

## 1. Printing rows of the Data

```
# Print the first 5 rows
display(data.head())
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value	Indus
0	2023	Level 1	99999	All industries	Dollars (millions)	H01	Total income	Financial performance	930995	At (exc
1	2023	Level 1	99999	All industries	Dollars (millions)	H04	Sales, government funding, grants and subsidies	Financial performance	821630	At (exc
2	2023	Level 1	99999	All industries	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	84354	At (exc
3	2023	Level 1	99999	All industries	Dollars (millions)	H07	Non-operating income	Financial performance	25010	At (exc
4	2023	Level 1	99999	All industries	Dollars (millions)	H08	Total expenditure	Financial performance	832964	At (exc

```
# Print the last 5 rows
display(data.tail())
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value
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

```
# Print a specific row by index
display(data.iloc[22])
```

```
Year                                2023
Industry_aggregation_NZSIOC        Level 1
Industry_code_NZSIOC                99999
Industry_name_NZSIOC               All industries
Units                             Dollars (millions)
Variable_code                      H32
Variable_name                      Current liabilities
Variable_category                  Financial position
Value                             1074693
Industry_code_ANZSIC06             ANZSIC06 divisions A-S (excluding classes K633...
Name: 22, dtype: object
```

```
display(data)
```

	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

## 2. Printing the column names of the DataFrame

```
print(data.columns)
print("\n")
print(list(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')
```

```
['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

```
print(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
```

```
None
```

### 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

```
tot_records = data.shape
print("Data Shape: ", tot_records)
```

Data Shape: (50985, 10)

```
tot_null = data.isnull().sum()
print("Total Null Value: ", tot_null)
```

```
Total Null Value: 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
```

```
tot_not_null = data.notnull().sum()
print("Total Null Value: ", tot_not_null)
```

```
Total Null Value: Year          50985
Industry_aggregation_NZSIOC    50985
Industry_code_NZSIOC           50985
Industry_name_NZSIOC           50985
Units                          50985
Variable_code                  50985
Variable_name                   50985
Variable_category              50985
Value                          50985
Industry_code_ANZSIC06         50985
dtype: int64
```

```
data = data.dropna()
```



## 6. Sorting DataFrame values

```
sorted_data_single = data.sort_values(by='Value')
display(sorted_data_single)
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value
27424	2018	Level 4	QQ111	Hospitals	Percentage	H40	Return on total assets	Financial ratios	-1
47238	2013	Level 4	CC411	Printing	Percentage	H40	Return on total assets	Financial ratios	-1
47202	2013	Level 3	CC41	Printing	Percentage	H40	Return on total assets	Financial ratios	-1
3270	2023	Level 4	KK121	Life Insurance	Percentage	H40	Return on total assets	Financial ratios	-1
32685	2016	Level 4	AA131	Dairy Cattle Farming	Percentage	H40	Return on total assets	Financial ratios	-1
...	...	...	...	...	...	...	...	...	...
49887	2013	Level 4	LL122	Non-Residential Property Operation	Dollars (millions)	H27	Additions to fixed assets	Financial position	S
31347	2017	Level 4	LL122	Non-Residential Property Operation	Dollars (millions)	H27	Additions to fixed assets	Financial position	S
40219	2015	Level 3	KK11	Finance	Dollars (millions)	H26	Fixed tangible assets	Financial position	S
40221	2015	Level 3	KK11	Finance	Dollars (millions)	H28	Disposals of fixed assets	Financial position	S
40618	2015	Level 4	LL122	Non-Residential Property Operation	Dollars (millions)	H28	Disposals of fixed assets	Financial position	S

Sort with multiple parameters go by order in which they specified

```
sort_multi = data.sort_values(by=['Year', 'Value', 'Units'], ascending=[True, True, False])
display(sort_multi)
```

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value
47202	2013	Level 3	CC41	Printing	Percentage	H40	Return on total assets	Financial ratios	-
47238	2013	Level 4	CC411	Printing	Percentage	H40	Return on total assets	Financial ratios	-
47623	2013	Level 3	CC71	Primary Metal and Metal Product Manufacturing	Dollars	H35	Surplus per employee count	Financial ratios	-131,30
47659	2013	Level 4	CC711	Primary Metal and Metal Product Manufacturing	Dollars	H35	Surplus per employee count	Financial ratios	-131,30
47129	2013	Level 3	CC32	Pulp, Paper and Converted Paper Product Manufa...	Percentage	H39	Return on equity	Financial ratios	-1
...	...	...	...	...	...	...	...	...	...
4455	2023	Level 4	RS113	Gambling Activities	Dollars (millions)	H25	Current assets	Financial position	0
4459	2023	Level 4	RS113	Gambling Activities	Dollars (millions)	H29	Other assets	Financial position	0
4602	2023	Level 3	ZZ11	Food Product Manufacturing	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	0
4603	2023	Level 3	ZZ11	Food Product Manufacturing	Dollars (millions)	H06	Government funding, grants and subsidies	Financial performance	0

## 7. Merge Data Frames

```
: df1 = data.copy()
  df2 = data.copy()
  df_merged = pd.merge(df1, df2)
```

## 8. Apply Function

```
def fun(value):
    if value>2019:
        return "COVID"
    else:
        return "No"

data_copy=data.copy()
data_copy['Year_mod'] = data_copy['Year'].apply(fun)
data_copy.head(-10)
```

_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value	Industry_code_ANZSIC06	Year_mod
Level 1	99999	All industries	Dollars (millions)	H01	Total income	Financial performance	930995	ANZSIC06 divisions A-S (excluding classes K633...	COVID
Level 1	99999	All industries	Dollars (millions)	H04	Sales, government funding, grants and subsidies	Financial performance	821630	ANZSIC06 divisions A-S (excluding classes K633...	COVID
Level 1	99999	All industries	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	84354	ANZSIC06 divisions A-S (excluding classes K633...	COVID
Level 1	99999	All industries	Dollars (millions)	H07	Non-operating income	Financial performance	25010	ANZSIC06 divisions A-S (excluding classes K633...	COVID
Level 1	99999	All industries	Dollars (millions)	H08	Total expenditure	Financial performance	832964	ANZSIC06 divisions A-S (excluding classes K633...	COVID
...	...	...	...	...	...	...	...	...	...
Level 3	ZZ11	Food product manufacturing	Dollars (millions)	H27	Additions to fixed assets	Financial position	1,339	ANZSIC06 groups C111, C112, C113, C114, C115, ...	No
Level 3	ZZ11	Food product manufacturing	Dollars (millions)	H28	Disposals of fixed assets	Financial position	375	ANZSIC06 groups C111, C112, C113, C114, C115, ...	No
Level 3	ZZ11	Food product manufacturing	Dollars (millions)	H29	Other assets	Financial position	5,434	ANZSIC06 groups C111, C112, C113, C114, C115, ...	No

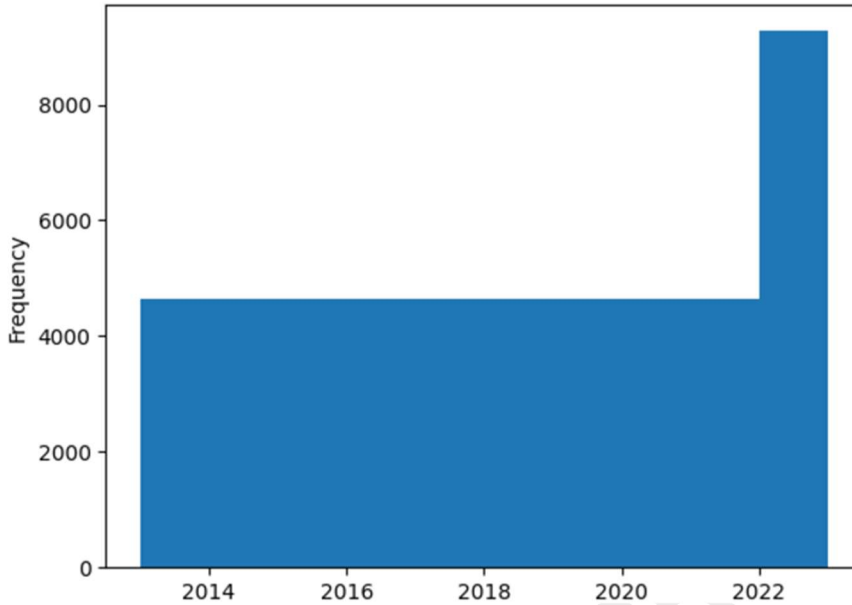
## 9. By using the lambda operator

```
data_copy = data.copy()
data_copy['Year'] = data_copy['Year'].apply(lambda x: x + 10)
display(data_copy)
```

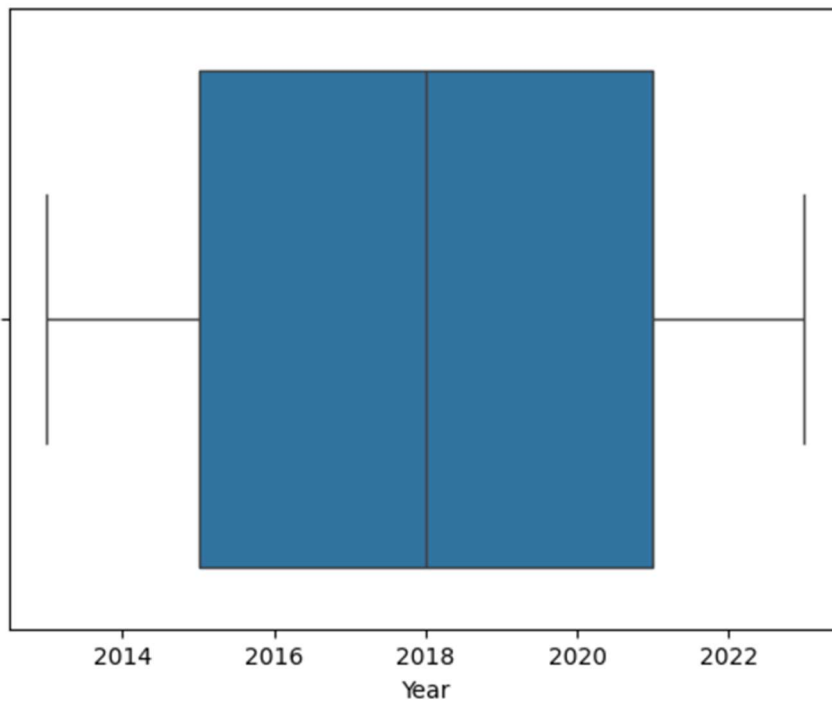
	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable_code	Variable_name	Variable_category	Value
0	2033	Level 1	99999	All industries	Dollars (millions)	H01	Total income	Financial performance	930995
1	2033	Level 1	99999	All industries	Dollars (millions)	H04	Sales, government funding, grants and subsidies	Financial performance	821630
2	2033	Level 1	99999	All industries	Dollars (millions)	H05	Interest, dividends and donations	Financial performance	84354
3	2033	Level 1	99999	All industries	Dollars (millions)	H07	Non-operating income	Financial performance	25010
4	2033	Level 1	99999	All industries	Dollars (millions)	H08	Total expenditure	Financial performance	832964
...	...	...	...	...	...	...	...	...	...
50980	2023	Level 3	ZZ11	Food product manufacturing	Percentage	H37	Quick ratio	Financial ratios	52
50981	2023	Level 3	ZZ11	Food product manufacturing	Percentage	H38	Margin on sales of goods for resale	Financial ratios	40
50982	2023	Level 3	ZZ11	Food product manufacturing	Percentage	H39	Return on equity	Financial ratios	12

## 10. Visualizing DataFrame

```
import matplotlib.pyplot as plt
data['Year'].plot(kind='hist')
plt.show()
```



```
sns.boxplot(x='Year', data=data)
plt.show()
```



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

```
print("Number of columns:", data.shape[1])
length = len(data.columns)
print("Number of columns:", length)
```

Number of columns: 10  
Number of columns: 10

## 12 print the name of all the columns.

```
print(data.columns)
print("\n")
print(list(data.columns))
print("\n")
list_val = list(data.columns)
for li in list_val:
    print(li)
```

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

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

```
Year
Industry_aggregation_NZSIOC
Industry_code_NZSIOC
Industry_name_NZSIOC
Units
Variable_code
Variable_name
Variable_category
Value
Industry_code_ANZSIC06
```

## 13 How is the dataset indexed?

```
print(data.index)
```

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

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

```
print(data.shape[0])
```

50985

```
tot_records = data.shape
print("Data Shape: ", tot_records)
```

Data Shape: (50985, 10)