Assignment no. 2

Title :- Find out missing data in dataset

Aim :- i) To understand and apply data-preprocessing concept

Software - Requirement :-

i) Ubuntu 16 +

ii) Python 3.9 +

iii) Anaconda Spider / Jupiter notebook

Theory :- Missing data can occur when no information is provided for one or more items or for a whole unit. Missing data is very big problem in real life scenarios. Missing data can also refer to as NA values in pandas. In data Missing data can also refer to dataset simply active missing data, either because it exists and was not collected or it never existed. For Example suppose different user being surveyed may choose not to share their income, some user may choose not to share address in this way many data-sets 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 is a special floating point value recognized by all systems that use standard IEEE floating point representation.

pandas treat none and NaN as essentially interchangable for indicating missing or null values. To facilate this convention there are several useful functions for detecting, removing and replacing null values in pandas data-frame.

* isnull()
* notnull()
* dropna()
* fillna()
* replace()
* interpolate()

checking for missing values isnull() and notnull()

In order to check missing values in pandas dataframe, we use a function is null() and 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 is null()

In order to check null values in pandas data-frame, we use isnull() function this function return data frame of boolean values which are true for NaN values.

conclusion:

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