

Coding Challenge Python

Name: Divya Sree Murali

Date: 15/11/2024

1. Printing rows of the Data

```
import pandas as pd
import numpy as np
```

```
###1.Printing rows of data
```

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

```
print(data) # print all rows
```

```
      Year  Industry_aggregation_NZSIOC  Industry_code_NZSIOC \
0      2023                        Level 1                99999
1      2023                        Level 1                99999
2      2023                        Level 1                99999
3      2023                        Level 1                99999
4      2023                        Level 1                99999
...      ...                        ...                    ...
50980  2013                        Level 3                 ZZ11
50981  2013                        Level 3                 ZZ11
50982  2013                        Level 3                 ZZ11
50983  2013                        Level 3                 ZZ11
50984  2013                        Level 3                 ZZ11
```

```
      Industry_name_NZSIOC  Units  Variable_code \
0      All industries  Dollars (millions)      H01
1      All industries  Dollars (millions)      H04
2      All industries  Dollars (millions)      H05
3      All industries  Dollars (millions)      H07
4      All industries  Dollars (millions)      H08
...      ...      ...      ...      ...
```

```
...      ...      ...      ...      ...
50980  Food product manufacturing  Percentage      H37
50981  Food product manufacturing  Percentage      H38
50982  Food product manufacturing  Percentage      H39
50983  Food product manufacturing  Percentage      H40
50984  Food product manufacturing  Percentage      H41
```

```
      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
...      ...      ...
50980      Quick ratio  Financial ratios
50981  Margin on sales of goods for resale  Financial ratios
50982      Return on equity  Financial ratios
50983      Return on total assets  Financial ratios
50984  Liabilities structure  Financial ratios
```

```
      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...
...      ...
50980  52  ANZSIC06 groups C111, C112, C113, C114, C115, ...
50981  40  ANZSIC06 groups C111, C112, C113, C114, C115, ...
50982  12  ANZSIC06 groups C111, C112, C113, C114, C115, ...
50983  5   ANZSIC06 groups C111, C112, C113, C114, C115, ...
50984  46  ANZSIC06 groups C111, C112, C113, C114, C115, ...
```

```
[50985 rows x 10 columns]
```

1.1. Printing first 5 rows

```
print(data.head()) #print first 5 rows
```

```
   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

```
In [13]: #---2.Printing columns of data
print(list(data.columns))
```

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

3. Summary of Data Frame

```
In [4]: #---3.Summary of the data
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. Descriptive Statistical Measures of a DataFrame

```
In [14]: #---4.Descriptive Statistical Measures of a DataFrame
print(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
```

5. Missing Data Handling

```
In [6]: #---5.Missing Data Handling
data.dropna()
```

Out[6]:

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value
0	2023	Level 1	99999	All industries	Dollars (millions)	H01	Total income	Financial performance	930995
1	2023	Level 1	99999	All industries	Dollars (millions)	H04	Sales, government funding, grants and subsidies	Financial performance	821630
2	2023	Level 1	99999	All industries	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	84354
3	2023	Level 1	99999	All industries	Dollars (millions)	H07	Non-operating income	Financial performance	25010
4	2023	Level 1	99999	All industries	Dollars (millions)	H08	Total expenditure	Financial performance	832964
...
50980	2013	Level 3	ZZ11	Food product manufacturing	Percentage	H37	Quick ratio	Financial ratios	52
50981	2013	Level 3	ZZ11	Food product manufacturing	Percentage	H38	Margin on sales of goods for resale	Financial ratios	40
50982	2013	Level 3	ZZ11	Food product manufacturing	Percentage	H39	Return on equity	Financial ratios	12
50983	2013	Level 3	ZZ11	Food product manufacturing	Percentage	H40	Return on total assets	Financial ratios	5
50984	2013	Level 3	ZZ11	Food product manufacturing	Percentage	H41	Liabilities structure	Financial ratios	46

50985 rows x 10 columns

For Counting missing values

```
In [7]: data.isnull().sum()
```

```
Out[7]: 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
```

6.Sorting DataFrame values

```
In [8]: #---6.Sorting DataFrame values
sorted_data = data.sort_values(by='Year', ascending=True)
print(sorted_data)
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	\
50984	2013	Level 3	ZZ11	
47889	2013	Level 4	CC822	
47890	2013	Level 4	CC822	
47891	2013	Level 4	CC822	
47892	2013	Level 4	CC822	
...	
3092	2023	Level 1	KK	
3093	2023	Level 1	KK	
3094	2023	Level 1	KK	
3042	2023	Level 3	JJ11	
0	2023	Level 1	99999	

	Industry_name_NZSIOC	Units	Variable_code	\
50984	Food product manufacturing	Percentage	H41	
47889	Machinery Manufacturing	Dollars (millions)	H09	
47890	Machinery Manufacturing	Dollars (millions)	H10	
47891	Machinery Manufacturing	Dollars (millions)	H11	
47892	Machinery Manufacturing	Dollars (millions)	H12	
...	
3092	Financial and Insurance Services	Dollars (millions)	H09	
3093	Financial and Insurance Services	Dollars (millions)	H10	
3094	Financial and Insurance Services	Dollars (millions)	H11	
3042	Information Media Services	Dollars (millions)	H30	
0	All industries	Dollars (millions)	H01	

	Variable_name	Variable_category	Value	\
50984	Liabilities structure	Financial ratios	46	
47889	Interest and donations	Financial performance	36	
47890	Indirect taxes	Financial performance	9	
47891	Depreciation	Financial performance	72	
47892	Salaries and wages paid	Financial performance	908	
...	
3092	Interest and donations	Financial performance	36369	
3093	Indirect taxes	Financial performance	131	
3094	Depreciation	Financial performance	1328	
3042	Total equity and liabilities	Financial position	5966	
0	Total income	Financial performance	930995	

	Industry_code_ANZSIC06
50984	ANZSIC06 groups C111, C112, C113, C114, C115, ...
47889	ANZSIC06 groups C245, C246, and C249
47890	ANZSIC06 groups C245, C246, and C249
47891	ANZSIC06 groups C245, C246, and C249
47892	ANZSIC06 groups C245, C246, and C249
...	...
3092	ANZSIC06 division K
3093	ANZSIC06 division K
3094	ANZSIC06 division K
3042	ANZSIC06 groups J541, J542, J551, J552, J561, ...
0	ANZSIC06 divisions A-S (excluding classes K633...

[50985 rows x 10 columns]

11.What is the number of columns in the dataset?

```
In [15]: #---11.what is the number of columns in the dataset?

df=pd.read_csv("annual-enterprise-survey-2023-financial-year-provisional.csv")
print("Number of columns :",len(data.columns))

Number of columns : 10
```

8. Apply Function

In the output newly created func_column is observed.

```
In [16]: #---8.Apply Function
def func(value):
    if value>2015:
        return True
    else:
        return False

data['Func_Column']=data['Year'].apply(func)
print(data)
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	\
0	2023	Level 1	99999	
1	2023	Level 1	99999	
2	2023	Level 1	99999	
3	2023	Level 1	99999	
4	2023	Level 1	99999	
...	
50980	2013	Level 3	ZZ11	
50981	2013	Level 3	ZZ11	
50982	2013	Level 3	ZZ11	
50983	2013	Level 3	ZZ11	
50984	2013	Level 3	ZZ11	

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

50980	Food product manufacturing	Percentage	H37
50981	Food product manufacturing	Percentage	H38
50982	Food product manufacturing	Percentage	H39
50983	Food product manufacturing	Percentage	H40
50984	Food product manufacturing	Percentage	H41

	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	
...	
50980	Quick ratio	Financial ratios	
50981	Margin on sales of goods for resale	Financial ratios	
50982	Return on equity	Financial ratios	
50983	Return on total assets	Financial ratios	
50984	Liabilities structure	Financial ratios	

	Value	Industry_code_ANZSIC06	Func_Column
0	930995	ANZSIC06 divisions A-S (excluding classes K633...	True
1	821630	ANZSIC06 divisions A-S (excluding classes K633...	True
2	84354	ANZSIC06 divisions A-S (excluding classes K633...	True
3	25010	ANZSIC06 divisions A-S (excluding classes K633...	True
4	832964	ANZSIC06 divisions A-S (excluding classes K633...	True
...
50980	52	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False
50981	40	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False
50982	12	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False
50983	5	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False
50984	46	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False

[50985 rows x 11 columns]

9.By using the lambda operator

Newly added lambda column is observed in the output

```
In [18]: #----9.By using the Lambda operator
data['Lambdacolumn']=data['Year'].apply(lambda x:x*10)
print(data)
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	
0	2023	Level 1	99999	
1	2023	Level 1	99999	
2	2023	Level 1	99999	
3	2023	Level 1	99999	
4	2023	Level 1	99999	
...	
50980	2013	Level 3	ZZ11	
50981	2013	Level 3	ZZ11	
50982	2013	Level 3	ZZ11	
50983	2013	Level 3	ZZ11	
50984	2013	Level 3	ZZ11	

	Industry_name_NZSIOC	Units	Variable_code	
0	All industries	Dollars (millions)	H01	
1	All industries	Dollars (millions)	H04	
2	All industries	Dollars (millions)	H05	
3	All industries	Dollars (millions)	H07	
4	All industries	Dollars (millions)	H08	
...	
50980	Food product manufacturing	Percentage	H37	
50981	Food product manufacturing	Percentage	H38	
50982	Food product manufacturing	Percentage	H39	
50983	Food product manufacturing	Percentage	H40	
50984	Food product manufacturing	Percentage	H41	

	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	
...	
50980	Quick ratio	Financial ratios	
50981	Margin on sales of goods for resale	Financial ratios	
50982	Return on equity	Financial ratios	
50983	Return on total assets	Financial ratios	
50984	Liabilities structure	Financial ratios	

	Value	Industry_code_ANZSIC06	Func_Column	
0	930995	ANZSIC06 divisions A-S (excluding classes K633...	True	
1	821630	ANZSIC06 divisions A-S (excluding classes K633...	True	
2	84354	ANZSIC06 divisions A-S (excluding classes K633...	True	
3	25010	ANZSIC06 divisions A-S (excluding classes K633...	True	
4	832964	ANZSIC06 divisions A-S (excluding classes K633...	True	
...	
50980	52	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False	
50981	40	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False	
50982	12	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False	
50983	5	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False	
50984	46	ANZSIC06 groups C111, C112, C113, C114, C115, ...	False	

	Lambdacolumn
0	2033
1	2033
2	2033
3	2033
4	2033
...	...
50980	2023
50981	2023
50982	2023
50983	2023
50984	2023

[50985 rows x 12 columns]

14. What is the number of observations in the dataset?

```
In [22]: #---14.What is the number of observations in the dataset?
print("Number of observations is: ",len(data))

Number of observations is: 50985
```

