

```
In [ ]: #PYTHON PANDAS# SYNAIDU
```

```
In [ ]: Pandas used for data analysis ,data manipulation ,cleaning and exploring data.
```

```
In [ ]: Pandas dataframes
1.series-1D
2.DataFrame-2D
3.Panel-3D
```

```
In [1]: pip install pandas

Requirement already satisfied: pandas in c:\pavan9\lib\site-packages (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\pavan9\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\pavan9\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in c:\pavan9\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\pavan9\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\pavan9\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: import pandas as pd
```

```
In [4]: print(pd.__version__)

2.0.2
```

```
In [ ]: How to download a Dataset#kaggle.com#
```

```
In [5]: d=pd.read_csv("C:\\\\Users\\\\synai\\\\Downloads\\\\business-financial-data-march-2023.csv")
df=pd.DataFrame(d)
```

```
In [6]: df
```

Out[6]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
------------------	--------	------------	------------	--------	-------	-----------

6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F	Dollars	6
------	--------------	---------	---------	-----	---	---------	---

6675 rows x 14 columns

In [12]: `print(df)`

Out[12]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 14 columns

In [13]: `import pandas as pd`In [21]: `df`

Out[21]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 14 columns

```
In [17]: import pandas as pd  
c = pd.read_excel("C:\\\\Users\\\\synai\\\\Downloads\\\\Orders-With Nulls.xlsx")
```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
File C:\pavan9\lib\site-packages\pandas\compat\_optional.py:142, in import_optional_dependency(name, extra, errors, min_version)
    141     try:
--> 142         module = importlib.import_module(name)
    143     except ImportError:

File C:\pavan9\lib\importlib\__init__.py:126, in import_module(name, package)
    125         level += 1
--> 126     return _bootstrap._gcd_import(name[level:], package, level)

File <frozen importlib._bootstrap>:1050, in _gcd_import(name, package, level)

File <frozen importlib._bootstrap>:1027, in _find_and_load(name, import_)

File <frozen importlib._bootstrap>:1004, in _find_and_load_unlocked(name, import_)

ModuleNotFoundError: No module named 'openpyxl'

```

During handling of the above exception, another exception occurred:

```

ImportError                                     Traceback (most recent call last)
Cell In[17], line 3
      1 import pandas as pd
--> 3 c = pd.read_excel("C:\\\\Users\\\\synai\\\\Downloads\\\\Orders-With Nulls.xlsx")

File C:\pavan9\lib\site-packages\pandas\io\excel\_base.py:478, in read_excel(io, sheet_name, header, names, index_col, usecols, dtype, engine, converters, true_values, false_values, skiprows, nrows, na_values, keep_default_na, na_filter, verbose, parse_dates, date_parser, date_format, thousands, decimal, comment, skipfooter, storage_options, dtype_backend)
    476 if not isinstance(io, ExcelFile):
    477     should_close = True
--> 478     io = ExcelFile(io, storage_options=storage_options, engine=engine)
    479 elif engine and engine != io.engine:
    480     raise ValueError(
        "Engine should not be specified when passing "
        "an ExcelFile - ExcelFile already has the engine set"
    481     )
    482
    483

File C:\pavan9\lib\site-packages\pandas\io\excel\_base.py:1513, in ExcelFile.__init__(self, path_or_buffer, engine, storage_options)
    1510 self.engine = engine
    1511 self.storage_options = storage_options
-> 1513 self._reader = self._engines[engine](self._io, storage_options=storage_options)

File C:\pavan9\lib\site-packages\pandas\io\excel\_openpyxl.py:548, in OpenpyxlReader.__init__(self, filepath_or_buffer, storage_options)
    533 @doc(storage_options=_shared_docs["storage_options"])
    534 def __init__(
    535     self,
    536     filepath_or_buffer: FilePath | ReadBuffer[bytes],
    537     storage_options: StorageOptions = None,
    538 ) -> None:
    539     """
    540     Reader using openpyxl engine.
    541

```

```
(...)  
546     {storage_options}  
547     """  
--> 548     import_optional_dependency("openpyxl")  
549     super().__init__(filepath_or_buffer, storage_options=storage_options)  
  
File C:\pavan9\lib\site-packages\pandas\compat\_optional.py:145, in import_optional_dependency(name, extra, errors, min_version)  
143 except ImportError:  
144     if errors == "raise":  
--> 145         raise ImportError(msg)  
146     return None  
148 # Handle submodules: if we have submodule, grab parent module from sys.modules  
  
ImportError: Missing optional dependency 'openpyxl'. Use pip or conda to install openpyxl.
```

In [19]: `pip install openpyxl`

```
Collecting openpyxl  
  Downloading openpyxl-3.1.2-py2.py3-none-any.whl (249 kB)  
          0.0/250.0 kB ? eta -----  
----- 81.9/250.0 kB 2.3 MB/s eta 0:00:01  
----- 235.5/250.0 kB 2.4 MB/s eta 0:00:01  
----- 245.8/250.0 kB 2.5 MB/s eta 0:00:01  
----- 250.0/250.0 kB 1.4 MB/s eta 0:00:00  
  
Collecting et-xmlfile (from openpyxl)  
  Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)  
Installing collected packages: et-xmlfile, openpyxl  
Successfully installed et-xmlfile-1.1.0 openpyxl-3.1.2  
Note: you may need to restart the kernel to use updated packages.
```

In [23]: `import pandas as pd`

```
c = pd.read_excel("C:\\\\Users\\\\synai\\\\Downloads\\\\Orders-With Nulls.xlsx")
```

In [24]: `dfe=pd.DataFrame(c)`

In [25]: `print(c)`

```

      Order ID Order Date   Order Quantity     Sales    Ship Mode \
0            3 2010-10-13             6  261.5400  Regular Air
1            6 2012-02-20             2   6.9300  Regular Air
2           32 2011-07-15            26 2808.0800  Regular Air
3           32 2011-07-15            24 1761.4000 Delivery Truck
4           32 2011-07-15            23 160.2335  Regular Air
...
1002        7171 2011-02-13            17  303.1865  Regular Air
1003        7174 2012-03-10            10 141.9200  Regular Air
1004        7175 2010-02-07            10  748.2500 Delivery Truck
1005        7203 2009-01-08            25 21752.0100  Regular Air
1006        7239 2011-06-29            50  6206.1600  Regular Air

```

```

          Profit Unit Price Customer Name Customer Segment \
0      -213.250      38.94 Muhammed MacIntyre  Small Business
1       -4.640       2.08 Ruben Dartt       Corporate
2      1054.820      107.53 Liz Pelletier       Corporate
3     -1748.560      70.89 Liz Pelletier       Corporate
4      -85.129       7.99 Liz Pelletier       Corporate
...
1002     92.592      20.99 Andy Gerbode      Consumer
1003    12.200      13.73 Thomas Thornton      Consumer
1004    -86.990      70.98 Helen Andreada       Corporate
1005   9296.348      896.99 Ruben Dartt       Corporate
1006   1416.270      120.33 Craig Carroll  Small Business

```

```

Product Category
0   Office Supplies
1   Office Supplies
2       Furniture
3       Furniture
4   Technology
...
1002      Technology
1003      Furniture
1004      Furniture
1005  Office Supplies
1006  Office Supplies

```

[1007 rows x 10 columns]

```
In [27]: di = {
    "Name": ["abc", "def", "ghi"],
    "Roll_no": [1, 2, 3],
    "percentage": [90, 89, 58]
}
```

```
In [28]: print(di)
```

```
{'Name': ['abc', 'def', 'ghi'], 'Roll_no': [1, 2, 3], 'percentage': [90, 89, 58]}
```

```
In [29]: dfd=pd.DataFrame(di)
```

```
In [31]: print(dfd)
```

	Name	Roll_no	percentage
0	abc	1	90
1	def	2	89
2	ghi	3	58

```
In [32]: l = [  
    ("abc", 10, 89.2),  
    ("def", 11, 96),  
    ("klm", 12, 69.9)  
]
```

```
In [33]: print(l)  
[('abc', 10, 89.2), ('def', 11, 96), ('klm', 12, 69.9)]
```

```
In [34]: df1=pd.DataFrame(l)
```

```
In [35]: print(df1)  
  
   0   1   2  
0 abc 10 89.2  
1 def 11 96.0  
2 klm 12 69.9
```

```
In [36]: df
```

Out[36]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F	Dollars	6
6675 rows x 14 columns							
In [37]:	df.head()						
Out[37]:	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
0	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
1	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
2	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
3	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
4	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
In [38]:	df.tail()						

Out[38]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6670	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
6671	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
6672	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
6673	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F	Dollars	6

In [39]: dfe.head()

Out[39]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Customer Segment
0	3	2010-10-13	6	261.5400	Regular Air	-213.250	38.94	Muhammed MacIntyre	Small Business
1	6	2012-02-20	2	6.9300	Regular Air	-4.640	2.08	Ruben Dartt	Corporate
2	32	2011-07-15	26	2808.0800	Regular Air	1054.820	107.53	Liz Pelletier	Corporate
3	32	2011-07-15	24	1761.4000	Delivery Truck	-1748.560	70.89	Liz Pelletier	Corporate
4	32	2011-07-15	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Corporate

In [41]: `dfe.tail()`

Out[41]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Customer Segment
<b>1002</b>	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Consumer
<b>1003</b>	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Consumer
<b>1004</b>	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Corporate
<b>1005</b>	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Corporate
<b>1006</b>	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Small Business

In [42]: `df.head(1)`

Out[42]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Significance	Buckets
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386		NaN	F	Dollars	6	Collaboration

In [43]: `dfe.tail(1)`

Out[43]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Customer Segment
<b>1006</b>	7239	2011-06-29	50	6206.16	Regular Air	1416.27	120.33	Craig Carroll	Small Business

In [44]: `dfe.tail(10)`

Out[44]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Customer Segm
997	7136	2009-03-22	9	47.2800	Regular Air	17.050	4.91	Lena Cacioppo	Holiday
998	7142	2011-07-21	36	132.8600	Regular Air	57.000	3.69	Cari Sayre	Corporate
999	7169	2010-03-26	22	446.7200	Regular Air	-39.000	20.28	Lela Donovan	Corporate
1000	7169	2010-03-26	30	1580.6005	Regular Air	303.525	65.99	Lela Donovan	Corporate
1001	7171	2011-02-13	28	1703.8505	Regular Air	316.062	65.99	Andy Gerbode	Consumer
1002	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Consumer
1003	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Consumer
1004	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Corporate
1005	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Corporate
1006	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Small Business

In [46]: `df.describe()`

Out[46]:

	Period	Data_value	Magnitude	Series_title_5
<b>count</b>	6675.000000	6120.000000	6675.0	0.0
<b>mean</b>	2019.344234	4824.513007	6.0	NaN
<b>std</b>	2.034234	7133.845971	0.0	NaN
<b>min</b>	2016.060000	-398.194000	6.0	NaN
<b>25%</b>	2018.030000	654.557500	6.0	NaN
<b>50%</b>	2019.120000	2120.532500	6.0	NaN
<b>75%</b>	2021.090000	4549.585000	6.0	NaN
<b>max</b>	2023.030000	40940.911000	6.0	NaN

In [48]: `df.shape`

Out[48]: (6675, 14)

In [49]: `df[0:10:2] #Indexing#start:stop:step`

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Si
0	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6	Bu Coll
2	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6	Bu Coll
4	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6	Bu Coll
6	BDCQ.SF1AA2CA	2017.12	1290.820	NaN	F	Dollars	6	Bu Coll
8	BDCQ.SF1AA2CA	2018.06	1488.055	NaN	F	Dollars	6	Bu Coll

In [50]: `df["Period"]#column_name`

Out[50]: 0 2016.06  
1 2016.09  
2 2016.12  
3 2017.03  
4 2017.06  
...  
6670 2022.03  
6671 2022.06  
6672 2022.09  
6673 2022.12  
6674 2023.03  
Name: Period, Length: 6675, dtype: float64

In [51]: `df[[ "Period", "UNITS" ]]#For Multiple columns Access create list first and give in`

Out[51]:

	Period	UNITS
0	2016.06	Dollars
1	2016.09	Dollars
2	2016.12	Dollars
3	2017.03	Dollars
4	2017.06	Dollars
...	...	...
6670	2022.03	Dollars
6671	2022.06	Dollars
6672	2022.09	Dollars
6673	2022.12	Dollars
6674	2023.03	Dollars

6675 rows × 2 columns

In [52]:

`df[["Period", "UNITS"]][0:10:2]`

Out[52]:

	Period	UNITS
0	2016.06	Dollars
2	2016.12	Dollars
4	2017.06	Dollars
6	2017.12	Dollars
8	2018.06	Dollars

In [54]:

`for record in dfd.iterrows():
 print(record)`

```
(0, Name      abc
Roll_no     1
percentage   90
Name: 0, dtype: object)
(1, Name      def
Roll_no     2
percentage   89
Name: 1, dtype: object)
(2, Name      ghi
Roll_no     3
percentage   58
Name: 2, dtype: object)
```

In [ ]:

`#In LOC STOP INDEX IS INCLUDE, Represented column_names#`

In [55]:

`df.loc[100]`

```
Out[55]: Series_reference          BDCQ.SF1AA3CA
Period                  2020.06
Data_value              1421.593
Suppressed             NaN
STATUS                  F
UNITS                  Dollars
Magnitude               6
Subject                Business Data Collection - BDC
Group                  Industry by financial variable (NZSIOC Level 2)
Series_title_1          Sales (operating income)
Series_title_2          Fishing, Aquaculture and Agriculture, Forestry...
Series_title_3          Current prices
Series_title_4          Unadjusted
Series_title_5          NaN
Name: 100, dtype: object
```

```
In [59]: df.loc[19, "UNITS"]
```

```
Out[59]: 'Dollars'
```

```
In [60]: df.loc[0:5]
```

```
Out[60]:    Series_reference  Period  Data_value  Suppressed  STATUS  UNITS  Magnitude  Su
```

0	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

1	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

2	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

3	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

4	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

5	BDCQ.SF1AA2CA	2017.09	1282.436	NaN	F	Dollars	6	Bu Coll
---	---------------	---------	----------	-----	---	---------	---	------------

```
In [62]: df.loc[0:5,"UNITS"]
```

```
Out[62]: 0    Dollars
1    Dollars
2    Dollars
3    Dollars
4    Dollars
5    Dollars
Name: UNITS, dtype: object
```

```
In [63]: df.loc[0:5,[ "UNITS", "STATUS"]]
```

Out[63]:

	UNITS	STATUS
0	Dollars	F
1	Dollars	F
2	Dollars	F
3	Dollars	F
4	Dollars	F
5	Dollars	F

In [65]: df.loc[0:4,"Period":"Group"]#Range for Both Index AND COLUMN Name#

Out[65]:

	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group
0	2016.06	1116.386	NaN	F	Dollars	6	Business Data Collection - BDC	Industry by financial variable (NZSIOC Level 2)
1	2016.09	1070.874	NaN	F	Dollars	6	Business Data Collection - BDC	Industry by financial variable (NZSIOC Level 2)
2	2016.12	1054.408	NaN	F	Dollars	6	Business Data Collection - BDC	Industry by financial variable (NZSIOC Level 2)
3	2017.03	1010.665	NaN	F	Dollars	6	Business Data Collection - BDC	Industry by financial variable (NZSIOC Level 2)
4	2017.06	1233.700	NaN	F	Dollars	6	Business Data Collection - BDC	Industry by financial variable (NZSIOC Level 2)

In [ ]: #In iloc stop index exclude and acces the data by column index number#

In [66]: df.iloc[2,2]#rownumber:columnnumber#

Out[66]: 1054.408

In [67]: `df.iloc[0:5,0:3]#Range#`

Out[67]:

	Series_reference	Period	Data_value
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700

In [68]: `df.iloc[0:5,2]#range, column index#`

Out[68]:

0	1116.386
1	1070.874
2	1054.408
3	1010.665
4	1233.700

Name: Data\_value, dtype: float64

In [69]: `df.iloc[[6,7,8],[1,2,5]]`

Out[69]:

	Period	Data_value	UNITS
<b>6</b>	2017.12	1290.820	Dollars
<b>7</b>	2018.03	1412.007	Dollars
<b>8</b>	2018.06	1488.055	Dollars

In [70]: `df.iloc[ : ,[1]]`

Out[70]:

	Period
<b>0</b>	2016.06
<b>1</b>	2016.09
<b>2</b>	2016.12
<b>3</b>	2017.03
<b>4</b>	2017.06
...	...
<b>6670</b>	2022.03
<b>6671</b>	2022.06
<b>6672</b>	2022.09
<b>6673</b>	2022.12
<b>6674</b>	2023.03

6675 rows × 1 columns

```
In [71]: df.sort_values("Period")
```

Out[71]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>616</b>	BDCQ.SF1CC3CS	2016.06	2211.608	NaN	R	Dollars	6
<b>4213</b>	BDCQ.SF2MN2CA	2016.06	2052.551	NaN	F	Dollars	6
<b>4241</b>	BDCQ.SF2MNCA	2016.06	6073.890	NaN	F	Dollars	6
<b>4269</b>	BDCQ.SF2PP1CA	2016.06	263.075	NaN	F	Dollars	6
...	...	...	...	...	...	...	...
<b>5233</b>	BDCQ.SF3JJCA	2023.03	697.332	NaN	F	Dollars	6
<b>3303</b>	BDCQ.SF1RSCA	2023.03	4508.136	NaN	F	Dollars	6
<b>5261</b>	BDCQ.SF3LL1CA	2023.03	596.688	NaN	F	Dollars	6
<b>3471</b>	BDCQ.SF2BB1CA	2023.03	789.992	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 14 columns

```
In [72]: df.sort_values("Period", ascending=False)
```

Out[72]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>6674</b>	BDCQ.SF8RSCA	2023.03	663.630	NaN	F	Dollars	6
<b>1735</b>	BDCQ.SF1FFCS	2023.03	38494.540	NaN	F	Dollars	6
<b>3163</b>	BDCQ.SF1RS1CS	2023.03	1477.596	NaN	F	Dollars	6
<b>1763</b>	BDCQ.SF1FFCT	2023.03	NaN	NaN	C	Dollars	6
<b>3191</b>	BDCQ.SF1RS1CT	2023.03	NaN	NaN	C	Dollars	6
...	...	...	...	...	...	...	...
<b>4633</b>	BDCQ.SF3CC2CA	2016.06	109.105	NaN	F	Dollars	6
<b>1792</b>	BDCQ.SF1GH1CS	2016.06	NaN	NaN	C	Dollars	6
<b>1820</b>	BDCQ.SF1GH1CT	2016.06	NaN	NaN	C	Dollars	6
<b>1848</b>	BDCQ.SF1GH2CA	2016.06	NaN	Y	C	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
------------------	--------	------------	------------	--------	-------	-----------

0	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
---	---------------	---------	----------	-----	---	---------	---

6675 rows x 14 columns

In [76]: `df.sort_values(["Period", "Subject"])`

Out[76]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>28</b>	BDCQ.SF1AA2CS	2016.06	1061.958	NaN	R	Dollars	6
<b>56</b>	BDCQ.SF1AA2CT	2016.06	1026.174	NaN	R	Dollars	6
<b>84</b>	BDCQ.SF1AA3CA	2016.06	1189.735	NaN	F	Dollars	6
<b>112</b>	BDCQ.SF1AA3CS	2016.06	1266.364	NaN	R	Dollars	6
...							
<b>6562</b>	BDCQ.SF8QQ1CA	2023.03	783.696	NaN	F	Dollars	6
<b>6590</b>	BDCQ.SF8QQCA	2023.03	783.696	NaN	F	Dollars	6
<b>6618</b>	BDCQ.SF8RS1CA	2023.03	487.585	NaN	F	Dollars	6
<b>6646</b>	BDCQ.SF8RS2CA	2023.03	176.045	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 14 columns

In [ ]: ##### Manipulating DataFrame #####

In [77]: df["Total"] = 0 #adding Total column#

In [78]: df

Out[78]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 15 columns

In [85]: df["Total"] = df["Data\_value"] + df["Magnitude"] #adding Total

In [86]: df

Out[86]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 15 columns

In [87]: df["per"]="P/F"

In [88]: df

Out[88]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 16 columns

In [90]: df.drop(columns="per")

Out[90]:		Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
	<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
	<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
	<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
	<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
	<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
	...	...	...	...	...	...	...	...
	<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
	<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
	<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
	<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
------------------	--------	------------	------------	--------	-------	-----------

6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6
------	--------------	---------	---------	-----	-----------	---

6675 rows x 15 columns

In [91]: df

Out[91]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F Dollars	6

6675 rows x 16 columns

```
In [92]: df.drop(columns='per',inplace=True)###Permenant delete of column#
```

```
In [93]: df
```

Out[93]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	NaN	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	NaN	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	NaN	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	NaN	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	NaN	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	NaN	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	NaN	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	NaN	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	NaN	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	NaN	F	Dollars

6675 rows × 15 columns

In [ ]:

```
###Removing the Duplicates###
```

In [98]:

```
import pandas as pd

g = pd.read_excel("C:\\\\Users\\\\synai\\\\Downloads\\\\Orders-With Nulls.xlsx", sheet_name='dfg')
dfg = pd.DataFrame(g)
```

In [99]:

```
dfg
```

Out[99]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Cus Se
0	3	2010-10-13	6	261.5400	Regular Air	-213.250	38.94	Muhammed MacIntyre	Buyer
1	6	2012-02-20	2	6.9300	Regular Air	-4.640	2.08	Ruben Dartt	Corporation
2	32	2011-07-15	26	2808.0800	Regular Air	1054.820	107.53	Liz Pelletier	Corporation
3	32	2011-07-15	24	1761.4000	Delivery Truck	-1748.560	70.89	Liz Pelletier	Corporation
4	32	2011-07-15	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Corporation
...	...	...	...	...	...	...	...	...	...
1002	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Corporation
1003	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Corporation
1004	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Corporation
1005	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Corporation
1006	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Buyer

1007 rows × 10 columns

In [101...]: `dfg.duplicated()#checking Duplicates#`

Out[101]:

```
0      False
1      False
2      False
3      False
4      False
...
1002   False
1003   False
1004   False
1005   False
1006   False
Length: 1007, dtype: bool
```

In [102...]: `dfg.drop_duplicates()`

Out[102]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Cus
0	3	2010-10-13	6	261.5400	Regular Air	-213.250	38.94	Muhammed MacIntyre	Bu
1	6	2012-02-20	2	6.9300	Regular Air	-4.640	2.08	Ruben Dartt	Cor
2	32	2011-07-15	26	2808.0800	Regular Air	1054.820	107.53	Liz Pelletier	Cor
3	32	2011-07-15	24	1761.4000	Delivery Truck	-1748.560	70.89	Liz Pelletier	Cor
4	32	2011-07-15	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Cor
...	...	...	...	...	...	...	...	...	...
1002	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Cor
1003	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Cor
1004	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Cor
1005	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Cor
1006	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Bu

1007 rows × 10 columns

In [ ]: `##Handling Missing Data##`

In [103...]: `df.fillna(80)`

Out[103]:

	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
<b>0</b>	BDCQ.SF1AA2CA	2016.06	1116.386	80	F	Dollars	6
<b>1</b>	BDCQ.SF1AA2CA	2016.09	1070.874	80	F	Dollars	6
<b>2</b>	BDCQ.SF1AA2CA	2016.12	1054.408	80	F	Dollars	6
<b>3</b>	BDCQ.SF1AA2CA	2017.03	1010.665	80	F	Dollars	6
<b>4</b>	BDCQ.SF1AA2CA	2017.06	1233.700	80	F	Dollars	6
...							
<b>6670</b>	BDCQ.SF8RSCA	2022.03	493.945	80	F	Dollars	6
<b>6671</b>	BDCQ.SF8RSCA	2022.06	579.955	80	F	Dollars	6
<b>6672</b>	BDCQ.SF8RSCA	2022.09	609.161	80	F	Dollars	6
<b>6673</b>	BDCQ.SF8RSCA	2022.12	518.615	80	F	Dollars	6

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude
6674	BDCQ.SF8RSCA	2023.03	663.630	80	F Dollars	6

6675 rows x 15 columns

In [104]: df.dropna()

Out[104]: Series\_reference Period Data\_value Suppressed STATUS UNITS Magnitude Subject

In [105]: df.dropna(inplace=True)###Permenant Remove###

In [106]: df

Out[106]: Series\_reference Period Data\_value Suppressed STATUS UNITS Magnitude Subject

In [107]: #####Data Filtering and Conditional Changes#####

In [109]: dfe

Out[109]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Cus Se
0	3	2010-10-13	6	261.5400	Regular Air	-213.250	38.94	Muhammed MacIntyre	Bu
1	6	2012-02-20	2	6.9300	Regular Air	-4.640	2.08	Ruben Dartt	Cor
2	32	2011-07-15	26	2808.0800	Regular Air	1054.820	107.53	Liz Pelletier	Cor
3	32	2011-07-15	24	1761.4000	Delivery Truck	-1748.560	70.89	Liz Pelletier	Cor
4	32	2011-07-15	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Cor
...	...	...	...	...	...	...	...	...	...
1002	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Cor
1003	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Cor
1004	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Cor
1005	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Cor
1006	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Bu

1007 rows × 10 columns

In [111...]

dfe.loc[dfe["Order Quantity"] &lt; 100]

Out[111]:

	Order ID	Order Date	Order Quantity	Sales	Ship Mode	Profit	Unit Price	Customer Name	Cus Se
0	3	2010-10-13	6	261.5400	Regular Air	-213.250	38.94	Muhammed MacIntyre	Bu
1	6	2012-02-20	2	6.9300	Regular Air	-4.640	2.08	Ruben Dartt	Cor
2	32	2011-07-15	26	2808.0800	Regular Air	1054.820	107.53	Liz Pelletier	Cor
3	32	2011-07-15	24	1761.4000	Delivery Truck	-1748.560	70.89	Liz Pelletier	Cor
4	32	2011-07-15	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Cor
...	...	...	...	...	...	...	...	...	...
1002	7171	2011-02-13	17	303.1865	Regular Air	92.592	20.99	Andy Gerbode	Cor
1003	7174	2012-03-10	10	141.9200	Regular Air	12.200	13.73	Thomas Thornton	Cor
1004	7175	2010-02-07	10	748.2500	Delivery Truck	-86.990	70.98	Helen Andreada	Cor
1005	7203	2009-01-08	25	21752.0100	Regular Air	9296.348	896.99	Ruben Dartt	Cor
1006	7239	2011-06-29	50	6206.1600	Regular Air	1416.270	120.33	Craig Carroll	Bu

1007 rows × 10 columns

In [ ]:

#####Export Data Frame#####

In [113...]

dfe.to\_excel("C:\\\\Users\\\\synai\\\\Downloads\\\\Modified\_Orders-With Nulls.xlsx", inc

In [ ]:

In [ ]: