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: Edson Ray E. San Juan	**Instructor**: Prof. Maria Rizette Sayo

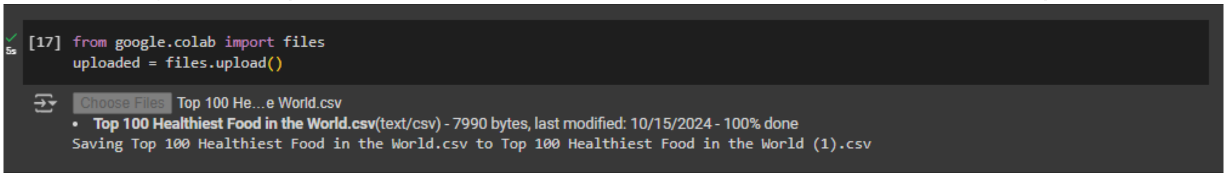
Intended Learning Outcomes (ILO):

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:

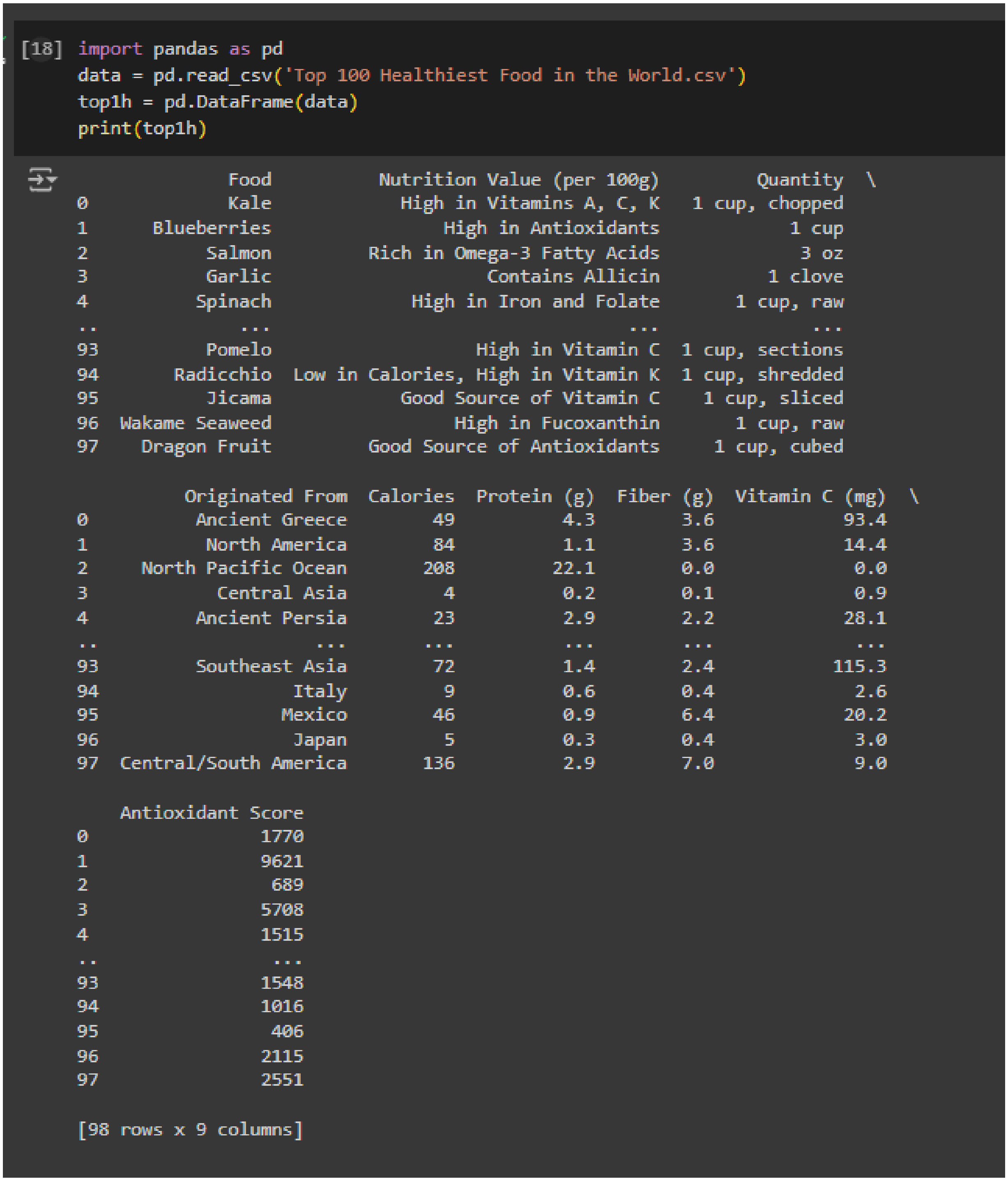
1. **Loading your dataset**: Refer back to your chosen dataset from the PRELIM period. Whether you downloaded it or stored it in your Google Drive, you are required to load it into the Google Colab. Watch this video to learn more about how to read CSV files in Google Colab.**(Take a screenshot to document successful execution.)**



```
[17] from google.colab import files
     uploaded = files.upload()
```

```
Top 100 He...e World.csv
• Top 100 Healthiest Food in the World.csv(text/csv) - 7990 bytes, last modified: 10/15/2024 - 100% done
Saving Top 100 Healthiest Food in the World.csv to Top 100 Healthiest Food in the World (1).csv
```

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.)**



```
[18] import pandas as pd
     data = pd.read_csv('Top 100 Healthiest Food in the World.csv')
     top1h = pd.DataFrame(data)
     print(top1h)
```

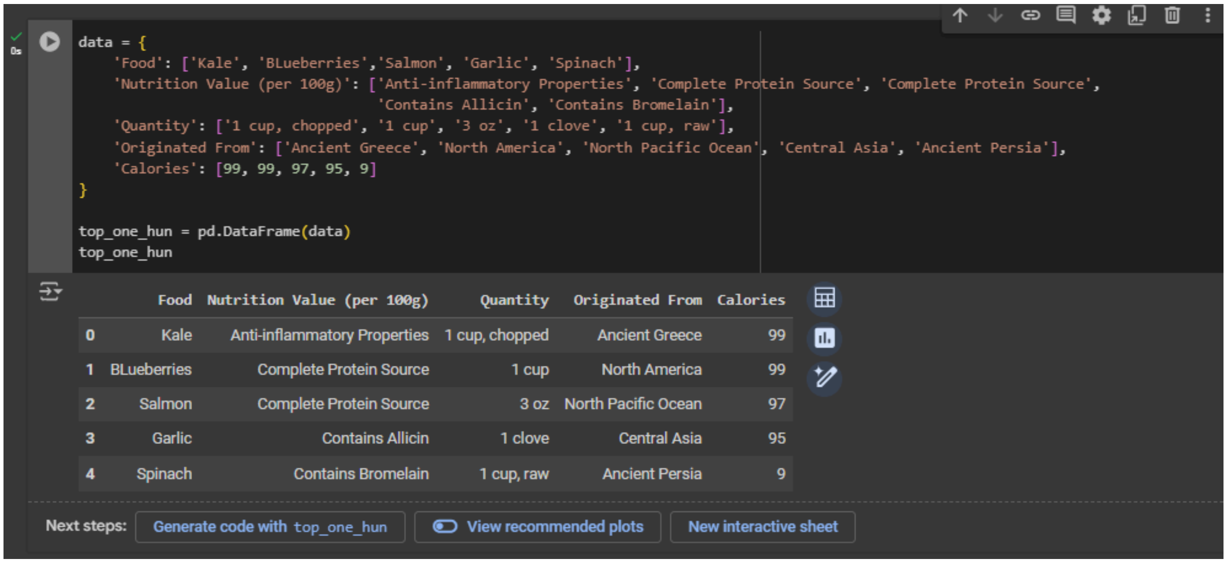
```
             Food            Nutrition Value (per 100g)        Quantity  \
0            Kale               High in Vitamins A, C, K   1 cup, chopped
1     Blueberries                  High in Antioxidants            1 cup
2          Salmon           Rich in Omega-3 Fatty Acids             3 oz
3          Garlic                       Contains Allicin          1 clove
4         Spinach                 High in Iron and Folate       1 cup, raw
..            ...                                    ...              ...
93         Pomelo                     High in Vitamin C   1 cup, sections
94      Radicchio   Low in Calories, High in Vitamin K   1 cup, shredded
95         Jicama                Good Source of Vitamin C    1 cup, sliced
96  Wakame Seaweed                  High in Fucoxanthin       1 cup, raw
97    Dragon Fruit            Good Source of Antioxidants     1 cup, cubed

         Originated From  Calories  Protein (g)  Fiber (g)  Vitamin C (mg)  \
0          Ancient Greece        49          4.3        3.6            93.4
1           North America        84          1.1        3.6            14.4
2     North Pacific Ocean       208         22.1        0.0             0.0
3            Central Asia         4          0.2        0.1             0.9
4          Ancient Persia        23          2.9        2.2            28.1
..                   ...       ...          ...        ...             ...
93         Southeast Asia        72          1.4        2.4           115.3
94                  Italy         9          0.6        0.4             2.6
95                 Mexico        46          0.9        6.4            20.2
96                  Japan         5          0.3        0.4             3.0
97  Central/South America       136          2.9        7.0             9.0

    Antioxidant Score
0                1770
1                9621
2                 689
3                5708
4                1515
..                ...
93               1548
94               1016
95                406
96               2115
97               2551

[98 rows x 9 columns]
```

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.)**

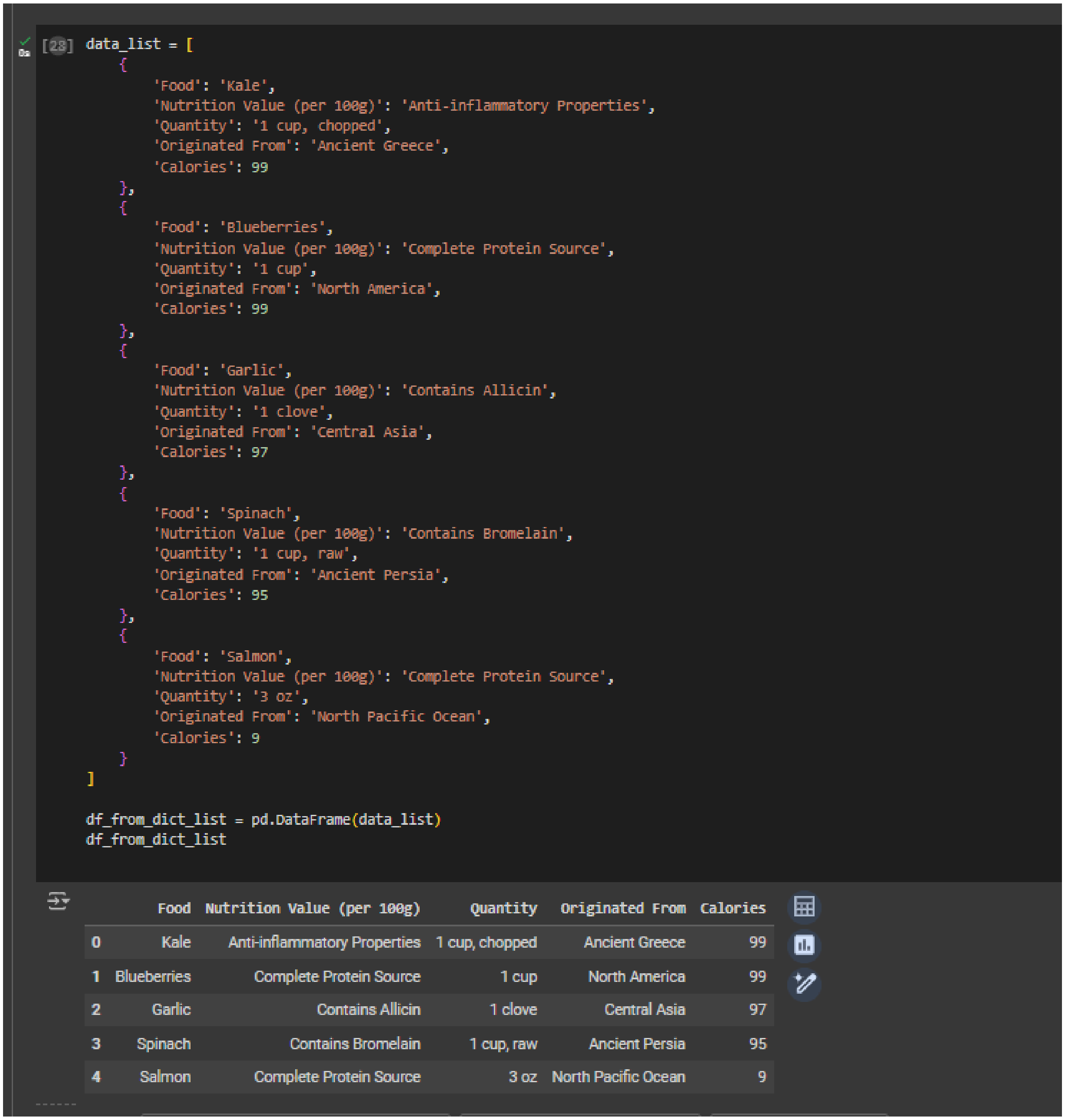


```python
data = {
    'Food': ['Kale', 'BLueberries','Salmon', 'Garlic', 'Spinach'],
    'Nutrition Value (per 100g)': ['Anti-inflammatory Properties', 'Complete Protein Source', 'Complete Protein Source',
                                   'Contains Allicin', 'Contains Bromelain'],
    'Quantity': ['1 cup, chopped', '1 cup', '3 oz', '1 clove', '1 cup, raw'],
    'Originated From': ['Ancient Greece', 'North America', 'North Pacific Ocean', 'Central Asia', 'Ancient Persia'],
    'Calories': [99, 99, 97, 95, 9]
}

top_one_hun = pd.DataFrame(data)
top_one_hun
```

```
          Food  Nutrition Value (per 100g)        Quantity      Originated From  Calories
0         Kale  Anti-inflammatory Properties  1 cup, chopped       Ancient Greece        99
1  BLueberries       Complete Protein Source           1 cup        North America        99
2       Salmon       Complete Protein Source            3 oz  North Pacific Ocean        97
3       Garlic              Contains Allicin         1 clove         Central Asia        95
4      Spinach            Contains Bromelain      1 cup, raw       Ancient Persia         9
```

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.)**

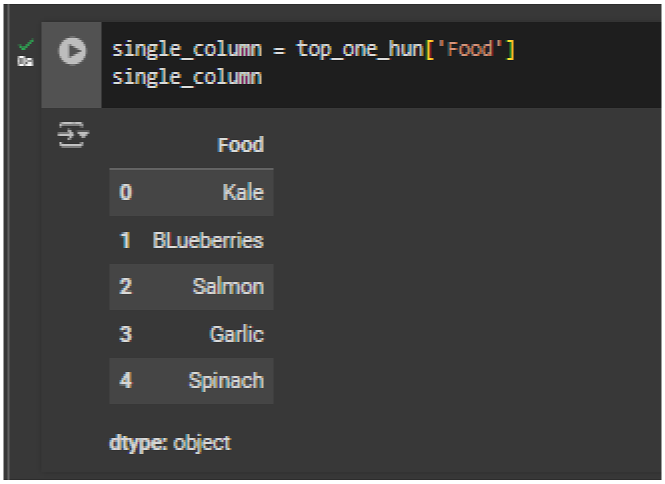


```
data_list = [
    {
        'Food': 'Kale',
        'Nutrition Value (per 100g)': 'Anti-inflammatory Properties',
        'Quantity': '1 cup, chopped',
        'Originated From': 'Ancient Greece',
        'Calories': 99
    },
    {
        'Food': 'Blueberries',
        'Nutrition Value (per 100g)': 'Complete Protein Source',
        'Quantity': '1 cup',
        'Originated From': 'North America',
        'Calories': 99
    },
    {
        'Food': 'Garlic',
        'Nutrition Value (per 100g)': 'Contains Allicin',
        'Quantity': '1 clove',
        'Originated From': 'Central Asia',
        'Calories': 97
    },
    {
        'Food': 'Spinach',
        'Nutrition Value (per 100g)': 'Contains Bromelain',
        'Quantity': '1 cup, raw',
        'Originated From': 'Ancient Persia',
        'Calories': 95
    },
    {
        'Food': 'Salmon',
        'Nutrition Value (per 100g)': 'Complete Protein Source',
        'Quantity': '3 oz',
        'Originated From': 'North Pacific Ocean',
        'Calories': 9
    }
]

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

	Food	Nutrition Value (per 100g)	Quantity	Originated From	Calories
0	Kale	Anti-inflammatory Properties	1 cup, chopped	Ancient Greece	99
1	Blueberries	Complete Protein Source	1 cup	North America	99
2	Garlic	Contains Allicin	1 clove	Central Asia	97
3	Spinach	Contains Bromelain	1 cup, raw	Ancient Persia	95
4	Salmon	Complete Protein Source	3 oz	North Pacific Ocean	9

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.)**
    a. Single Dataframe Columns:

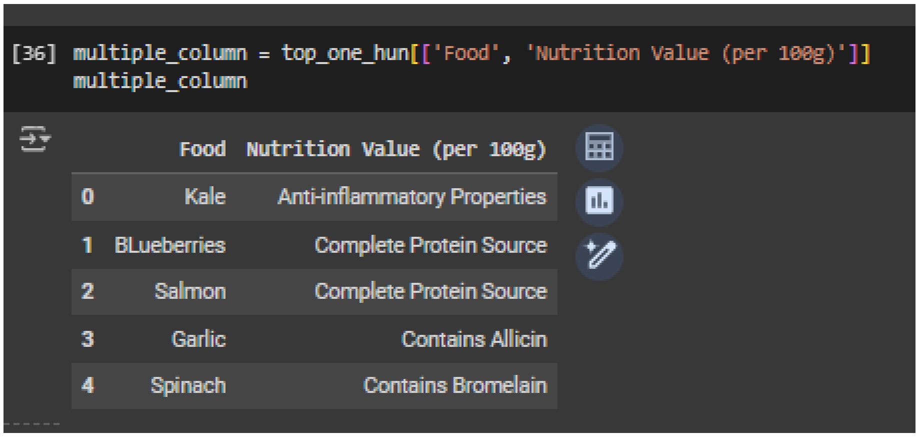


```
single_column = top_one_hun['Food']
single_column
```

```
          Food
0         Kale
1  BLueberries
2       Salmon
3       Garlic
4      Spinach

dtype: object
```

b. Multiple Dataframe Columns:

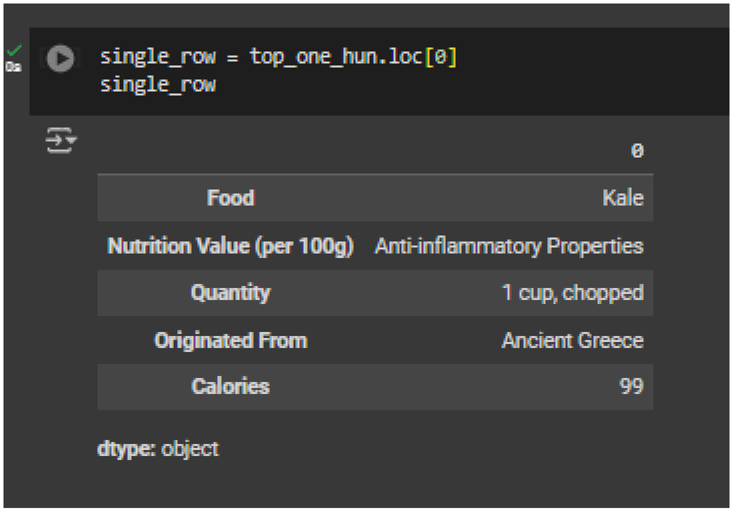


```
[36] multiple_column = top_one_hun[['Food', 'Nutrition Value (per 100g)']]
     multiple_column
```

```
          Food  Nutrition Value (per 100g)
0         Kale  Anti-inflammatory Properties
1  BLueberries      Complete Protein Source
2       Salmon      Complete Protein Source
3       Garlic             Contains Allicin
4      Spinach          Contains Bromelain
```

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.

   a. Single Dataframe Rows:

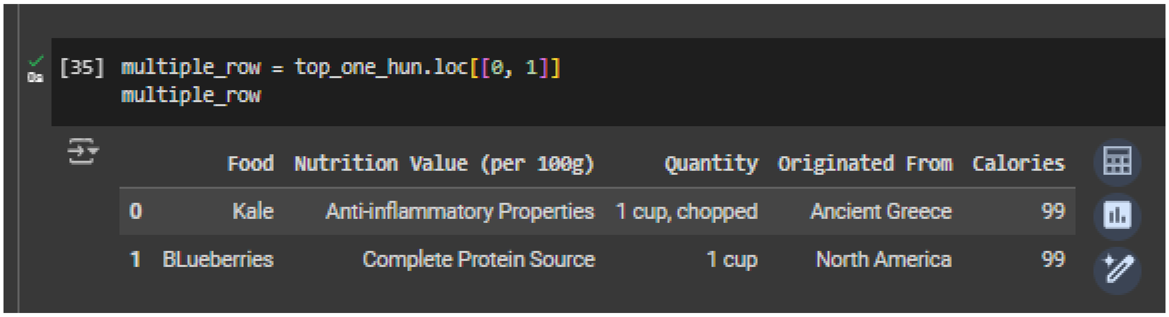


```
single_row = top_one_hun.loc[0]
single_row
```

```
                                                       0
                      Food                          Kale
Nutrition Value (per 100g)  Anti-inflammatory Properties
                  Quantity               1 cup, chopped
            Originated From               Ancient Greece
                  Calories                           99

dtype: object
```

b. Multiple Dataframe Rows:



```
multiple_row = top_one_hun.loc[[0, 1]]
multiple_row
```

	Food	Nutrition Value (per 100g)	Quantity	Originated From	Calories
0	Kale	Anti-inflammatory Properties	1 cup, chopped	Ancient Greece	99
1	BLueberries	Complete Protein Source	1 cup	North America	99

Output:

In this lab activity, we learned how to create a DataFrame in Python using two different methods: one from a dictionary of lists and another from a list of dictionaries. We also practiced selecting specific columns, whether just one or several, from a DataFrame. Additionally, we used the .loc method to select certain rows by their index numbers. Overall, this lab helped us build important skills for handling data using pandas, making it easier to organize, explore, and prepare data for analysis.