

```
[1]: import numpy as np
import pandas as pd
df=pd.read_csv(r"C:\Users\niveth\OneDrive\Documents\Downloads\Hotel_Dataset - Hotel_Dataset.csv")
df
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary	Age Group.1
0	1	20-25	4	Ibis	veg	1300	2	40000	20-25
1	2	30-35	5	LemonTree	Non-Veg	2000	3	59000	30-35
2	3	25-30	6	RedFox	Veg	1322	2	30000	25-30
3	4	20-25	1	LemonTree	Veg	1234	2	120000	20-25
4	5	35+	3	Ibis	Vegetarian	989	2	45000	35+
5	6	35+	3	Ibys	Non-Veg	1909	2	122220	35+
6	7	35+	4	RedFox	Vegetarian	1000	-1	21122	35+
7	8	20-25	7	LemonTree	Veg	2999	-10	345673	20-25
8	9	25-30	2	Ibis	Non-Veg	3456	3	-99999	25-30
9	9	25-30	2	Ibis	Non-Veg	3456	3	-99999	25-30
10	10	30-35	5	RedFox	non-Veg	6755	4	87777	30-35

```
[2]: df.duplicated()
```

```
[2]: 0    False
1    False
2    False
3    False
4    False
5    False
6    False
7    False
8    False
9    True
10   False
dtype: bool
```

```
[3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   CustomerID  11 non-null    int64  
 1   Age_Group   11 non-null    object 
 2   Rating(1-5) 11 non-null    int64  
 3   Hotel        11 non-null    object 
 4   FoodPreference 11 non-null    object 
 5   Bill         11 non-null    int64  
 6   NoOfPax     11 non-null    int64  
 7   EstimatedSalary 11 non-null    int64  
 8   Age_Group.1  11 non-null    object 
dtypes: int64(5), object(4)
memory usage: 924.0+ bytes
```

```
[4]: df.drop_duplicates(inplace=True)
df
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary	Age Group.1
0	1	20-25	4	Ibis	veg	1300	2	40000	20-25
1	2	30-35	5	LemonTree	Non-Veg	2000	3	59000	30-35
2	3	25-30	6	RedFox	Veg	1322	2	30000	25-30
3	4	20-25	-1	LemonTree	Veg	1234	2	120000	20-25
4	5	35+	3	Ibis	Vegetarian	989	2	45000	35+
5	6	35+	3	Ibys	Non-Veg	1909	2	122220	35+
6	7	35+	4	RedFox	Vegetarian	1000	-1	21122	35+
7	8	20-25	7	LemonTree	Veg	2999	-10	345673	20-25
8	9	25-30	2	Ibis	Non-Veg	3456	3	-99999	25-30
10	10	30-35	5	RedFox	non-Veg	-6755	4	87777	30-35

```
[5]: len(df)
```

```
[5]: 10
```

```
[6]: index=np.array(list(range(0,len(df))))
df.set_index(index,inplace=True)
index
```

```
[6]: array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
```

```
[7]: df
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary	Age Group.1
0	1	20-25	4	Ibis	veg	1300	2	40000	20-25
1	2	30-35	5	LemonTree	Non-Veg	2000	3	59000	30-35
2	3	25-30	6	RedFox	Veg	1322	2	30000	25-30
3	4	20-25	-1	LemonTree	Veg	1234	2	120000	20-25
4	5	35+	3	Ibis	Vegetarian	989	2	45000	35+
5	6	35+	3	Ibys	Non-Veg	1909	2	122220	35+
6	7	35+	4	RedFox	Vegetarian	1000	-1	21122	35+
7	8	20-25	7	LemonTree	Veg	2999	-10	345673	20-25
8	9	25-30	2	Ibis	Non-Veg	3456	3	-99999	25-30
9	10	30-35	5	RedFox	non-Veg	-6755	4	87777	30-35

```
[8]: df.drop(['Age_Group.1'],axis=1,inplace=True)
df
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary
0	1	20-25	4	Ibis	veg	1300	2	40000
1	2	30-35	5	LemonTree	Non-Veg	2000	3	59000
2	3	25-30	6	RedFox	Veg	1322	2	30000
3	4	20-25	-1	LemonTree	Veg	1234	2	120000
4	5	35+	3	Ibis	Vegetarian	989	2	45000
5	6	35+	3	Ibys	Non-Veg	1909	2	122220
6	7	35+	4	RedFox	Vegetarian	1000	-1	21122
7	8	20-25	7	LemonTree	Veg	2999	-10	345673

```
[9]: df.CustomerID.loc[df.CustomerID<0]=np.nan  
df.Bill.loc[df.Bill<0]=np.nan  
df.EstimatedSalary.loc[df.EstimatedSalary<0]=np.nan  
df
```

```
C:\Users\nivet\AppData\Local\Temp\ipykernel_21372\1423144776.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df.CustomerID.loc[df.CustomerID<0]=np.nan
```

```
C:\Users\nivet\AppData\Local\Temp\ipykernel_21372\1423144776.py:2: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df.Bill.loc[df.Bill<0]=np.nan
```

```
C:\Users\nivet\AppData\Local\Temp\ipykernel_21372\1423144776.py:3: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df.EstimatedSalary.loc[df.EstimatedSalary<0]=np.nan
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary
0	1.0	20-25	4	Ibis	veg	1300.0	2	40000.0
1	2.0	30-35	5	LemonTree	Non-Veg	2000.0	3	59000.0
2	3.0	25-30	6	RedFox	Veg	1322.0	2	30000.0
3	4.0	20-25	-1	LemonTree	Veg	1234.0	2	120000.0
4	5.0	35+	3	Ibis	Vegetarian	989.0	2	45000.0
5	6.0	35+	3	Ibys	Non-Veg	1909.0	2	122220.0
6	7.0	35+	4	RedFox	Vegetarian	1000.0	-1	21122.0
7	8.0	20-25	7	LemonTree	Veg	2999.0	-10	345673.0
8	9.0	25-30	2	Ibis	Non-Veg	3456.0	3	NaN
9	10.0	30-35	5	RedFox	non-Veg	NaN	4	87777.0

```
[10]: df['NoOfPax'].loc[(df['NoOfPax']<1) | (df['NoOfPax']>28)]=np.nan  
df
```

```
C:\Users\nivet\AppData\Local\Temp\ipykernel_21372\80558596.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df['NoOfPax'].loc[(df['NoOfPax']<1) | (df['NoOfPax']>28)]=np.nan
```

	CustomerID	Age Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary
0	1.0	20-25	4	Ibis	veg	1300.0	2.0	40000.0
1	2.0	30-35	5	LemonTree	Non-Veg	2000.0	3.0	59000.0
2	3.0	25-30	6	RedFox	Veg	1322.0	2.0	30000.0
3	4.0	20-25	-1	LemonTree	Veg	1234.0	2.0	120000.0
4	5.0	35+	3	Ibis	Vegetarian	989.0	2.0	45000.0
5	6.0	35+	3	Ibys	Non-Veg	1909.0	2.0	122220.0
6	7.0	35+	4	RedFox	Vegetarian	1000.0	NaN	21122.0
7	8.0	20-25	7	LemonTree	Veg	2999.0	NaN	345673.0
8	9.0	25-30	2	Ibis	Non-Veg	3456.0	3.0	NaN
9	10.0	30-35	5	RedFox	non-Veg	NaN	4.0	87777.0

```
[11]: df.Age_Group.unique()
```

```
[11]: array(['20-25', '30-35', '25-30', '35+'], dtype=object)

[12]: df.Hotel.unique()

[13]: array(['Ibis', 'LemonTree', 'RedFox', 'ibys'], dtype=object)

[13]: df.Hotel.replace(['ibys'], 'ibis', inplace=True)
df.FoodPreference.unique()

[13]: <bound method Series.unique of 8>
      veg
      1    Non-Veg
      2        Veg
      3        Veg
      4  Vegetarian
      5    Non-Veg
      6  Vegetarian
      7        Veg
      8    Non-Veg
      9    non-Veg
Name: FoodPreference, dtype: object>

[14]: df.FoodPreference.replace(['Vegetarian', 'veg'], 'Veg', inplace=True)
df.FoodPreference.replace(['non-Veg'], 'Non-Veg', inplace=True)

[15]: df.EstimatedSalary.fillna(round(df.EstimatedSalary.mean()), inplace=True)
df.NoOfPax.fillna(round(df.NoOfPax.median()), inplace=True)
df['Rating(1-5)'].fillna(round(df['Rating(1-5)'].median()), inplace=True)
df.Bill.fillna(round(df.Bill.mean()), inplace=True)
df
```

	CustomerID	Age	Group	Rating(1-5)	Hotel	FoodPreference	Bill	NoOfPax	EstimatedSalary
0	1.0	20-25		4	Ibis	Veg	1300.0	2.0	40000.0
1	2.0	30-35		5	LemonTree	Non-Veg	2000.0	3.0	59000.0
2	3.0	25-30		6	RedFox	Veg	1322.0	2.0	30000.0
3	4.0	20-25		-1	LemonTree	Veg	1234.0	2.0	120000.0
4	5.0	35+		3	Ibis	Veg	989.0	2.0	45000.0
5	6.0	35+		3	Ibis	Non-Veg	1909.0	2.0	122220.0
6	7.0	35+		4	RedFox	Veg	1000.0	2.0	21122.0
7	8.0	20-25		7	LemonTree	Veg	2999.0	2.0	345673.0
8	9.0	25-30		2	Ibis	Non-Veg	3456.0	3.0	96755.0
9	10.0	30-35		5	RedFox	Non-Veg	1801.0	4.0	87777.0