Example -1:

- https://github.com/sanjeev-kallepalli/training.git

Pandas

- Installing Pandas using pip
- '!' mark runs the command in terminal.
- we are installing pandas library here by using pip.

```
!pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\hp\appdata\local\
programs\python\python310\lib\site-packages (2.2.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\hp\appdata\
local\programs\python\python310\lib\site-packages (from pandas)
(2024.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\hp\
appdata\roaming\python\python310\site-packages (from pandas)
(2.9.0.post0)
Requirement already satisfied: numpy>=1.22.4 in c:\users\hp\appdata\
local\programs\python\python310\lib\site-packages (from pandas)
(1.26.4)
Requirement already satisfied: tzdata>=2022.7 in c:\users\hp\appdata\
local\programs\python\python310\lib\site-packages (from pandas)
(2024.1)
Requirement already satisfied: six>=1.5 in c:\users\hp\appdata\
roaming\python\python310\site-packages (from python-dateutil>=2.8.2-
>pandas) (1.16.0)

WARNING: You are using pip version 22.0.4; however, version 24.3.1 is
available.
You should consider upgrading via the 'C:\Users\HP\AppData\Local\
Programs\Python\Python310\python.exe -m pip install --upgrade pip'
command.
```

```
# import
import pandas as pd
```

Step-1: Creating an "Empty DataFrame"

- How we are going to create an empty dataframe means by using 'pd.Dataframe()'
- Create a dataframe with "no columns" and "no rows"

```
df = pd.DataFrame()
print(df)
```

```
Empty DataFrame
Columns: []
Index: []
```

Structured data:

- Means it is having defined no. of columns and defined no. of rows.
- i,e Equal number of "rows" and equal number of "columns"

```
type(df)

pandas.core.frame.DataFrame
```

Step-2: Empty Dataframe with 3 columns names [Headings]

```
df = pd.DataFrame(columns=['Col1', 'Col2', 'Col3'])
print(df)

Empty DataFrame
Columns: [Col1, Col2, Col3]
Index: []

# In pandas No. of Columns = "pandas Series"
k = pd.Series([1, 2, 3])
type(k)

pandas.core.series.Series
```

Step-3: Inserting 'data' into the respective 'columns'

Dataframe with 3 columns and respective values/Data.

- Here we are inserting values/data in to 3 columns
- Whenever we are creating a dataframe we should have to keep in mind that "all the lists should be of equal size."
  - i,e Col1= 3 values, Col2= 3 values, Col3= 3 values
- If u give 2 values in any column means then it will through "error" bacause it is "structured data"
- Here we are inserting the values in "Key:Value" pairs only.
  - Ex: Col1:[1,2,3]
- Notice that even though it is the "dictinory", all the "values" are represented in "list"
- All the "keys" will be converted in to "column" names
  - Indexs will created by default
  - Index will always start with "0"

```
data = {'Col1': [1, 2, 3],
        'Col2': ['Raj', 'Sameera', 'Ketan'],
        'Col3': [20, 25, 22]}
df = pd.DataFrame(data=data)
print(df)
```

```
   Col1     Col2  Col3
0     1      Raj    20
1     2  Sameera    25
2     3    Ketan    22
```

```
type(df)
```

```
pandas.core.frame.DataFrame
```

[a] If u want to look at "1 column" in a data frame, then see below code:

- [] = means "no column name" will display in o/p
- [] = "pandas Series"

```
df['Col1']
```

```
0    1
1    2
2    3
Name: Col1, dtype: int64
```

```
df['Col3']
```

```
0    20
1    25
2    22
Name: Col3, dtype: int64
```

[b] Below is the dataframe as it contains 2 square brackets '[[]]' and contains "column name as Col3'

- [[]] = means "column name" will display in o/p
- [[]] = "Data Frame"

```
df[['Col3']]
```

```
   Col3
0    20
1    25
2    22
```

```
type(df['Col1'])
```

```
pandas.core.series.Series
```

Datatypes in pandas

- Every column has its own "data type"
  - [Q] How can we see all the data types at once?
  - Ans: By using "df.dtypes"

```
df.dtypes
```

```
Col1     int64
Col2    object
Col3     int64
dtype: object
```

Shape of dataframe

```
df.shape
```

```
(3, 3)
```

Modifying data

- [i] When the right side value is a static value i,e 1, then that value is applies to all rows.

```
df['Newcol'] = 1
df.head()
```

```
   Col1     Col2  Col3  Newcol
0     1      Raj    20       1
1     2  Sameera    25       1
2     3    Ketan    22       1
```

- [ii] When the right side value is not static value, make sure that the shapes match the lenght of list on right should be same as number of rows.

```
df['OneMoreCol'] = [1, 2, 3]
df.head()
```

```
   Col1     Col2  Col3  Newcol  OneMoreCol
0     1      Raj    20       1           1
1     2  Sameera    25       1           2
2     3    Ketan    22       1           3
```

Errors:

Error-1:

- "Commom error" which we will see in "pandas" is "ValueError: Length of values (4) does not match length of index (3)"

```
# what happens when the lenghts are not same.
df['OneMoreCol'] = [1, 2, 3, 4]
df.head()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[97], line 2
      1 # what happens when the lenghts are not same.
----> 2 df['OneMoreCol'] = [1, 2, 3, 4]
      3 df.head()

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\pandas\core\frame.py:4311, in DataFrame.__setitem__(self, key, value)
   4308     self._setitem_array([key], value)
   4309 else:
   4310     # set column
-> 4311     self._set_item(key, value)

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\pandas\core\frame.py:4524, in DataFrame._set_item(self, key, value)
   4514 def _set_item(self, key, value) -> None:
   4515     """
   4516     Add series to DataFrame in specified column.
   4517
   (...)
   4522     ensure homogeneity.
   4523     """
-> 4524     value, refs = self._sanitize_column(value)
   4526     if (
   4527         key in self.columns
   4528         and value.ndim == 1
   4529         and not isinstance(value.dtype, ExtensionDtype)
   4530     ):
   4531         # broadcast across multiple columns if necessary
   4532         if not self.columns.is_unique or isinstance(self.columns, MultiIndex):

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\pandas\core\frame.py:5266, in DataFrame._sanitize_column(self, value)
   5263     return _reindex_for_setitem(value, self.index)
   5265 if is_list_like(value):
-> 5266     com.require_length_match(value, self.index)
   5267 arr = sanitize_array(value, self.index, copy=True, allow_2d=True)
   5268 if (
   5269     isinstance(value, Index)
   5270     and value.dtype == "object"
   (...)
   5273     # TODO: Remove kludge in sanitize_array for string mode
```

```
when enforcing
   5274     # this deprecation

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\common.py:573, in require_length_match(data,
index)
    569 """
    570 Check the length of data matches the length of the index.
    571 """
    572 if len(data) != len(index):
--> 573     raise ValueError(
    574         "Length of values "
    575         f"({len(data)}) "
    576         "does not match length of index "
    577         f"({len(index)})"
    578     )

ValueError: Length of values (4) does not match length of index (3)
```

```
# Check size of dataframe
df.shape
```

```
(3, 5)
```

Step-4: Reading a "csv file"

- write a program, which accepts the "path of csv file" and return a pandas dataframe.
- Now here we are using "csv file", but later on we will see "pickle file", "JSON lip file", "Xl file" etc. through which u can store "Data Frame"
- [1] Writing a function

```
def read_file_frm_source(pth, ext):
    if ext == 'csv':
        return pd.read_csv(pth)
    elif ext == 'excel':
        return pd.read_excel(pth)

# note \ is an escape character and we will have to replace them
with / or \\
ibm_hr = read_file_frm_source("data/ibm_hr.csv", ext='csv')
```

[or]

- [2] Without writing a function

```
ibm_hr = pd.read_csv("data/ibm_hr.csv")

type(ibm_hr)

pandas.core.frame.DataFrame
```

```
# To see the shape of dataframe
# Shape shows how many records and columns are present in dataframe.
ibm_hr.shape
```

```
(1470, 35)
```

```
# To see dataframe top 5 records
ibm_hr.head()
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                  8                      0               1
6
1                 10                      3               3
10
2                  7                      3               3
0
3                  8                      3               3
```

```
8
4                   6                        3                3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

```
ibm_hr.head(10)
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development

5   32        No  Travel_Frequently       1005  Research & Development

6   59        No      Travel_Rarely       1324  Research & Development

7   30        No      Travel_Rarely       1358  Research & Development

8   38        No  Travel_Frequently        216  Research & Development

9   36        No      Travel_Rarely       1299  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7
5                 2          2  Life Sciences              1
```

```
8
6                  3           3        Medical                1
10
7                 24           1  Life Sciences                1
11
8                 23           3  Life Sciences                1
12
9                 27           3        Medical                1
13

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1
5  ...                         3            80                 0
6  ...                         1            80                 3
7  ...                         2            80                 1
8  ...                         2            80                 0
9  ...                         2            80                 2

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                  8                      0                1
6
1                 10                      3                3
10
2                  7                      3                3
0
3                  8                      3                3
8
4                  6                      3                3
2
5                  8                      2                2
7
6                 12                      3                2
1
7                  1                      2                3
1
8                 10                      2                3
9
9                 17                      3                2
7

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                    5
1                  7                        1                    7
2                  0                        0                    0
3                  7                        3                    0
```

```
4                    2                         2                         2
5                    7                         3                         6
6                    0                         0                         0
7                    0                         0                         0
8                    7                         1                         8
9                    7                         7                         7

[10 rows x 35 columns]
```

```
# Too see all columns
ibm_hr.columns
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate',
'Department',
       'DistanceFromHome', 'Education', 'EducationField',
'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender',
'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours',
'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear',
'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole',
'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

```
# To see indexes of the dataframe
# To see name of record
ibm_hr.index
```

```
RangeIndex(start=0, stop=1470, step=1)
```

```
[x for x in range(0, 10, 1)]
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

```
[x for x in range(0, 10, 2)]
```

```
[0, 2, 4, 6, 8]
```

random:-

- We will get "1" random number.
- Every time it will give some different number as o/p in between the 2 numbers (i,e 100 - 999)

```
import random
random.randint(100, 999)
```

```
543
```

```
ibm_hr.shape[0]
```

```
1470
```

To rename the index:

- "Index number" need not be just continuous number.
- We no need to change it, but in some cases if we want to change means, we can change like below.
- Generally it is not required to change "index" name all times.
  - Ex: present range = (1, 1470)----------index = (0, 1469)
  - Ex" Now changed to range = (1, 10000)----------index = (0,9999)

```
import random
ibm_hr.index = [random.randint(1, 10000) for i in
range(ibm_hr.shape[0])]
ibm_hr.head()
```

```
      Age Attrition     BusinessTravel  DailyRate
Department  \
4631   41       Yes      Travel_Rarely       1102
Sales
4977   49        No  Travel_Frequently        279  Research &
Development
221    37       Yes      Travel_Rarely       1373  Research &
Development
4695   33        No  Travel_Frequently       1392  Research &
Development
8358   27        No      Travel_Rarely        591  Research &
Development

      DistanceFromHome  Education EducationField  EmployeeCount  \
4631                 1          2  Life Sciences              1
4977                 8          1  Life Sciences              1
221                  2          2          Other              1
4695                 3          4  Life Sciences              1
8358                 2          1        Medical              1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
4631               1  ...                         1            80
4977               2  ...                         4            80
221                4  ...                         2            80
4695               5  ...                         3            80
8358               7  ...                         4            80
```

```
      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
4631                 0                  8                      0
4977                 1                 10                      3
221                  0                  7                      3
4695                 0                  8                      3
8358                 1                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
4631               1               6                  4
4977               3              10                  7
221                3               0                  0
4695               3               8                  7
8358               3               2                  2

      YearsSinceLastPromotion  YearsWithCurrManager
4631                        0                     5
4977                        1                     7
221                         0                     0
4695                        3                     0
8358                        2                     2

[5 rows x 35 columns]
```

On the other hand, we can also set "one of the column" as "index":

- Here we had taken "EducationField" column and used as "index"
- Why we are doing this "index changing" means, in case if we want to do any "future subsetting" in future.
  - Note: "Index" can also be a "categorical datatype".

    "drop=True":
- We used this because, what ever index is there avalable previously i donot want to store it.
- I want to just drop it, so we use drop=True.
  - Ex: previous index = 6824, 5348, 6927, 1210, 1037
  - Ex: Current index = Life Sciences, Life Sciences, Other, Life Sciences, Medical

```
ibm_hr2 = ibm_hr.set_index('EducationField', drop=True)
ibm_hr2.head()
```

```
                 Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences     41       Yes      Travel_Rarely       1102
Life Sciences     49        No  Travel_Frequently        279
Other             37       Yes      Travel_Rarely       1373
Life Sciences     33        No  Travel_Frequently       1392
Medical           27        No      Travel_Rarely        591

                            Department  DistanceFromHome  Education  \
```

```
EducationField
Life Sciences                      Sales                 1          2
Life Sciences   Research & Development                   8          1
Other           Research & Development                   2          2
Life Sciences   Research & Development                   3          4
Medical         Research & Development                   2          1

                EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences               1               1                        2
...
Life Sciences               1               2                        3
...
Other                       1               4                        4
...
Life Sciences               1               5                        4
...
Medical                     1               7                        1
...

               RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                         1             80
0
Life Sciences                         4             80
1
Other                                 2             80
0
Life Sciences                         3             80
0
Medical                               4             80
1

                TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance  \
EducationField

Life Sciences                    8                    0
1
Life Sciences                   10                    3
3
Other                            7                    3
3
Life Sciences                    8                    3
3
Medical                          6                    3
```

```
3

               YearsAtCompany  YearsInCurrentRole
YearsSinceLastPromotion  \
EducationField

Life Sciences                6                   4
0
Life Sciences               10                   7
1
Other                        0                   0
0
Life Sciences                8                   7
3
Medical                      2                   2
2

                YearsWithCurrManager
EducationField
Life Sciences                      5
Life Sciences                      7
Other                              0
Life Sciences                      0
Medical                            2

[5 rows x 34 columns]

# To see list of all columns
ibm_hr.columns

Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate',
'Department',
       'DistanceFromHome', 'Education', 'EducationField',
'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender',
'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours',
'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear',
'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole',
'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

To rename all columns:

- If u put empty list i,e [] then it will match with number of columns
- Now i am not doing this changes.

```
# ibm_hr.columns = []
```

To see all "data types"

```
ibm_hr.dtypes
```

```
Age                         int64
Attrition                  object
BusinessTravel             object
DailyRate                   int64
Department                 object
DistanceFromHome            int64
Education                   int64
EducationField             object
EmployeeCount               int64
EmployeeNumber              int64
EnvironmentSatisfaction     int64
Gender                     object
HourlyRate                  int64
JobInvolvement              int64
JobLevel                    int64
JobRole                    object
JobSatisfaction             int64
MaritalStatus              object
MonthlyIncome               int64
MonthlyRate                 int64
NumCompaniesWorked          int64
Over18                     object
OverTime                   object
PercentSalaryHike           int64
PerformanceRating           int64
RelationshipSatisfaction    int64
StandardHours               int64
StockOptionLevel            int64
TotalWorkingYears           int64
TrainingTimesLastYear       int64
WorkLifeBalance             int64
YearsAtCompany              int64
YearsInCurrentRole          int64
YearsSinceLastPromotion     int64
YearsWithCurrManager        int64
dtype: object
```

Subsetting will be done by below methods:

[1] Subsetting of data based on "slicing"

- slicing

[2] Subsetting of data based on "index"

- index

[3] Subsetting of data based on "value" (or) "column condition"

- value

[1] Subsetting of data based on "slicing"

- Slicing:
  - We have a big "data frame" and if u want to take a "chunk of it" means
  - Fetching only first 2 records by using "slicing" operator with out considering the index.

```
ibm_hr[:2]
```

```
      Age Attrition     BusinessTravel  DailyRate
Department  \
4631   41       Yes      Travel_Rarely       1102
Sales
4977   49        No  Travel_Frequently        279  Research &
Development

      DistanceFromHome  Education EducationField  EmployeeCount  \
4631                 1          2  Life Sciences              1
4977                 8          1  Life Sciences              1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
4631               1  ...                         1            80
4977               2  ...                         4            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
4631                 0                  8                      0
4977                 1                 10                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
4631               1               6                  4
4977               3              10                  7

      YearsSinceLastPromotion  YearsWithCurrManager
4631                        0                     5
4977                        1                     7

[2 rows x 35 columns]
```

```
ibm_hr[2]
```

```
---------------------------------------------------------------------
-----
KeyError                                  Traceback (most recent call
last)
File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:3805, in Index.get_loc(self, key)
   3804 try:
-> 3805     return self._engine.get_loc(casted_key)
   3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

File pandas\\_libs\\hashtable_class_helper.pxi:7081, in
pandas._libs.hashtable.PyObjectHashTable.get_item()

File pandas\\_libs\\hashtable_class_helper.pxi:7089, in
pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 2

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call
last)
Cell In[117], line 1
----> 1 ibm_hr[2]

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\frame.py:4102, in DataFrame.__getitem__(self,
key)
   4100 if self.columns.nlevels > 1:
   4101     return self._getitem_multilevel(key)
-> 4102 indexer = self.columns.get_loc(key)
   4103 if is_integer(indexer):
   4104     indexer = [indexer]

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:3812, in Index.get_loc(self, key)
   3807     if isinstance(casted_key, slice) or (
   3808         isinstance(casted_key, abc.Iterable)
   3809         and any(isinstance(x, slice) for x in casted_key)
   3810     ):
   3811         raise InvalidIndexError(key)
-> 3812     raise KeyError(key) from err
   3813 except TypeError:
   3814     # If we have a listlike key, _check_indexing_error will
raise
   3815     #  InvalidIndexError. Otherwise we fall through and re-
raise
```

```
   3816     #  the TypeError.
   3817     self._check_indexing_error(key)

KeyError: 2
```

[2] [A] Subsetting of data based on "index"

- .iloc:
  - Usage of ".loc" to subset matching rows.
  - It is done based on "index"
  - Here "index" we had given is "Life sciences"

```
ibm_hr2.loc['Life Sciences']
```

```
                 Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences     41       Yes      Travel_Rarely       1102
Life Sciences     49        No  Travel_Frequently        279
Life Sciences     33        No  Travel_Frequently       1392
Life Sciences     32        No  Travel_Frequently       1005
Life Sciences     30        No      Travel_Rarely       1358
...              ...       ...                ...        ...
Life Sciences     45        No      Travel_Rarely        374
Life Sciences     40        No      Travel_Rarely       1322
Life Sciences     35        No  Travel_Frequently       1199
Life Sciences     35        No      Travel_Rarely        287
Life Sciences     27        No      Travel_Rarely        155

                            Department  DistanceFromHome  Education  \
EducationField
Life Sciences                    Sales                 1          2
Life Sciences   Research & Development                 8          1
Life Sciences   Research & Development                 3          4
Life Sciences   Research & Development                 2          2
Life Sciences   Research & Development                24          1
...                                ...               ...        ...
Life Sciences                    Sales                20          3
Life Sciences   Research & Development                 2          4
Life Sciences   Research & Development                18          4
Life Sciences   Research & Development                 1          4
Life Sciences   Research & Development                 4          3

                 EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences                1               1                        2
...
Life Sciences                1               2                        3
...
```

```
Life Sciences                1              5                        4
...
Life Sciences                1              8                        4
...
Life Sciences                1             11                        4
...
...                        ...            ...                      ...
...
Life Sciences                1           2046                        4
...
Life Sciences                1           2048                        3
...
Life Sciences                1           2049                        3
...
Life Sciences                1           2052                        3
...
Life Sciences                1           2064                        2
...

                RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                          1             80
0
Life Sciences                          4             80
1
Life Sciences                          3             80
0
Life Sciences                          3             80
0
Life Sciences                          2             80
1
...                                  ...            ...                .
..
Life Sciences                          3             80
0
Life Sciences                          4             80
0
Life Sciences                          4             80
2
Life Sciences                          4             80
1
Life Sciences                          2             80
1

                 TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance  \
EducationField
```

```
Life Sciences                    8                     0
1
Life Sciences                   10                     3
3
Life Sciences                    8                     3
3
Life Sciences                    8                     2
2
Life Sciences                    1                     2
3
...                            ...                   ...                .
..
Life Sciences                    8                     3
3
Life Sciences                    8                     2
3
Life Sciences                   10                     2
4
Life Sciences                    4                     5
3
Life Sciences                    6                     0
3

                YearsAtCompany  YearsInCurrentRole
YearsSinceLastPromotion  \
EducationField

Life Sciences                6                     4
0
Life Sciences               10                     7
1
Life Sciences                8                     7
3
Life Sciences                7                     7
3
Life Sciences                1                     0
0
...                        ...                   ...
...
Life Sciences                5                     3
0
Life Sciences                2                     2
2
Life Sciences               10                     2
0
Life Sciences                4                     3
1
Life Sciences                6                     2
0
```

```
                 YearsWithCurrManager
EducationField
Life Sciences                       5
Life Sciences                       7
Life Sciences                       0
Life Sciences                       6
Life Sciences                       0
...                               ...
Life Sciences                       1
Life Sciences                       2
Life Sciences                       2
Life Sciences                       1
Life Sciences                       3

[606 rows x 34 columns]
```

```
ibm_hr2.loc['Life Sciences'].shape
```

```
(606, 34)
```

[3] [A] Subsetting of data based on "value" (or) "column condition"

- .iloc:
  - We can also subset based on "value" (or) "column condition"
    - With "1 column condition"
      - i,e ['Department']=='Sales']

```
ibm_hr2[ibm_hr2['Department']=='Sales']
```

```
                 Age Attrition     BusinessTravel  DailyRate Department
\
EducationField

Life Sciences     41       Yes      Travel_Rarely       1102      Sales

Life Sciences     53        No      Travel_Rarely       1219      Sales

Life Sciences     36       Yes      Travel_Rarely       1218      Sales

Marketing         42        No      Travel_Rarely        691      Sales

Marketing         46        No      Travel_Rarely        705      Sales

...              ...       ...                ...        ...        ...

Life Sciences     45        No      Travel_Rarely        374      Sales

Marketing         50       Yes      Travel_Rarely        410      Sales

Marketing         39        No      Travel_Rarely        722      Sales
```

```
Other             26         No      Travel_Rarely        1167       Sales

Medical           49         No  Travel_Frequently        1023       Sales


                  DistanceFromHome  Education  EmployeeCount
EmployeeNumber  \
EducationField

Life Sciences                    1          2              1
1
Life Sciences                    2          4              1
23
Life Sciences                    9          4              1
27
Marketing                        8          4              1
35
Marketing                        2          4              1
38
...                            ...        ...            ...
...
Life Sciences                   20          3              1
2046
Marketing                       28          3              1
2055
Marketing                       24          1              1
2056
Other                            5          3              1
2060
Medical                          2          3              1
2065

                  EnvironmentSatisfaction  ... RelationshipSatisfaction
\
EducationField                             ...

Life Sciences                           2  ...                         1

Life Sciences                           1  ...                         3

Life Sciences                           3  ...                         2

Marketing                               3  ...                         4

Marketing                               2  ...                         4

...                                   ...  ...                       ...

Life Sciences                           4  ...                         3
```

```
Marketing                              4  ...                          2

Marketing                              2  ...                          1

Other                                  4  ...                          4

Medical                                4  ...                          4


                 StandardHours  StockOptionLevel  TotalWorkingYears  \
EducationField
Life Sciences               80                 0                  8
Life Sciences               80                 0                 31
Life Sciences               80                 0                 10
Marketing                   80                 1                 10
Marketing                   80                 0                 22
...                        ...               ...                ...
Life Sciences               80                 0                  8
Marketing                   80                 1                 20
Marketing                   80                 1                 21
Other                       80                 0                  5
Medical                     80                 0                 17

                TrainingTimesLastYear  WorkLifeBalance
YearsAtCompany  \
EducationField

Life Sciences                       0                1              6

Life Sciences                       3                3             25

Life Sciences                       4                3              5

Marketing                           2                3              9

Marketing                           2                2              2

...                               ...              ...            ...

Life Sciences                       3                3              5

Marketing                           3                3              3

Marketing                           2                2             20

Other                               2                3              4

Medical                             3                2              9


                 YearsInCurrentRole  YearsSinceLastPromotion  \
```

```
EducationField
Life Sciences                     4                          0
Life Sciences                     8                          3
Life Sciences                     3                          0
Marketing                         7                          4
Marketing                         2                          2
...                             ...                        ...
Life Sciences                     3                          0
Marketing                         2                          2
Marketing                         9                          9
Other                             2                          0
Medical                           6                          0

                YearsWithCurrManager
EducationField
Life Sciences                      5
Life Sciences                      7
Life Sciences                      3
Marketing                          2
Marketing                          1
...                              ...
Life Sciences                      1
Marketing                          0
Marketing                          6
Other                              0
Medical                            8

[446 rows x 34 columns]
```

[3] [B] Subsetting of data based on "multiple values" (or) "multiple column condition"

- .isin:
  - We can also subset based on "multiple values" (or) "multiple column condition"
    - With "multiple column condition"
      - i,e 'Department'].isin(['Sales', 'Research & Development']

```
ibm_hr2[ibm_hr2['Department'].isin(['Sales', 'Research &
Development'])]
```

```
                Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences    41       Yes      Travel_Rarely       1102
Life Sciences    49        No  Travel_Frequently        279
Other            37       Yes      Travel_Rarely       1373
Life Sciences    33        No  Travel_Frequently       1392
Medical          27        No      Travel_Rarely        591
...             ...       ...                 ...        ...
Medical          36        No  Travel_Frequently        884
Medical          39        No      Travel_Rarely        613
```

```
Life Sciences    27        No      Travel_Rarely        155
Medical          49        No  Travel_Frequently       1023
Medical          34        No      Travel_Rarely        628

                             Department  DistanceFromHome  Education  \
EducationField
Life Sciences                     Sales                 1          2
Life Sciences   Research & Development                 8          1
Other           Research & Development                 2          2
Life Sciences   Research & Development                 3          4
Medical         Research & Development                 2          1
...                                ...               ...        ...
Medical         Research & Development                23          2
Medical         Research & Development                 6          1
Life Sciences   Research & Development                 4          3
Medical                           Sales                 2          3
Medical         Research & Development                 8          3

                EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences               1               1                        2
...
Life Sciences               1               2                        3
...
Other                       1               4                        4
...
Life Sciences               1               5                        4
...
Medical                     1               7                        1
...
...                       ...             ...                      ...
...
Medical                     1            2061                        3
...
Medical                     1            2062                        4
...
Life Sciences               1            2064                        2
...
Medical                     1            2065                        4
...
Medical                     1            2068                        2
...

               RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                         1             80
```

```
0
Life Sciences                            4              80
1
Other                                    2              80
0
Life Sciences                            3              80
0
Medical                                  4              80
1
...                                    ...             ...                .
..
Medical                                  3              80
1
Medical                                  1              80
1
Life Sciences                            2              80
1
Medical                                  4              80
0
Medical                                  1              80
0

                  TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance  \
EducationField

Life Sciences                     8                     0
1
Life Sciences                    10                     3
3
Other                             7                     3
3
Life Sciences                     8                     3
3
Medical                           6                     3
3
...                             ...                   ...                .
..
Medical                          17                     3
3
Medical                           9                     5
3
Life Sciences                     6                     0
3
Medical                          17                     3
2
Medical                           6                     3
4

                YearsAtCompany  YearsInCurrentRole
```

```
YearsSinceLastPromotion  \
EducationField

Life Sciences                6                    4
0
Life Sciences               10                    7
1
Other                        0                    0
0
Life Sciences                8                    7
3
Medical                      2                    2
2
...                        ...                  ...
...
Medical                      5                    2
0
Medical                      7                    7
1
Life Sciences                6                    2
0
Medical                      9                    6
0
Medical                      4                    3
1

                YearsWithCurrManager
EducationField
Life Sciences                      5
Life Sciences                      7
Other                              0
Life Sciences                      0
Medical                            2
...                              ...
Medical                            3
Medical                            7
Life Sciences                      3
Medical                            8
Medical                            2

[1407 rows x 34 columns]
```

[3] [C] Subsetting of data based on "multiple values" (or) "multiple column condition"

- Tilda ("~"):
  - For Not executing multiple conditions

```
ibm_hr2[~ibm_hr2['Department'].isin(['Sales', 'Research &
Development'])]
```

```
                  Age Attrition     BusinessTravel  DailyRate
Department  \
EducationField

Medical            46        No      Travel_Rarely        945  Human
Resources
Human Resources    37       Yes      Travel_Rarely        807  Human
Resources
Human Resources    59        No         Non-Travel       1420  Human
Resources
Human Resources    54        No         Non-Travel        142  Human
Resources
Life Sciences      26        No      Travel_Rarely       1355  Human
Resources
...               ...       ...                ...        ...
...
Human Resources    27       Yes  Travel_Frequently       1337  Human
Resources
Other              38        No  Travel_Frequently       1444  Human
Resources
Human Resources    55        No      Travel_Rarely        189  Human
Resources
Human Resources    25        No      Travel_Rarely        309  Human
Resources
Life Sciences      35        No      Travel_Rarely       1146  Human
Resources

                  DistanceFromHome  Education  EmployeeCount
EmployeeNumber  \
EducationField

Medical                          5          2              1
103
Human Resources                  6          4              1
133
Human Resources                  2          4              1
140
Human Resources                 26          3              1
148
Life Sciences                   25          1              1
177
...                            ...        ...            ...
...
Human Resources                 22          3              1
1944
Other                            1          4              1
1972
Human Resources                 26          4              1
1973
Human Resources                  2          3              1
```

```
1987
Life Sciences                    26          4               1
2040

                  EnvironmentSatisfaction  ... RelationshipSatisfaction
\
EducationField                             ...

Medical                                  2  ...                         4

Human Resources                          3  ...                         4

Human Resources                          3  ...                         4

Human Resources                          4  ...                         3

Life Sciences                            3  ...                         4

...                                    ...  ...                       ...

Human Resources                          1  ...                         1

Other                                    4  ...                         2

Human Resources                          3  ...                         1

Human Resources                          3  ...                         3

Life Sciences                            3  ...                         3


                  StandardHours  StockOptionLevel  TotalWorkingYears  \
EducationField
Medical                      80                1                 16
Human Resources              80                0                  7
Human Resources              80                1                 30
Human Resources              80                0                 23
Life Sciences                80                1                  8
...                         ...              ...                ...
Human Resources              80                0                  1
Other                        80                1                  7
Human Resources              80                1                 35
Human Resources              80                0                  6
Life Sciences                80                0                  9

                 TrainingTimesLastYear  WorkLifeBalance YearsAtCompany
\
EducationField

Medical                              2                3              4
```

```
Human Resources                          3                  3                 3

Human Resources                          3                  3                 3

Human Resources                          3                  3                 5

Life Sciences                            3                  3                 8

...                                    ...                ...               ...

Human Resources                          2                  3                 1

Other                                    2                  3                 6

Human Resources                          0                  3                10

Human Resources                          3                  3                 2

Life Sciences                            2                  3                 9


                   YearsInCurrentRole  YearsSinceLastPromotion  \
EducationField
Medical                             2                        0
Human Resources                     2                        0
Human Resources                     2                        2
Human Resources                     3                        4
Life Sciences                       7                        5
...                               ...                      ...
Human Resources                     0                        0
Other                               2                        1
Human Resources                     9                        1
Human Resources                     0                        1
Life Sciences                       0                        1

                   YearsWithCurrManager
EducationField
Medical                               2
Human Resources                       2
Human Resources                       2
Human Resources                       4
Life Sciences                         7
...                                 ...
Human Resources                       0
Other                                 2
Human Resources                       4
Human Resources                       2
Life Sciences                         7

[63 rows x 34 columns]
```

[2] [B] Subsetting of data based on "multiple indexs" for "multiple values"

- .iloc:
  - we could also subset "multiple indexs" using ".loc"
  - lov = List Of Values

```
lov = ['Life Sciences', 'Marketing']
ibm_hr2.loc[lov]
```

```
                 Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences     41       Yes      Travel_Rarely       1102
Life Sciences     49        No  Travel_Frequently        279
Life Sciences     33        No  Travel_Frequently       1392
Life Sciences     32        No  Travel_Frequently       1005
Life Sciences     30        No      Travel_Rarely       1358
...              ...       ...                ...        ...
Marketing         34        No      Travel_Rarely        704
Marketing         36        No         Non-Travel        301
Marketing         36        No      Travel_Rarely       1120
Marketing         50       Yes      Travel_Rarely        410
Marketing         39        No      Travel_Rarely        722

                             Department  DistanceFromHome  Education  \
EducationField
Life Sciences                     Sales                 1          2
Life Sciences   Research & Development                  8          1
Life Sciences   Research & Development                  3          4
Life Sciences   Research & Development                  2          2
Life Sciences   Research & Development                 24          1
...                                ...               ...        ...
Marketing                         Sales                28          3
Marketing                         Sales                15          4
Marketing                         Sales                11          4
Marketing                         Sales                28          3
Marketing                         Sales                24          1

                 EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences                1               1                        2
...
Life Sciences                1               2                        3
...
Life Sciences                1               5                        4
...
Life Sciences                1               8                        4
...
Life Sciences                1              11                        4
...
```

```
...                          ...              ...                          ...
...
Marketing                      1             2035                           4
...
Marketing                      1             2036                           4
...
Marketing                      1             2045                           2
...
Marketing                      1             2055                           4
...
Marketing                      1             2056                           2
...

                RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                          1              80
0
Life Sciences                          4              80
1
Life Sciences                          3              80
0
Life Sciences                          3              80
0
Life Sciences                          2              80
1
...                                  ...             ...                 .
..
Marketing                              4              80
2
Marketing                              1              80
1
Marketing                              1              80
1
Marketing                              2              80
1
Marketing                              1              80
1

                 TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance  \
EducationField

Life Sciences                    8                     0
1
Life Sciences                   10                     3
3
Life Sciences                    8                     3
3
```

```
Life Sciences                    8                       2
2
Life Sciences                    1                       2
3
...                            ...                     ...                 .
..
Marketing                        8                       2
3
Marketing                       15                       4
2
Marketing                        8                       2
2
Marketing                       20                       3
3
Marketing                       21                       2
2

                YearsAtCompany  YearsInCurrentRole
YearsSinceLastPromotion  \
EducationField

Life Sciences                    6                       4
0
Life Sciences                   10                       7
1
Life Sciences                    8                       7
3
Life Sciences                    7                       7
3
Life Sciences                    1                       0
0
...                            ...                     ...
...
Marketing                        8                       7
1
Marketing                       15                      12
11
Marketing                        6                       3
0
Marketing                        3                       2
2
Marketing                       20                       9
9

                 YearsWithCurrManager
EducationField
Life Sciences                       5
Life Sciences                       7
Life Sciences                       0
Life Sciences                       6
```

```
Life Sciences                          0
...                                  ...
Marketing                              7
Marketing                             11
Marketing                              0
Marketing                              0
Marketing                              6

[765 rows x 34 columns]
```

[4] [A] Subsetting of "indices" and "columns"

- Syntax:
  - ibm_hr2.loc[indexs , columns]
  - ibm_hr2.loc['Life Sciences', 'Age']
- For only "1 index" & "1 column"

```
ibm_hr2.loc['Life Sciences', 'Age']

EducationField
Life Sciences    41
Life Sciences    49
Life Sciences    33
Life Sciences    32
Life Sciences    30
                 ..
Life Sciences    45
Life Sciences    40
Life Sciences    35
Life Sciences    35
Life Sciences    27
Name: Age, Length: 606, dtype: int64
```

[4] [B] Subsetting of "multiple indices" and "multiple columns"

- Syntax:
  - ibm_hr2.loc[indexs , columns]
  - ibm_hr2.loc[['Life Sciences', 'Marketing'], ['Age', 'Attrition']]
- For only "multiple indexs" & "multiple columns"
- Here we are sending "multiple indexs" & "multiple columns" in "List" format

```
ibm_hr2.loc[['Life Sciences', 'Marketing'], ['Age', 'Attrition']]

                 Age Attrition
EducationField
Life Sciences    41       Yes
Life Sciences    49        No
Life Sciences    33        No
Life Sciences    32        No
Life Sciences    30        No
```

```
...              ...       ...
Marketing         34        No
Marketing         36        No
Marketing         36        No
Marketing         50       Yes
Marketing         39        No

[765 rows x 2 columns]
```

To see all rows and columns using ".loc"

```
ibm_hr2.loc[:, :]
```

```
                 Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences     41       Yes      Travel_Rarely       1102
Life Sciences     49        No  Travel_Frequently        279
Other             37       Yes      Travel_Rarely       1373
Life Sciences     33        No  Travel_Frequently       1392
Medical           27        No      Travel_Rarely        591
...              ...       ...                ...        ...
Medical           36        No  Travel_Frequently        884
Medical           39        No      Travel_Rarely        613
Life Sciences     27        No      Travel_Rarely        155
Medical           49        No  Travel_Frequently       1023
Medical           34        No      Travel_Rarely        628

                            Department  DistanceFromHome  Education  \
EducationField
Life Sciences                    Sales                 1          2
Life Sciences   Research & Development                 8          1
Other           Research & Development                 2          2
Life Sciences   Research & Development                 3          4
Medical         Research & Development                 2          1
...                                ...               ...        ...
Medical         Research & Development                23          2
Medical         Research & Development                 6          1
Life Sciences   Research & Development                 4          3
Medical                          Sales                 2          3
Medical         Research & Development                 8          3

                 EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences                1               1                        2
...
Life Sciences                1               2                        3
...
Other                        1               4                        4
```

```
...
Life Sciences                1              5                          4
...
Medical                      1              7                          1
...
...                        ...            ...                        ...
...
Medical                      1           2061                          3
...
Medical                      1           2062                          4
...
Life Sciences                1           2064                          2
...
Medical                      1           2065                          4
...
Medical                      1           2068                          2
...

                RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                          1             80
0
Life Sciences                          4             80
1
Other                                  2             80
0
Life Sciences                          3             80
0
Medical                                4             80
1
...                                  ...            ...                .
..
Medical                                3             80
1
Medical                                1             80
1
Life Sciences                          2             80
1
Medical                                4             80
0
Medical                                1             80
0

                 TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance  \
EducationField

Life Sciences                    8                     0
```

```
1
Life Sciences                    10                       3
3
Other                             7                       3
3
Life Sciences                     8                       3
3
Medical                           6                       3
3
...                             ...                     ...                .
..
Medical                          17                       3
3
Medical                           9                       5
3
Life Sciences                     6                       0
3
Medical                          17                       3
2
Medical                           6                       3
4

                 YearsAtCompany  YearsInCurrentRole
YearsSinceLastPromotion  \
EducationField

Life Sciences                6                       4
0
Life Sciences               10                       7
1
Other                        0                       0
0
Life Sciences                8                       7
3
Medical                      2                       2
2
...                        ...                     ...
...
Medical                      5                       2
0
Medical                      7                       7
1
Life Sciences                6                       2
0
Medical                      9                       6
0
Medical                      4                       3
1

                 YearsWithCurrManager
```

```
EducationField
Life Sciences                        5
Life Sciences                        7
Other                                0
Life Sciences                        0
Medical                              2
...                                ...
Medical                              3
Medical                              7
Life Sciences                        3
Medical                              8
Medical                              2

[1470 rows x 34 columns]
```

```
ibm_hr2.loc[:,  ['Age', 'Attrition']]
```

```
                 Age Attrition
EducationField
Life Sciences     41       Yes
Life Sciences     49        No
Other             37       Yes
Life Sciences     33        No
Medical           27        No
...              ...       ...
Medical           36        No
Medical           39        No
Life Sciences     27        No
Medical           49        No
Medical           34        No

[1470 rows x 2 columns]
```

Example -2:

```
temp_df = pd.DataFrame({
    'animal' : ['cat', 'dog', 'sheep', 'goat', 'cow'],
    'count'  : ['3', '5', '10', '15', '10'],
    'location'  : [ 'home',  'home',  'farm',  'farm',  'farm'],
    'id'     : ['I101', 'I102', 'I103', 'I104', 'I105']
})
temp_df
```

```
  animal count location    id
0    cat     3     home  I101
1    dog     5     home  I102
2  sheep    10     farm  I103
```

```
3   goat    15     farm  I104
4    cow    10     farm  I105
```

```
# all columns all rows
temp_df.loc[:, :]
```

```
  animal count location    id
0    cat     3     home  I101
1    dog     5     home  I102
2  sheep    10     farm  I103
3   goat    15     farm  I104
4    cow    10     farm  I105
```

```
# [[]]
temp_df.loc[[2, 4]]
```

```
  animal count location    id
2  sheep    10     farm  I103
4    cow    10     farm  I105
```

[or]

```
# index 2 and 4 rows with all columns
temp_df.loc[[2, 4], :]
```

```
  animal count location    id
2  sheep    10     farm  I103
4    cow    10     farm  I105
```

```
# index 1 and 3 rows with few columns
temp_df.loc[[1, 3], ['animal', 'id']]
```

```
  animal    id
1    dog  I102
3   goat  I104
```

```
# we could also use slicing as below
temp_df.loc[2:4]
```

```
  animal count location    id
2  sheep    10     farm  I103
3   goat    15     farm  I104
4    cow    10     farm  I105
```

NOTE:

- .loc:
  - It uses "index"
  - Ex: [2:4] means 2,3,4
- .iloc:
  - It wonot uses "index"

- Ex: [2:4] means 2,3
- .iloc refers to "Row count"

```
temp_df.iloc[2:4]
```

```
  animal count location    id
2  sheep    10     farm  I103
3   goat    15     farm  I104
```

```
# important note -- notice the difference between ".loc" and ".iloc"
in below scenario
# let us sort the dataframe based on "animal" column
# It will sort in "alphabetical" manner

temp_df.sort_values(by=['animal'], ascending=True, inplace=True)
temp_df.head()
```

```
  animal count location    id
0    cat     3     home  I101
4    cow    10     farm  I105
1    dog     5     home  I102
3   goat    15     farm  I104
2  sheep    10     farm  I103
```

```
# this still fetches the same set of records even after sorting.
# loc actually looks for values 0 to 4 in index and fetches them.
# once the 4th record is fetched, it  does not fetch the missing
records.

# .iloc refers to "Row count"
# .loc refers to "Index"
temp_df.loc[0:4]
```

```
  animal count location    id
0    cat     3     home  I101
4    cow    10     farm  I105
```

```
temp_df.loc[1:2]
```

```
  animal count location    id
1    dog     5     home  I102
3   goat    15     farm  I104
2  sheep    10     farm  I103
```

```
# .iloc specifically locates the actual record irrestive of index
# notice how 2nd record (row=2) and 3rd row (row=3) are picked though
index is something else.
temp_df.iloc[1:3]
```

```
  animal count location    id
4    cow    10     farm  I105
1    dog     5     home  I102
```

```
# set id as index
temp_df.set_index('id', inplace=True, drop=True)
temp_df
```

```
     animal count location
id
I101    cat     3     home
I105    cow    10     farm
I102    dog     5     home
I104   goat    15     farm
I103  sheep    10     farm
```

```
# notice how .iloc still fetches records 2, and 3
temp_df.iloc[1:3]
```

```
     animal count location
id
I105    cow    10     farm
I102    dog     5     home
```

```
# as .loc depends on index we will have to use only those values are
available in index.
# for ex. the below code fails.
temp_df.loc[2:4]
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call
last)
Cell In[141], line 3
      1 # as .loc depends on index we will have to use only those
values are available in index.
      2 # for ex. the below code fails.
----> 3 temp_df.loc[2:4]

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexing.py:1191, in
_LocationIndexer.__getitem__(self, key)
   1189 maybe_callable = com.apply_if_callable(key, self.obj)
   1190 maybe_callable = self._check_deprecated_callable_usage(key,
maybe_callable)
-> 1191 return self._getitem_axis(maybe_callable, axis=axis)

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexing.py:1411, in
_LocIndexer._getitem_axis(self, key, axis)
   1409 if isinstance(key, slice):
   1410     self._validate_key(key, axis)
-> 1411     return self._get_slice_axis(key, axis=axis)
   1412 elif com.is_bool_indexer(key):
   1413     return self._getbool_axis(key, axis=axis)
```

```
File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexing.py:1443, in
_LocIndexer._get_slice_axis(self, slice_obj, axis)
   1440     return obj.copy(deep=False)
   1442 labels = obj._get_axis(axis)
-> 1443 indexer = labels.slice_indexer(slice_obj.start,
slice_obj.stop, slice_obj.step)
   1445 if isinstance(indexer, slice):
   1446     return self.obj._slice(indexer, axis=axis)

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:6662, in
Index.slice_indexer(self, start, end, step)
   6618 def slice_indexer(
   6619     self,
   6620     start: Hashable | None = None,
   6621     end: Hashable | None = None,
   6622     step: int | None = None,
   6623 ) -> slice:
   6624     """
   6625     Compute the slice indexer for input labels and step.
   6626
   (...)
   6660     slice(1, 3, None)
   6661     """
-> 6662     start_slice, end_slice = self.slice_locs(start, end,
step=step)
   6664     # return a slice
   6665     if not is_scalar(start_slice):

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:6879, in Index.slice_locs(self,
start, end, step)
   6877 start_slice = None
   6878 if start is not None:
-> 6879     start_slice = self.get_slice_bound(start, "left")
   6880 if start_slice is None:
   6881     start_slice = 0

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:6794, in
Index.get_slice_bound(self, label, side)
   6790 original_label = label
   6792 # For datetime indices label may be a string that has to be
converted
   6793 # to datetime boundary according to its resolution.
-> 6794 label = self._maybe_cast_slice_bound(label, side)
   6796 # we need to look up the label
   6797 try:
```

```
File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:6727, in
Index._maybe_cast_slice_bound(self, label, side)
   6725 # reject them, if index does not contain label
   6726 if (is_float(label) or is_integer(label)) and label not in
self:
-> 6727     self._raise_invalid_indexer("slice", label)
   6729 return label

File c:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-
packages\pandas\core\indexes\base.py:4301, in
Index._raise_invalid_indexer(self, form, key, reraise)
   4299 if reraise is not lib.no_default:
   4300     raise TypeError(msg) from reraise
-> 4301 raise TypeError(msg)

TypeError: cannot do slice indexing on Index with these indexers [2]
of type int
```

```
temp_df.loc['I101':'I102']
```

```
     animal count location
id
I101    cat     3     home
I105    cow    10     farm
I102    dog     5     home
```

- If i want reset my index then use ".reset_index"

```
temp_df.reset_index(drop=True)
```

```
  animal count location
0    cat     3     home
1    cow    10     farm
2    dog     5     home
3   goat    15     farm
4  sheep    10     farm
```

```
temp_df
```

```
     animal count location
id
I101    cat     3     home
I105    cow    10     farm
I102    dog     5     home
I104   goat    15     farm
I103  sheep    10     farm
```

```
# After runnig above dataframe, we oberved that the "index" were not
reset.
```

```
# To rest it we have to use below code.
# We have to use compusary "inplace=True"

temp_df.reset_index(drop=False, inplace=True)
temp_df
```

```
     id animal count location
0  I101    cat     3     home
1  I105    cow    10     farm
2  I102    dog     5     home
3  I104   goat    15     farm
4  I103  sheep    10     farm
```

```
# we can reset index and set records from 0 again.
ibm_hr.reset_index(drop=True, inplace=True)
ibm_hr.head()
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
```

```
YearsAtCompany  \
0                    8                          0                 1
6
1                   10                          3                 3
10
2                    7                          3                 3
0
3                    8                          3                 3
8
4                    6                          3                 3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                         0                     5
1                  7                         1                     7
2                  0                         0                     0
3                  7                         3                     0
4                  2                         2                     2

[5 rows x 35 columns]
```

Subsetting based on position

```
# fetch first 10 records
ibm_hr[:10]

   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development

5   32        No  Travel_Frequently       1005  Research & Development

6   59        No      Travel_Rarely       1324  Research & Development

7   30        No      Travel_Rarely       1358  Research & Development

8   38        No  Travel_Frequently        216  Research & Development

9   36        No      Travel_Rarely       1299  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
```

```
0                  1          2  Life Sciences              1
1
1                  8          1  Life Sciences              1
2
2                  2          2          Other              1
4
3                  3          4  Life Sciences              1
5
4                  2          1        Medical              1
7
5                  2          2  Life Sciences              1
8
6                  3          3        Medical              1
10
7                 24          1  Life Sciences              1
11
8                 23          3  Life Sciences              1
12
9                 27          3        Medical              1
13

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1
5  ...                         3            80                 0
6  ...                         1            80                 3
7  ...                         2            80                 1
8  ...                         2            80                 0
9  ...                         2            80                 2

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                  8                      0               1
6
1                 10                      3               3
10
2                  7                      3               3
0
3                  8                      3               3
8
4                  6                      3               3
2
5                  8                      2               2
7
6                 12                      3               2
1
```

```
7                    1                         2                 3
1
8                   10                         2                 3
9
9                   17                         3                 2
7

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2
5                  7                        3                     6
6                  0                        0                     0
7                  0                        0                     0
8                  7                        1                     8
9                  7                        7                     7

[10 rows x 35 columns]
```

```
ibm_hr.iloc[[1, 5, 7]]
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
1   49        No  Travel_Frequently        279  Research & Development

5   32        No  Travel_Frequently       1005  Research & Development

7   30        No      Travel_Rarely       1358  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
1                 8          1  Life Sciences              1
2
5                 2          2  Life Sciences              1
8
7                24          1  Life Sciences              1
11

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
1  ...                         4            80                 1
5  ...                         3            80                 0
7  ...                         2            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
1                 10                      3               3
10
```

```
5                    8                         2                 2
7
7                    1                         2                 3
1

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
1                  7                        1                     7
5                  7                        3                     6
7                  0                        0                     0

[3 rows x 35 columns]
```

```
# keep in mind that .iloc does not consider index
ibm_hr2.iloc[[1, 5, 7]]
```

```
                 Age Attrition     BusinessTravel  DailyRate  \
EducationField
Life Sciences     49        No  Travel_Frequently        279
Life Sciences     32        No  Travel_Frequently       1005
Life Sciences     30        No      Travel_Rarely       1358

                             Department  DistanceFromHome  Education  \
EducationField
Life Sciences   Research & Development                 8          1
Life Sciences   Research & Development                 2          2
Life Sciences   Research & Development                24          1

                 EmployeeCount  EmployeeNumber  EnvironmentSatisfaction
...  \
EducationField
...
Life Sciences                1               2                        3
...
Life Sciences                1               8                        4
...
Life Sciences                1              11                        4
...

               RelationshipSatisfaction  StandardHours
StockOptionLevel  \
EducationField

Life Sciences                         4             80
1
Life Sciences                         3             80
0
Life Sciences                         2             80
1

                 TotalWorkingYears TrainingTimesLastYear
```

```
WorkLifeBalance  \
EducationField

Life Sciences                    10                        3
3
Life Sciences                     8                        2
2
Life Sciences                     1                        2
3

                YearsAtCompany  YearsInCurrentRole
YearsSinceLastPromotion  \
EducationField

Life Sciences               10                     7
1
Life Sciences                7                     7
3
Life Sciences                1                     0
0

                 YearsWithCurrManager
EducationField
Life Sciences                       7
Life Sciences                       6
Life Sciences                       0

[3 rows x 34 columns]

ibm_hr.columns

Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate',
'Department',
       'DistanceFromHome', 'Education', 'EducationField',
'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender',
'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours',
'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear',
'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole',
'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

```
# same goes with selecting columns using .iloc
ibm_hr.iloc[[1,4,5,2],[1,3,5]]
```

```
  Attrition  DailyRate  DistanceFromHome
1        No        279                 8
4        No        591                 2
5        No       1005                 2
2       Yes       1373                 2
```

```
ibm_hr2.iloc[[1,4,5,2],[1,3,5]]
```

```
                Attrition  DailyRate  DistanceFromHome
EducationField
Life Sciences          No        279                 8
Medical                No        591                 2
Life Sciences          No       1005                 2
Other                 Yes       1373                 2
```

```
ibm_hr['Age'].iloc[0]
```

```
41
```

```
ibm_hr['Age'].iloc[1]
```

```
49
```

```
ibm_hr['Age'].iloc[3]
```

```
33
```

To see "all null values" in dataframe

```
ibm_hr.isnull().sum()
```

```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
```

```
MonthlyIncome                0
MonthlyRate                  0
NumCompaniesWorked           0
Over18                       0
OverTime                     0
PercentSalaryHike            0
PerformanceRating            0
RelationshipSatisfaction     0
StandardHours                0
StockOptionLevel             0
TotalWorkingYears            0
TrainingTimesLastYear        0
WorkLifeBalance              0
YearsAtCompany               0
YearsInCurrentRole           0
YearsSinceLastPromotion      0
YearsWithCurrManager         0
dtype: int64
```

Subsetting based on condition (and & or)

- [1] and
- [2] or

```
ibm_hr.head()
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
```

```
7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                  8                      0               1
6
1                 10                      3               3
10
2                  7                      3               3
0
3                  8                      3               3
8
4                  6                      3               3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                       0                     5
1                  7                       1                     7
2                  0                       0                     0
3                  7                       3                     0
4                  2                       2                     2

[5 rows x 35 columns]
```

When we want both/all conditions to be satisfied, this is easiest way

- Syntax:
  - ibm_hr[() & () & ()]

```
ibm_hr[() & () & () ]

---------------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
Cell In[158], line 3
      1 # when we want both/all conditions to be satisfied, this is
easiest way
      2 # Syntax is below one
----> 3 ibm_hr[() & () & () ]

TypeError: unsupported operand type(s) for &: 'tuple' and 'tuple'
```

When we want atleast one condition to be satisfied, this is easiest way

- Syntax:
  - ibm_hr[() | () | ()]

```
ibm_hr[() | () | ()]
```

```
---------------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
Cell In[159], line 3
      1 # when we want atleast one condition to be satisfied, this is
easiest way
      2 # Syntax is below one
----> 3 ibm_hr[() | () | ()]

TypeError: unsupported operand type(s) for |: 'tuple' and 'tuple'
```

```
# Subsetting with "single column"
ibm_hr[(ibm_hr['BusinessTravel']=='Travel_Rarely')]
```

```
      Age Attrition BusinessTravel  DailyRate              Department
\
0      41       Yes  Travel_Rarely       1102                   Sales

2      37       Yes  Travel_Rarely       1373  Research & Development

4      27        No  Travel_Rarely        591  Research & Development

6      59        No  Travel_Rarely       1324  Research & Development

7      30        No  Travel_Rarely       1358  Research & Development

...   ...       ...            ...        ...                     ...

1462   39        No  Travel_Rarely        722                   Sales

1464   26        No  Travel_Rarely       1167                   Sales

1466   39        No  Travel_Rarely        613  Research & Development

1467   27        No  Travel_Rarely        155  Research & Development

1469   34        No  Travel_Rarely        628  Research & Development


      DistanceFromHome  Education EducationField  EmployeeCount  \
0                    1          2  Life Sciences              1
2                    2          2          Other              1
4                    2          1        Medical              1
6                    3          3        Medical              1
```

```
7                   24          1  Life Sciences              1
...                ...        ...            ...            ...
1462                24          1     Marketing               1
1464                 5          3         Other               1
1466                 6          1       Medical               1
1467                 4          3  Life Sciences              1
1469                 8          3       Medical               1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
0                  1  ...                         1            80
2                  4  ...                         2            80
4                  7  ...                         4            80
6                 10  ...                         1            80
7                 11  ...                         2            80
...              ...  ...                       ...           ...
1462            2056  ...                         1            80
1464            2060  ...                         4            80
1466            2062  ...                         1            80
1467            2064  ...                         2            80
1469            2068  ...                         1            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                    0                  8                      0
2                    0                  7                      3
4                    1                  6                      3
6                    3                 12                      3
7                    1                  1                      2
...                ...                ...                    ...
1462                 1                 21                      2
1464                 0                  5                      2
1466                 1                  9                      5
1467                 1                  6                      0
1469                 0                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0                  1               6                  4
2                  3               0                  0
4                  3               2                  2
6                  2               1                  0
7                  3               1                  0
...              ...             ...                ...
1462               2              20                  9
1464               3               4                  2
1466               3               7                  7
1467               3               6                  2
1469               4               4                  3

      YearsSinceLastPromotion  YearsWithCurrManager
0                           0                     5
2                           0                     0
```

```
4                            2                     2
6                            0                     0
7                            0                     0
...                        ...                   ...
1462                         9                     6
1464                         0                     0
1466                         1                     7
1467                         0                     3
1469                         1                     2

[1043 rows x 35 columns]
```

```
# Subset using "two conditions" and fetch "all columns"
ibm_hr[(ibm_hr['BusinessTravel']=='Travel_Rarely') &
(ibm_hr['EducationField']=='Other')]
```

```
      Age Attrition BusinessTravel  DailyRate              Department
\
2      37       Yes  Travel_Rarely       1373  Research & Development

25     53        No  Travel_Rarely       1282  Research & Development

31     44        No  Travel_Rarely       1459  Research & Development

40     35        No  Travel_Rarely        464  Research & Development

77     45        No  Travel_Rarely        193  Research & Development

...   ...       ...            ...        ...                     ...

1408   23        No  Travel_Rarely        571  Research & Development

1413   25        No  Travel_Rarely        977  Research & Development

1433   25        No  Travel_Rarely       1382                   Sales

1459   29        No  Travel_Rarely       1378  Research & Development

1464   26        No  Travel_Rarely       1167                   Sales


      DistanceFromHome  Education EducationField  EmployeeCount  \
2                    2          2          Other              1
25                   5          3          Other              1
31                  10          4          Other              1
40                   4          2          Other              1
77                   6          4          Other              1
...                ...        ...            ...            ...
1408                12          2          Other              1
1413                 2          1          Other              1
1433                 8          2          Other              1
```

```
1459                13          2          Other              1
1464                 5          3          Other              1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
2                  4  ...                         2            80
25                32  ...                         4            80
31                40  ...                         4            80
40                53  ...                         3            80
77               101  ...                         2            80
...              ...  ...                       ...           ...
1408            1982  ...                         3            80
1413            1992  ...                         3            80
1433            2018  ...                         2            80
1459            2053  ...                         1            80
1464            2060  ...                         4            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
2                    0                  7                      3
25                   1                 26                      3
31                   0                  9                      5
40                   1                  1                      3
77                   0                 17                      3
...                ...                ...                    ...
1408                 0                  5                      6
1413                 1                  7                      2
1433                 1                  6                      3
1459                 1                 10                      2
1464                 0                  5                      2

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
2                  3               0                  0
25                 2              14                 13
31                 4               4                  2
40                 3               1                  0
77                 4               0                  0
...              ...             ...                ...
1408               4               5                  2
1413               2               2                  2
1433               2               5                  3
1459               3               4                  3
1464               3               4                  2

      YearsSinceLastPromotion  YearsWithCurrManager
2                           0                     0
25                          4                     8
31                          1                     3
40                          0                     0
77                          0                     0
...                       ...                   ...
1408                        1                     4
```

```
1413                          0                       2
1433                          0                       4
1459                          0                       3
1464                          0                       0

[61 rows x 35 columns]
```

```
# Fetch a particular "1 column" (i,e StandardHours) with "2
conditions"
ibm_hr.loc[(ibm_hr['BusinessTravel']=='Travel_Rarely') &
(ibm_hr['EducationField']=='Other'), 'StandardHours']
```

```
2       80
25      80
31      80
40      80
77      80
        ..
1408    80
1413    80
1433    80
1459    80
1464    80
Name: StandardHours, Length: 61, dtype: int64
```

```
# To see all null values
ibm_hr.isnull().sum()
```

```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
```

```
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

```
ibm_hr['BusinessTravel'].unique()
```

```
array(['Travel_Rarely', 'Travel_Frequently', 'Non-Travel'],
dtype=object)
```

```
array(['Travel_Rarely', 'Travel_Frequently', 'Non-Travel'],
dtype=object)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
Cell In[165], line 1
----> 1 array(['Travel_Rarely', 'Travel_Frequently', 'Non-Travel'],
dtype=object)

NameError: name 'array' is not defined
```

```
# to replace values using subsetting.
ibm_hr[['BusinessTravel']][(ibm_hr['BusinessTravel']=='Non-Travel')]
```

```
     BusinessTravel
17       Non-Travel
20       Non-Travel
46       Non-Travel
53       Non-Travel
83       Non-Travel
...             ...
1434     Non-Travel
1437     Non-Travel
1441     Non-Travel
1447     Non-Travel
1463     Non-Travel

[150 rows x 1 columns]
```

```
ibm_hr.loc[(ibm_hr['BusinessTravel']=='Non-Travel'), 'BusinessTravel']
```

```
17      Non-Travel
20      Non-Travel
46      Non-Travel
53      Non-Travel
83      Non-Travel
          ...
1434    Non-Travel
1437    Non-Travel
1441    Non-Travel
1447    Non-Travel
1463    Non-Travel
Name: BusinessTravel, Length: 150, dtype: object
```

```
# note that the right side of below code is statis value (i,e No
Travel)
ibm_hr.loc[(ibm_hr['BusinessTravel']=='Non-Travel'), 'BusinessTravel']
= 'No Travel'
```

```
ibm_hr.loc[(ibm_hr['BusinessTravel']=='Non Travel'), 'BusinessTravel']
```

```
Series([], Name: BusinessTravel, dtype: object)
```

```
ibm_hr.loc[(ibm_hr['BusinessTravel']=='No Travel'), 'BusinessTravel']
```

```
17      No Travel
20      No Travel
46      No Travel
53      No Travel
83      No Travel
          ...
1434    No Travel
1437    No Travel
1441    No Travel
1447    No Travel
1463    No Travel
Name: BusinessTravel, Length: 150, dtype: object
```

```
# how to get values from different column where the conditions match
# When ever u copt the values from 1 column to another column the
conditions on both sides should be same.
# Only the column name is changed
```

```
ibm_hr.loc[(ibm_hr['BusinessTravel']=='No Travel'), 'BusinessTravel']
= ibm_hr.loc[(ibm_hr['BusinessTravel']=='No Travel'),
'EducationField']
```

```
ibm_hr['BusinessTravel'].unique()
```

```
array(['Travel_Rarely', 'Travel_Frequently', 'Medical', 'Other',
       'Marketing', 'Life Sciences', 'Human Resources',
       'Technical Degree'], dtype=object)
```

Some more cocepts are there like below

- groupby,
- merge/join,
- concat