

## ✓ IMPORTING LIBRARIES

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import scipy as sp
import plotly.express as px
import re
```



## ✓ LOADING DATASET

```
data=pd.read_csv('/content/part1.csv')
```



## ✓ DATA ANONYMIZATION

```
data.head(1)
```



	HOTEL NAME	YEAR OF SALES	OWNER	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES
0	LOWRY HOTEL	2021	ROSE KELLER	HULL	53.76775	-0.335827	MARRIAGE

```
# drop columns 'OWNER' and 'LICENSE NUMBER'
data = data.drop(['OWNER', 'LICENSE NUMBER'], axis=1)

print(data)
```



HOTEL NAME    YEAR OF SALES    HOTEL LOCATION    LATTITUDE    LC

0	LOWRY HOTEL	2021	HULL	53.767750	-6
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1
6	EDEN HOTEL	2021	DURHAM	54.776100	-1
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2
8	ROYAL HOTEL	2021	PERTH	56.396999	-3
9	PALACE HOTEL	2021	BATH	51.380001	-2
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-6
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1
15	HAVEN HOTEL	2021	TRURO	50.259998	-5
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	6
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
21	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
22	LOWRY HOTEL	2022	BELFAST	53.767750	-6
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1
28	EDEN HOTEL	2022	DURHAM	54.776100	-1
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-6
35	THE BARN HOTEL	2022	KENT	53.483002	-2
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	6
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2

	SPECIALITY FOCUSES	SEASON FOCUSES	TOURISTS	REVIEW	\
0	MARRIAGE	SUMMER	INDIA	5.50	
1	BIRTHDAY	WINTER	ROME	6.00	
2	DINING	SUMMER	QATAR	7.00	
3	KITTY PARTY	SUMMER	IRAN	10.00	
4	RECEPTION	WINTER	PAKISTAN	5.00	
5	ENGAGEMENT	AUTUMN	GERMANY	8.00	
6	FAMILY GET TOGETHER	AUTUMN	DUBAI	8.00	
7	GUEST ACCOMODATION ONLY	WINTER	AUSTRALIA	10.00	

8	BIRTHDAY	AUTUMN	UNITED STATES	9.00
9	GUEST ACCOMODATION ONLY	SUMMER	EGYPT	10.00
10	MARRIAGE	SUMMER	SRILANKA	6.70
11	DINING	WINTER	BELGIUM	8.80
12	FAMILY GET TOGETHER	AUTUMN	ITALY	9.90
13	RECEPTION	WINTER	NEWZELAND	5.00
14	BIRTHDAY	SUMMER	CANADA	10.00

## ✓ CLEANING DATA

## ✓ HANDLING MISSING DATA

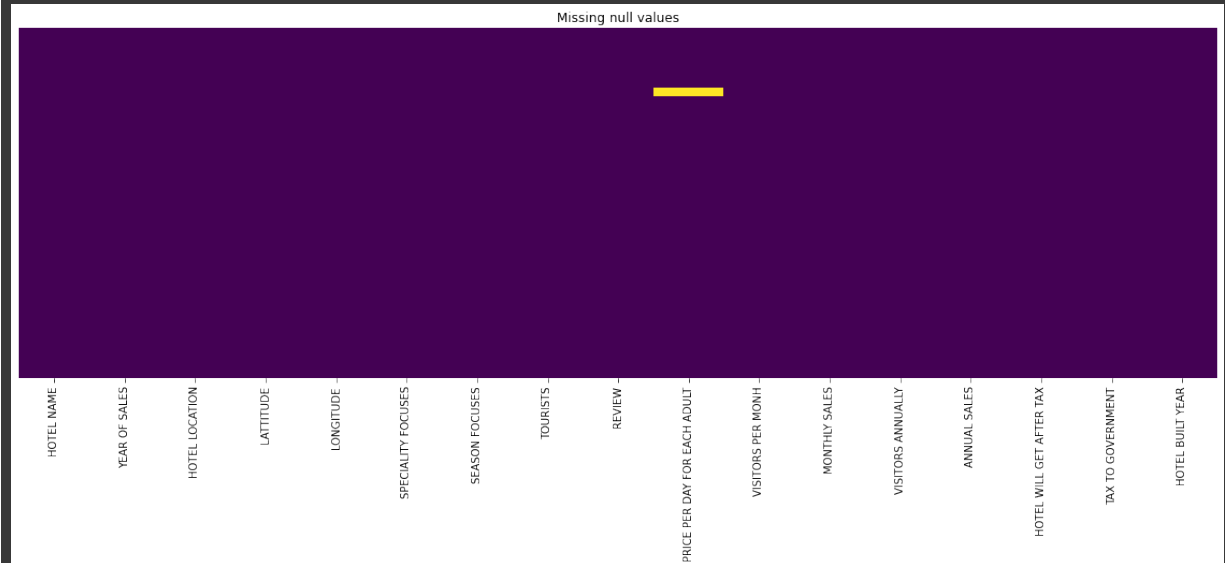
### CHECKING MISSING DATA

```
data.isnull().sum().sort_values(ascending=False)
```

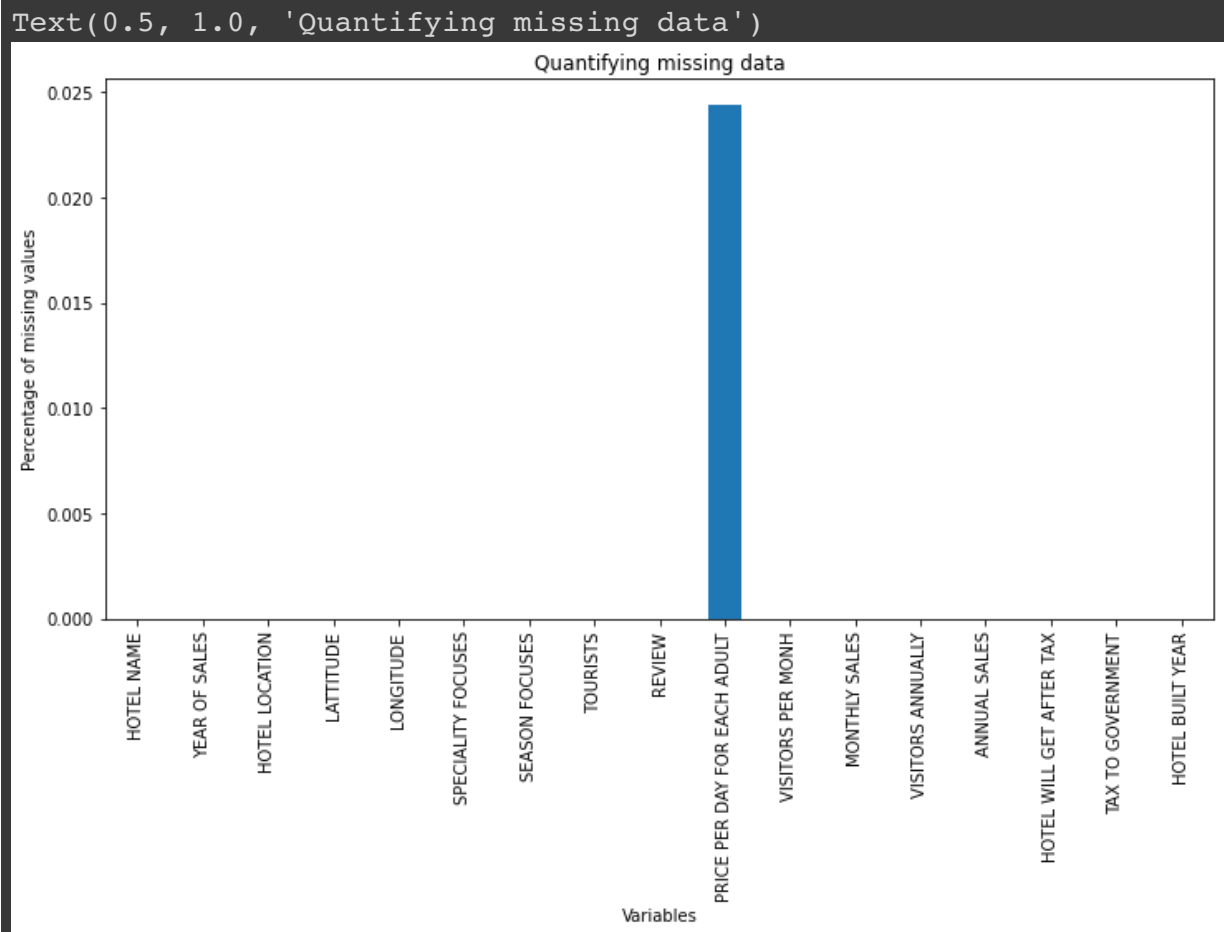
```
PRICE PER DAY FOR EACH ADULT    1
HOTEL NAME                      0
TAX TO GOVERNMENT                0
HOTEL WILL GET AFTER TAX         0
ANNUAL SALES                     0
VISITORS ANNUALLY                0
MONTHLY SALES                    0
VISITORS PER MONH                0
REVIEW                           0
YEAR OF SALES                    0
TOURISTS                         0
SEASON FOCUSES                   0
SPECIALITY FOCUSES               0
LONGITUDE                        0
LATTITUDE                        0
HOTEL LOCATION                   0
HOTEL BUILT YEAR                 0
dtype: int64
```

```
import matplotlib
matplotlib.rcParams['figure.figsize'] = (20,6)
sns.heatmap(data.isnull(),yticklabels = False, cbar = False , cmap = 'v.
plt.title("Missing null values")
```

Text(0.5, 1.0, 'Missing null values')



```
data.isnull().mean().plot.bar(figsize=(12,6))
plt.ylabel('Percentage of missing values')
plt.xlabel('Variables')
plt.title('Quantifying missing data')
```



## RENAME COLUMNS NAMES FIRST

```
data.rename({'YEAR OF SALES':'SALES YEAR'},axis=1)
```

```
HOTEL NAME    SALES YEAR    HOTEL LOCATION    LATTITUDE    LONGITUDE    SPE
```

0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MA
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BI
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1.466667	
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITT
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	REC
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAG
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAM TO
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	ACCOMO
8	ROYAL HOTEL	2021	PERTH	56.396999	-3.437000	BI
9	PALACE HOTEL	2021	BATH	51.380001	-2.360000	ACCOMO
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1.510556	MA
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3.179090	
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-0.118092	FAM TO
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2.293100	REC
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1.381130	BI
15	HAVEN HOTEL	2021	TRURO	50.259998	-5.051000	MA
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0.119167	ENGAG
	STONE					

17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1.257677	KITT
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
21	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
22	LOWRY HOTEL	2022	BELFAST	53.767750	-0.335827	MA
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2.244644	BI
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1.466667	
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1.133333	KITT
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4.251433	REC
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1.600000	ENGAG
28	EDEN HOTEL	2022	DURHAM	54.776100	-1.573300	FAM TO
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2.983333	ACCOMO
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3.437000	BI
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2.360000	ACCOMO
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1.510556	MA
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3.179090	
34	HOWTOWN	2022	LONDON	51.509865	-0.118092	FAM

35	THE BARN HOTEL	2022	KENT	53.483002	-2.293100	REC
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1.381130	BL
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5.051000	MA
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0.119167	ENGA
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1.257677	KITT
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2.587910	

## REPLACING SYMBOLS WITH NAN

```
#show all special characters
import re
def highlight_special_characters(val):
    if isinstance(val, str) and re.search(r'^A-Za-z0-9', val):
        return 'background-color: blue'
    return ''

data.style.applymap(highlight_special_characters)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPE
0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MA
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BL
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1.466667	
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITT



4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	REC
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAG
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAM TO
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	ACCOMO
8	ROYAL HOTEL	2021	PERTH	56.396999	-3.437000	BI
9	PALACE HOTEL	2021	BATH	51.380001	-2.360000	ACCOMO
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1.510556	MA
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3.179090	
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-0.118092	FAM TO
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2.293100	REC
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1.381130	BI
15	HAVEN HOTEL	2021	TRURO	50.259998	-5.051000	MA
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0.119167	ENGAG
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1.257677	KITT
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	

21	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
22	LOWRY HOTEL	2022	BELFAST	53.767750	-0.335827	MA
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2.244644	BL
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1.466667	
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1.133333	KITT
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4.251433	REC
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1.600000	ENGAG
28	EDEN HOTEL	2022	DURHAM	54.776100	-1.573300	FAM TO
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2.983333	ACCOMO
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3.437000	BL
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2.360000	ACCOMO
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1.510556	MA
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3.179090	
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-0.118092	FAM TO
35	THE BARN HOTEL	2022	KENT	53.483002	-2.293100	REC
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1.381130	BL
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5.051000	MA

38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0.119167	ENGAGE
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1.257677	KITT
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2.587910	

```
#select the symbol column which i want
data[data['VISITORS PER MONH'].str.contains(r'^A-Za-z0-9')] ]
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES	S
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	GUEST ACCOMODATION ONLY	V

## REPLACING SPECIFIC BLANK CELLS WITH NAN

```
#replace with nan if you don't know the input
df = data.replace({'VISITORS PER MONH': {'***': pd.np.nan}})

print(df)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LC
0	LOWRY HOTEL	2021	HULL	53.767750	-0
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1
6	EDEN HOTEL	2021	DURHAM	54.776100	-1
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2
8	ROYAL HOTEL	2021	PERTH	56.396999	-3
9	PALACE HOTEL	2021	BATH	51.380001	-2
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3

12	HOWTOWN HOTEL	2021	LONDON	51.509865	-6
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1
15	HAVEN HOTEL	2021	TRURO	50.259998	-5
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
21	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
22	LOWRY HOTEL	2022	BELFAST	53.767750	-6
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1
28	EDEN HOTEL	2022	DURHAM	54.776100	-1
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-6
35	THE BARN HOTEL	2022	KENT	53.483002	-2
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2

	SPECIALITY FOCUSES	SEASON FOCUSES	TOURISTS	REVIEW	\
0	MARRIAGE	SUMMER	INDIA	5.50	
1	BIRTHDAY	WINTER	ROME	6.00	
2	DINING	SUMMER	QATAR	7.00	
3	KITTY PARTY	SUMMER	IRAN	10.00	
4	RECEPTION	WINTER	PAKISTAN	5.00	
5	ENGAGEMENT	AUTUMN	GERMANY	8.00	
6	FAMILY GET TOGETHER	AUTUMN	DUBAI	8.00	
7	GUEST ACCOMODATION ONLY	WINTER	AUSTRALIA	10.00	
8	BIRTHDAY	AUTUMN	UNITED STATES	9.00	
9	GUEST ACCOMODATION ONLY	SUMMER	EGYPT	10.00	
10	MARRIAGE	SUMMER	SRILANKA	6.70	
11	DINING	WINTER	BELGIUM	8.80	
12	FAMILY GET TOGETHER	AUTUMN	ITALY	9.90	
13	RECEPTION	WINTER	NEWZELAND	5.00	
14	BIRTHDAY	SUMMER	CANADA	10.00	

```
#show all special characters
```

```
import re
```

```
def highlight_special_characters(val):
```

```
    if isinstance(val, str) and re.search(r'^[A-Za-z0-9]', val):
```

```

return 'background-color: blue'
return ''

```

```
df.style.applymap(highlight_special_characters)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPE
0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MA
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BI
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1.466667	
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITT
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	REC
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAG
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAM TO
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	ACCOMO
8	ROYAL HOTEL	2021	PERTH	56.396999	-3.437000	BI
9	PALACE HOTEL	2021	BATH	51.380001	-2.360000	ACCOMO
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1.510556	MA
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3.179090	
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-0.118092	FAM TO
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2.293100	REC

14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1.381130	BL
15	HAVEN HOTEL	2021	TRURO	50.259998	-5.051000	MA
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0.119167	ENGAC
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1.257677	KITT
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
21	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
22	LOWRY HOTEL	2022	BELFAST	53.767750	-0.335827	MA
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2.244644	BL
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1.466667	
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1.133333	KITT
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4.251433	REC
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1.600000	ENGAC
28	EDEN HOTEL	2022	DURHAM	54.776100	-1.573300	FAM TO
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2.983333	ACCOMO
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3.437000	BL
	PALACE					

31	PALACE HOTEL	2022	STOCKTON	51.380001	-2.360000	ACCOMO
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1.510556	MA
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3.179090	
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-0.118092	FAM TO
35	THE BARN HOTEL	2022	KENT	53.483002	-2.293100	REC
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1.381130	BI
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5.051000	MA
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0.119167	ENGAC
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1.257677	KITT
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2.587910	

df.info

```
<bound method DataFrame.info of
SALES HOTEL LOCATION LATTITUDE LONGITUDE \
0 LOWRY HOTEL 2021 HULL 53.767750
-0.335827
1 GORING HOTEL 2021 MANCHESTER 53.483959
-2.244644
2 HOTEL RUSSELL 2021 DERBY 52.916668
-1.466667
3 CARLTON HOTEL 2021 LEICESTER 52.633331
-1.133333
4 BLUE MOUNT 2021 GLASGOW 55.860916
-4.251433
5 FEATHERS HOTEL 2021 NEWCASTLE 54.966667
-1.600000
6 EDEN HOTEL 2021 DURHAM 54.776100
-1.573300
7 CAPITAL HOTEL 2021 LIVERPOOL 53.400002
```

-2.983333				
8	ROYAL HOTEL	2021	PERTH	56.396999
-3.437000				
9	PALACE HOTEL	2021	BATH	51.380001
-2.360000				
10	THE BULL HOTEL	2021	COVENTRY	52.408054
-1.510556				
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583
-3.179090				
12	HOWTOWN HOTEL	2021	LONDON	51.509865
-0.118092				
13	THE BARN HOTEL	2021	SALFORD	53.483002
-2.293100				
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101
-1.381130				
15	HAVEN HOTEL	2021	TRURO	50.259998
-5.051000				
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276
0.119167				
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022
-1.257677				
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514
-2.587910				
19	CLAREMONT HOTEL	2021	BRISTOL	51.454514
-2.587910				
20	CLAREMONT HOTEL	2021	BRISTOL	51.454514
-2.587910				
21	CLAREMONT HOTEL	2021	BRISTOL	51.454514
-2.587910				
22	LOWRY HOTEL	2022	BELFAST	53.767750
-0.335827				
23	GORING HOTEL	2022	MANCHESTER	53.483959
-2.244644				
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668
-1.466667				
25	CARLTON HOTEL	2022	LEICESTER	52.633331
-1.133333				
26	BLUE MOUNT	2022	GLASGOW	55.860916
-4.251433				
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667
-1.600000				
28	EDEN HOTEL	2022	DURHAM	54.776100

## CHECKING NULL VALUES



```
df.isnull().sum().sort_values(ascending=False)
```

PRICE PER DAY FOR EACH ADULT	1
VISITORS PER MONH	1
HOTEL NAME	0
TAX TO GOVERNMENT	0
HOTEL WILL GET AFTER TAX	0
ANNUAL SALES	0
VISITORS ANNUALLY	0
MONTHLY SALES	0
REVIEW	0
YEAR OF SALES	0
TOURISTS	0
SEASON FOCUSES	0
SPECIALITY FOCUSES	0
LONGITUDE	0
LATTITUDE	0
HOTEL LOCATION	0
HOTEL BUILT YEAR	0

dtype: int64

```
df.isna().any()
```

HOTEL NAME	False
YEAR OF SALES	False
HOTEL LOCATION	False
LATTITUDE	False
LONGITUDE	False
SPECIALITY FOCUSES	False
SEASON FOCUSES	False
TOURISTS	False
REVIEW	False
PRICE PER DAY FOR EACH ADULT	True
VISITORS PER MONH	True
MONTHLY SALES	False
VISITORS ANNUALLY	False
ANNUAL SALES	False
HOTEL WILL GET AFTER TAX	False
TAX TO GOVERNMENT	False
HOTEL BUILT YEAR	False

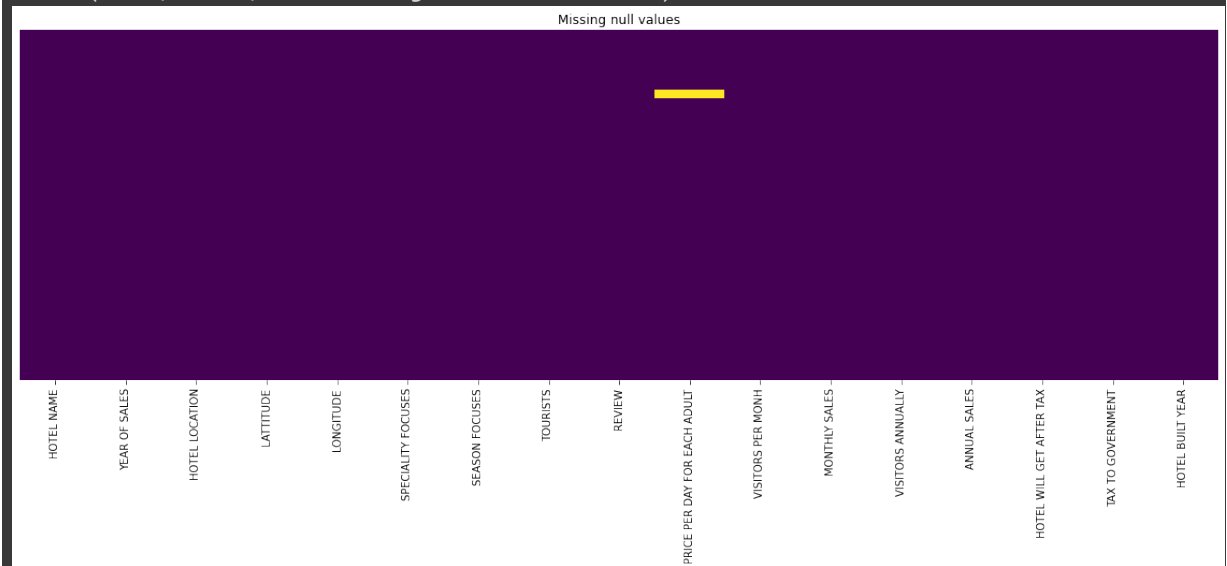
dtype: bool

```
df.isna().sum()
```

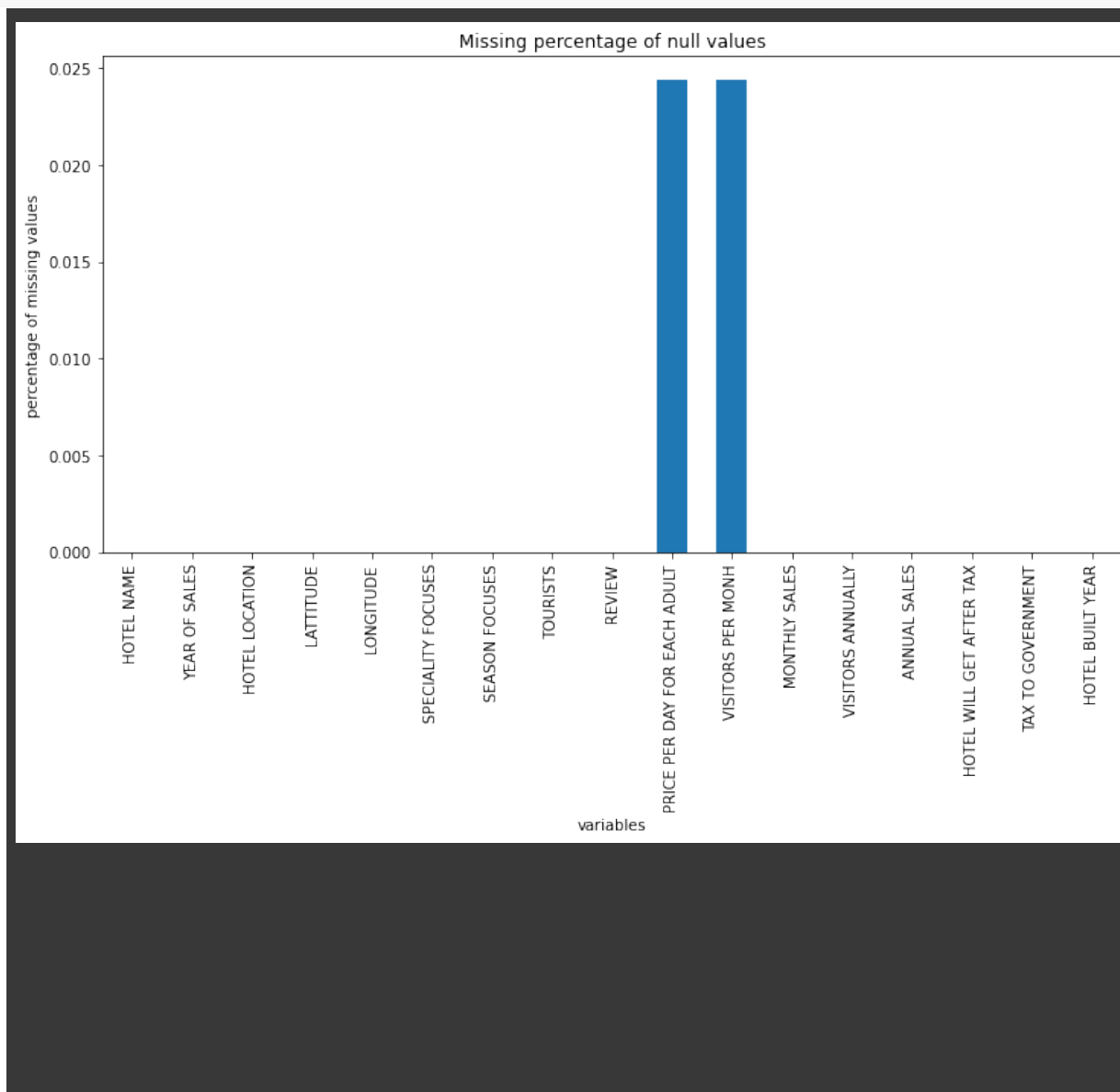
```
HOTEL NAME          0
YEAR OF SALES       0
HOTEL LOCATION      0
LATTITUDE           0
LONGITUDE           0
SPECIALITY FOCUSES  0
SEASON FOCUSES      0
TOURISTS            0
REVIEW              0
PRICE PER DAY FOR EACH ADULT  1
VISITORS PER MONH   1
MONTHLY SALES       0
VISITORS ANNUALLY   0
ANNUAL SALES        0
HOTEL WILL GET AFTER TAX  0
TAX TO GOVERNMENT    0
HOTEL BUILT YEAR     0
dtype: int64
```

```
import matplotlib
matplotlib.rcParams['figure.figsize'] = (20,6)
sns.heatmap(data.isnull(),yticklabels = False, cbar = False , cmap = 'v.
plt.title("Missing null values")
```

```
Text(0.5, 1.0, 'Missing null values')
```



```
df.isnull().mean().plot.bar(figsize=(12,6))
plt.ylabel('percentage of missing values')
plt.xlabel('variables')
plt.title("Missing percentage of null values")
plt.show()
```



**Observation :**

Dealing with the null values

**The columns having highest percentage of null values are :**

Ans: Price per day for each adult

(If these two are not that much helpful if we want we can drop these columns and in the further steps we drop and add some features in the dataset after the Analysis)

**1. we can drop the small null values for the columns:**

Ans: No

**2. good to fill/replace the null values are:**

Ans: Yes. Visitors per month

**3. replace with values '0' are "NaN":**

Ans: No

**4. median transformation values are:**

Ans: Yes. Price per day for each adult

**METHOD 2-FILL VALUES**

```
#In the VISITORS PER MONH column
#replace with an input if you know the value
df.iloc[7,11]=np.nan
df=df.fillna(900)
df.head(8)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALIT FOCUSE
0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MARRIAG
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BIRTHDA
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1.466667	DININ
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITTY PART
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	RECEPTIO
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAGEMENT
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAMILY GE TOGETHE
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	GUES ACCOMODATIO ONL

## METHOD 4- MISSING VALUE TRANSFORMATION

### NUMERICAL-MEDIAN

```
#in the PRICE PER DAY FOR EACH ADULT column
```

```
# Calculate the median of the 'column_name' column  
median = df['PRICE PER DAY FOR EACH ADULT'].median()
```

```
# Fill the missing values in the 'column_name' column with the median value  
df.loc[:, 'PRICE PER DAY FOR EACH ADULT'].fillna(median, inplace=True)
```

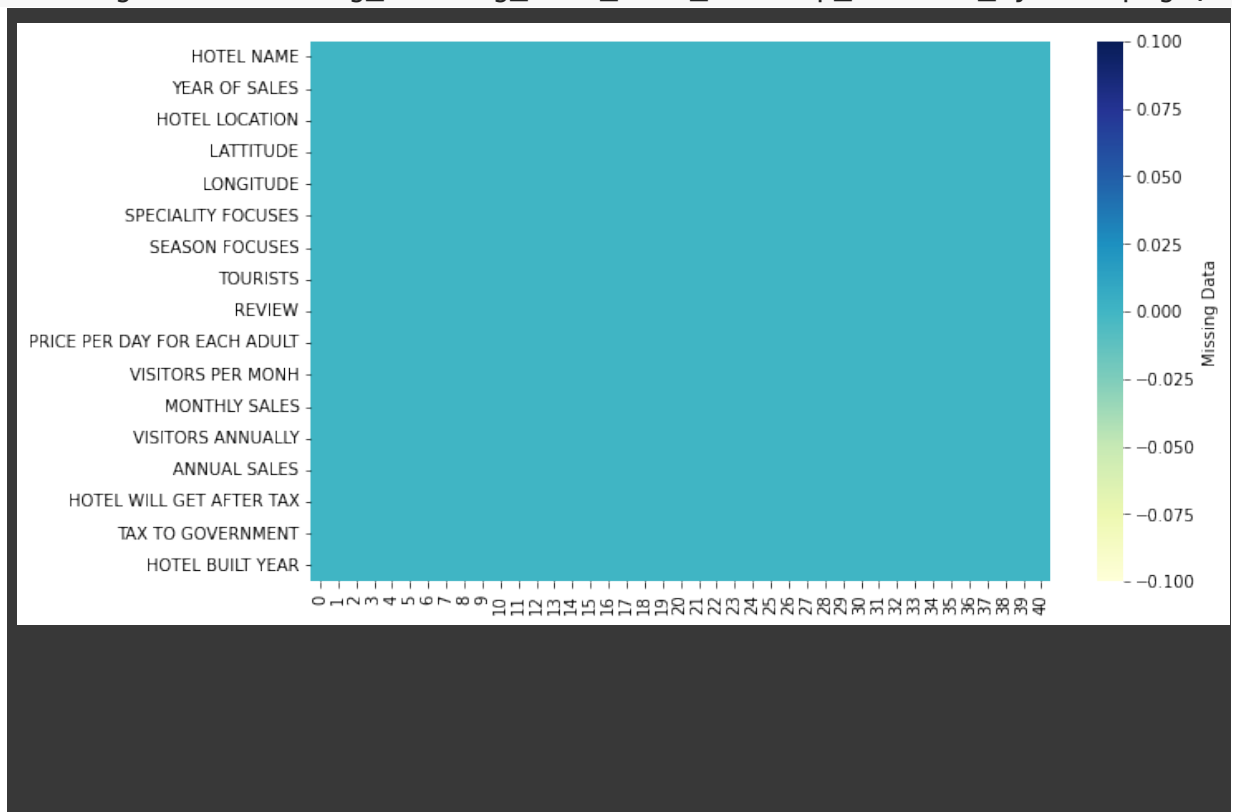
```
df.head(8)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUS
0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MARRIAGE
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY
2	HOTEL RUSSELL	2021	DERBY	52.916668	-1.466667	DINING
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITTY PARTY
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	RECEPTION
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAGEMENT
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAMILY GET TOGETHER
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	GUEST ACCOMODATION ONLY

```
df.isnull().sum().sort_values(ascending=False)
```

```
HOTEL NAME      0
PRICE PER DAY FOR EACH ADULT  0
TAX TO GOVERNMENT  0
HOTEL WILL GET AFTER TAX  0
ANNUAL SALES    0
VISITORS ANNUALLY  0
MONTHLY SALES   0
VISITORS PER MONH  0
REVIEW          0
YEAR OF SALES   0
TOURISTS        0
SEASON FOCUSES  0
SPECIALITY FOCUSES  0
LONGITUDE       0
LATTITUDE       0
HOTEL LOCATION  0
HOTEL BUILT YEAR  0
dtype: int64
```

```
plt.figure(figsize=(10,6))
sns.heatmap(df.isna().transpose(),
            cmap="YlGnBu",
            cbar_kws={'label': 'Missing Data'})
plt.savefig("visualizing_missing_data_with_heatmap_Seaborn_Python.png",
```



## ✓ CHANGING DATA TYPES

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41 entries, 0 to 40
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   HOTEL NAME                           41 non-null     object
1   YEAR OF SALES                        41 non-null     int64
2   HOTEL LOCATION                       41 non-null     object
3   LATITUDE                            41 non-null     float64
4   LONGITUDE                           41 non-null     float64
5   SPECIALITY FOCUSES                  41 non-null     object
6   SEASON FOCUSES                      41 non-null     object
7   TOURISTS                           41 non-null     object
8   REVIEW                             41 non-null     float64
9   PRICE PER DAY FOR EACH ADULT        41 non-null     float64
10  VISITORS PER MONTH                  41 non-null     object
11  MONTHLY SALES                      41 non-null     float64
12  VISITORS ANNUALLY                  41 non-null     int64
13  ANNUAL SALES                      41 non-null     int64
14  HOTEL WILL GET AFTER TAX            41 non-null     float64
15  TAX TO GOVERNMENT                  41 non-null     float64
16  HOTEL BUILT YEAR                   41 non-null     int64
dtypes: float64(7), int64(4), object(6)
memory usage: 5.6+ KB
```

```
#float to number
df["PRICE PER DAY FOR EACH ADULT"]=df["PRICE PER DAY FOR EACH ADULT"].a

df["HOTEL WILL GET AFTER TAX"]=df["HOTEL WILL GET AFTER TAX"].astype(in

df["TAX TO GOVERNMENT"]=df["TAX TO GOVERNMENT"].astype(int)

#object to int

df["VISITORS PER MONTH"]=df["VISITORS PER MONTH"].astype(int)
```



```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41 entries, 0 to 40
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   HOTEL NAME                            41 non-null     object
1   YEAR OF SALES                         41 non-null     int64
2   HOTEL LOCATION                        41 non-null     object
3   LATTITUDE                             41 non-null     float64
4   LONGITUDE                             41 non-null     float64
5   SPECIALITY FOCUSES                    41 non-null     object
6   SEASON FOCUSES                        41 non-null     object
7   TOURISTS                              41 non-null     object
8   REVIEW                                41 non-null     float64
9   PRICE PER DAY FOR EACH ADULT          41 non-null     int64
10  VISITORS PER MONH                     41 non-null     int64
11  MONTHLY SALES                         41 non-null     float64
12  VISITORS ANNUALLY                     41 non-null     int64
13  ANNUAL SALES                          41 non-null     int64
14  HOTEL WILL GET AFTER TAX              41 non-null     int64
15  TAX TO GOVERNMENT                     41 non-null     int64
16  HOTEL BUILT YEAR                      41 non-null     int64
dtypes: float64(4), int64(8), object(5)
memory usage: 5.6+ KB
```

## ✓ HANDLING DUPLICATES

```
df.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
5      False
6      False
7      False
8      False
9      False
10     False
11     False
12     False
13     False
14     False
15     False
16     False
17     False
18     False
19      True
20      True
21      True
22     False
23     False
24     False
25     False
26     False
27     False
28     False
29     False
30     False
31     False
32     False
33     False
34     False
35     False
36     False
37     False
38     False
39     False
40     False
dtype: bool
```

```
new1=df.drop_duplicates()
print(new1)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE
0	LOWRY HOTEL	2021	HULL	53.767750	-0.108000
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644

2	HOTEL RUSSELL	2021	DERBY	52.916668	-1
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1
6	EDEN HOTEL	2021	DURHAM	54.776100	-1
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2
8	ROYAL HOTEL	2021	PERTH	56.396999	-3
9	PALACE HOTEL	2021	BATH	51.380001	-2
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-0
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1
15	HAVEN HOTEL	2021	TRURO	50.259998	-5
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2
22	LOWRY HOTEL	2022	BELFAST	53.767750	-0
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1
28	EDEN HOTEL	2022	DURHAM	54.776100	-1
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-0
35	THE BARN HOTEL	2022	KENT	53.483002	-2
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2

	SPECIALITY FOCUSES	SEASON FOCUSES	TOURISTS	REVIEW	\
0	MARRIAGE	SUMMER	INDIA	5.50	
1	BIRTHDAY	WINTER	ROME	6.00	
2	DINING	SUMMER	QATAR	7.00	
3	KITTY PARTY	SUMMER	IRAN	10.00	
4	RECEPTION	WINTER	PAKISTAN	5.00	
5	ENGAGEMENT	AUTUMN	GERMANY	8.00	
6	FAMILY GET TOGETHER	AUTUMN	DUBAI	8.00	
7	GUEST ACCOMODATION ONLY	WINTER	AUSTRALIA	10.00	
8	BIRTHDAY	AUTUMN	UNITED STATES	9.00	
9	GUEST ACCOMODATION ONLY	SUMMER	EGYPT	10.00	
10	MARRIAGE	SUMMER	SRILANKA	6.70	
11	DINING	WINTER	BELGIUM	8.80	
12	FAMILY GET TOGETHER	AUTUMN	ITALY	9.90	

13	RECEPTION	WINTER	NEWZELAND	5.00
14	BIRTHDAY	SUMMER	CANADA	10.00
15	MARRIAGE	WINTER	NETHERLAND	7.00
16	ENGAGEMENT	AUTUMN	SWITZERLAND	9.00
17	KITTY PARTY	WINTER	INDIA	5.00

```
new1.head(1)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES	SEASON FOCUSES
0	LOWRY HOTEL	2021	HULL	53.76775	-0.335827	MARRIAGE	SUMMER

## ✓ ROWS OPERATIONS

### RENAME VARIABLES IN ROWS

```
new1['TOURISTS'].replace({"UNITED STATES":"US","IRAN":"ISRAEL"},inplace=True)
/usr/local/lib/python3.8/dist-packages/pandas/core/generic.py:6619:
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/10min/5min\\_modifying.html](https://pandas.pydata.org/pandas-docs/stable/10min/5min_modifying.html)

```
return self._update_inplace(result)
```

```
new1
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES
0	LOWRY HOTEL	2021	HULL	53.767750	-0.335827	MARRIAGE
1	GORING HOTEL	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY
2	HOTEL	2021	DERBY	52.916668	-1.466667	RECEPTION

2	RUSSELL	2021	DERBY	52.916667	-1.466667	
3	CARLTON HOTEL	2021	LEICESTER	52.633331	-1.133333	KITT
4	BLUE MOUNT	2021	GLASGOW	55.860916	-4.251433	REC
5	FEATHERS HOTEL	2021	NEWCASTLE	54.966667	-1.600000	ENGAG
6	EDEN HOTEL	2021	DURHAM	54.776100	-1.573300	FAM TO
7	CAPITAL HOTEL	2021	LIVERPOOL	53.400002	-2.983333	ACCOMO
8	ROYAL HOTEL	2021	PERTH	56.396999	-3.437000	BI
9	PALACE HOTEL	2021	BATH	51.380001	-2.360000	ACCOMO
10	THE BULL HOTEL	2021	COVENTRY	52.408054	-1.510556	MA
11	CARBIS BAY HOTEL	2021	CARDIFF	51.481583	-3.179090	
12	HOWTOWN HOTEL	2021	LONDON	51.509865	-0.118092	FAM TO
13	THE BARN HOTEL	2021	SALFORD	53.483002	-2.293100	REC
14	CORBYN HEAD HOTEL	2021	SUNDERLAND	54.906101	-1.381130	BI
15	HAVEN HOTEL	2021	TRURO	50.259998	-5.051000	MA
16	BEDFORD HOTEL	2021	CAMBRIDGE	52.205276	0.119167	ENGAG
17	STONE HOUSE HOTEL	2021	OXFORD	51.752022	-1.257677	KITT
18	CLAREMONT HOTEL	2021	BRISTOL	51.454514	-2.587910	
19	LOWRY	2021	BRISTOL	51.454514	-2.587910	MA

22	HOTEL	2022	BELFAST	53.767750	-0.335827	MA
23	GORING HOTEL	2022	MANCHESTER	53.483959	-2.244644	BI
24	HOTEL RUSSELL	2022	SHEFFIELD	52.916668	-1.466667	
25	CARLTON HOTEL	2022	LEICESTER	52.633331	-1.133333	KITT
26	BLUE MOUNT	2022	GLASGOW	55.860916	-4.251433	REC
27	FEATHERS HOTEL	2022	NEWCASTLE	54.966667	-1.600000	ENGA
28	EDEN HOTEL	2022	DURHAM	54.776100	-1.573300	FAM TO
29	CAPITAL HOTEL	2022	MIDDLESBROUGH	53.400002	-2.983333	ACCOMO
30	ROYAL HOTEL	2022	STROKE ON TRENT	56.396999	-3.437000	BI
31	PALACE HOTEL	2022	STOCKTON	51.380001	-2.360000	ACCOMO
32	THE BULL HOTEL	2022	COVENTRY	52.408054	-1.510556	MA
33	CARBIS BAY HOTEL	2022	CARDIFF	51.481583	-3.179090	
34	HOWTOWN HOTEL	2022	LONDON	51.509865	-0.118092	FAM TO
35	THE BARN HOTEL	2022	KENT	53.483002	-2.293100	REC
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1.381130	BI
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5.051000	MA
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0.119167	ENGA
39	STONE HOUSE	2022	OXFORD	51.752022	-1.257677	KITT

	HOTEL	YEAR	LOCATION	SAL	LONG
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2.587910

## ✓ COLUMN OPERATIONS

### RENAME COLUMNS

```
df_new=new1.rename(columns={'LOWRYA HOTEL':'LOWRY HOTEL'})
print(df_new)
```

	HOTEL NAME	YEAR OF	SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE
0	LOWRY HOTEL	2021		HULL	53.767750	-0.132778
1	GORING HOTEL	2021		MANCHESTER	53.483959	-2.244644
2	HOTEL RUSSELL	2021		DERBY	52.916668	-1.082272
3	CARLTON HOTEL	2021		LEICESTER	52.633331	-1.131211
4	BLUE MOUNT	2021		GLASGOW	55.860916	-4.251806
5	FEATHERS HOTEL	2021		NEWCASTLE	54.966667	-1.082272
6	EDEN HOTEL	2021		DURHAM	54.776100	-1.577412
7	CAPITAL HOTEL	2021		LIVERPOOL	53.400002	-2.990684
8	ROYAL HOTEL	2021		PERTH	56.396999	-3.207693
9	PALACE HOTEL	2021		BATH	51.380001	-2.359084
10	THE BULL HOTEL	2021		COVENTRY	52.408054	-1.501714
11	CARBIS BAY HOTEL	2021		CARDIFF	51.481583	-3.174632
12	HOWTOWN HOTEL	2021		LONDON	51.509865	-0.127758
13	THE BARN HOTEL	2021		SALFORD	53.483002	-2.244644
14	CORBYN HEAD HOTEL	2021		SUNDERLAND	54.906101	-1.082272
15	HAVEN HOTEL	2021		TRURO	50.259998	-5.087403
16	BEDFORD HOTEL	2021		CAMBRIDGE	52.205276	0.122028
17	STONE HOUSE HOTEL	2021		OXFORD	51.752022	-1.257625
18	CLAREMONT HOTEL	2021		BRISTOL	51.454514	-2.587910
22	LOWRY HOTEL	2022		BELFAST	53.767750	-0.132778
23	GORING HOTEL	2022		MANCHESTER	53.483959	-2.244644
24	HOTEL RUSSELL	2022		SHEFFIELD	52.916668	-1.082272
25	CARLTON HOTEL	2022		LEICESTER	52.633331	-1.131211
26	BLUE MOUNT	2022		GLASGOW	55.860916	-4.251806
27	FEATHERS HOTEL	2022		NEWCASTLE	54.966667	-1.082272
28	EDEN HOTEL	2022		DURHAM	54.776100	-1.577412
29	CAPITAL HOTEL	2022		MIDDLESBROUGH	53.400002	-2.990684
30	ROYAL HOTEL	2022		STROKE ON TRENT	56.396999	-3.207693
31	PALACE HOTEL	2022		STOCKTON	51.380001	-2.359084
32	THE BULL HOTEL	2022		COVENTRY	52.408054	-1.501714
33	CARBIS BAY HOTEL	2022		CARDIFF	51.481583	-3.174632
34	HOWTOWN HOTEL	2022		LONDON	51.509865	-0.127758

35	THE BARN HOTEL	2022	KENT	53.483002	-2
36	CORBYN HEAD HOTEL	2022	SUNDERLAND	54.906101	-1
37	HAVEN HOTEL	2022	WASHINGTON	50.259998	-5
38	BEDFORD HOTEL	2022	CAMBRIDGE	52.205276	0
39	STONE HOUSE HOTEL	2022	OXFORD	51.752022	-1
40	CLAREMONT HOTEL	2022	BRISTOL	51.454514	-2

	SPECIALITY FOCUSES	SEASON	FOCUSES	TOURISTS	REVIEW	\
0	MARRIAGE		SUMMER	INDIA	5.50	
1	BIRTHDAY		WINTER	ROME	6.00	
2	DINING		SUMMER	QATAR	7.00	
3	KITTY PARTY		SUMMER	ISRAEL	10.00	
4	RECEPTION		WINTER	PAKISTAN	5.00	
5	ENGAGEMENT		AUTUMN	GERMANY	8.00	
6	FAMILY GET TOGETHER		AUTUMN	DUBAI	8.00	
7	GUEST ACCOMODATION ONLY		WINTER	AUSTRALIA	10.00	
8	BIRTHDAY		AUTUMN	US	9.00	
9	GUEST ACCOMODATION ONLY		SUMMER	EGYPT	10.00	
10	MARRIAGE		SUMMER	SRILANKA	6.70	
11	DINING		WINTER	BELGIUM	8.80	
12	FAMILY GET TOGETHER		AUTUMN	ITALY	9.90	
13	RECEPTION		WINTER	NEWZELAND	5.00	
14	BIRTHDAY		SUMMER	CANADA	10.00	
15	MARRIAGE		WINTER	NETHERLAND	7.00	
16	ENGAGEMENT		AUTUMN	SWITZERLAND	9.00	
17	KITTY PARTY		WINTER	BANGALADESH	5.00	

## REMOVE COLUMN

```
df_new.drop('HOTEL BUILT YEAR',inplace=True,axis=1)
```

```
df_new.head(0)
```

Hotel Name	Year of Sales	Hotel Location	Latitude	Longitude	Speciality Focuses	Season Focuses	Time

## ✓ TEXT OPERATIONS



## CHANGE TO LOWER CASE

```
df_new['HOTEL NAME']=df_new['HOTEL NAME'].str.lower()
```

df\_new

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIAL FOCUS
0	lowry hotel	2021	HULL	53.767750	-0.335827	MARRIAGE
1	goring hotel	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY
2	hotel russell	2021	DERBY	52.916668	-1.466667	DINING
3	carlton hotel	2021	LEICESTER	52.633331	-1.133333	KITTY PANTS
4	blue mount	2021	GLASGOW	55.860916	-4.251433	RECEPTION
5	feathers hotel	2021	NEWCASTLE	54.966667	-1.600000	ENGAGEMENT
6	eden hotel	2021	DURHAM	54.776100	-1.573300	FAMILY TOGETHER
7	capital hotel	2021	LIVERPOOL	53.400002	-2.983333	GUINNESS ACCOMODATION
8	royal hotel	2021	PERTH	56.396999	-3.437000	BIRTHDAY
9	palace hotel	2021	BATH	51.380001	-2.360000	GUINNESS ACCOMODATION
10	the bull hotel	2021	COVENTRY	52.408054	-1.510556	MARRIAGE
11	carbis bay hotel	2021	CARDIFF	51.481583	-3.179090	DINING

12	howtown hotel	2021	LONDON	51.509865	-0.118092	FAMILY TOGET
13	the barn hotel	2021	SALFORD	53.483002	-2.293100	RECEPT
14	corbyn head hotel	2021	SUNDERLAND	54.906101	-1.381130	BIRTH
15	haven hotel	2021	TRURO	50.259998	-5.051000	MARRI
16	bedford hotel	2021	CAMBRIDGE	52.205276	0.119167	ENGAGEM
17	stone house hotel	2021	OXFORD	51.752022	-1.257677	KITTY PA
18	claremont hotel	2021	BRISTOL	51.454514	-2.587910	DIN
22	lowry hotel	2022	BELFAST	53.767750	-0.335827	MARRI
23	goring hotel	2022	MANCHESTER	53.483959	-2.244644	BIRTH
24	hotel russell	2022	SHEFFIELD	52.916668	-1.466667	DIN
25	carlton hotel	2022	LEICESTER	52.633331	-1.133333	KITTY PA
26	blue mount	2022	GLASGOW	55.860916	-4.251433	RECEPT
27	feathers hotel	2022	NEWCASTLE	54.966667	-1.600000	ENGAGEM
28	eden hotel	2022	DURHAM	54.776100	-1.573300	FAMILY TOGET
29	capital hotel	2022	MIDDLESBROUGH	53.400002	-2.983333	GU ACCOMODAT C
30	royal hotel	2022	STROKE ON TRENT	56.396999	-3.437000	BIRTH
31	palace hotel	2022	STOCKTON	51.380001	-2.360000	GU ACCOMODAT C



SALES						
0	lowry hotel	2021	HULL	53.767750	-0.335827	MARRIOTT
1	goring hotel	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY
2	hotel russell	2021	DERBY	52.916668	-1.466667	DINING
3	carlton hotel	2021	LEICESTER	52.633331	-1.133333	KITTY PARK
4	blue mount	2021	GLASGOW	55.860916	-4.251433	RECEPTION
5	feathers hotel	2021	NEWCASTLE	54.966667	-1.600000	ENGAGEMENT
6	eden hotel	2021	DURHAM	54.776100	-1.573300	FAMILY TOGETHER
7	capital hotel	2021	LIVERPOOL	53.400002	-2.983333	GUINNESS ACCOMODATION
8	royal hotel	2021	PERTH	56.396999	-3.437000	BIRTHDAY
9	palace hotel	2021	BATH	51.380001	-2.360000	GUINNESS ACCOMODATION
10	the bull hotel	2021	COVENTRY	52.408054	-1.510556	MARRIOTT
11	carbis bay hotel	2021	CARDIFF	51.481583	-3.179090	DINING
12	howtown hotel	2021	LONDON	51.509865	-0.118092	FAMILY TOGETHER
13	the barn hotel	2021	SALFORD	53.483002	-2.293100	RECEPTION
14	corbyn head hotel	2021	SUNDERLAND	54.906101	-1.381130	BIRTHDAY
15	haven hotel	2021	TRURO	50.259998	-5.051000	MARRIOTT
16	bedford	2021	CHURCH	50.259998	-5.051000	MARRIOTT

16	stone house hotel	2021	CAMBRIDGE	52.205276	0.119167	ENGAGEMENT
17	stone house hotel	2021	OXFORD	51.752022	-1.257677	KITTY PA
18	claremont hotel	2021	BRISTOL	51.454514	-2.587910	DIN
22	lowry hotel	2022	BELFAST	53.767750	-0.335827	MARRI
23	goring hotel	2022	MANCHESTER	53.483959	-2.244644	BIRTH
24	hotel russell	2022	SHEFFIELD	52.916668	-1.466667	DIN
25	carlton hotel	2022	LEICESTER	52.633331	-1.133333	KITTY PA
26	blue mount	2022	GLASGOW	55.860916	-4.251433	RECEPT
27	feathers hotel	2022	NEWCASTLE	54.966667	-1.600000	ENGAGEMENT
28	eden hotel	2022	DURHAM	54.776100	-1.573300	FAMILY TOGET
29	capital hotel	2022	MIDDLESBROUGH	53.400002	-2.983333	GU ACCOMODAT C
30	royal hotel	2022	STROKE ON TRENT	56.396999	-3.437000	BIRTH
31	palace hotel	2022	STOCKTON	51.380001	-2.360000	GU ACCOMODAT C
32	the bull hotel	2022	COVENTRY	52.408054	-1.510556	MARRI
33	carbis bay hotel	2022	CARDIFF	51.481583	-3.179090	DIN
34	howtown hotel	2022	LONDON	51.509865	-0.118092	FAMILY TOGET
35	the barn hotel	2022	KENT	53.483002	-2.293100	RECEPT
36	corbyn head	2022	SUNDERLAND	54.906101	-1.381130	BIRTH

36	head hotel	2022	GONDERLAND	54.966161	-1.561150	BIRTHDAY
37	haven hotel	2022	WASHINGTON	50.259998	-5.051000	MARRIAGE
38	bedford hotel	2022	CAMBRIDGE	52.205276	0.119167	ENGAGEMENT
39	stone house hotel	2022	OXFORD	51.752022	-1.257677	KITTY PARTY
40	claremont hotel	2022	BRISTOL	51.454514	-2.587910	DINING

## FINDING SUM AND FORMATION A NEW COLUMN

```
#subtract discount value from PRICE PER DAY FOR EACH ADULT
# and form a new column named as PREMIUM CUSTOMERS CHARGES PER DAY
df_new['PREMIUM CUSTOMERS PRICE PER DAY']=df_new['PRICE PER DAY FOR EACH ADULT']
```

```
df_new.head()
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES	SEASONAL FOCUS
0	lowry hotel	2021	HULL	53.767750	-0.335827	MARRIAGE	SUMMER
1	goring hotel	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY	WINTER
2	hotel russell	2021	DERBY	52.916668	-1.466667	DINING	SUMMER
3	carlton hotel	2021	LEICESTER	52.633331	-1.133333	KITTY PARTY	SUMMER
4	blue mount	2021	GLASGOW	55.860916	-4.251433	RECEPTION	WINTER

```
df_new.insert(12, 'PREMIUM CUSTOMERS PRICE PER DAY', df_new.pop('PREMIUM CUSTOMERS PRICE PER DAY'))
```

```
df_new.head(5)
```

	HOTEL NAME	YEAR OF SALES	HOTEL LOCATION	LATTITUDE	LONGITUDE	SPECIALITY FOCUSES	SEASON FOCUS
0	lowry hotel	2021	HULL	53.767750	-0.335827	MARRIAGE	SUMMER
1	goring hotel	2021	MANCHESTER	53.483959	-2.244644	BIRTHDAY	WINTER
2	hotel russell	2021	DERBY	52.916668	-1.466667	DINING	SUMMER
3	carlton hotel	2021	LEICESTER	52.633331	-1.133333	KITTY PARTY	SUMMER
4	blue mount	2021	GLASGOW	55.860916	-4.251433	RECEPTION	WINTER

✓ **FINAL CHECKS**

```
df_new.dtypes
```



```
HOTEL NAME          object
YEAR OF SALES        int64
HOTEL LOCATION       object
LATTITUDE            float64
LONGITUDE            float64
SPECIALITY FOCUSES   object
SEASON FOCUSES       object
TOURISTS             object
REVIEW              float64
PRICE PER DAY FOR EACH ADULT  int64
VISITORS PER MONH    int64
DISCOUNT            int64
PREMIUM CUSTOMERS PRICE PER DAY  int64
MONTHLY SALES        float64
VISITORS ANNUALLY    int64
ANNUAL SALES         int64
HOTEL WILL GET AFTER TAX  int64
TAX TO GOVERNMENT     int64
dtype: object
```

## ✓ COPY OF CLEANED DATASET

```
df_new.to_csv('part2.csv',index=False)
```



