From the raw data below create a data frame: 'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'], 'last_name': ['Miller', 'Jacobson', ".", 'Milner', 'Cooze'], 'age': [42, 52, 36, 24, 73], 'preTestScore': [4, 24, 31, ".", "."],'postTestScore': ["25,000", "94,000", 57, 62, 70] Objective: Perform data processing on raw data: Save the data frame into a csv file as project.csv Read the project.csv and print the data frame Read the project.csv without column heading Read the project.csv and make the index columns as 'First Name' and 'Last Name' Print the data frame in a Boolean form as True or False. True for Null/ NaN values and false for non null values Read the data frame by skipping first 3 rows and print the data frame

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
#Initialise data of lists (https://www.geeksforgeeks.org/different-ways-to-create-pandas-dataframe/)
data = {'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'], 'last_name': ['Miller', 'Jacobson', ".", 'Milner', 'Cooze'], 'age': [42, 52, 36, 24, 73], 'preTestScore': [4, 24, 31, ".", "."],'postTestScore': ["25,000", "94,000", 57, 62, 70]}
```

In [3]:
```
#create dataframe
df = pd.DataFrame(data)
df
```

Out[3]:

	first_name	last_name	age	preTestScore	postTestScore
0	Jason	Miller	42	4	25,000
1	Molly	Jacobson	52	24	94,000
2	Tina	.	36	31	57
3	Jake	Milner	24	.	62
4	Amy	Cooze	73	.	70

In [4]:
```
df.columns=['first_name','last_name','age','preTestScore','postTestScore']
```

In [5]:
```
#Save the data frame into a csv file as project.csv
df.to_csv(r'C:\Users\ctoqu\Desktop\people.csv')
```

In [6]:
```
#Read the project.csv and print the data frame
df1 = pd.read_csv(r'C:\Users\ctoqu\Desktop\people.csv')
```

In [7]:
```
df1
```

Out[7]:

	Unnamed: 0	first_name	last_name	age	preTestScore	postTestScore
0	0	Jason	Miller	42	4	25,000
1	1	Molly	Jacobson	52	24	94,000
2	2	Tina	.	36	31	57
3	3	Jake	Milner	24	.	62
4	4	Amy	Cooze	73	.	70

In [8]:
```
#Read the project.csv without column heading
df1 = pd.read_csv(r'C:\Users\ctoqu\Desktop\people.csv', header =None)
```

In [9]:
```
df1
```

Out[9]:

	0	1	2	3	4	5
0	NaN	first_name	last_name	age	preTestScore	postTestScore
1	0.0	Jason	Miller	42	4	25,000
2	1.0	Molly	Jacobson	52	24	94,000
3	2.0	Tina	.	36	31	57
4	3.0	Jake	Milner	24	.	62
5	4.0	Amy	Cooze	73	.	70

In [14]:
```
#Read the project.csv and make the index columns as 'First Name' and 'Last Nam
e'
FirstLast = df.set_index(['first_name','last_name'])
FirstLast
```

Out[14]:

		Unnamed: 0	age	preTestScore	postTestScore
first_name	**last_name**				
Jason	**Miller**	0	42	4.0	25,000
Molly	**Jacobson**	1	52	24.0	94,000
Tina	**NaN**	2	36	31.0	57
Jake	**Milner**	3	24	NaN	62
Amy	**Cooze**	4	73	NaN	70

In [12]:
```
#Print the data frame in a Boolean form as True or False. True for Null/ NaN v
alues and false for non null values
df =pd.read_csv(r'C:\Users\ctoqu\Desktop\people.csv', na_values=["."])
print(pd.isnull(df))
```

```
   Unnamed: 0  first_name  last_name    age  preTestScore  postTestScore
0       False       False      False  False         False          False
1       False       False      False  False         False          False
2       False       False       True  False         False          False
3       False       False      False  False          True          False
4       False       False      False  False          True          False
```

In [13]:
```
#Read the data frame by skipping first 3 rows and print the data frame

df3 = pd.read_csv(r'C:\Users\ctoqu\Desktop\people.csv', skiprows=[1,2,3])
df3
```

Out[13]:

	Unnamed: 0	first_name	last_name	age	preTestScore	postTestScore
0	3	Jake	Milner	24	.	62
1	4	Amy	Cooze	73	.	70

In [ ]:
```

```