import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt

task=pd.read\_excel("TASK -- convert raw data - clean data - Copy.xlsx") # Read excel into python jupyter  
task

Unnamed: 0 Unnamed: 1 \  
0 NaN THIS IS NOT CLEAN DATA   
1 NaN NaN   
2 NaN NaN   
3 NaN NaN   
  
 Unnamed: 2 Unnamed: 3 \  
0 NaN NaN   
1 ADDR DT   
2 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00   
3 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00   
  
 Unnamed: 4 Unnamed: 5   
0 NaN NaN   
1 NAME time   
2 abc 8:30:21   
3 dec 5:30:21

task=pd.read\_excel("TASK -- convert raw data - clean data - Copy.xlsx",header=None) # Assigning header as none  
task

0 1 2 \  
0 NaN NaN NaN   
1 NaN THIS IS NOT CLEAN DATA NaN   
2 NaN NaN ADDR   
3 NaN NaN 45 rd, kenith street, btm, bangalore 500038   
4 NaN NaN 45 rd, street, hitech, Hd 500038   
  
 3 4 5   
0 NaN NaN NaN   
1 NaN NaN NaN   
2 DT NAME time   
3 2022-04-03 00:00:00 abc 8:30:21   
4 2022-04-02 00:00:00 dec 5:30:21

task.columns

Int64Index([0, 1, 2, 3, 4, 5], dtype='int64')

task.columns=["no data", "no data1","Adress","Date","Name", "Time"] # Renaming the columns  
task

no data no data1 \  
0 NaN NaN   
1 NaN THIS IS NOT CLEAN DATA   
2 NaN NaN   
3 NaN NaN   
4 NaN NaN   
  
 Adress Date Name \  
0 NaN NaN NaN   
1 NaN NaN NaN   
2 ADDR DT NAME   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time   
0 NaN   
1 NaN   
2 time   
3 8:30:21   
4 5:30:21

task.index.name="S.No" # Adding serial number  
task

no data no data1 \  
S.No   
0 NaN NaN   
1 NaN THIS IS NOT CLEAN DATA   
2 NaN NaN   
3 NaN NaN   
4 NaN NaN   
  
 Adress Date Name \  
S.No   
0 NaN NaN NaN   
1 NaN NaN NaN   
2 ADDR DT NAME   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time   
S.No   
0 NaN   
1 NaN   
2 time   
3 8:30:21   
4 5:30:21

new\_task=pd.DataFrame(task)  
new\_task

no data no data1 \  
S.No   
0 NaN NaN   
1 NaN THIS IS NOT CLEAN DATA   
2 NaN NaN   
3 NaN NaN   
4 NaN NaN   
  
 Adress Date Name \  
S.No   
0 NaN NaN NaN   
1 NaN NaN NaN   
2 ADDR DT NAME   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time   
S.No   
0 NaN   
1 NaN   
2 time   
3 8:30:21   
4 5:30:21

new\_task1=new\_task.drop(["no data","no data1"],axis=1) #Dropping unnecessary cols

new\_task1

Adress Date Name \  
S.No   
0 NaN NaN NaN   
1 NaN NaN NaN   
2 ADDR DT NAME   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time   
S.No   
0 NaN   
1 NaN   
2 time   
3 8:30:21   
4 5:30:21

new\_task2=new\_task1.drop([0,1,2]) # Dropping unwanted first three rows

new\_task2

Adress Date Name \  
S.No   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time   
S.No   
3 8:30:21   
4 5:30:21

new\_task2.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2 entries, 3 to 4  
Data columns (total 4 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Adress 2 non-null object  
 1 Date 2 non-null object  
 2 Name 2 non-null object  
 3 Time 2 non-null object  
dtypes: object(4)  
memory usage: 196.0+ bytes

new\_task2[["House Add","Street","City","City\_Fullname"]]=new\_task2.Adress.str.split(",",expand=True)  
new\_task2

Adress Date Name \  
S.No   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time House Add Street City City\_Fullname   
S.No   
3 8:30:21 45 rd kenith street btm bangalore 500038   
4 5:30:21 45 rd street hitech Hd 500038

new\_task2

Adress Date Name \  
S.No   
3 45 rd, kenith street, btm, bangalore 500038 2022-04-03 00:00:00 abc   
4 45 rd, street, hitech, Hd 500038 2022-04-02 00:00:00 dec   
  
 Time House Add Street City City\_Fullname   
S.No   
3 8:30:21 45 rd kenith street btm bangalore 500038   
4 5:30:21 45 rd street hitech Hd 500038

new\_task3=new\_task2.drop(['Adress'],axis=True)  
new\_task3

Date Name Time House Add Street City \  
S.No   
3 2022-04-03 00:00:00 abc 8:30:21 45 rd kenith street btm   
4 2022-04-02 00:00:00 dec 5:30:21 45 rd street hitech   
  
 City\_Fullname   
S.No   
3 bangalore 500038   
4 Hd 500038

new\_task3

Date Name Time House Add Street City \  
S.No   
3 2022-04-03 00:00:00 abc 8:30:21 45 rd kenith street btm   
4 2022-04-02 00:00:00 dec 5:30:21 45 rd street hitech   
  
 City\_Fullname   
S.No   
3 bangalore 500038   
4 Hd 500038

new\_task3.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2 entries, 3 to 4  
Data columns (total 7 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Date 2 non-null object  
 1 Name 2 non-null object  
 2 Time 2 non-null object  
 3 House Add 2 non-null object  
 4 Street 2 non-null object  
 5 City 2 non-null object  
 6 City\_Fullname 2 non-null object  
dtypes: object(7)  
memory usage: 244.0+ bytes

new\_task3['Date'] = pd.to\_datetime(new\_task3['Date']) #converting object to date data type

new\_task3.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2 entries, 3 to 4  
Data columns (total 10 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Date 2 non-null datetime64[ns]  
 1 Name 2 non-null object   
 2 Time 2 non-null object   
 3 House Add 2 non-null object   
 4 Street 2 non-null object   
 5 City 2 non-null object   
 6 City\_Fullname 2 non-null object   
 7 Year 2 non-null int64   
 8 Month 2 non-null int64   
 9 DayOfMonth 2 non-null int64   
dtypes: datetime64[ns](1), int64(3), object(6)  
memory usage: 292.0+ bytes

new\_task3["Year"]=pd.DatetimeIndex(new\_task3['Date']).year  
  
new\_task3["Month"]=pd.DatetimeIndex(new\_task3['Date']).month  
new\_task3["DayOfMonth"]=pd.DatetimeIndex(new\_task3['Date']).day

new\_task3["Year"]=pd.DatetimeIndex(new\_task3['Date']).year  
new\_task3

Date Name Time House Add Street City \  
S.No   
3 2022-04-03 abc 8:30:21 45 rd kenith street btm   
4 2022-04-02 dec 5:30:21 45 rd street hitech   
  
 City\_Fullname Year Month DayOfMonth   
S.No   
3 bangalore 500038 2022 4 3   
4 Hd 500038 2022 4 2

new\_task3[["Hrs","Mins","Secs"]]=new\_task3.Time.str.split(":",expand=True)  
new\_task3

Date Name Time House Add Street City \  
S.No   
3 2022-04-03 abc 8:30:21 45 rd kenith street btm   
4 2022-04-02 dec 5:30:21 45 rd street hitech   
  
 City\_Fullname Year Month DayOfMonth Hrs Mins Secs   
S.No   
3 bangalore 500038 2022 4 3 8 30 21   
4 Hd 500038 2022 4 2 5 30 21

new\_task4=new\_task3.drop(['Time'],axis=True)

new\_task5=new\_task4.drop(['Date'],axis=True)  
new\_task5

Name House Add Street City City\_Fullname Year Month \  
S.No   
3 abc 45 rd kenith street btm bangalore 500038 2022 4   
4 dec 45 rd street hitech Hd 500038 2022 4   
  
 DayOfMonth Hrs Mins Secs   
S.No   
3 3 8 30 21   
4 2 5 30 21