```
1 #Read the data with help of google drive
2 from google.colab import drive
```

```
1 drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
1 import pandas as pd
```

```
1 hoteldata= pd.read_csv('/content/drive/MyDrive/hotel_bookings.csv')
```

```
1 from google.colab import drive
2 drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remou
```

```
1 hoteldata
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_d
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	
...	...	...	...	...	...	...	
119385	City Hotel	0	23	2017	August	35	
	City						

```
1 #Read the from local file
2 hoteldata1= pd.read_csv('/content/hotel_bookings.csv')
```

119387	Hotel	0	34	2017	August	35	

```
1 #Display first 20 rows
2 hoteldata.head(20)
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_o
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	
5	Resort Hotel	0	14	2015	July	27	
6	Resort Hotel	0	0	2015	July	27	
7	Resort Hotel	0	9	2015	July	27	
8	Resort Hotel	1	85	2015	July	27	
9	Resort Hotel	1	75	2015	July	27	
10	Resort Hotel	1	23	2015	July	27	
11	Resort Hotel	0	35	2015	July	27	
12	Resort Hotel	0	68	2015	July	27	
13	Resort Hotel	0	18	2015	July	27	
	Resort						

14	Resort Hotel	0	37	2015	July	27
15	Resort Hotel	0	68	2015	July	27
16	Resort Hotel	0	37	2015	July	27
17	Resort	0	12	2015	July	27

```
#2. Display the last 10 rows
hoteldata.tail(10)
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_da
119380	City Hotel	0	44	2017	August	35	
119381	City	0	188	2017	August	35	

```
1 len(hoteldata)
```

```
119390
```

119383	Hotel	0	164	2017	August	35	

```
1 #List of total no of rows and columns
2 hoteldata.shape
```

```
(119390, 32)
```

Hotel

```
1 #4. Use the describe() function
2 hoteldata['lead_time'].describe()
```

```
count    119390.000000
mean        104.011416
std         106.863097
min           0.000000
25%          18.000000
50%          69.000000
75%         160.000000
max         737.000000
Name: lead_time, dtype: float64
```

```
1 hoteldata.describe()
```

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_w
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	
mean	0.370416	104.011416	2016.156554	27.165173	15.798241	
std	0.482918	106.863097	0.707476	13.605138	8.780829	
min	0.000000	0.000000	2015.000000	1.000000	1.000000	
25%	0.000000	18.000000	2016.000000	16.000000	8.000000	
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	
75%	1.000000	160.000000	2017.000000	38.000000	23.000000	

```
1 #find all the null value in term of binary
2 hoteldata.isna()
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_da
0	False	False	False	False	False	False	

```
#4. Use the describe() function
hoteldata['adr'].describe()
```

```
count    119390.000000
mean        101.831122
std          50.535790
min          -6.380000
25%          69.290000
50%          94.575000
75%         126.000000
max        5400.000000
Name: adr, dtype: float64
```

119387	False	False	False	False	False	False

```
#find all the null value in term of total sum
hoteldata.isna().sum()
```

```
hotel                               0
is_canceled                         0
lead_time                           0
arrival_date_year                   0
arrival_date_month                  0
arrival_date_week_number            0
arrival_date_day_of_month           0
stays_in_weekend_nights             0
stays_in_week_nights                0
adults                              0
children                            4
babies                              0
meal                                0
country                           488
market_segment                      0
distribution_channel                0
is_repeated_guest                   0
previous_cancellations              0
previous_bookings_not_canceled      0
reserved_room_type                  0
```

```
assigned_room_type                  0
booking_changes                     0
deposit_type                        0
agent                           16340
company                        112593
days_in_waiting_list                0
customer_type                       0
adr                                 0
required_car_parking_spaces         0
total_of_special_requests           0
reservation_status                  0
reservation_status_date             0
dtype: int64
```

```
def clean_data(pd):
  pd.fillna(0,inplace=True)
  print(pd.isnull().sum())
```

```
clean_data(hoteldata)
```

```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        0
babies                          0
meal                            0
country                         0
market_segment                  0
distribution_channel            0
is_repeated_guest               0
previous_cancellations          0
previous_bookings_not_canceled  0
reserved_room_type              0
```

```
assigned_room_type              0
booking_changes                 0
deposit_type                    0
agent                           0
company                         0
days_in_waiting_list            0
customer_type                   0
adr                             0
required_car_parking_spaces     0
total_of_special_requests       0
reservation_status              0
reservation_status_date         0
dtype: int64
```

```
#3. Type pd.columns() in another code block and observe the columns in the data
hoteldata.columns
```

```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
       'arrival_date_month', 'arrival_date_week_number',
       'arrival_date_day_of_month', 'stays_in_weekend_nights',
       'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
       'country', 'market_segment', 'distribution_channel',
       'is_repeated_guest', 'previous_cancellations',
       'previous_bookings_not_canceled', 'reserved_room_type',
       'assigned_room_type', 'booking_changes', 'deposit_type', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

```
hoteldata.head()
```

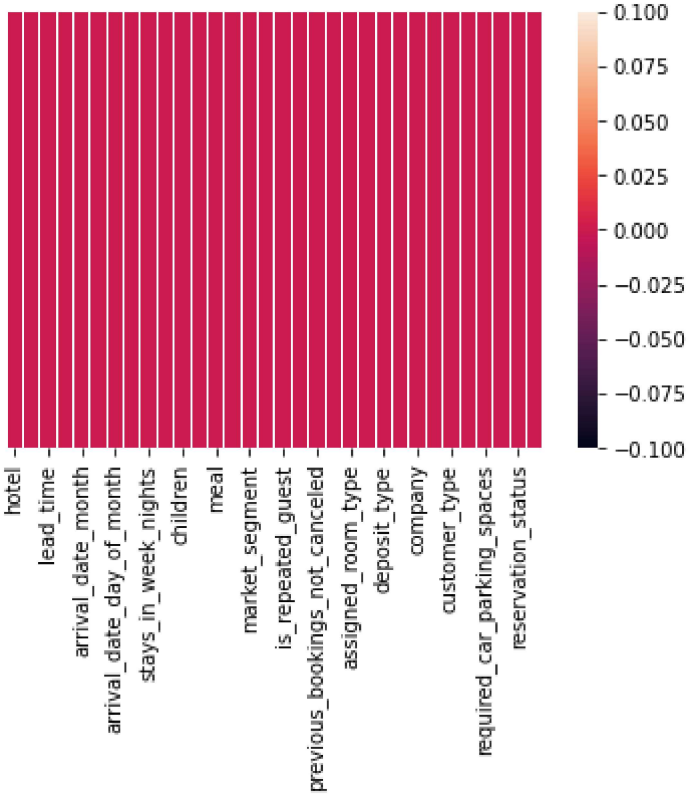
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	

```
1 import seaborn as sns
```

```
1 #heatmap without cleaning null values
2 sns.heatmap(hoteldata1.isnull(), yticklabels=False);
```


```
1 #heatmap after cleaning the null values
2 sns.heatmap(hoteldata.isnull(), yticklabels=False)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1653e9bd0>
```




```
1 HD= hoteldata.drop(hoteldata.index[[-50]])
```

```
1 HD
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_d
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	
...	...	...	...	...	...	...	...
119385	City Hotel	0	23	2017	August	35	
119386	City Hotel	0	102	2017	August	35	
119387	City Hotel	0	34	2017	August	35	
119388	City Hotel	0	109	2017	August	35	
119389	City Hotel	0	205	2017	August	35	

119389 rows x 32 columns

```
#Data Visulisations

HD.drop(HD.tail(50).index,inplace=True)
```

Q-1

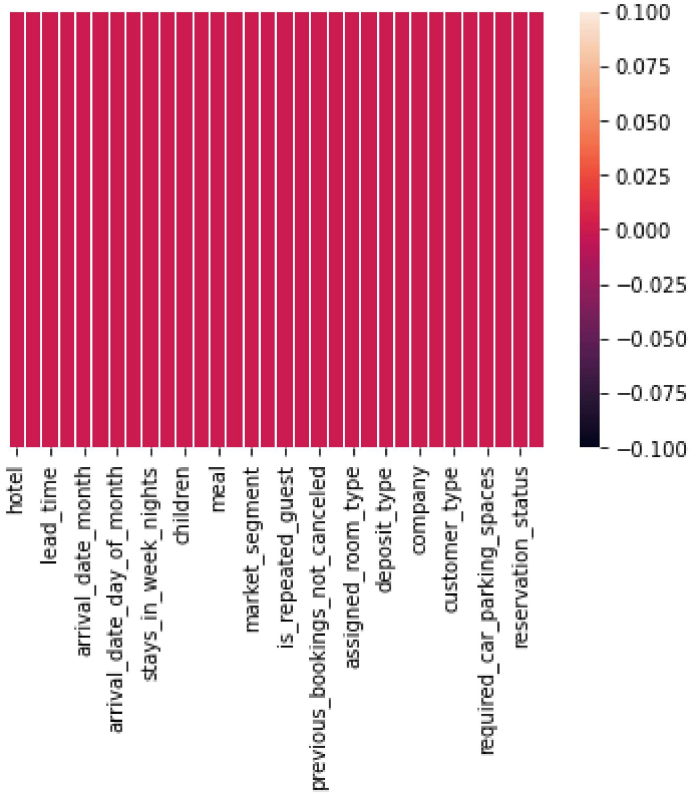
```
1 HD
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_d
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	
...	...	...	...	...	...	...	...
119334	City Hotel	0	325	2017	August	35	
119335	City Hotel	0	63	2017	August	35	
119336	City Hotel	0	103	2017	August	35	
119337	City Hotel	0	107	2017	August	35	
119338	City Hotel	0	137	2017	August	35	

119339 rows × 32 columns

```
1 sns.heatmap(HD.isnull(), yticklabels=False)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa164eaad10>
```



Colab paid products - Cancel contracts here