Python-Coding Challenge

Submitted By-

Aniroop Gupta
DE Batch1

1. Printing rows of the Data

- DataFrame rows are displayed using `.head()` or `.tail()` for quick inspection.
- Helps preview the structure and understand the data.

```
import pandas as pd

data = pd.read_csv("annual-enterprise-survey-2023-financial-year-provisional.csv")

print(data.head())
```

```
   Year Industry_aggregation_NZSIOC Industry_code_NZSIOC Industry_name_NZSIOC  \
0  2023                     Level 1                99999       All industries
1  2023                     Level 1                99999       All industries
2  2023                     Level 1                99999       All industries
3  2023                     Level 1                99999       All industries
4  2023                     Level 1                99999       All industries

              Units Variable_code  \
0  Dollars (millions)           H01
1  Dollars (millions)           H04
2  Dollars (millions)           H05
3  Dollars (millions)           H07
4  Dollars (millions)           H08

                                  Variable_name      Variable_category  \
0                                  Total income  Financial performance
1  Sales, government funding, grants and subsidies  Financial performance
2               Interest, dividends and donations  Financial performance
3                           Non-operating income  Financial performance
4                              Total expenditure  Financial performance

    Value                            Industry_code_ANZSIC06
0  930995  ANZSIC06 divisions A-S (excluding classes K633...
1  821630  ANZSIC06 divisions A-S (excluding classes K633...
2   84354  ANZSIC06 divisions A-S (excluding classes K633...
3   25010  ANZSIC06 divisions A-S (excluding classes K633...
4  832964  ANZSIC06 divisions A-S (excluding classes K633...
```

2. Printing the column names of the DataFrame

- The `.columns` attribute lists all column names in the DataFrame.
- Useful for understanding the dataset's schema.

```
data.columns
```

```
Index(['Year', 'Industry_aggregation_NZSIOC', 'Industry_code_NZSIOC',
       'Industry_name_NZSIOC', 'Units', 'Variable_code', 'Variable_name',
       'Variable_category', 'Value', 'Industry_code_ANZSIC06'],
      dtype='object')
```

3. Summary of Data Frame

- The `.info()` method provides a concise summary of the DataFrame.
- Displays data types, non-null counts, and memory usage.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50985 entries, 0 to 50984
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype 
---  ------                      --------------  ----- 
 0   Year                        50985 non-null  int64 
 1   Industry_aggregation_NZSIOC  50985 non-null  object
 2   Industry_code_NZSIOC        50985 non-null  object
 3   Industry_name_NZSIOC        50985 non-null  object
 4   Units                       50985 non-null  object
 5   Variable_code               50985 non-null  object
 6   Variable_name               50985 non-null  object
 7   Variable_category           50985 non-null  object
 8   Value                       50985 non-null  object
 9   Industry_code_ANZSIC06      50985 non-null  object
dtypes: int64(1), object(9)
memory usage: 3.9+ MB
```

4. Number of columns in the dataset

- Use `len(data.columns)` to count the columns.
- Provides an overview of the dataset's dimensionality.

```
size = len(data.columns)
print("Number of columns:", size)
```

```
Number of columns: 10
```

5. Descriptive Statistical Measures of a DataFrame

- The `.describe()` method gives summary statistics for numerical columns.
- Includes metrics like mean, median, and standard deviation.

```
data.describe()
```

	Year
count	50985.000000
mean	2018.000000
std	3.162309
min	2013.000000
25%	2015.000000
50%	2018.000000
75%	2021.000000
max	2023.000000

6. Print the name of all the columns.

- Using `data.columns.tolist()` converts column names into a list.
- Helps when column names need to be iterated or modified.

```
print(data.columns.tolist())
```

```
['Year', 'Industry_aggregation_NZSIOC', 'Industry_code_NZSIOC', 'Industry_name_NZSIOC', 'Units', 'Variable_code', 'Variable_name', 'Variable_category',
'Value', 'Industry_code_ANZSIC06']
```

7. Number of observations in the dataset.

- len`(data)` returns the total number of rows (observations).
- Helps determine dataset size.

```
size_obj = len(data)
print("Number of Observation", size_obj)
```

```
Number of Observation 50985
```

8. How is the dataset indexed?

- data.`index` reveals the index structure (e.g., range or custom index).
- Key for understanding how rows are identified.

```
data.index
```

```
RangeIndex(start=0, stop=50985, step=1)
```

9. Sorting DataFrame values.

•. sort_values(by='column_name') organizes data by a specific column.

• Enables ranking or prioritizing rows based on criteria.

```
data_sorted = data.sort_values(by='Value', ascending=False)
print(data_sorted.head(5))
```

```
       Year Industry_aggregation_NZSIOC Industry_code_NZSIOC  \
40287  2015                     Level 4               KK112
40219  2015                     Level 3                KK11
22078  2019                     Level 4               LL122
31348  2017                     Level 4               LL122
31347  2017                     Level 4               LL122

                  Industry_name_NZSIOC              Units Variable_code  \
40287          Financial Asset Investing  Dollars (millions)          H26
40219                            Finance  Dollars (millions)          H26
22078  Non-Residential Property Operation  Dollars (millions)          H28
31348  Non-Residential Property Operation  Dollars (millions)          H28
31347  Non-Residential Property Operation  Dollars (millions)          H27

                 Variable_name   Variable_category Value  \
40287     Fixed tangible assets  Financial position     S
40219     Fixed tangible assets  Financial position     S
22078  Disposals of fixed assets  Financial position     S
31348  Disposals of fixed assets  Financial position     S
31347  Additions to fixed assets  Financial position     S

                        Industry_code_ANZSIC06
40287                      ANZSIC06 group K624
40219  ANZSIC06 groups K621, K622, K623, and K624
22078                    ANZSIC06 class L671200
31348                    ANZSIC06 class L671200
31347                    ANZSIC06 class L671200
```

10. Missing Data Handing.

- . isnull`().sum()` identifies missing values in columns.
- Allows for cleaning or imputing missing data.

```
data.isnull().sum()
```

```
Year                          0
Industry_aggregation_NZSIOC   0
Industry_code_NZSIOC          0
Industry_name_NZSIOC          0
Units                         0
Variable_code                 0
Variable_name                 0
Variable_category             0
Value                         0
Industry_code_ANZSIC06        0
dtype: int64
```