Name : Tejaswini Sunil Mahale
Practical No.4
Roll No.26 Sub:-DV

Program/Notebook to deal with missing values, Data Reshping, Filtering, Filtering Data, Merging Data

1) Dealing with missing values

```
#Importing Required Libraries
import pandas as pd
import numpy as np

#Reding excel File
df = pd.read_excel("Fees_Data.xlsx")
df.head(8)
```

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	File Name
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	EN23204195_SHINDE GAURAV PRAKASH
1	2	EN23146043	Chaudhari Prasad mahesh	IT	NaN	NaN	2023-04-08 00:00:00	UPI	11.40 am	EN23146043_Chaudhari Prasad mahesh
2	3	EN23135942	Borse Gunwant Ashok	NaN	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	EN23135942_Borse Gunwant Ashok
3	4	EN23135942	Borse Gunwant Ashok	Computer	NaN	25000	2023-04-08 00:00:00	Credit Card	12:00:00	EN23135942_Borse Gunwant Ashok
4	5	EN23119584	Hire Tejas	Computer	53778.0	53778	2023-04-08	UPI	12:10:00	EN23119584_Hire Tejas

```
#Replacing NaN Values with other
df['Fees Paid'].fillna(10000, inplace=True)
print(df['Fees Paid'].head(4))
```

```
0    10056
1    10000
2    28778
3    25000
Name: Fees Paid, dtype: object
```

Data Reshaping

```
#Importing Required Libraries
import pandas as pd
data = pd.read_excel('Fees_Data.xlsx')
data.head()
```

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	NaN	NaN	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	NaN	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No

```
data['Status'] = data['Status'].map({'Yes':1, 'No':0}).astype(int)
#data['out'] = data['out'].map({'Yes': 1, 'No': 0}).astype(int)
data.head()
```

				CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	1
	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH							
1	2	EN23146043	Chaudhari Prasad mahesh	IT	NaN	NaN	2023-04-08 00:00:00	UPI	11.40 am	1
							2023-04-08		11.50	

Filtering Data

```
import pandas as pd

df = pd.read_excel('Cricket.xlsx')
df.head()
```

	id	name	age	Runs	out
0	1	Virat	34	45	Yes
1	2	Rohit	36	87	Yes
2	3	Hardik	30	40	Yes
3	4	Bumrah	39	12	Yes
4	5	Jadeja	32	24	Yes

```
#Filtering Top runs Scroring Players
df = df[df['Runs'] >= 40]
df
```

	id	name	age	Runs	out
0	1	Virat	34	45	Yes
1	2	Rohit	36	87	Yes
2	3	Hardik	30	40	Yes
5	6	Gill	23	56	No
6	7	Kishan	24	46	Yes
9	10	Iyer	29	54	No

```
#Removing age Column
df = df.drop(['age'], axis=1)
df
```

	id	name	Runs	out
0	1	Virat	45	Yes
1	2	Rohit	87	Yes
2	3	Hardik	40	Yes
5	6	Gill	56	No
6	7	Kishan	46	Yes
9	10	Iyer	54	No

Merging Data

```
import pandas as pd

Data1 = pd.read_excel('Cricket.xlsx')
Data1.head()
```

	id	name	age	Runs	out
0	1	Virat	34	45	Yes
1	2	Rohit	36	87	Yes
2	3	Hardik	30	40	Yes
3	4	Bumrah	39	12	Yes
4	5	Jadeja	32	24	Yes

```
Data2 =pd.read_excel('Cricket2.xlsx')
Data2.head()
```

	id	Sixes	Fours
0	1	3	5
1	2	4	4
2	3	1	3
3	4	2	2
4	5	1	1

```
Data3 = pd.merge(Data1, Data2, on = 'id')
Data3.head()
```

	id	name	age	Runs	out	Sixes	Fours
0	1	Virat	34	45	Yes	3	5
1	2	Rohit	36	87	Yes	4	4
2	3	Hardik	30	40	Yes	1	3
3	4	Bumrah	39	12	Yes	2	2
4	5	Jadeja	32	24	Yes	1	1

Subsetting DataFrames in Pandas

```
#Import Reuquired Libraries
import pandas as pd
```

```
df = pd.read_excel("Book1.xlsx")
df
```

	Sr.	EnNo.	Name of Student	CA-I (10marks)	CA-II (10marks)	Midterm (20marks)	UT1	OEA	Midsem split
0	1	2054491246001	AAKANKSHA ANIL SALUNKE	8	8	18	6	8	6
1	2	2054491246002	ABHINASH KAILASH JOSHI	8	9	18	7	8	4
2	3	2054491246003	ADITYA JITENDRA MALI	6	8	14	6	8	4
3	4	2054491246004	AISHWARYA AVINASH PATIL	9	9	19	8	8	5
4	5	2054491246005	AKSHAY AVINASH PATIL	3	3	9	8	8	5
...	...	...	...	...	...	...	...	...	...
62	63	2154491246502	LATIKA PRAVIN BHANDARI	8	8	15	7	10	6
63	64	2154491246503	CHETAN VILAS KHAIRNAR	8	8	19	6	8	5
64	65	2154491246504	DIVYA LALCHAND CHAUDHARI	8	9	18	7	8	5
65	66	2154491246505	HARSHIT ARUN GUJARATHI	10	7	18	5	8	6
66	67	2154491246506	PRANAV SWAPNIL SONAWANE	6	7	11	7	8	3

67 rows × 9 columns

#Creating subset of 1st to 5th rows and columnSr, Enrollment No. , Name of Student
df.iloc[1:5, 0:3]

	Sr.	Enrollment No.	Name of Student
1	2	2054491246002	ABHINASH KAILASH JOSHI
2	3	2054491246003	ADITYA JITENDRA MALI
3	4	2054491246004	AISHWARYA AVINASH PATIL
4	5	2054491246005	AKSHAY AVINASH PATIL

#Creating subset dataframe using indexing operator
df[['Sr.', 'Enrollment No.', 'OEA']].head()

	Sr.	Enrollment No.	OEA
0	1	2054491246001	8
1	2	2054491246002	8
2	3	2054491246003	8
3	4	2054491246004	8
4	5	2054491246005	8

#Creating a subset o columns EnNo. and Name of Student
#Using Filter method()
df.filter(like=En)

Reshaping the Data and Pivot

The pivot(index, columns, values) function produces pivot table based on 3 columns of the DataFrame

import pandas as pd

df = pd.read_excel('Fees_Data.xlsx')
df

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	NaN	NaN	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	NaN	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No
3	4	EN23135942	Borse Gunwant Ashok	Computer	NaN	25000	2023-04-08 00:00:00	Credit Card	12:00:00	No
4	5	EN23119584	Hire Tejas Ravindra	Computer	53778.0	53778	2023-04-08 00:00:00	UPI	12:10:00	Yes
...	...	...	...	...	...	...	...	...	...	...
66	67	EN23273933	Ahirrao Leeleshwar Subhash	Civil	53778.0	40000	16/8/2015	UPI	17:00:00	No
67	68	EN23200937	Gujar Mohit Yatin	Civil	53778.0	25000	17/8/2015	UPI	11.58 am	No

df.pivot(index='EN No.', columns= 'Branch', values= 'Fees Paid')

Branch	NaN	CIVIL	Civil	Computer	EE	Electrical	IT	Mechanical
EN No.								
EN23102805	NaN	NaN	NaN	NaN	NaN	53778	NaN	NaN
EN23103954	NaN	NaN	NaN	43778+10000	NaN	NaN	NaN	NaN
EN23117836	NaN	NaN	NaN	NaN	NaN	40000	NaN	NaN
EN23118753	NaN	NaN	NaN	NaN	NaN	43778+10000	NaN	NaN
EN23119584	NaN	NaN	NaN	53778	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
EN23254832	NaN	NaN	NaN	NaN	NaN	NaN	NaN	53778
EN23268026	NaN	NaN	10056	NaN	NaN	NaN	NaN	NaN

Backfill & Forwardfill

70 rows × 8 columns

```
import pandas as pd
df = pd.read_excel('Fees_Data.xlsx')
df
```

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	NaN	NaN	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	NaN	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No
3	4	EN23135942	Borse Gunwant Ashok	Computer	NaN	25000	2023-04-08 00:00:00	Credit Card	12:00:00	No
4	5	EN23119584	Hire Tejas Ravindra	Computer	53778.0	53778	2023-04-08 00:00:00	UPI	12:10:00	Yes
...	...	...	...	...	...	...	...	...	...	...
66	67	EN23273933	Ahirrao Leeleshwar Subhash	Civil	53778.0	40000	16/8/2015	UPI	17:00:00	No
67	68	EN23200937	Gujar Mohit Yatin	Civil	53778.0	25000	17/8/2015	UPI	11.58 am	No

```
df.bfill(axis='rows').head()
```

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	53778.0	28778	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	Computer	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No

```
df.ffill(axis='rows').head()
```

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	10056.0	10056	2023-04-08 00:00:00	UPI	11.40 am	Yes

2	3	EN23135942	Borse Gunwant Ashok	IT	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No	

df.bfill(axis='columns').head()

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	2023-04-08 00:00:00	2023-04-08 00:00:00	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	53778.0	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No

df.ffill(axis='columns').head()

	Sr. No	EN No.	Name of the student	Branch	Total Fees	Fees Paid	Date of payment	Mode of Payment	Time	Status
0	1	EN23204195	SHINDE GAURAV PRAKASH	CIVIL	10056.0	10056	27/7/23	UPI	2.45 pm	Yes
1	2	EN23146043	Chaudhari Prasad mahesh	IT	IT	IT	2023-04-08 00:00:00	UPI	11.40 am	Yes
2	3	EN23135942	Borse Gunwant Ashok	Borse Gunwant Ashok	53778.0	28778	2023-04-08 00:00:00	Credit Card	11.50 am	No