Handle Missing Values

Date	17 November 2022
Team ID	PNT2022TMID45545
Project name	MACHINE LEARNING BASED VEHICLE PERFORMANCE ANALAYZER

Data pre-processing is a process of cleaning the raw data i.e. the data is collected in the real world and is converted to a clean data set. In other words, whenever the data is gathered from different sources it is collected in a raw format and this data isn't feasible for the analysis. Therefore, certain steps are executed to convert the data into a small clean data set, this part of the process is called as data pre-processing Follow the following steps to process your Data

- Import the Libraries
- Importing the dataset
- Taking care of Missing Data
- Label encoding
- One Hot Encoding
- Feature Scaling
- Splitting Data into Train and Test

Step1: Importing the libraries

First step is usually importing the libraries that will be needed in the program. Import the pandas library and give a shortcut name as pd

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Step 2: Import the Dataset

We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program) and read it using a method called read_csv which can be found in the library called pandas. Here we are using a data set which you can find in the below link:
https://thesmartbridge.com/documents/spsaimldocs/Data.csv

```
dataset = pd.read_csv(
r'C:\Users\Hari Chandan\Desktop\Data_Preprocessing\Data.csv')
```

Step 3: Taking Care of missing Data

Sometimes you may find some data are missing in the dataset. We need to be equipped to handle the problem when we come across them. Obviously you could remove the entire line of data but what if you are unknowingly removing crucial information? Of course we would not want to do that. One of the most common ideas to handle the problem is to take a mean of all the values of the same column and have it to replace the missing data. We will be suing dataset.isnull().any() method to see which column has missing value

```
dataset.isnull().any()
```

```
Country      False
Age           True
Salary        True
Purchased    False
dtype: bool
```