ex - 3a

October 31, 2025

```python
[ ]: import numpy as np
import pandas as pd
df = pd.read_csv("Hotel_Dataset.csv")
print("Original Data:")
print(df)
print("\nDuplicate rows:", df.duplicated().sum())
df.drop_duplicates(inplace=True)
print("\nAfter removing duplicates:")
print(df)
df.reset_index(drop=True, inplace=True)
print("\nAfter resetting index:")
print(df)
if 'Age_Group.1' in df.columns:
    df.drop(['Age_Group.1'], axis=1, inplace=True)
df.loc[df['CustomerID'] < 0, 'CustomerID'] = np.nan
df.loc[df['Bill'] < 0, 'Bill'] = np.nan
df.loc[df['EstimatedSalary'] < 0, 'EstimatedSalary'] = np.nan
df.loc[(df['NoOfPax'] < 1) | (df['NoOfPax'] > 20), 'NoOfPax'] = np.nan
df['Hotel'] = df['Hotel'].replace({'Ibys': 'Ibis'})
df['FoodPreference'] = df['FoodPreference'].replace(
    {'Vegetarian': 'Veg', 'veg': 'Veg', 'non-Veg': 'Non-Veg'}
)
df['EstimatedSalary'] = df['EstimatedSalary'].
 ↪fillna(round(df['EstimatedSalary'].mean()))
df['NoOfPax'] = df['NoOfPax'].fillna(round(df['NoOfPax'].median()))
df['Bill'] = df['Bill'].fillna(round(df['Bill'].mean()))
df.loc[(df['Rating(1-5)'] < 1) | (df['Rating(1-5)'] > 5), 'Rating(1-5)'] = np.
 ↪nan
df['Rating(1-5)'] = df['Rating(1-5)'].fillna(round(df['Rating(1-5)'].median()))
df = df.astype({
    'CustomerID': 'Int64',
    'NoOfPax': 'Int64',
    'Bill': 'Int64',
    'EstimatedSalary': 'Int64',
    'Rating(1-5)': 'Int64'
})
print("\nFinal Cleaned DataFrame:")
```

```python
print(df)
print("\nCleaned DataFrame Info:")
print(df.info())
```