**Experiment No : 4 Pandas DataFrames**

**Consider Sample Python dictionary data and list labels:**

**exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],**

**'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],**

**'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],**

**'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}**

**labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']**

**i)Write a Pandas program to create and display a DataFrame from a specified dictionary data which has the index labels.**

**Aim :**  To write a Pandas program to create and display a DataFrame from a specified dictionary data which has the index labels.

**Description :** Creating a DataFrame from Dictionary with user-defined indexes.

**Program:**

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily',

'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df=pd.DataFrame(exam\_data,index=labels)

print(df)

**Output:**

name score attempts qualify

a Anastasia 12.5 1 yes

b Dima 9.0 3 no

c Katherine 16.5 2 yes

d James NaN 3 no

e Emily 9.0 2 no

f Michael 20.0 3 yes

g Matthew 14.5 1 yes

h Laura NaN 1 no

i Kevin 8.0 2 no

j Jonas 19.0 1 yes

​

**4.ii)Write a Pandas program to change the name 'James' to 'Suresh' in name column of the DataFrame.**

**Aim :**  To write a Pandas program to change the name 'James' to 'Suresh' in name column of the DataFrame.

**Description:**

Pandas dataframe.replace() function is used to replace a string, regex, list, dictionary, series, number etc. from a dataframe. This is a very rich function as it has many variations. The most powerful thing about this function is that it can work with Python regex (regular expressions).

Syntax: DataFrame.replace(to\_replace=None, value=None, inplace=False, limit=None, regex=False, method=’pad’, axis=None)

**Program:**

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily',

'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df=pd.DataFrame(exam\_data,index=labels)

print("\n Changing name James to Suresh")

df['name']=df['name'].replace('James','Suresh')

print(df)

**Output:**

Changing name James to Suresh

name score attempts qualify

a Anastasia 12.5 1 yes

b Dima 9.0 3 no

c Katherine 16.5 2 yes

d Suresh NaN 3 no

e Emily 9.0 2 no

f Michael 20.0 3 yes

g Matthew 14.5 1 yes

h Laura NaN 1 no

i Kevin 8.0 2 no

j Jonas 19.0 1 yes

​

**iii)Write a Pandas program to insert a new column in existing DataFrame.**

**Aim :**  To write a Pandas program to insert a new column in existing DataFrame.

**Description :** By declaring a new list as a column.

**Program:**

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily',

'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df=pd.DataFrame(exam\_data,index=labels)

color = ['Red','Blue','Orange','Red','White','White','Blue','Green','Green','Red']

print("Method 1: By declaring a new list as a column. ")

df['color']=color

print(df)

**output :**

Method 1: By declaring a new list as a column.

name score attempts qualify color

a Anastasia 12.5 1 yes Red

b Dima 9.0 3 no Blue

c Katherine 16.5 2 yes Orange

d James NaN 3 no Red

e Emily 9.0 2 no White

f Michael 20.0 3 yes White

g Matthew 14.5 1 yes Blue

h Laura NaN 1 no Green

i Kevin 8.0 2 no Green

j Jonas 19.0 1 yes Red

**iv)Write a Pandas program to get list from DataFrame column headers.**

**Aim :** To write a Pandas program to get list from DataFrame column headers.

**Description :**  column.values method returs an array of index.

**Program:**

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily',

'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df=pd.DataFrame(exam\_data,index=labels)

print(df)

print(list(df.columns.values))

**Output:**

name score attempts qualify

a Anastasia 12.5 1 yes

b Dima 9.0 3 no

c Katherine 16.5 2 yes

d James NaN 3 no

e Emily 9.0 2 no

f Michael 20.0 3 yes

g Matthew 14.5 1 yes

h Laura NaN 1 no

i Kevin 8.0 2 no

j Jonas 19.0 1 yes

['name', 'score', 'attempts', 'qualify']