Assignment 2

Title - Find out missing data in Dataset.

Aim :-
1. To understand and apply that Data Pre-processing concept
2. To study detailed Data Pre-processing concept in Python

Software Requirements:
1. Ubuntu 16+
2. Python 3.9+
3. Anaconda Spider / Jupiter Notebook.

Theory:
Missing Data can occur when no information is provided for one or more items or for a whole until. Missing Data is a very big problem in real-life scenarios Missing Data can also refer to as NA(not Available) values in pandas. In Data frame Sometimes many data sets simply arrives with missing data, either because it exists and was not collected or it never existed For ex, suppose different users being surveyed may choose not to share different users their income, Some users many choose not to share their address in this way many datasets went missing.
In Pandas missing data is represented by two value:
- None: None is a python singleton object that is often used for missing data in Python code.
- NaN - NaN can acornym for Not a number, is a special floating point value recognized by all systems that use the standazed IEEE floating point representation Pandas treat None & NaN as essentially interchargeable for indicating missing or null values. To facilitate this convention, there are several useful functions for detecting, removing & replacing null values in Pandas Data Frame.

- isnull()
- notnull()
- dropna()
- Fillna()
- replace()
- interpolate()

Checking for missing values is null() & notnull() In order to check missing values in Pandas DataFrame, we use a Function isnull() & notnull(). Both Function help in checking whether a value is NaN or not These function can also be used in Pandas series in order to find null values in a series.

Checking for missing values using isnull()
In order to check null values in Pandas DataFrame, we use isnull() function return dataFrame of Boolean values which are true for NaN values.

Conculsion :-
Thus we have studied different methods to replace missing data in dataset.

Checking for missing values using isnull()

In [1]:
```
#Import all the libraries

import pandas as pd
import numpy as np

# dictionary of lists
dict={'First Score':[100, 90, np.nan, 95],
        'Second Score': [30, 45, 56, np.nan],
        'Third Score':[np.nan, 40, 80, 98]}

# creating a dataframe using dictionary
df =pd.DataFrame(dict)

# using notnull() function
df.notnull()
```

Out[1]:

	First Score	Second Score	Third Score
0	True	True	False
1	True	True	True
2	False	True	True
3	True	False	True

In [ ]:
```

```