

In [1]: `import pandas as pd`

In [3]: `data = pd.read_excel(r"C:\Users\AKSHAY\Downloads\TASK -- convert raw data - clea`

In [5]: `data`

Out[5]:

	ADDR	DT	NAME	time
0	45 rd, kenith street, btm, bangalore, 500038	2022-04-03	abc	8:30:21
1	45 rd, street, hitech, Hd, 500038	2022-04-02	dec	5:30:21

CLEANING THE ADDR VALUES

In [8]: `data['ADDR']`

Out[8]:

```
0    45 rd, kenith street, btm, bangalore, 500038
1              45 rd, street, hitech, Hd, 500038
Name: ADDR, dtype: object
```

In [12]: *# WE CANNOT DIRECTLY PERFORM THE SPLIT OPERATION TO THE DATA, SO WE NEED TO USE*

In [22]: `addr_split = data['ADDR'].str.split(',', expand = True)`

In [24]: `addr_split`

Out[24]:

	0	1	2	3	4
0	45 rd	kenith street	btm	bangalore	500038
1	45 rd	street	hitech	Hd	500038

In [51]:

```
data['House Address'] = addr_split[0].str.strip()
data['Street'] = addr_split[1].str.strip()
data['City'] = addr_split[2].str.strip().str.strip()
data['State'] = addr_split[3].str.strip()
data['Pincode'] = addr_split[4].str.strip()

data = data.drop('ADDR', axis = 1)
```

In [53]: `data`

Out[53]:

	DT	NAME	time	House Address	Street	City	State	Pincode
0	2022-04-03	abc	8:30:21	45 rd	kenith street	btm	bangalore	500038
1	2022-04-02	dec	5:30:21	45 rd	street	hitech	Hd	500038

CLEANING THE DATE VALUES

```
In [56]: data['DT']
```

```
Out[56]: 0    2022-04-03  
         1    2022-04-02  
         Name: DT, dtype: datetime64[ns]
```

```
In [58]: data['DT'] = pd.to_datetime(data['DT'])
```

```
In [62]: data['Year'] = data['DT'].dt.year  
         data['Month'] = data['DT'].dt.month  
         data['Day'] = data['DT'].dt.day  
  
         data = data.drop('DT', axis = 1)
```

```
In [64]: data
```

```
Out[64]:
```

	NAME	time	House Address	Street	City	State	Pincode	Year	Month	Day
0	abc	8:30:21	45 rd	kenith street	btm	bangalore	500038	2022	4	3
1	dec	5:30:21	45 rd	street	hitech	Hd	500038	2022	4	2

CLEANING THE TIME VALUES

```
In [67]: data['time']
```

```
Out[67]: 0    8:30:21  
         1    5:30:21  
         Name: time, dtype: object
```

```
In [69]: data['time'] = pd.to_timedelta(data['time'])
```

```
In [71]: data['Hours'] = data['time'].dt.components.hours  
         data['Minutes'] = data['time'].dt.components.minutes  
         data['Seconds'] = data['time'].dt.components.seconds  
  
         data = data.drop('time', axis = 1)
```

```
In [73]: data
```

```
Out[73]:
```

	NAME	House Address	Street	City	State	Pincode	Year	Month	Day	Hours	Min
0	abc	45 rd	kenith street	btm	bangalore	500038	2022	4	3	8	
1	dec	45 rd	street	hitech	Hd	500038	2022	4	2	5	

In [75]: `data.columns`

Out[75]: Index(['NAME', 'House Address', 'Street', 'City', 'State', 'Pincode', 'Year', 'Month', 'Day', 'Hours', 'Minutes', 'Seconds'], dtype='object')

In [79]: `data = data[['House Address', 'Street', 'City', 'State', 'Pincode', 'NAME', 'Year', 'Month', 'Day', 'Hours', 'Minutes', 'Seconds']]`

Out[79]:

	House Address	Street	City	State	Pincode	NAME	Year	Month	Day	Hours	Min
--	---------------	--------	------	-------	---------	------	------	-------	-----	-------	-----

0	45 rd	kenith street	btm	bangalore	500038	abc	2022	4	3	8
---	-------	---------------	-----	-----------	--------	-----	------	---	---	---

1	45 rd	street	hitech	Hd	500038	dec	2022	4	2	5
---	-------	--------	--------	----	--------	-----	------	---	---	---

