EXPERIMENT 3B

Pandas Library – Handling Missing Values

Aim:

Demonstrate an experiment to handle missing data and inappropriate data in a dataset using python pandas library for data preprocessing.

Algorithm:

1. Import the pandas library.

2. Read the hotel dataset using pd.read\_csv() and display the first few rows.

3. Check and display the number of missing values in each column.

4. Identify and display invalid ratings that are not between 1 and 5.

5. Clean the FoodPreference column by removing spaces, converting to lowercase, and standardizing text values (e.g., “veg” → “Vegetarian”).

6. Fill missing values in the Bill column with its mean.

7. Fill missing values in the FoodPreference column with its mode.

8. Replace invalid ratings (outside 1–5) with the median rating of valid entries.

9. Remove duplicate or unnecessary column Age\_Group.1 if it exists.

10. Display the cleaned dataset and confirm that no missing values remain.

Program:

import pandas as pd

df = pd.read\_csv("C:\Hotel\_Dataset.csv")

print(df.head())

print("\n")

print(df.isnull().sum())

print("\n")

print("\nInvalid Ratings (outside 1-5):")

print(df[(df['Rating(1-5)'] > 5) | (df['Rating(1-5)'] < 1)])

print("\n")

df['FoodPreference'] = df['FoodPreference'].str.strip().str.lower()

df['FoodPreference'] = df['FoodPreference'].replace({

'veg': 'Vegetarian',

'vegetarian': 'Vegetarian',

'veg.': 'Vegetarian',

'non-veg': 'Non-Vegetarian',

'non vegetarian': 'Non-Vegetarian',

'non-veg.': 'Non-Vegetarian'

})

df['Bill'] = df['Bill'].fillna(df['Bill'].mean())

df['FoodPreference'] = df['FoodPreference'].fillna(df['FoodPreference'].mode()[0])

valid\_ratings = df[(df['Rating(1-5)'] >= 1) & (df['Rating(1-5)'] <= 5)]

median\_rating = valid\_ratings['Rating(1-5)'].median()

df.loc[(df['Rating(1-5)'] > 5) | (df['Rating(1-5)'] < 1), 'Rating(1-5)'] = median\_rating

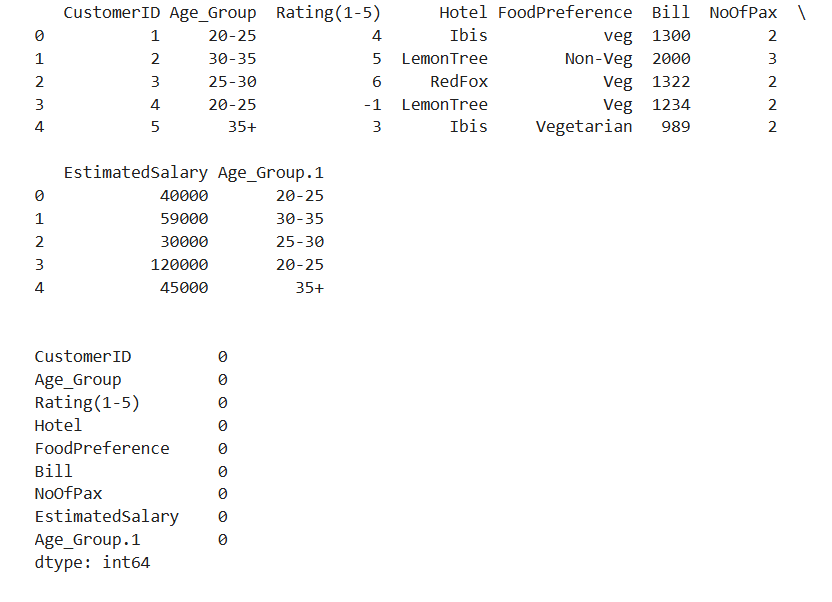
if 'Age\_Group.1' in df.columns:

df = df.drop(columns=['Age\_Group.1'])

print(df.head())

print("\n")

print(df.isnull().sum())

Output:

A screenshot of a computer

AI-generated content may be incorrect.

Result:

Hence the python program to handle missing appropriate values is written and executed successfully.