28 1959
```

## Selecting dataframe columns

```
# Selecting a single column (Player Name)
print("===Single Column Selection: Player Name===")
print(df['Player Name'].head(5))
print("\n")

# Selecting multiple columns (Player Name, Team, PTS)
print("===Multiple Column Selection: Player Name, Team, PTS===")
print(df[['Player Name', 'Team', 'PTS']].head(5))
print("\n")
```

```
===Single Column Selection: Player Name===
0      Jayson Tatum
1      Joel Embiid
2      Luka Doncic
3  Shai Gilgeous-Alexander
4  Giannis Antetokounmpo
Name: Player Name, dtype: object

===Multiple Column Selection: Player Name, Team, PTS===
   Player Name Team  PTS
0   Jayson Tatum  BOS 2225
1   Joel Embiid  PHI 2183
2     Luka Doncic  DAL 2138
3 Shai Gilgeous-Alexander  OKC 2135
4  Giannis Antetokounmpo  MIL 1959
```

## Selecting dataframe rows

✓  
0s



```
# Selecting a single row using Pandas indexing (first row)
print("====Single Row Selection (First Observation)====")
print(df.iloc[0])
print("\n")

# Selecting multiple rows using Python slicing (first 3 rows)
print("====Multiple Row Selection (First 3 Observations)====")
print(df.iloc[0:3])
print("\n")
```



```
====Single Row Selection (First Observation)====
```

|             |              |
|-------------|--------------|
| Player Name | Jayson Tatum |
| POS         | SF           |
| Team        | BOS          |
| Age         | 25           |
| GP          | 74           |
| W           | 52           |
| L           | 22           |
| Min         | 2732.2       |
| PTS         | 2225         |

Name: 0, dtype: object

```
====Multiple Row Selection (First 3 Observations)====
```

|   | Player Name  | POS | Team | Age | GP | W  | L  | Min    | PTS  |
|---|--------------|-----|------|-----|----|----|----|--------|------|
| 0 | Jayson Tatum | SF  | BOS  | 25  | 74 | 52 | 22 | 2732.2 | 2225 |
| 1 | Joel Embiid  | C   | PHI  | 29  | 66 | 43 | 23 | 2284.1 | 2183 |
| 2 | Luka Doncic  | PG  | DAL  | 24  | 66 | 33 | 33 | 2390.5 | 2138 |