# 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)
Requirement already satisfied: tzdata>=2022.7 in c:\users\hp\appdata\
local\programs\python\python310\lib\site-packages (from pandas)
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"

```
df = pd.DataFrame(data=data)
print(df)
  Col1
         Col2 Col3
0
         Raj
               20
    1
    2
      Sameera
               25
1
    3
               22
        Ketan
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

Coll 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
      2
                               1
1
                     25
         Sameera
2
      3
                     22
                               1
           Ketan
```

• [ii] When the right side value is not static value, make sure that the shapes match the length 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
      2
                               1
                                            2
1
         Sameera
                     25
2
      3
                     22
                                            3
           Ketan
```

### **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
            self. set item(key, value)
-> 4311
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
   (\ldots)
   4522
            ensure homogeneity.
   4523
-> 4524
            value, refs = self. sanitize column(value)
   4526
            if (
                key in self.columns
   4527
   4528
                and value.ndim == 1
                and not isinstance(value.dtype, ExtensionDtype)
   4529
   4530
            ):
                # broadcast across multiple columns if necessary
   4531
   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):
            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"
   (\ldots)
            # TODO: Remove kludge in sanitize array for string mode
   5273
```

```
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)}) "
                "does not match length of index "
    576
    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()
                     BusinessTravel DailyRate
                                                               Department
   Age Attrition
0
    41
                       Travel Rarely
             Yes
                                            1102
                                                                    Sales
    49
              No
                  Travel_Frequently
                                             279
                                                  Research & Development
2
    37
             Yes
                       Travel Rarely
                                            1373
                                                  Research & Development
3
    33
              No
                  Travel Frequently
                                            1392
                                                  Research & Development
    27
              No
                       Travel Rarely
                                             591
                                                  Research & Development
   DistanceFromHome
                     Education EducationField EmployeeCount
EmployeeNumber
0
                                 Life Sciences
1
1
                                 Life Sciences
                              1
                                                              1
2
2
                   2
                              2
                                          0ther
                                                              1
4
3
                                 Life Sciences
5
4
                                       Medical
                                                              1
7
        RelationshipSatisfaction StandardHours
                                                  StockOptionLevel
0
                                1
                                              80
                                                                  0
1
                                4
                                              80
                                                                  1
                                2
2
                                              80
                                                                  0
3
                                3
                                              80
                                                                  0
                                4
                                              80
                                                                  1
   TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
YearsAtCompany
                    8
0
                                            0
                                                             1
6
1
                   10
                                            3
                                                             3
10
2
                    7
                                                             3
                                            3
0
3
                    8
                                            3
                                                            3
```

8 4 2			6				3			3	
9 0 1 2 3 4	'earsInC	urrentRo	ole 4 7 0 7 2	YearsSi	Inc	eLast	Promotion 0 1 0 3 2	) - ) }	YearsWithCu	urrMan	ager 5 7 0 0 2
[5	rows x 3	35 colur	nns]								
ibm	_hr.head	d( <mark>10</mark> )									
\	Age Att	rition	В	Business	Tr	avel	DailyRat	:e		Dep	artment
0	41	Yes		Travel_	Ra	rely	110	)2			Sales
1	49	No	Trav	el_Fred	que	ntly	27	9	Research 8	Deve	lopment
2	37	Yes		Travel_	Ra	rely	137	'3	Research 8	Deve	lopment
3	33	No	Trav	el_Fred	que	ntly	139	2	Research &	Deve	lopment
4	27	No		Travel_	_Ra	rely	59	1	Research &	Deve	lopment
5	32	No	Trav	el_Fred	que	ntly	100	)5	Research &	Deve	lopment
6	59	No		Travel_	Ra	rely	132	24	Research &	Deve	lopment
7	30	No		Travel_	Ra	rely	135	8	Research &	Deve	lopment
8	38	No	Trav	el_Fred	que	ntly	21	.6	Research &	Deve	lopment
9	36	No		Travel_	Ra	rely	129	9	Research &	Deve	lopment
Emp	Distanco loyeeNur		ne E	ducatio	n	Educa <sup>.</sup>	tionField	l	EmployeeCo	unt	
0 1	redycental	iibei (	1		2	Life	Sciences	5		1	
1			8		1	Life	Sciences	5		1	
2			2		2		0ther	•		1	
4 3			3		4	Life	Sciences	<b>.</b>		1	
5 4			2		1		Medical			1	
7 5			2		2	Life	Sciences	;		1	

8				_
6 10	3	3	Medical	1
	24	1 Life	Sciences	1
8	23	3 Life	Sciences	1
12 9	27	3	Medical	1
13				
Relations 0 1 2 3 4 5 6 7 8 9	hipSatisfac	tion Star 1 4 2 3 4 3 1 2 2 2	80 80 80 80 80 80 80 80 80 80	StockOptionLevel \
TotalWorkingYe YearsAtCompany \ 0		ngTimesLa	astYear Wor	kLifeBalance 1
6 1	10		3	3
10 2	7		3	3
0				
3 8	8		3	3
4 2	6		3	3
5 7	8		2	2
6	12		3	2
1 7	1		2	3
1	10		2	3
9 9	17		3	2
9 7			_	
YearsInCurrentR 0 1 2	ole YearsS 4 7 0 7	inceLastF	Promotion 0 1 0 3	YearsWithCurrManager 5 7 0 0

```
4
                   2
                                             2
5
                   7
                                             3
                                                                    6
6
                   0
                                             0
                                                                    0
7
                                                                    0
                   0
                                             0
                                                                    8
8
                   7
                                             1
9
[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
                         Travel Rarely
       41
                Yes
                                               1102
Sales
4977
       49
                 No Travel Frequently
                                                279
                                                     Research &
Development
221
       37
                Yes
                          Travel Rarely
                                               1373 Research &
Development
4695
       33
                    Travel Frequently
                                               1392
                                                     Research &
                 No
Development
8358
       27
                 No
                          Travel Rarely
                                                591 Research &
Development
      DistanceFromHome
                         Education EducationField
                                                    EmployeeCount
4631
                      1
                                 2 Life Sciences
                                                                1
                     8
4977
                                 1 Life Sciences
                                                                1
                      2
221
                                 2
                                                                1
                                            0ther
                                    Life Sciences
4695
                     3
                                 4
                                                                1
                     2
8358
                                 1
                                          Medical
                                                                 1
      EmployeeNumber
                            RelationshipSatisfaction StandardHours
                       . . .
4631
                    1
                                                    1
                                                                  80
4977
                    2
                                                    4
                                                                 80
                       . . .
                                                    2
221
                                                                  80
                    4
                    5
                                                    3
4695
                                                                  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
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 available 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
                                   Travel Rarely
Life Sciences
                 41
                          Yes
                                                        1102
Life Sciences
                 49
                           No
                               Travel Frequently
                                                         279
                                   Travel Rarely
0ther
                 37
                          Yes
                                                        1373
Life Sciences
                 33
                               Travel Frequently
                           No
                                                        1392
Medical
                 27
                                   Travel Rarely
                                                         591
                           No
                            Department DistanceFromHome Education \
```

EducationField					
Life Sciences Life Sciences Other Life Sciences Medical	Research & Dev Research & Dev Research & Dev Research & Dev	velopment velopment		1 8 2 3 2	2 1 2 4 1
	EmployeeCount	EmployeeN	umber	EnvironmentSat	isfaction
\ EducationField					
Life Sciences	1		1		2
Life Sciences	1		2		3
 Other	1		4		4
 Life Sciences	1		5		4
 Medical	1		7		1
StockOptionLeve EducationField	RelationshipSa <sup>r</sup> l \	tisfaction	Standa	rdHours	
Life Sciences		1		80	
0 Life Sciences		4		80	
1 Other O		2		80	
Life Sciences		3		80	
Medical		4		80	
1					
WorkLifeBalance EducationField	TotalWorkingYe	ears Traini	ngTimes	LastYear	
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
                                                 4
                             6
Life Sciences
                           10
                                                 7
0ther
                                                 0
                                                 7
Life Sciences
                                                 2
Medical
                             2
                YearsWithCurrManager
EducationField
Life Sciences
                                    5
                                    7
Life Sciences
Other
                                    0
                                    0
Life Sciences
Medical
[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"

To see all data types	
ibm_hr.dtypes	
Ago	in+61
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 \
                          Travel Rarely
4631
       41
                Yes
                                               1102
Sales
4977
                 No Travel Frequently
                                                     Research &
       49
                                                279
Development
      DistanceFromHome
                         Education EducationField
                                                    EmployeeCount
4631
                                 2 Life Sciences
                      8
                                                                 1
4977
                                 1 Life Sciences
                            RelationshipSatisfaction StandardHours \
      EmployeeNumber
4631
                    1
                                                    1
                                                                  80
                                                    4
4977
                    2
                                                                  80
      StockOptionLevel TotalWorkingYears
                                            TrainingTimesLastYear
4631
                      0
                                         8
                                                                  0
4977
                                        10
                                                                  3
                      YearsAtCompany YearsInCurrentRole
     WorkLifeBalance
4631
                    1
4977
                    3
                                   10
                                                        7
      YearsSinceLastPromotion YearsWithCurrManager
4631
                             0
                                                    5
                                                    7
4977
[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:
            return self. engine.get loc(casted key)
-> 3805
   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:
            return self. getitem multilevel(key)
-> 4102 indexer = self.columns.get loc(key)
   4103 if is integer(indexer):
            indexer = [indexer]
   4104
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:
            # If we have a listlike key, _check_indexing_error will
   3814
raise
            # InvalidIndexError. Otherwise we fall through and re-
   3815
raise
```

```
# the TypeError.
self._check_indexing_error(key)
   3816
   3817
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.	locf	'Life	Sciences'	1
_~~				0020000	

<pre>ibm_hr2.loc['Life Sciences']</pre>								
	Age Attritio	on Busin	essTravel	DailyRate	\			
EducationField Life Sciences Life Sciences Life Sciences Life Sciences Life Sciences	49 M 33 M 32 M 30 M	No Travel_F No Travel_F No Travel_F No Trav	rel_Rarely requently requently requently rel_Rarely	1102 279 1392 1005 1358				
Life Sciences Life Sciences Life Sciences Life Sciences Life Sciences	45 M 40 M 35 M	No Trav No Travel_F No Trav	rel_Rarely rel_Rarely requently rel_Rarely rel_Rarely	374 1322 1199 287 155				
Education Field		Department	DistanceF	romHome Ed	ducation \			
EducationField Life Sciences Life Sciences Life Sciences Life Sciences Life Sciences	Research & I Research & I Research & I Research & I	Development Development		1 8 3 2 24	2 1 4 2 1			
Life Sciences Life Sciences Life Sciences Life Sciences Life Sciences	Research & I Research & I Research & I Research & I	Development Development		20 2 18 1 4	3 4 4 4 3			
,	EmployeeCour	nt Employee	Number En	vironmentSa	atisfaction			
\ 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
		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
		_	2001		_
	RelationshipS	Satisfaction	StandardHo	urs	
StockOptionLev EducationField	el \				
Life Sciences		1		80	
0 Life Sciences		4		80	
1					
Life Sciences 0		3		80	
Life Sciences		3		80	
0 Life Sciences		2		80	
1		_			
Life Sciences		3		80	
0 Life Sciences		4		80	
0					
Life Sciences 2		4		80	
Life Sciences		4		80	
1 Life Sciences		2		80	
1		2			
	TotalWorking	yYears Traini	ngTimesLast	Year	
WorkLifeBalanc EducationField	e \	,10013 110111			

Life 1	Sciences		8	0	
Life	Sciences		10	3	
	Sciences		8	3	
	Sciences		8	2	
2 Life	Sciences		1	2	
3					
Life 3	Sciences		8	3	
	Sciences		8	2	
Life	Sciences		10	2	
	Sciences		4	5	
3	Caionaga		6	0	
3	Sciences		0	U	
	sSinceLast ationField	Promotion \	YearsInCurrentRol	Le	
Life 0	Sciences	6		4	
Life	Sciences	10		7	
1 Life 3	Sciences	8		7	
Life	Sciences	7		7	
3 Life 0	Sciences	1		0	
• • • •					
	Sciences	5		3	
0 Life	Sciences	2		2	
2	Sciences	10		2	
0					
Life 1	Sciences	4		3	
Life 0	Sciences	6		2	

ger
5
7
0
6
0
1
2
2
1
3
9

# [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']

<pre>ibm_hr2[ibm_hr2['Department']=='Sales']</pre>									
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	Trave	l_Rar	rely	1167	Sales
Medical	49	No T	ravel_Fr	eauer	ntlv	1023	Sales
	-		_	- 1	,		
EmployeeNumber EducationField	DistanceF \	romHome	Educat	ion	EmployeeC	ount	
Life Sciences		1		2		1	
1 Life Sciences 23		2		4		1	
Life Sciences		9		4		1	
27 Marketing		8		4		1	
35 Marketing		2		4		1	
38							
		20					
Life Sciences 2046		20		3		1	
Marketing 2055		28		3		1	
Marketing 2056		24		1		1	
Other		5		3		1	
2060 Medical 2065		2		3		1	
	Environme	entSatis	faction		Relations	hipSatis	faction
\ 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	StockOntion	Lovol Total	.WorkingYears \
EducationField Life Sciences Life Sciences Life Sciences Marketing Marketing Life Sciences Marketing Marketing Other Medical	80 80 80 80  80 80 80 80		0 0 1 0  0 1 1 0	8 31 10 10 22 8 20 21 5 17
YearsAtCompany EducationField	TrainingTimesLas	stYear Work	LifeBalance	
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
	YearsInCurrent	Role YearsS	inceLastProm	notion \

EducationField Life Sciences Life Sciences Life Sciences Marketing Marketing Life Sciences Marketing Marketing Marketing Marketing Marketing Other Medical	4 8 3 7 2  3 2 9 2	0 3 0 4 2  0 2 9 0
Υ	/earsWithCurrManager	
EducationField	g	
Life Sciences	5	
	3 7	
Life Sciences	7	
Life Sciences	3 2	
Marketing	2	
Marketing	1	
Life Sciences	1	
Marketing	Θ	
Marketing	6	
0ther	9	
Medical	8	
	J	
[446 rows x 34 co	olumns]	

# [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	

Medical   34 No Travel_Rarely   628	Life Sciences Medical	27 49	No No	Travel_F		ly 10	L55 )23	
EducationField Life Sciences Life Sciences Other Research & Development Research & Developm	Medical	34	No	Trav	el_Rare	ly 6	528	
Life Sciences Cife Sciences Ci	Education Et al d		De	partment	Distan	ceFromHome	Education	\
Other Life Sciences Research & Development Research & Development 3 4 4 Medical Research & Development 2 1 1  Medical Research & Development 23 2 2 1 1  Medical Research & Development 23 2 2 Medical Research & Development 6 1 1 1 4 3 3 3 3 4 Medical Research & Development 4 3 3 3 2 Medical Research & Development 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3				Sales		1	2	
Life Sciences Research & Development Research Research & Development Research Res						8		
MedicalResearch & Development21MedicalResearch & Development232MedicalResearch & Development61Life SciencesResearch & Development43MedicalResearch & Development83MedicalEmployeeCountEmployeeNumberEnvironmentSatisfactionEmployeeCountEmployeeNumberEnvironmentSatisfaction2333Other1444454Medical171Medical120613Medical120624Medical120654Medical120654RelationshipSatisfactionStandardHoursStockOptionLevel\text{EducationField}						3		
Medical Medical Medical Medical Medical Medical Medical Medical Medical Medical Medical Medical Medical 	Medical	Research &	Dev	elopment		2	1	
Medical Life Sciences Medical<		Research &	Dev	elopment				
EmployeeCount EmployeeNumber EnvironmentSatisfaction  \ EducationField Life Sciences		Research &	Dev	elopment			1	
EmployeeCount EmployeeNumber EnvironmentSatisfaction  \ EducationField Life Sciences		Research &	νev				3	
EducationField Life Sciences 1 1 2 Life Sciences 1 2 3 Other 1 4 4 4 Life Sciences 1 5 4 Medical 1 7 1 1 Medical 1 2061 3 Medical 1 2062 4 Medical 1 2065 4 Medical 1 2068 2 Medical 1 2068 2 Medical 1 Sciences 5 Medical 1 Sciences 6 Medical 1 Sciences 7 StandardHours StockOptionLevel \ EducationField	Medical	Research &	Dev	elopment			3	
EducationField Life Sciences 1 1 2 Life Sciences 1 2 3 Other 1 4 4 4 Life Sciences 1 5 4 Medical 1 7 1 1 Medical 1 2061 3 Medical 1 2062 4 Medical 1 2065 4 Medical 1 2068 2 RelationshipSatisfaction StandardHours StockOptionLevel \ EducationField		EmployeeCo	unt	Employee	Number	Environmer	ntSatisfact	ion
Life Sciences 1 1 2 3 3  Other 1 4 4 4 4  Life Sciences 1 5 4  Medical 1 7 1 1  Medical 1 2061 3  Medical 1 2062 4  Life Sciences 1 2064 2  Medical 1 2065 4  Medical 1 2068 2  RelationshipSatisfaction StandardHours	EducationField							
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 Medical 1 Sciences 5 Medical 1 Sciences 4 Medical 5 Sciences 5 Medical 6 StandardHours StockOptionLevel \ EducationField			1		1			2
Other       1       4       4          Life Sciences       1       5       4            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       StockOptionLevel \ EducationField}			1		2			3
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			1		4			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 Medical StandardHours StockOptionLevel \ EducationField								
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			1		5			4
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       \ EducationField       StandardHours	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       \ EducationField       StandardHours								
Medical 1 2062 4 Life Sciences 1 2064 2 Medical 1 2065 4 Medical 1 2068 2 RelationshipSatisfaction StandardHours StockOptionLevel \ EducationField			1		2061			3
Life Sciences 1 2064 2  Medical 1 2065 4  Medical 1 2068 2  RelationshipSatisfaction StandardHours  StockOptionLevel \ EducationField								
Medical 1 2065 4  Medical 1 2068 2  RelationshipSatisfaction StandardHours  StockOptionLevel \ EducationField	Medical		1		2062			4
Medical 1 2065 4  Medical 1 2068 2  RelationshipSatisfaction StandardHours StockOptionLevel \ EducationField			1		2064			2
Medical 1 2068 2  RelationshipSatisfaction StandardHours StockOptionLevel \ EducationField			1		2065			4
RelationshipSatisfaction StandardHours StockOptionLevel \ EducationField			1		2068			2
StockOptionLevel \ EducationField								
Life Sciences 1 80	StockOptionLeve		pSat	isfaction	Stand	ardHours		
	Life Sciences			1		80		

0			
Life Sciences		4	80
1 Other		2	80
0		Z	80
Life Sciences		3	80
0			
Medical		4	80
1			
			• • • •
Medical		3	80
1		<b>J</b>	
Medical		1	80
1			
Life Sciences		2	80
1 Medical		4	80
0		4	80
Medical		1	80
0			
	T . 11/		
WanklifaDalana	TotalWorkingYears		mesLastYear
WorkLifeBalanc EducationField			
Life Sciences	8		0
1			
Life Sciences	10		3
3 Other	7	•	3
3	1		3
Life Sciences	8		3
3			
Medical	6		3
3			
			• • •
Medical	17	•	3
3			_
Medical	9		5
3			2
Life Sciences	6		0
3 Medical	17		3
2	17		3
Medical	6		3
4			
	V		+Dala
	YearsAtCompany Ye	arsincurren	trote

```
YearsSinceLastPromotion \
EducationField
Life Sciences
                             6
                                                   4
Life Sciences
                             10
                                                   7
0ther
                                                   0
Life Sciences
                                                   7
                                                   2
Medical
2
Medical
                                                   2
Medical
                                                   7
Life Sciences
                                                   2
Medical
                                                   6
Medical
                                                   3
                 YearsWithCurrManager
EducationField
Life Sciences
                                     5
                                     7
Life Sciences
0ther
                                     0
Life Sciences
                                     0
                                     2
Medical
Medical
                                     3
Medical
                                     7
                                     3
Life Sciences
                                     8
Medical
Medical
[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'])]
```

EducationField         Medical         46         No         Travel_Rarely         945         Human Resources           Resources         37         Yes         Travel_Rarely         807         Human Resources           Resources         59         No         Non-Travel         1420         Human Resources           Resources         Human Resources         54         No         Non-Travel         142         Human Resources           Resources         26         No         Travel_Rarely         1355         Human Resources           Resources         27         Yes         Travel_Frequently         1337         Human Resources           Resources         38         No         Travel_Frequently         1444         Human Resources           Resources         55         No         Travel_Rarely         189         Human Resources           Resources         1         No         Travel_Rarely         309         Human Resources           Life Sciences         35         No         Travel_Rarely         146         Human Resources           103         Human         Education EmployeeCount         Human Resources         2         4         1           133         Human         1							
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	Department \ EducationField	Age	Attrition	BusinessTr	avel	DailyRate	
Human Resources   37	Medical Resources	46	No	Travel_Ra	rely	945	Human
Human Resources   59	Human Resources	37	Yes	Travel_Ra	rely	807	Human
Human Resources         54         No         Non-Travel         142         Human Resources           Life Sciences         26         No         Travel_Rarely         1355         Human Resources           No               Human Resources         27         Yes Travel_Frequently         1337         Human Resources           Resources         38         No Travel_Frequently         1444         Human Resources           Human Resources         55         No Travel_Rarely         189         Human Resources           Resources         25         No Travel_Rarely         146         Human Resources           Life Sciences         35         No Travel_Rarely         1146         Human Resources           Life Sciences         35         No Travel_Rarely         1146         Human Resources           Life Sciences         35         No Travel_Rarely         1146         Human Resources           Life Sciences         2         4         1         1           Light         14         1         1         1           Human Resources         26         3         1         1           Life Sciences         25         1	Human Resources	59	No	Non-Tr	avel	1420	Human
Life Sciences Resources	Human Resources	54	No	Non-Tr	avel	142	Human
Human Resources 27 Yes Travel_Frequently 1337 Human Resources 0ther 38 No Travel_Frequently 1444 Human Resources Human Resources 55 No Travel_Rarely 189 Human Resources 25 No Travel_Rarely 309 Human Resources 25 No Travel_Rarely 1146 Human Resources 26 Travel_Rarely 1146 Human Resources 27 Travel_Rarely 1146 Human Resources 28 Travel_Rarely 1146 Human Resources 29 Travel_Rarely 1146 Human Resources 20 Travel_Rarely 1146 Human Resources 21 Travel_Rarely 1146 Human Resources 22 Travel_Rarely 1146 Human Resources 24 Travel_Rarely 1146 Human Resources 25 Travel_Rarely 1146 Human Resources 2	Life Sciences Resources	26	No	Travel_Ra	rely	1355	Human
Resources Other							
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 1972 Human Resources 26 4 1 1973	Human Resources Resources	27	Yes	Travel_Freque	ntly	1337	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 2 5 3 1 148 Life Sciences 25 1 1 177 Human Resources 22 3 1 1972 Human Resources 26 4 1 1972 Human Resources 26 4 1 1973	Other Resources	38	No	Travel_Freque	ntly	1444	Human
Resources Life Sciences 35 No Travel_Rarely 1146 Human Resources    DistanceFromHome   Education   EmployeeCount	Human Resources Resources	55	No	Travel_Ra	rely	189	Human
DistanceFromHome   Education   EmployeeCount	Human Resources Resources	25	No	Travel_Ra	rely		Human
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	Life Sciences Resources	35	No	Travel_Ra	rely	1146	Human
103         Human Resources       6       4       1         133       1         Human Resources       2       4       1         140       1       1         Human Resources       26       3       1         156       1       1       1         177       1       1       1         1944       1       1       4       1         1972       1       4       1       1         1973       1       4       1       1         1973       26       4       1       1	EmployeeNumber EducationField	Dist \	anceFromHo	me Education	Emplo	oyeeCount	
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	Medical			5 2		1	
Human Resources       2       4       1         140       1       1         Human Resources       26       3       1         148       1       1         Life Sciences       25       1       1         177            Human Resources       22       3       1         1944       1       4       1         1972       1       4       1         Human Resources       26       4       1         1973       26       4       1	Human Resources			6 4		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 140			2 4		1	
177 Human Resources 22 3 1 1944 Other 1 4 1 1972 Human Resources 26 4 1 1973	Human Resources 148			26 3		1	
1944 Other 1 4 1 1972 Human Resources 26 4 1 1973	Life Sciences 177			25 1			
Other 1 4 1 1972 Human Resources 26 4 1 1973	Human Resources			22 3		1	
Human Resources 26 4 1 1973	0ther			1 4		1	
	Human Resources		;	26 4		1	
	Human Resources			2 3		1	

1987 Life Sciences	26	4	1
2040	EnvironmentCaticfac	rtion Dol	lationshinCatisfaction
\ EducationField	ENVITORMENTSACISTAC	itton Ke	lationshipSatisfaction
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
EducationField	StandardHours Stoc	ckOptionLevel	TotalWorkingYears \
Medical	80	1	16
Human Resources Human Resources	80 80	0 1	7 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 Life Sciences	80 80	0 0	6 9
	TrainingTimesLastYea	ar WorkLifeBa	alance 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
EducationField	YearsInCurrentRole	YearsSinceLastPi	romotion	\
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 Human Resources	9		1	
Life Sciences	0 0		1 1	
Life Sciences	U		Τ.	
	YearsWithCurrManage	er		
EducationField		2		
Medical Human Resources		2		
Human Resources		2		
Human Resources		4		
Life Sciences		7		
Human Resources		0		
Other _		2		
Human Resources		4		
Human Resources Life Sciences		2 7		
rile actelices		1		
[63 rows x 34 co	lumns]			

# [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

- tov -	LIST OF Values						
<pre>lov = ['Life Sc ibm_hr2.loc[lov</pre>		Marke	eting']				
EducationField	Age Attri	tion	Busin	essTrave	el DailyRa	te \	
Life Sciences Life Sciences	41 49	Yes No	Trav Travel F	el_Rarel		02 79	
Life Sciences	33	No	Travel_F	requentl	.y 13	92	
Life Sciences Life Sciences	32 30	No No	Travel_F Trav	requentl el_Rarel			
 Marketing Marketing	 34 36	No No		el_Rarel on-Trave	y 7	 04 01	
Marketing Marketing	36 50	No	Trav	el_Rarel el Rarel	y 11		
Marketing	39	Yes No		el_Rarel		22	
		De	partment	Distanc	ceFromHome	Education	\
EducationField Life Sciences Life Sciences Life Sciences Life Sciences	Research Research Research	& Dev	elopment		1 8 3 2	2 1 4 2	
Life Sciences	Research		•		24	1	
Marketing Marketing Marketing Marketing Marketing Marketing			Sales Sales Sales Sales Sales		28 15 11 28 24	3 4 4 3 1	
,	EmployeeC	ount	Employee	Number	Environmen	tSatisfacti	.on
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

Manufaction	1			4
Marketing	1	203	5	4
Marketing	1	203	6	4
Marketing	1	204	.5	2
Marketing	1	205	5	4
 Marketing	1	205	6	2
StockOptionLevel EducationField	RelationshipSatisfa L \	action Sta	ndardHours	
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 2		4	80	
Marketing		1	80	
1 Marketing		1	80	
1		2	80	
Marketing 1		Z		
Marketing 1		1	80	
-	Tatallyla which may you may	TanininaTi		
WorkLifeBalance EducationField	TotalWorkingYears \	rrainingri	mestastrear	
Life Sciences	8		0	
1 Life Sciences	10		3	
3 Life Sciences	8		3	
3	8		3	

Life Sciences 2	8		2	
Life Sciences	1		2	
3				
			2	
Marketing 3	8		2	
Marketing	15		4	
2	10			
Marketing	8		2	
2	2.0		2	
Marketing 3	20		3	
Marketing	21		2	
2	21		2	
YearsSinceLastPro		arsInCurrentRole		
EducationField	IIOCTOII /			
Laded Clotti le ca				
Life Sciences	6	4		
0	10	_		
Life Sciences	10	7		
1 Life Sciences	8	7		
3	O	,		
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	3	2		
Marketing	20	9		
9				
V	aarchitthCurrMana	ao r		
EducationField	earsWithCurrMana	ger		
Life Sciences		5		
Life Sciences		5 7		
Life Sciences		0		
Life Sciences		6		

```
Life Sciences 0
... 7
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
                 35
Life Sciences
                 27
Life Sciences
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 Marketing Marketing Marketing Marketing Marketing	34 36 36 50 39	No No No Yes No
[765 rows x	2 columns]	

To see all rows and columns using ".loc" ibm hr2.loc[:, :] BusinessTravel DailyRate \ Age Attrition EducationField Yes Life Sciences 41 Travel\_Rarely 1102 Life Sciences 49 No Travel Frequently 279 37 Travel Rarely 0ther Yes 1373 Life Sciences 33 No Travel\_Frequently 1392 Medical 27 No Travel Rarely 591 . . . . . . . . . . . . Medical 36 Travel\_Frequently No 884 Travel\_Rarely Medical 39 No 613 Life Sciences 27 No Travel Rarely 155 49 Travel Frequently Medical No 1023 Medical 34 No Travel Rarely 628 DistanceFromHome Department Education \ EducationField Life Sciences Sales 2 Life Sciences Research & Development 8 1 2 0ther Research & Development 2 Research & Development 3 Life Sciences 4 2 Medical Research & Development 1 . . . Research & Development Medical 23 2 Medical Research & Development 6 1 3 Life Sciences Research & Development 4 3 2 Sales Medical Medical Research & Development 8 3 EmployeeCount EmployeeNumber EnvironmentSatisfaction EducationField Life Sciences 2 Life Sciences 3 0ther 1 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
Re StockOptionLevel EducationField	elationshipSatisfa \	action Stand	ardHours	
Life Sciences		1	80	
0 Life Sciences		4	80	
1 Other		2	80	
0 Life Sciences		3	80	
0				
Medical 1		4	80	
Medical		3	80	
1 Medical		1	80	
1 Life Sciences		2	80	
1 Medical		4	80	
0		4		
Medical 0		1	80	
TotalWorkingYears TrainingTimesLastYear				
	\	rrainingrime	stastyear	
Life Sciences	8		0	

1		
Life Sciences 3	10	3
Other	7	3
3	•	J
Life Sciences	8	3
3 Medical	6	3
3	Ü	3
 Medical	17	3
3	17	3
Medical	9	5
3	6	0
Life Sciences 3	6	0
Medical	17	3
2 Madiaal		2
Medical 4	6	3
YearsSinceLast EducationField		urrentRole
Life Sciences	6	4
0 Life Sciences	10	7
1	10	,
Other	0	0
0 Life Sciences	8	7
3	0	,
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
                                     5
Life Sciences
                                     7
Life Sciences
0ther
                                     0
                                     0
Life Sciences
                                     2
Medical
. . .
Medical
                                     3
                                     7
Medical
                                     3
Life Sciences
                                     8
Medical
                                     2
Medical
[1470 rows x 34 columns]
ibm hr2.loc[:, ['Age', 'Attrition']]
                 Age Attrition
EducationField
Life Sciences
                  41
                           Yes
Life Sciences
                  49
                            No
0ther
                  37
                            Yes
                  33
Life Sciences
                             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'],
     'id' : ['I101', 'I102', 'I103', 'I104', 'I105']
})
temp df
  animal count location
                                  id
                3
0
      cat
                        home I101
                5
1
      dog
                        home I102
2 sheep
               10
                        farm I103
```

```
3
           15
                  farm I104
   goat
4 cow
           10
                  farm I105
# all columns all rows
temp df.loc[:, :]
  animal count location
                          id
            3
                  home I101
    cat
            5
1
    dog
                  home I102
2
  sheep
           10
                  farm I103
3
   goat
           15
                  farm I104
                  farm I105
    COW
           10
# [[]]
temp_df.loc[[2, 4]]
  animal count location
                          id
2 sheep
           10
                  farm I103
4
           10
                  farm I105
    COW
```

### [or]

```
# index 2 and 4 rows with all columns
temp_df.loc[[2, 4], :]
 animal count location
                           id
2 sheep
           10
                  farm I103
           10
4 cow
                  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
           10
                   farm I105
    COW
```

### NOTE:

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

```
temp_df.iloc[2:4]
 animal count location
                           id
            10
                   farm I103
   sheep
            15
                   farm I104
    goat
# 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
0
     cat
            3
                   home I101
            10
                   farm I105
4
     COW
1
             5
                   home I102
     doa
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
             3
                   home
                         I101
     cat
            10
     COW
                   farm I105
temp df.loc[1:2]
  animal count location
                           id
1
     dog
             5
                   home I102
            15
                   farm I104
    goat
2 sheep
                   farm I103
            10
# .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
            10
                   farm I105
     COW
1
             5
                   home I102
     dog
```

Ex: [2:4] means 2,3

.iloc refers to "Row count"

```
# set id as index
temp df.set index('id', inplace=True, drop=True)
temp df
     animal count location
id
I101
               3
                      home
        cat
I105
                      farm
        COW
               10
I102
        dog
               5
                      home
I104
               15
                      farm
       goat
I103 sheep
               10
                      farm
# notice how .iloc still fetches records 2, and 3
temp df.iloc[1:3]
     animal count location
id
I105
               10
                      farm
        COW
I102
             5
                      home
        dog
# 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):
            self. validate key(key, axis)
   1410
-> 1411
            return self. get slice axis(key, axis=axis)
   1412 elif com.is bool indexer(key):
            return self. getbool axis(key, axis=axis)
   1413
```

```
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):
            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
   (\ldots)
   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:
            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:
            self. raise invalid indexer("slice", label)
-> 6727
   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:
            raise TypeError(msg) from reraise
   4300
-> 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
                3
                      home
        cat
I105
        COW
               10
                      farm
I102
                5
        dog
                      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
                    farm
            10
     COW
2
             5
                    home
     dog
3
            15
                    farm
    goat
   sheep
            10
                    farm
temp df
     animal count location
id
I101
                 3
                       home
        cat
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
   I101
           cat
                   3
                          home
                          farm
1
  I105
           COW
                  10
  I102
           doa
                   5
                          home
                  15
3
                          farm
  I104
          goat
  I103
                  10
                          farm
         sheep
# 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
                      Travel Rarely
                                                                   Sales
             Yes
                                           1102
1
    49
              No
                  Travel Frequently
                                            279
                                                 Research & Development
                      Travel Rarely
                                                 Research & Development
    37
             Yes
                                           1373
    33
              No
                  Travel Frequently
                                           1392
                                                 Research & Development
                                                 Research & Development
    27
              No
                      Travel Rarely
                                            591
   DistanceFromHome
                     Education EducationField
                                                EmployeeCount
EmployeeNumber
                                 Life Sciences
0
                  1
1
1
                  8
                                 Life Sciences
2
2
                              2
                                                             1
                                         0ther
4
3
                                 Life Sciences
5
4
                                       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	_	
7	3	3
0	2	2
3 8	3	3
8	2	2
4 6 2	3	3
2		
YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0 4	0	5
1 7	1	7
2 0 7	0	0
3 7	3	Θ
4 2	2	2
[5 rows x 35 columns]		

### Subsetting based on position

# fetch first 10 records
ibm\_hr[:10]
 Age Attrition BusinessTravel DailyR

	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	1	2	Life	Sciences	1	
1	8	1	Life	Sciences	1	
2 2 4	2	2		0ther	1	
	3	4	life	Sciences	1	
3 5 4 7						
	2	1		Medical	1	
5 8 6	2	2	Life	Sciences	1	
	3	3		Medical	1	
10 7	24	1	Life	Sciences	1	
11 8	23	3	Life	Sciences	1	
12					1	
9 13	27	3		Medical	T	
0 1 2 3 4 5 6 7 8 9	RelationshipS	atisfactio	on Star 1 4 2 3 4 3 1 2 2	80 80 80 80 80 80 80 80 80 80	StockOptionLe	evel \     0     1     0     0     1     0     3     1     0     2
	TotalWorkingYears arsAtCompany \	TrainingT	imesLa		kLifeBalance	
0 6 1	8			0	1	
1 10	10			3	3	
	7			3	3	
3	8			3	3	
8 4	6			3	3	
2 0 3 8 4 2 5 7	8			2	2	
7						
6	12			3	2	

7 1	1		2	3
8	10		2	3
9	17		3	2
7				
YearsInCurr 0 1 2 3	rentRole Year 4 7 0 7	sSinceLastPro	omotion Yea 0 1 0 3	rsWithCurrManager 5 7 0 0
4 5 6 7 8	2 7 0 0 7		2 3 0 0 1	2 6 0 0
9	7		7	7
[10 rows x 35	_			
ibm_hr.iloc[[				
Age Attrit	ion Busin	essTravel Da	ailyRate	Department
1 49	No Travel_F	requently	279 Re	esearch & Development
5 32	No Travel_F	requently	1005 Re	esearch & Development
7 30	No Trav	el_Rarely	1358 Re	esearch & Development
DistanceFr EmployeeNumbe	er \	tion Educatio	·	loyeeCount
1 2	8	1 Life So	ciences	1
5 8	2	2 Life So	ciences	1
7 11	24	1 Life So	ciences	1
Relat 1 5 7	ionshipSatisf	action Standa 4 3 2	ardHours St 80 80 80	cockOptionLevel \ 1 0 1
TotalWorki YearsAtCompar 1		ningTimesLast	Year WorkLi 3	feBalance 3
10				

5 7	8	2	2			
7 7 1	1	2	3			
YearsInCurren	utRole YearsSin	nceLastPromotion	YearsWithCurr	Manager		
1 5 7	7 7 0	1 3 0		7 6 0		
[3 rows x 35 co	lumns]					
<pre># keep in mind ibm_hr2.iloc[[1</pre>		not consider in	dex			
EducationField	Age Attrition	BusinessTrav	el DailyRate	\		
Life Sciences Life Sciences Life Sciences	49 No 32 No 30 No	Travel_Frequent Travel_Frequent Travel_Rare	ly 1005			
Education Et al.d.	De	partment Distan	ceFromHome Ed	ucation \		
EducationField Life Sciences Life Sciences Life Sciences	Research & Dev Research & Dev Research & Dev	elopment	8 2 24	1 2 1		
	EmployeeCount	EmployeeNumber	EnvironmentSa	tisfaction		
\ 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 1		2	80			
	TotalWorkingYe	ears TrainingTime	sLastYear			

```
WorkLifeBalance \
EducationField
Life Sciences
                                10
                                                       3
Life Sciences
                                                       2
Life Sciences
                                                       2
               YearsAtCompany YearsInCurrentRole
YearsSinceLastPromotion \
EducationField
Life Sciences
                           10
                                                 7
Life Sciences
                                                 7
Life Sciences
                                                 0
                YearsWithCurrManager
EducationField
Life Sciences
                                    7
Life Sciences
                                    6
                                    0
Life Sciences
[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
4
                   591
                                        2
         No
5
                                        2
         No
                   1005
2
        Yes
                  1373
                                        2
ibm_hr2.iloc[[1,4,5,2],[1,3,5]]
               Attrition DailyRate DistanceFromHome
EducationField
Life Sciences
                       No
                                 279
                                                      8
                                                      2
Medical
                       No
                                 591
                                                      2
Life Sciences
                       No
                                1005
                                                      2
0ther
                      Yes
                                1373
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
                              0
Department
                              0
DistanceFromHome
                              0
Education
EducationField
                              0
                              0
EmployeeCount
                              0
EmployeeNumber
EnvironmentSatisfaction
                              0
                              0
Gender
                              0
HourlyRate
JobInvolvement
                              0
                              0
JobLevel
                              0
JobRole
JobSatisfaction
                              0
MaritalStatus
                              0
```

	_
MonthlyIncome	0
MonthlyRate	0
NumCompaniesWorked	0
Over18	_
	0
0verTime	0
PercentSalaryHike	0
PerformanceRating	0
RelationshipSatisfaction	0
StandardHours	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	Õ
	•
YearsInCurrentRole	0
YearsSinceLastPromotion	0
YearsWithCurrManager	0
dtype: int64	
acyper incor	

# Subsetting based on condition (and & or)

[1] and [2] or

<pre>ibm_hr.head()</pre>								
	Age /	Attrition	Busi	inessTr	avel	DailyRate		Department
0	41	Yes	Tra	avel_Ra	rely	1102		Sales
1	49	No	Travel_	_Freque	ntly	279	Research	& Development
2	37	Yes	Tra	avel_Ra	rely	1373	Research	& Development
3	33	No	Travel_	_Freque	ntly	1392	Research	& Development
4	27	No	Tra	avel_Ra	rely	591	Research	& Development
Emr		anceFromHom eNumber \	ne Educ	ation	Educat	tionField	EmployeeCo	ount
0			1	2	Life	Sciences		1
1 1 2			8	1	Life	Sciences		1
2			2	2		0ther		1
4 3 5			3	4	Life	Sciences		1
4			2	1		Medical		1

7		
RelationshipS 0 1 2 3 4	atisfaction StandardHours  1 80 4 80 2 80 3 80 4 80	0 1 0 0
TotalWorkingYears YearsAtCompany \	TrainingTimesLastYear Wo	rkLifeBalance
9 8 6	0	1
1 10	3	3
10 2 7	3	3
0 3 8	3	3
8 4 6 2	3	3
YearsInCurrentRole 0 4 1 7 2 0 3 7 4 2	YearsSinceLastPromotion  0  1  0  3 2	YearsWithCurrManager 5 7 0 0 2
[5 rows x 35 columns]		

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

• Syntax:

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

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

• Syntax:

```
ibm_hr[() | () | ()]
ibm_hr[() | () | ()]
                                           Traceback (most recent call
TypeError
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
                                          1373 Research & Development
       37
                Yes Travel Rarely
       27
                 No Travel Rarely
                                           591
                                                Research & Development
       59
                    Travel Rarely
                                          1324
                                                Research & Development
                 No
       30
                 No Travel Rarely
                                          1358
                                                Research & Development
1462
       39
                 No
                    Travel Rarely
                                           722
                                                                  Sales
1464
                                                                  Sales
       26
                    Travel Rarely
                                          1167
                 No
1466
       39
                 No Travel Rarely
                                           613 Research & Development
1467
       27
                     Travel Rarely
                                           155
                                                Research & Development
                 No
1469
       34
                    Travel Rarely
                                           628
                                                Research & Development
                 No
      DistanceFromHome
                        Education EducationField
                                                   EmployeeCount
                                   Life Sciences
                     1
                                 2
                                                               1
2
                     2
                                 2
                                            0ther
                                                               1
4
                     2
                                 1
                                          Medical
                                                               1
6
                     3
                                 3
                                          Medical
                                                               1
```

7 1462 1464 1466 1467 1469	24  24 5 6 4	1 Life Sciences 1 Marketing 3 Other 1 Medical 3 Life Sciences 3 Medical	1 1
0 2 4 6 7  1462 1464 1466 1467 1469	EmployeeNumber 1 4 7 10 11 2056 2060 2062 2064 2068	RelationshipSatisfact	ion StandardHours \ 1     80 2     80 4     80 1     80 2     80     1     80 4     80 1     80 2     80 1     80 1     80 1     80 2     80 1     80
0 2 4 6 7  1462 1464 1466 1467 1469	StockOptionLevel Tot 0 0 1 1 3 1 1 0 1 0 0	Train 8 7 6 12 1 21 5 9 6 6	ingTimesLastYear \
0 2 4 6 7  1462 1464 1466 1467 1469	WorkLifeBalance Years 1 3 3 2 3 2 3 3 4	AtCompany YearsInCurr 6 0 2 1 20 4 7 6 4	entRole \
0 2	YearsSinceLastPromoti	on YearsWithCurrMana 0 0	ger 5 0

```
4
                              2
                                                     2
6
                              0
                                                     0
7
                              0
                                                     0
1462
                             9
                                                     6
1464
                              0
                                                     0
                              1
                                                     7
1466
1467
                              0
                                                     3
1469
                              1
[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
                      Travel Rarely
                                           1282
                                                  Research & Development
                  No
31
       44
                  No
                      Travel Rarely
                                           1459
                                                  Research & Development
40
                     Travel Rarely
                                                  Research & Development
       35
                  No
                                            464
77
       45
                  No
                      Travel Rarely
                                            193
                                                  Research & Development
1408
       23
                     Travel_Rarely
                                            571
                                                  Research & Development
                  No
                                                  Research & Development
1413
       25
                  No
                      Travel Rarely
                                            977
1433
       25
                  No
                     Travel Rarely
                                           1382
                                                                    Sales
1459
       29
                  No
                      Travel Rarely
                                           1378
                                                  Research & Development
1464
       26
                      Travel Rarely
                                                                    Sales
                  No
                                            1167
      DistanceFromHome
                         Education EducationField
                                                     EmployeeCount
2
                                  2
                                              0ther
                                                                  1
25
                      5
                                  3
                                              0ther
                                                                  1
31
                     10
                                  4
                                              0ther
                                                                  1
                                  2
                                                                  1
40
                      4
                                              0ther
                                  4
                                                                  1
77
                      6
                                              0ther
. . .
                                  2
1408
                     12
                                              0ther
                                                                  1
1413
                      2
                                  1
                                              0ther
                                                                  1
                                  2
1433
                      8
                                              0ther
                                                                  1
```

1459 1464	13 5	2 3	Other Other		1 1
2 25 31 40 77  1408 1413 1433 1459 1464	EmployeeNumber 4 32 40 53 101 1982 1992 2018 2053 2060	Relationshi	ipSatisfacti	ion Standard 2 4 4 3 2  3 3 2 1 4	Hours \     80     80     80     80      80     80     80     80     80     80     80     80
2 25 31 40 77  1408 1413 1433 1459 1464	StockOptionLevel To 0 1 0 1 0 0 1 1 1 1 1	talWorkingYe	ears Traini 7 26 9 1 17  5 7 6 10	ingTimesLast	Year \     3     3     5     3      6     2     3     2     2
2 25 31 40 77  1408 1413 1433 1459 1464	orkLifeBalance Year 3 2 4 3 4 4 2 2 3 3 3	sAtCompany \\	⁄earsInCurre	entRole \     0     13     2     0     0      2     2     3     3     2	
2 25 31 40 77  1408	YearsSinceLastPromot	ion YearsWi 0 4 1 0 0	ithCurrManag	ger 0 8 3 0 0	

```
1413
                             0
                                                     2
                                                     4
1433
                             0
                                                     3
1459
                             0
                                                     0
1464
                             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()
                             0
Age
                             0
Attrition
BusinessTravel
                             0
                             0
DailyRate
                             0
Department
                             0
DistanceFromHome
                             0
Education
                             0
EducationField
EmployeeCount
                             0
EmployeeNumber
                             0
                             0
EnvironmentSatisfaction
Gender
                             0
                             0
HourlyRate
JobInvolvement
                             0
                             0
JobLevel
JobRole
                             0
JobSatisfaction
                             0
                             0
MaritalStatus
                             0
MonthlyIncome
                             0
MonthlyRate
NumCompaniesWorked
                             0
0ver18
                             0
```

```
0verTime
                             0
PercentSalaryHike
                             0
PerformanceRating
                             0
RelationshipSatisfaction
                             0
                             0
StandardHours
StockOptionLevel
                             0
                             0
TotalWorkingYears
TrainingTimesLastYear
                             0
WorkLifeBalance
                             0
                             0
YearsAtCompany
YearsInCurrentRole
                             0
                             0
YearsSinceLastPromotion
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)
                                           Traceback (most recent call
NameError
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
         Non-Travel
1441
         Non-Travel
1447
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
        Non-Travel
1441
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
        No Travel
20
        No Travel
46
        No Travel
53
83
        No Travel
1434
        No Travel
1437
        No Travel
        No Travel
1441
1447
        No Travel
        No Travel
1463
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'l
ibm hr['BusinessTravel'].unique()
```

Some more cocepts are there like below

- groupby,
- merge/join,
- concat