## Split the data from Raw data to Clean data using Pandas

```
import pandas as pd
In [1]:
         import numpy as np
In [2]:
         data = pd.read_excel(r'C:\Users\Gopi Reddy\Downloads\Address1 (1).xlsx') # Read
Out[2]:
                                         Address
                                                          Date
                                                                         Time
              12thblock, Hyderabad, Telengana, 500085
                                                  15th jan 2025 6hr 55min 7sec
                                                  24th feb 1992 8hr 30min 4sec
            29throad, Secundrabad, Telengana, 500014
         2
              15thblock, hyderabad, Telengana, 500048
                                                  16th jan 2025 6hr 55min 7sec
         3
                                                  25th feb 1992 8hr 30min 4sec
                  6thblock,srnagar,Telengana,508242
         4
                    19thblock,jntu,Telengana,546128 17th jan 2025 6hr 55min 7sec
         print(type(data)) # prints the which type of data
       <class 'pandas.core.frame.DataFrame'>
In [4]: data.isnull() # Shows the any null values are present in the data
Out[4]:
            Address Date Time
         0
               False
                      False
                            False
                            False
               False
                      False
         2
                            False
               False
                      False
         3
               False
                      False
                            False
               False
                     False
                            False
         data.shape # show the dimensions of the data
Out[5]: (5, 3)
In [6]:
         len(data) # length of data
Out[6]: 5
         data.columns # shows the attributes of the data
Out[7]: Index(['Address', 'Date', 'Time'], dtype='object')
In [8]: data.dtypes
```

```
object
 Out[8]: Address
          Date
                      object
          Time
                      object
          dtype: object
 In [9]: data.info() # it will show the total information about the data
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 5 entries, 0 to 4
        Data columns (total 3 columns):
              Column
                      Non-Null Count Dtype
                       -----
              Address 5 non-null
         a
                                        object
          1
              Date
                       5 non-null
                                        object
         2
              Time
                       5 non-null
                                        object
        dtypes: object(3)
        memory usage: 252.0+ bytes
In [10]:
         data.info
Out[10]:
          <bound method DataFrame.info of</pre>
                                                                                Address
          Date
              12thblock, Hyderabad, Telengana, 500085 15th jan 2025 6hr 55min 7sec
          0
             29throad, Secundrabad, Telengana, 500014 24th feb 1992 8hr 30min 4sec
              15thblock, hyderabad, Telengana, 500048
                                                       16th jan 2025
                                                                       6hr 55min 7sec
          3
                  6thblock, srnagar, Telengana, 508242
                                                       25th feb 1992 8hr 30min 4sec
          4
                    19thblock, jntu, Telengana, 546128
                                                       17th jan 2025 6hr 55min 7sec>
          data.isnull().sum()
In [11]:
Out[11]:
          Address
                      0
          Date
                      0
                      0
          Time
          dtype: int64
In [12]:
          data
Out[12]:
                                         Address
                                                          Date
                                                                         Time
          0
              12thblock, Hyderabad, Telengana, 500085
                                                   15th jan 2025
                                                                6hr 55min 7sec
             29throad, Secundrabad, Telengana, 500014
                                                  24th feb 1992
                                                                8hr 30min 4sec
          2
               15thblock, hyderabad, Telengana, 500048
                                                   16th jan 2025
                                                                6hr 55min 7sec
          3
                  6thblock,srnagar,Telengana,508242
                                                  25th feb 1992
                                                                8hr 30min 4sec
                    19thblock,jntu,Telengana,546128
                                                   17th jan 2025 6hr 55min 7sec
In [13]:
          data[:1] # slicing
Out[13]:
                                        Address
                                                         Date
                                                                       Time
          0 12thblock, Hyderabad, Telengana, 500085 15th jan 2025 6hr 55min 7sec
```

## .iloc[]:

It stands for "integer location" and is used to select data based on integer position.

You use it to select rows and columns by their index positions (i.e., integer-based indexing).

Example: df.iloc[0, 1] will select the element in the first row and second column of the DataFrame.

```
In [14]: data.iloc[:,1] # Gives total information about the Date column
Out[14]: 0
                15th jan 2025
          1
                24th feb 1992
               16th jan 2025
               25th feb 1992
               17th jan 2025
          Name: Date, dtype: object
In [15]: data.iloc[:,2]
                6hr 55min 7sec
Out[15]: 0
                8hr 30min 4sec
               6hr 55min 7sec
               8hr 30min 4sec
               6hr 55min 7sec
          Name: Time, dtype: object
In [16]:
          data
Out[16]:
                                          Address
                                                           Date
                                                                          Time
          0
               12thblock, Hyderabad, Telengana, 500085
                                                   15th jan 2025 6hr 55min 7sec
             29throad, Secundrabad, Telengana, 500014
                                                   24th feb 1992 8hr 30min 4sec
          2
               15thblock, hyderabad, Telengana, 500048
                                                   16th jan 2025 6hr 55min 7sec
          3
                   6thblock,srnagar,Telengana,508242
                                                   25th feb 1992 8hr 30min 4sec
                     19thblock, intu, Telengana, 546128 17th jan 2025 6hr 55min 7sec
          4
          data[['hours','mins','sec']]=data['Time'].str.split('[]',expand=True) # here sp
          data.head()
```

Out[17]:		Address	[	Date	Time	hours	mir	ns s	sec
	0	12 th block, Hyderabad, Telengana, 500085		n jan 2025	6hr 55mir 7sec	bni	55mi	in 7s	sec
	1	29throad, Secundrabad, Telengana, 500014		n feb 1992	8hr 30mir 4sec	8hi	30mi	n 4s	sec
	2	15thblock, hyderabad, Telengana, 500048		n jan 2025	6hr 55mir 7sec	bni	55mi	in 7s	sec
	3	6thblock,srnagar,Telengana,508242		n feb 1992	8hr 30mir 4sec	۸nı	30mi	n 4s	sec
	4	19thblock, jntu, Telengana, 546128		n jan 2025	6hr 55mir 7sec	bni	55mi	in 7s	sec
In [18]:		ta[['day','month','year']]=data['Da ta.head()	te'].s	str.spl	.it('[ ]'	,expand	l=True)	) # H	ere sp
Out[18]:		Address	Date	Time	hours	mins	sec	day	month
	0	12thblock, Hyderabad, Telengana, 500085	15th jan 2025	6hr 55min 7sec	6hr	55min	7sec	15th	jar
	1	29throad, Secundrabad, Telengana, 500014	24th feb 1992	8hr 30min 4sec	8hr	30min	4sec	24th	fek
	2	15thblock, hyderabad, Telengana, 500048	16th jan 2025	6hr 55min 7sec	6hr	55min	7sec	16th	jar
	3	6thblock, srnagar, Telengana, 508242	25th feb 1992	8hr 30min 4sec	8hr	30min	4sec	25th	fek
	4	19thblock, jntu, Telengana, 546128	17th jan 2025	6hr 55min 7sec	6hr	55min	7sec	17th	jar
	4								<b>&gt;</b>
In [19]:	da	ta[['block','city', 'state', 'pinco	de']]=	data['	Address'	].str.s	plit(r	٠'[,	]',exp
In [20]:	da	ta.head()							

Out[20]:		Address	Date	Time	hours	mins	sec	day	month
	0	12 th block, Hyderabad, Telengana, 500085	15th jan 2025	6hr 55min 7sec	6hr	55min	7sec	15th	jar
	1	29throad, Secundrabad, Telengana, 500014	24th feb 1992	8hr 30min 4sec	8hr	30min	4sec	24th	fek
	2	15 th block, hyderabad, Telengana, 500048	16th jan 2025	6hr 55min 7sec	6hr	55min	7sec	16th	jar
	3	6th block, srnagar, Telengana, 508242	25th feb 1992	8hr 30min 4sec	8hr	30min	4sec	25th	fek
	4	19thblock, jntu, Telengana, 546128	17th jan 2025	6hr 55min 7sec	6hr	55min	7sec	17th	jar
	4								•
In [21]:	da	ta							
Out[21]:		Address	Date	Time	hours	mins	sec	day	month
Out[21]:	0	Address  12thblock, Hyderabad, Telengana, 500085	15th jan 2025	6hr 55min 7sec		mins 55min			jar
Out[21]:	0		15th jan	6hr 55min	6hr		7sec	15th	
Out[21]:		12thblock, Hyderabad, Telengana, 500085	15th jan 2025 24th feb	6hr 55min 7sec 8hr 30min	6hr 8hr	55min	7sec 4sec	15th 24th	jar
Out[21]:	1	12thblock, Hyderabad, Telengana, 500085 29throad, Secundrabad, Telengana, 500014	15th jan 2025 24th feb 1992 16th jan	6hr 55min 7sec 8hr 30min 4sec 6hr 55min	6hr 8hr 6hr	55min 30min	7sec 4sec 7sec	15th 24th 16th	jar fek
Out[21]:	1 2	12thblock, Hyderabad, Telengana, 500085 29throad, Secundrabad, Telengana, 500014 15thblock, hyderabad, Telengana, 500048	15th jan 2025 24th feb 1992 16th jan 2025 25th feb	6hr 55min 7sec 8hr 30min 4sec 6hr 55min 7sec 8hr 30min	6hr 8hr 6hr	55min 30min 55min	7sec 4sec 4sec 4sec	15th 24th 16th	jar fek jar
Out[21]:	1 2 3	12thblock, Hyderabad, Telengana, 500085 29throad, Secundrabad, Telengana, 500014 15thblock, hyderabad, Telengana, 500048 6thblock, srnagar, Telengana, 508242	15th jan 2025 24th feb 1992 16th jan 2025 25th feb 1992 17th jan	6hr 55min 7sec 8hr 30min 4sec 6hr 55min 7sec 8hr 30min 4sec 6hr 55min	6hr 8hr 6hr	55min 30min 55min 30min	7sec 4sec 4sec 4sec	15th 24th 16th	jar fek jar

Out[22]:		block	city	state	pincode	day	month	year	hours	mins	sec
	0	12thblock	Hyderabad	Telengana	500085	15th	jan	2025	6hr	55min	7sec
	1	29throad	Secundrabad	Telengana	500014	24th	feb	1992	8hr	30min	4sec
	2	15thblock	hyderabad	Telengana	500048	16th	jan	2025	6hr	55min	7sec
	3	6thblock	srnagar	Telengana	508242	25th	feb	1992	8hr	30min	4sec
	4	19thblock	jntu	Telengana	546128	17th	jan	2025	6hr	55min	7sec
	4										<b></b>
In [23]:	ne	w_data.to_	csv("clean_d	ata.csv")	# convert	the	new data	ı to a	csv fi	Le	