Creating and Accessing Pandas DataFrames

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: October 15, 2024
Section: CPE21S4	**Date Submitted**: October 15, 2024
Name: Don Eleazar T. Fernandez	**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.
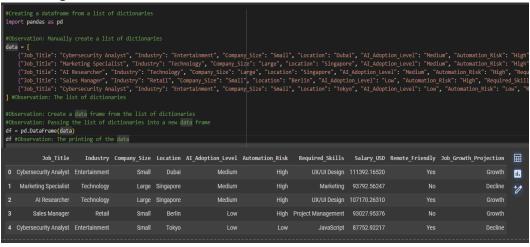
Output:

1. **Loading your dataset**

```
[37] #Loading your dataset
     from google.colab import files
     uploaded = files.upload()
```

```
Choose Files  ai_job_mar... Insights.csv
• ai_job_market_insights - (Tab 2) AI Job Market Insights.csv(text/csv) - 44723 bytes, last modified: 10/15/2024 - 100% done
Saving ai_job_market_insights - (Tab 2) AI Job Market Insights.csv to ai_job_market_insights - (Tab 2) AI Job Market Insights.csv
```

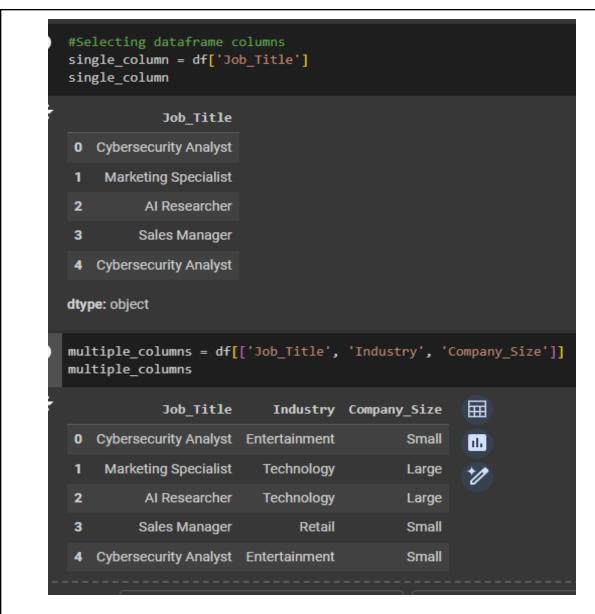2. **Creating a dataframe from your CSV file**

```python
#Creating a dataframe from your CSV file
data = pd.read_csv('ai_job_market_insights - (Tab 2) AI Job Market Insights.csv')
print(data)
```

```
                   Job_Title          Industry Company_Size   Location  \
0      Cybersecurity Analyst     Entertainment        Small      Dubai
1       Marketing Specialist        Technology        Large  Singapore
2              AI Researcher        Technology        Large  Singapore
3              Sales Manager            Retail        Small     Berlin
4      Cybersecurity Analyst     Entertainment        Small      Tokyo
..                       ...               ...          ...        ...
495            Data Scientist  Telecommunications     Medium     Berlin
496     Cybersecurity Analyst  Telecommunications      Small     London
497     Cybersecurity Analyst            Energy        Large      Dubai
498         Operations Manager        Healthcare        Large      Paris
499                 HR Manager     Entertainment       Medium     Berlin

    AI_Adoption_Level Automation_Risk       Required_Skills    Salary_USD  \
0              Medium            High         UX/UI Design   111392.16520
1              Medium            High            Marketing    93792.56247
2              Medium            High         UX/UI Design   107170.26310
3                 Low            High    Project Management    93027.95376
4                 Low             Low            JavaScript    87752.92217
..                ...             ...                   ...           ...
495               Low          Medium      Machine Learning  105821.39400
496               Low            High         UX/UI Design   119794.99210
497              High             Low         UX/UI Design    79644.93310
498              High             Low                Python    77642.15063
499            Medium            High    Project Management    68764.37892

    Remote_Friendly Job_Growth_Projection
0               Yes                Growth
1                No               Decline
2               Yes                Growth
3                No                Growth
4               Yes               Decline
..              ...                   ...
495             Yes                Stable
496              No               Decline
497             Yes                Stable
498             Yes                Stable
499             Yes               Decline

[500 rows x 10 columns]
```

3. **Creating a dataframe from a dictionary of lists**

```
#Creating a dataframe from a dictionary of lists
data_dict = { #Observation: With the tolist() function, the data on each variable has been listed according to their variable
    'Job_Title': df['Job_Title'].tolist(),
             #Observation: The Job Title data is listed
    'Industry': df['Industry'].tolist(),
             #Observation: The Industry data is listed
    'Company_Size': df['Company_Size'].tolist(),
    'Location': df['Location'].tolist(),
    'AI_Adoption_Level': df['AI_Adoption_Level'].tolist(),
    'Automation_Risk': df['Automation_Risk'].tolist(),
    'Required_Skills': df['Required_Skills'].tolist(),
    'Salary_USD': df['Salary_USD'].tolist(),
    'Remote_Friendly': df['Remote_Friendly'].tolist(),
    'Job_Growth_Projection': df['Job_Growth_Projection'].tolist(),
} #Observation: The data of each variable has been listed to the dictionary

data_2 = pd.DataFrame(data_dict) #Observation: The data on the dictionary has been placed to a data frame
```

4. Creating a dataframe from a list of dictionaries

```
#Creating a dataframe from a list of dictionaries
import pandas as pd

#Observation: Manually create a list of dictionaries
data = [
    {"Job_Title": "Cybersecurity Analyst", "Industry": "Entertainment", "Company_Size": "Small", "Location": "Dubai", "AI_Adoption_Level": "Medium", "Automation_Risk": "High"
    {"Job_Title": "Marketing Specialist", "Industry": "Technology", "Company_Size": "Large", "Location": "Singapore", "AI_Adoption_Level": "Medium", "Automation_Risk": "High"
    {"Job_Title": "AI Researcher", "Industry": "Technology", "Company_Size": "Large", "Location": "Singapore", "AI_Adoption_Level": "Medium", "Automation_Risk": "High", "Requ
    {"Job_Title": "Sales Manager", "Industry": "Retail", "Company_Size": "Small", "Location": "Berlin", "AI_Adoption_Level": "Low", "Automation_Risk": "High", "Required_Skill
    {"Job_Title": "Cybersecurity Analyst", "Industry": "Entertainment", "Company_Size": "Small", "Location": "Tokyo", "AI_Adoption_Level": "Low", "Automation_Risk": "Low", "R
] #Observation: The list of dictionaries

#Observation: Create a data frame from the list of dictionaries
#Observation: Passing the list of dictionaries into a new data frame
df = pd.DataFrame(data)
df #Observation: The printing of the data
```

	Job_Title	Industry	Company_Size	Location	AI_Adoption_Level	Automation_Risk	Required_Skills	Salary_USD	Remote_Friendly	Job_Growth_Projection
0	Cybersecurity Analyst	Entertainment	Small	Dubai	Medium	High	UX/UI Design	111392.16520	Yes	Growth
1	Marketing Specialist	Technology	Large	Singapore	Medium	High	Marketing	93792.56247	No	Decline
2	AI Researcher	Technology	Large	Singapore	Medium	High	UX/UI Design	107170.26310	Yes	Growth
3	Sales Manager	Retail	Small	Berlin	Low	High	Project Management	93027.95376	No	Growth
4	Cybersecurity Analyst	Entertainment	Small	Tokyo	Low	Low	JavaScript	87752.92217	Yes	Decline

5. Selecting dataframe columns

```
#Selecting dataframe columns
single_column = df['Job_Title']
single_column
```

	Job_Title
0	Cybersecurity Analyst
1	Marketing Specialist
2	AI Researcher
3	Sales Manager
4	Cybersecurity Analyst

dtype: object

```
multiple_columns = df[['Job_Title', 'Industry', 'Company_Size']]
multiple_columns
```

	Job_Title	Industry	Company_Size
0	Cybersecurity Analyst	Entertainment	Small
1	Marketing Specialist	Technology	Large
2	AI Researcher	Technology	Large
3	Sales Manager	Retail	Small
4	Cybersecurity Analyst	Entertainment	Small

6. **Selecting dataframe rows**

```python
#Selecting dataframe rows (Using Pandas Indexing)
single_row = df.loc[0]
single_row
```

	0
Job_Title	Cybersecurity Analyst
Industry	Entertainment
Company_Size	Small
Location	Dubai
AI_Adoption_Level	Medium
Automation_Risk	High
Required_Skills	UX/UI Design
Salary_USD	111392.1652
Remote_Friendly	Yes
Job_Growth_Projection	Growth

dtype: object

```python
# (Using Pandas Indexing)
multiple_row = df.loc[0:1]
multiple_row
```

	Job_Title	Industry	Company_Size	Location	AI_Adoption_Level	Automation_Risk	Required_Skills	Salary_USD	Remote_Friendly	Job_Growth_Projection
0	Cybersecurity Analyst	Entertainment	Small	Dubai	Medium	High	UX/UI Design	111392.16520	Yes	Growth
1	Marketing Specialist	Technology	Large	Singapore	Medium	High	Marketing	93792.56247	No	Decline

```python
# (Using Python Indexing)
single_row = df[0:1]
single_row
```

	Job_Title	Industry	Company_Size	Location	AI_Adoption_Level	Automation_Risk	Required_Skills	Salary_USD	Remote_Friendly	Job_Growth_Projection
0	Cybersecurity Analyst	Entertainment	Small	Dubai	Medium	High	UX/UI Design	111392.1652	Yes	Growth

```python
# (Using Python Indexing)
multiple_row = df[0:2]
multiple_row
```

	Job_Title	Industry	Company_Size	Location	AI_Adoption_Level	Automation_Risk	Required_Skills	Salary_USD	Remote_Friendly	Job_Growth_Projection
0	Cybersecurity Analyst	Entertainment	Small	Dubai	Medium	High	UX/UI Design	111392.16520	Yes	Growth
1	Marketing Specialist	Technology	Large	Singapore	Medium	High	Marketing	93792.56247	No	Decline