Creating and Accessing Pandas DataFrames

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: Oct 15, 2024
Section: CPE21S4	**Date Submitted**:
Name: Danica T. Guariño	**Instructor**: 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.)**

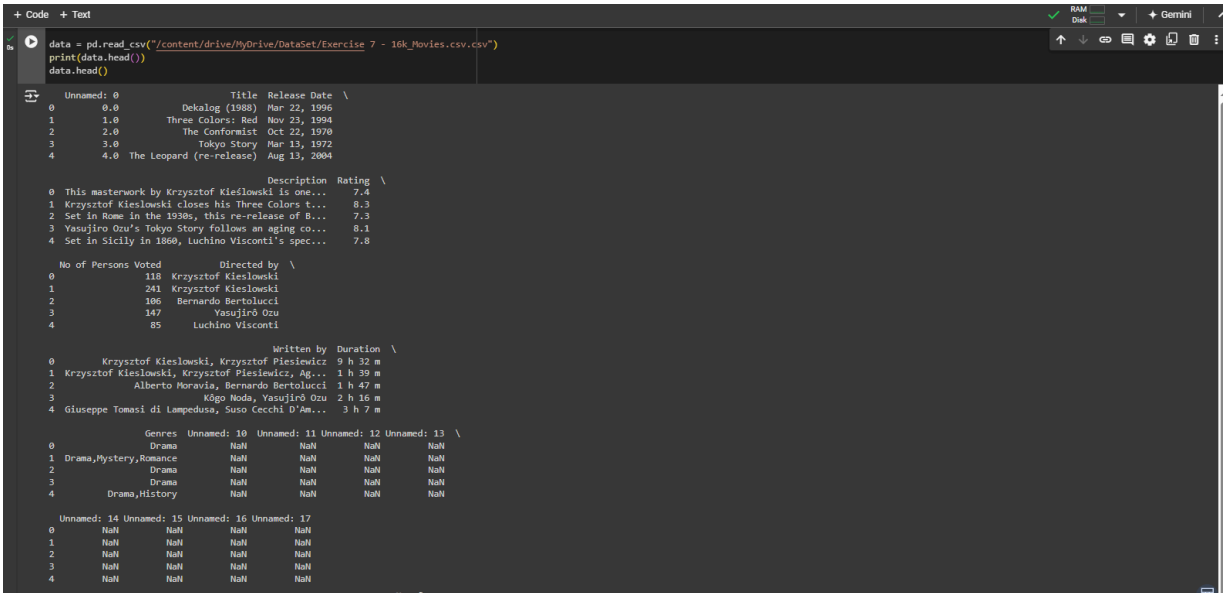
4. **Creating a dataframe from a list of dictionaries:** Manually create a list of dictionaries from your original dataset, then pass it into a dataframe, before printing it. You are required to provide at least five (5) observations in your list. **(Take a screenshot to document successful execution.)**

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.

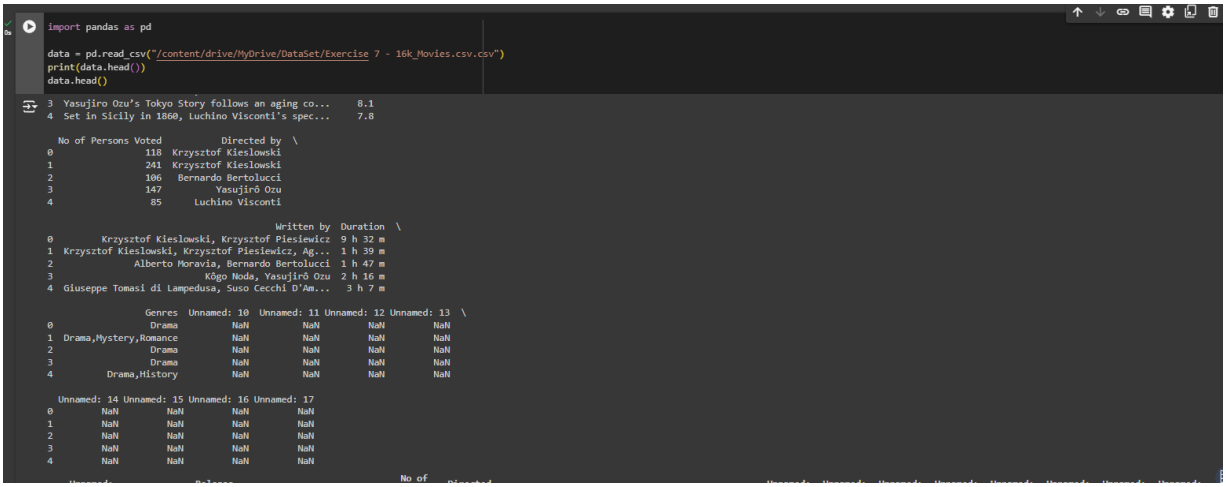
Output:

1. Loading your dataset:



```
+ Code + Text                                                                    RAM  Disk   + Gemini

data = pd.read_csv("/content/drive/MyDrive/DataSet/Exercise 7 - 16k_Movies.csv.csv")
print(data.head())
data.head()

   Unnamed: 0                   Title  Release Date  \
0         0.0          Dekalog (1988)  Mar 22, 1996
1         1.0  Three Colors: Red (1994)  Nov 23, 1994
2         2.0     The Conformist (1970)  Oct 22, 1970
3         3.0             Tokyo Story (1953)  Mar 13, 1972
4         4.0  The Leopard (re-release)  Aug 13, 2004

                                         Description  Rating  \
0  This masterwork by Krzysztof Kieślowski is one...     7.4
1  Krzysztof Kieslowski closes his Three Colors t...     8.3
2  Set in Rome in the 1930s, this re-release of B...     7.3
3  Yasujiro Ozu's Tokyo Story follows an aging co...     8.1
4  Set in Sicily in 1860, Luchino Visconti's spec...     7.8

  No of Persons Voted          Directed by  \
0                 118  Krzysztof Kieslowski
1                 241  Krzysztof Kieslowski
2                 106   Bernardo Bertolucci
3                 147         Yasujirô Ozu
4                  85      Luchino Visconti

                                      Written by  Duration  \
0        Krzysztof Kieslowski, Krzysztof Piesiewicz  9 h 32 m
1  Krzysztof Kieslowski, Krzysztof Piesiewicz, Ag...  1 h 39 m
2             Alberto Moravia, Bernardo Bertolucci  1 h 47 m
3                          Kôgo Noda, Yasujirô Ozu  2 h 16 m
4  Giuseppe Tomasi di Lampedusa, Suso Cecchi D'Am...   3 h 7 m

                Genres  Unnamed: 10  Unnamed: 11 Unnamed: 12 Unnamed: 13  \
0                Drama          NaN          NaN         NaN         NaN
1  Drama,Mystery,Romance          NaN          NaN         NaN         NaN
2                Drama          NaN          NaN         NaN         NaN
3                Drama          NaN          NaN         NaN         NaN
4         Drama,History          NaN          NaN         NaN         NaN

  Unnamed: 14 Unnamed: 15 Unnamed: 16 Unnamed: 17
0         NaN         NaN         NaN         NaN
1         NaN         NaN         NaN         NaN
2         NaN         NaN         NaN         NaN
3         NaN         NaN         NaN         NaN
4         NaN         NaN         NaN         NaN
```

2. Creating a dataframe from your CSV file:



```
import pandas as pd

data = pd.read_csv("/content/drive/MyDrive/DataSet/Exercise 7 - 16k_Movies.csv.csv")
print(data.head())
data.head()

3  Yasujiro Ozu's Tokyo Story follows an aging co...     8.1
4  Set in Sicily in 1860, Luchino Visconti's spec...     7.8

  No of Persons Voted          Directed by  \
0                 118  Krzysztof Kieslowski
1                 241  Krzysztof Kieslowski
2                 106   Bernardo Bertolucci
3                 147         Yasujirô Ozu
4                  85      Luchino Visconti

                                      Written by  Duration  \
0        Krzysztof Kieslowski, Krzysztof Piesiewicz  9 h 32 m
1  Krzysztof Kieslowski, Krzysztof Piesiewicz, Ag...  1 h 39 m
2             Alberto Moravia, Bernardo Bertolucci  1 h 47 m
3                          Kôgo Noda, Yasujirô Ozu  2 h 16 m
4  Giuseppe Tomasi di Lampedusa, Suso Cecchi D'Am...   3 h 7 m

                Genres  Unnamed: 10  Unnamed: 11 Unnamed: 12 Unnamed: 13  \
0                Drama          NaN          NaN         NaN         NaN
1  Drama,Mystery,Romance          NaN          NaN         NaN         NaN
2                Drama          NaN          NaN         NaN         NaN
3                Drama          NaN          NaN         NaN         NaN
4         Drama,History          NaN          NaN         NaN         NaN

  Unnamed: 14 Unnamed: 15 Unnamed: 16 Unnamed: 17
0         NaN         NaN         NaN         NaN
1         NaN         NaN         NaN         NaN
2         NaN         NaN         NaN         NaN
3         NaN         NaN         NaN         NaN
4         NaN         NaN         NaN         NaN
```

	Unnamed: 0	Title	Release Date	Description	Rating	No of Persons Voted	Directed by	Written by	Duration	Genres	Unnamed: 10	Unnamed: 11	Unnamed: 12	Unnamed: 13	Unnamed: 14	Unnamed: 15	Unnamed: 16	Unnamed: 17
0	0.0	Dekalog (1988)	Mar 22, 1996	This masterwork by Krzysztof Kieślowski is one...	7.4	118	Krzysztof Kieslowski	Krzysztof Kieslowski, Krzysztof Piesiewicz	9 h 32 m	Drama	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1.0	Three Colors: Red	Nov 23, 1994	Krzysztof Kieslowski closes his Three Colors t...	8.3	241	Krzysztof Kieslowski	Krzysztof Kieslowski, Krzysztof Piesiewicz, Ag...	1 h 39 m	Drama,Mystery,Romance	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2.0	The Conformist	Oct 22, 1970	Set in Rome in the 1930s, this re-release of B...	7.3	106	Bernardo Bertolucci	Alberto Moravia, Bernardo Bertolucci	1 h 47 m	Drama	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	3.0	Tokyo Story	Mar 13, 1972	Yasujiro Ozu's Tokyo Story follows an aging co...	8.1	147	Yasujirô Ozu	Kôgo Noda, Yasujirô Ozu	2 h 16 m	Drama	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	4.0	The Leopard (re-release)	Aug 13, 2004	Set in Sicily in 1860, Luchino Visconti's spec...	7.8	85	Luchino Visconti	Giuseppe Tomasi di Lampedusa, Suso Cecchi D'Am...	3 h 7 m	Drama,History	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

3. Creating a dataframe from a dictionary of lists:

```
import pandas as pd

path ="/content/drive/MyDrive/DataSet/Exercise 7 - 16k_Movies.csv.csv"
data = {
    'Title': ['Movie1', 'Movie2', 'Movie3', 'Movie4', 'Movie5'],
    'Release Date': [2020, 2021, 2022, 2023, 2024],
    'Description': ['Description of Movie1', 'Description of Movie2', 'Description of Movie3', 'Description of Movie4', 'Description of Movie5'],
    'Rating': [8.5, 9.0, 8.0, 7.5, 9.2],
    'No. of Persons Voted': [15000, 12000, 13000, 14000, 11000],
    'Directed by': ['Director1', 'Director2', 'Director3', 'Director4', 'Director5'],
    'Written by': ['Writer1', 'Writer2', 'Writer3', 'Writer4', 'Writer5'],
    'Duration': [120, 135, 150, 160, 145],
    'Genres': ['Action', 'Comedy', 'Drama', 'Horror', 'Thriller']
}

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

```
    Title  Release Date           Description  Rating  No. of Persons Voted  \
0  Movie1          2020  Description of Movie1     8.5                 15000
1  Movie2          2021  Description of Movie2     9.0                 12000
2  Movie3          2022  Description of Movie3     8.0                 13000
3  Movie4          2023  Description of Movie4     7.5                 14000
4  Movie5          2024  Description of Movie5     9.2                 11000

  Directed by Written by  Duration    Genres
0   Director1    Writer1       120    Action
1   Director2    Writer2       135    Comedy
2   Director3    Writer3       150     Drama
3   Director4    Writer4       160    Horror
4   Director5    Writer5       145  Thriller
```

4. Creating a dataframe from a list of dictionaries:

```
import pandas as pd

path ="/content/drive/MyDrive/DataSet/Exercise 7 - 16k_Movies.csv.csv"
data = [
    {'Title': 'Inception', 'Release Date': 2010, 'Description': 'A thief with the ability to enter dreams and steal secrets.', 'Rating': 8.8, 'No. of Persons Voted': 2000000, 'Directed by': 'Christopher Nolan', 'Written
    {'Title': 'Interstellar', 'Release Date': 2014, 'Description': 'A team of explorers travel through a wormhole in space.', 'Rating': 8.6, 'No. of Persons Voted': 1500000, 'Directed by': 'Christopher Nolan', 'Written b
    {'Title': 'The Dark Knight', 'Release Date': 2008, 'Description': 'Batman sets out to dismantle the remaining criminal organizations.', 'Rating': 9.0, 'No. of Persons Voted': 2300000, 'Directed by': 'Christopher Nola
    {'Title': 'Dunkirk', 'Release Date': 2017, 'Description': 'Allied soldiers are surrounded by the German Army, and evacuated.', 'Rating': 7.9, 'No. of Persons Voted': 600000, 'Directed by': 'Christopher Nolan', 'Writt
    {'Title': 'Tenet', 'Release Date': 2020, 'Description': 'A protagonist is armed with only one word, Tenet, to save the world.', 'Rating': 7.4, 'No. of Persons Voted': 500000, 'Directed by': 'Christopher Nolan', 'Writ
]

# Load the list into a DataFrame
df = pd.DataFrame(data)

# Print the DataFrame
print(df)
```

```
           Title  Release Date  \
0      Inception          2010
1   Interstellar          2014
2  The Dark Knight          2008
3        Dunkirk          2017
4          Tenet          2020

                                         Description  Rating  \
0  A thief with the ability to enter dreams and s...     8.8
1  A team of explorers travel through a wormhole ...     8.6
2  Batman sets out to dismantle the remaining cri...     9.0
3  Allied soldiers are surrounded by the German A...     7.9
4  A protagonist is armed with only one word, Ten...     7.4

   No. of Persons Voted        Directed by                    Written by  \
0               2000000  Christopher Nolan             Christopher Nolan
1               1500000  Christopher Nolan  Jonathan Nolan, Christopher Nolan
2               2300000  Christopher Nolan  Jonathan Nolan, Christopher Nolan
3                600000  Christopher Nolan             Christopher Nolan
4                500000  Christopher Nolan             Christopher Nolan

   Duration                  Genres
0       148  Action, Sci-Fi, Thriller
1       169  Adventure, Drama, Sci-Fi
2       152      Action, Crime, Drama
3       106     Action, Drama, History
4       150  Action, Sci-Fi, Thriller
```

5. Selecting dataframe columns:

```
# Selecting a single column
single_column = df['Title']
print("Single Column (Title):")
print(single_column)

# Selecting multiple columns
multiple_columns = df[['Title', 'Rating', 'Genres']]
print("\nMultiple Columns (Title, Rating, Genres):")
print(multiple_columns)
```

```
Single Column (Title):
0          Inception
1       Interstellar
2    The Dark Knight
3            Dunkirk
4              Tenet
Name: Title, dtype: object

Multiple Columns (Title, Rating, Genres):
              Title  Rating                    Genres
0         Inception     8.8  Action, Sci-Fi, Thriller
1     Interstellar      8.6  Adventure, Drama, Sci-Fi
2  The Dark Knight      9.0       Action, Crime, Drama
3           Dunkirk     7.9     Action, Drama, History
4             Tenet     7.4  Action, Sci-Fi, Thriller
```

6. Selecting dataframe rows:

Single Row (.loc and .iloc)

```
# Select a single row using .loc
single_row_loc = df.loc[0]
print("\nSingle row using .loc:")
print(single_row_loc)

# Select a single row using .iloc
single_row_iloc = df.iloc[0]
print("\nSingle row using .iloc:")
print(single_row_iloc)
```

```
Single row using .loc:
Title                                                 Inception
Release Date                                               2010
Description           A thief with the ability to enter dreams and s...
Rating                                                      8.8
No. of Persons Voted                                    2000000
Directed by                                      Christopher Nolan
Written by                                       Christopher Nolan
Duration                                                    148
Genres                                  Action, Sci-Fi, Thriller
Name: 0, dtype: object

Single row using .iloc:
Title                                                 Inception
Release Date                                               2010
Description           A thief with the ability to enter dreams and s...
Rating                                                      8.8
No. of Persons Voted                                    2000000
Directed by                                      Christopher Nolan
Written by                                       Christopher Nolan
Duration                                                    148
Genres                                  Action, Sci-Fi, Thriller
Name: 0, dtype: object
```

Multiple Row (.loc and .iloc)

```python
# Select multiple rows using .loc
multiple_rows_loc = df.loc[0:2]
print("\nMultiple rows using .loc:")
print(multiple_rows_loc)

# Select multiple rows using .iloc
multiple_rows_iloc = df.iloc[0:3]
print("\nMultiple rows using .iloc:")
print(multiple_rows_iloc)
```

```
Multiple rows using .loc:
             Title  Release Date  \
0        Inception          2010
1     Interstellar          2014
2  The Dark Knight          2008

                                      Description  Rating  \
0  A thief with the ability to enter dreams and s...     8.8
1  A team of explorers travel through a wormhole ...     8.6
2  Batman sets out to dismantle the remaining cri...     9.0

   No. of Persons Voted        Directed by                       Written by  \
0               2000000  Christopher Nolan                Christopher Nolan
1               1500000  Christopher Nolan  Jonathan Nolan, Christopher Nolan
2               2300000  Christopher Nolan  Jonathan Nolan, Christopher Nolan

   Duration                   Genres
0       148  Action, Sci-Fi, Thriller
1       169  Adventure, Drama, Sci-Fi
2       152      Action, Crime, Drama

Multiple rows using .iloc:
             Title  Release Date  \
0        Inception          2010
1     Interstellar          2014
2  The Dark Knight          2008

                                      Description  Rating  \
0  A thief with the ability to enter dreams and s...     8.8
1  A team of explorers travel through a wormhole ...     8.6
2  Batman sets out to dismantle the remaining cri...     9.0

   No. of Persons Voted        Directed by                       Written by  \
0               2000000  Christopher Nolan                Christopher Nolan
1               1500000  Christopher Nolan  Jonathan Nolan, Christopher Nolan
2               2300000  Christopher Nolan  Jonathan Nolan, Christopher Nolan

   Duration                   Genres
0       148  Action, Sci-Fi, Thriller
1       169  Adventure, Drama, Sci-Fi
2       152      Action, Crime, Drama
```