Creating and Accessing Pandas DataFrames

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10 / 15 / 2024
Section: CPE21S4	**Date Submitted**: 10 / 15 / 2024
Name: ROALLOS, Jean Gabriel Vincent G.	**Instructor**: Prof. Maria Rizette 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.)**

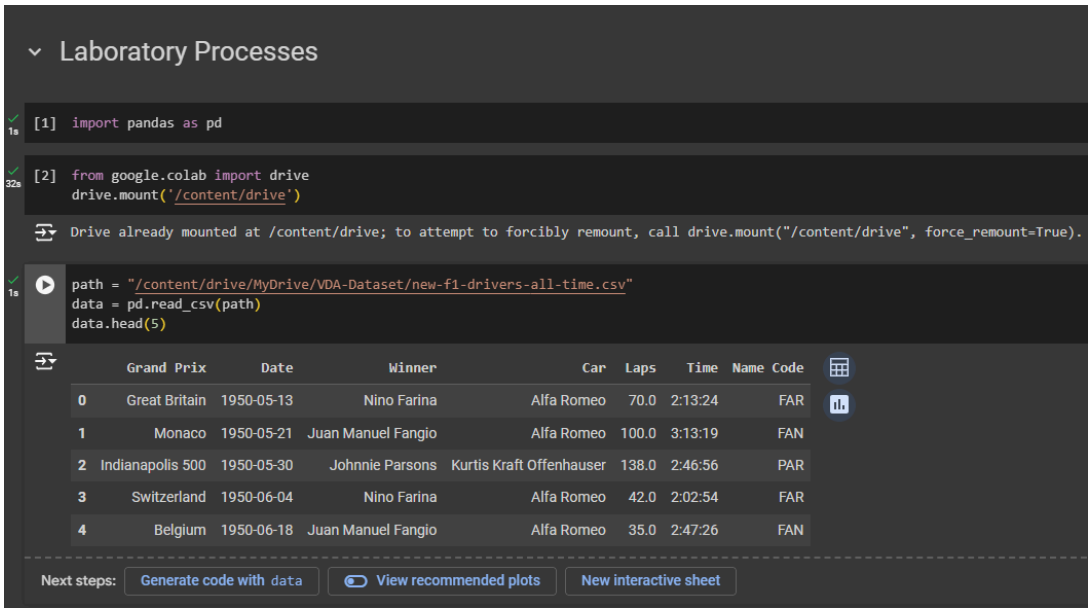
3. **Creating a dataframe from a dictionary of lists:** Manually create a dictionary where each value is composed of a list from your original dataset, then load it into a dataframe, before printing it. You are required to provide at least five (5) observations in your list. **(Take a screenshot to document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



Laboratory Processes

```
[1] import pandas as pd
```

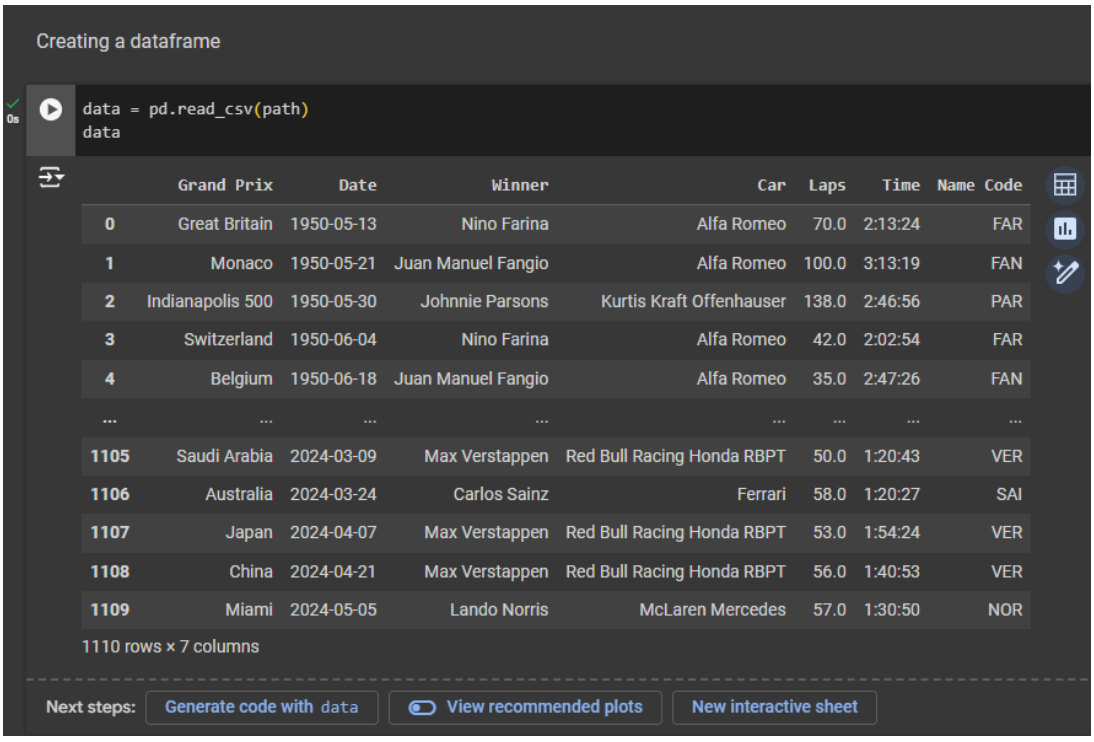
```
[2] from google.colab import drive
    drive.mount('/content/drive')
```

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

```
path = "/content/drive/MyDrive/VDA-Dataset/new-f1-drivers-all-time.csv"
data = pd.read_csv(path)
data.head(5)
```

	Grand Prix	Date	Winner	Car	Laps	Time	Name Code
0	Great Britain	1950-05-13	Nino Farina	Alfa Romeo	70.0	2:13:24	FAR
1	Monaco	1950-05-21	Juan Manuel Fangio	Alfa Romeo	100.0	3:13:19	FAN
2	Indianapolis 500	1950-05-30	Johnnie Parsons	Kurtis Kraft Offenhauser	138.0	2:46:56	PAR
3	Switzerland	1950-06-04	Nino Farina	Alfa Romeo	42.0	2:02:54	FAR
4	Belgium	1950-06-18	Juan Manuel Fangio	Alfa Romeo	35.0	2:47:26	FAN

Next steps: Generate code with data | View recommended plots | New interactive sheet



Creating a dataframe

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

	Grand Prix	Date	Winner	Car	Laps	Time	Name Code
0	Great Britain	1950-05-13	Nino Farina	Alfa Romeo	70.0	2:13:24	FAR
1	Monaco	1950-05-21	Juan Manuel Fangio	Alfa Romeo	100.0	3:13:19	FAN
2	Indianapolis 500	1950-05-30	Johnnie Parsons	Kurtis Kraft Offenhauser	138.0	2:46:56	PAR
3	Switzerland	1950-06-04	Nino Farina	Alfa Romeo	42.0	2:02:54	FAR
4	Belgium	1950-06-18	Juan Manuel Fangio	Alfa Romeo	35.0	2:47:26	FAN
...	...	...	...	...	...	...	...
1105	Saudi Arabia	2024-03-09	Max Verstappen	Red Bull Racing Honda RBPT	50.0	1:20:43	VER
1106	Australia	2024-03-24	Carlos Sainz	Ferrari	58.0	1:20:27	SAI
1107	Japan	2024-04-07	Max Verstappen	Red Bull Racing Honda RBPT	53.0	1:54:24	VER
1108	China	2024-04-21	Max Verstappen	Red Bull Racing Honda RBPT	56.0	1:40:53	VER
1109	Miami	2024-05-05	Lando Norris	McLaren Mercedes	57.0	1:30:50	NOR

1110 rows × 7 columns

Next steps: Generate code with data | View recommended plots | New interactive sheet

Creating a dataframe from a dictionary of lists

```
data_dictionary = {
    'Grand Prix': ['Bahrain', 'Saudi Arabia', 'Australia', 'Japan', 'China', 'Miami'],
    'Date': ['2024-03-02', '2024-03-09', '2024-03-24', '2024-04-07', '2024-04-21', '2024-05-05'],
    'Winner': ['Max Verstappen', 'Max Verstappen', 'Carlos Sainz', 'Max Verstappen', 'Max Verstappen', 'Lando Norris'],
    'Car': ['Red Bull Racing', 'Red Bull Racing', 'Ferrari', 'Red Bull Racing', 'Red Bull Racing', 'Mclaren Mercedes'],
    'Laps': [57, 50, 58, 53, 56, 57],
    'Time': ['1:31:45', '1:20:43', '1:20:27', '1:54:24', '1:40:53', '1:30:50']
}

# Grand Prix: Country of the F1 Championship it was held on
# Date: Date of the F1 Championship occured (all held in early 2024)
# Winner: The winner of the F1 Championship
# Car: Name of the F1 participant that won the F1 Championship
# Laps: Total number of laps done within the race
# Time: Lap time set by the winner

data_list = pd.DataFrame(data_dictionary)
data_list
```

	Grand Prix	Date	Winner	Car	Laps	Time
0	Bahrain	2024-03-02	Max Verstappen	Red Bull Racing	57	1:31:45
1	Saudi Arabia	2024-03-09	Max Verstappen	Red Bull Racing	50	1:20:43
2	Australia	2024-03-24	Carlos Sainz	Ferrari	58	1:20:27
3	Japan	2024-04-07	Max Verstappen	Red Bull Racing	53	1:54:24
4	China	2024-04-21	Max Verstappen	Red Bull Racing	56	1:40:53
5	Miami	2024-05-05	Lando Norris	Mclaren Mercedes	57	1:30:50

Next steps: Generate code with data_list | View recommended plots | New interactive sheet

Creating a dataframe from a list of dictionaries

```
data_list = [
    {'Grand Prix': 'Bahrain', 'Date': '2024-03-02', 'Winner': 'Max Verstappen', 'Car': 'Red Bull Racing', 'Laps': 57, 'Time': '1:31:45'},
    {'Grand Prix': 'Saudi Arabia', 'Date': '2024-03-09', 'Winner': 'Max Verstappen', 'Car': 'Red Bull Racing', 'Laps': 50, 'Time': '1:20:43'},
    {'Grand Prix': 'Australia', 'Date': '2024-03-24', 'Winner': 'Carlos Sainz', 'Car': 'Ferrari', 'Laps': 58, 'Time': '1:20:27'},
    {'Grand Prix': 'Japan', 'Date': '2024-04-07', 'Winner': 'Max Verstappen', 'Car': 'Red Bull Racing', 'Laps': 53, 'Time': '1:54:24'},
    {'Grand Prix': 'China', 'Date': '2024-04-21', 'Winner': 'Max Verstappen', 'Car': 'Red Bull Racing', 'Laps': 56, 'Time': '1:40:53'},
    {'Grand Prix': 'Miami', 'Date': '2024-05-05', 'Winner': 'Lando Norris', 'Car': 'Mclaren Mercedes', 'Laps': 57, 'Time': '1:30:50'}
]

data_dict = pd.DataFrame(data_list)
data_dict
```

	Grand Prix	Date	Winner	Car	Laps	Time
0	Bahrain	2024-03-02	Max Verstappen	Red Bull Racing	57	1:31:45
1	Saudi Arabia	2024-03-09	Max Verstappen	Red Bull Racing	50	1:20:43
2	Australia	2024-03-24	Carlos Sainz	Ferrari	58	1:20:27
3	Japan	2024-04-07	Max Verstappen	Red Bull Racing	53	1:54:24
4	China	2024-04-21	Max Verstappen	Red Bull Racing	56	1:40:53
5	Miami	2024-05-05	Lando Norris	Mclaren Mercedes	57	1:30:50

Next steps: Generate code with data_dict | View recommended plots | New interactive sheet

Selecting dataframe columns

```
print("Column Selection: Time")
data_dict['Time']
```

```
Column Selection: Time
```

	Time
0	1:31:45
1	1:20:43
2	1:20:27
3	1:54:24
4	1:40:53
5	1:30:50

dtype: object

Selecting dataframe rows

```
[17] print("Single Row Selection:")
     data_dict.iloc[0]
```

```
Single Row Selection:
```

	0
Grand Prix	Bahrain
Date	2024-03-02
Winner	Max Verstappen
Car	Red Bull Racing
Laps	57
Time	1:31:45

dtype: object

```
print("Single Row Selection:")
data_dict.iloc[:4]
```

```
Single Row Selection:
```

	Grand Prix	Date	Winner	Car	Laps	Time
0	Bahrain	2024-03-02	Max Verstappen	Red Bull Racing	57	1:31:45
1	Saudi Arabia	2024-03-09	Max Verstappen	Red Bull Racing	50	1:20:43
2	Australia	2024-03-24	Carlos Sainz	Ferrari	58	1:20:27
3	Japan	2024-04-07	Max Verstappen	Red Bull Racing	53	1:54:24