importing libraries

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
```

Loading the dataset

```
df = pd.read_csv('City_Hotel.csv')
df.head(3)
```

Displaying the data analysis and data frames

```
df.shape
```

	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
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	0.0	0	BB	PRT	Direct	Direct	0	0	0	A	C	3	No Deposit	NaN	NaN	0	Transient	0.0	0	0	Check-Out	2015-07-01
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	0.0	0	BB	PRT	Direct	Direct	0	0	0	A	C	4	No Deposit	NaN	NaN	0	Transient	0.0	0	0	Check-Out	2015-07-01
2	Resort Hotel	0	7	2015	July	27	1	0	1	1	0.0	0	BB	GBR	Direct	Direct	0	0	0	A	A	0	No Deposit	NaN	NaN	0	Transient	75.00	0	0	Check-Out	2015-07-02

3 rows × 32 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32):
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  object
dtype: float64, int64, object
memory usage: 29.1 MB
```

Displaying the data analysis and data frames

```
df.tail(3)
```

	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
119387	City Hotel	0	34	2017	August	35	30	2	4	2	0.0	0	BB	PRT	Online TA	TA/TO	0	0	0	A	A	3	No Deposit	250.0	NaN	0	Transient	96.14	0	0	Check-Out	2017-09-04
119388	City Hotel	0	109	2017	August	35	31	2	5	3	0.0	0	BB	GBR	Online TA	TA/TO	0	0	0	A	A	0	No Deposit	250.0	NaN	0	Transient	80.75	0	1	Check-Out	2017-09-05
119389	City Hotel	0	205	2017	August	35	29	2	7	2	0.0	0	HB	DEU	Online TA	TA/TO	0	0	0	A	A	8	No Deposit	250.0	NaN	0	Transient	157.80	0	2	Check-Out	2017-09-07

3 rows × 32 columns

```
df.shape
```

```
(119390, 32)
```

```
df.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')
```

Checking the data types of the columns and the shape of the dataset

```
df.dtypes
```

```
hotel                           object
is_canceled                      int64
lead_time                        int64
arrival_date_year                int64
arrival_date_month              object
arrival_date_week_number         int64
arrival_date_day_of_month        int64
stays_in_weekend_nights          int64
stays_in_week_nights             int64
adults                           int64
children                       float64
babies                           int64
meal                            object
country                         object
market_segment                  object
distribution_channel            object
is_repeated_guest               int64
previous_cancellations          int64
previous_bookings_not_canceled  int64
reserved_room_type              object
assigned_room_type              object
booking_changes                 int64
deposit_type                    object
agent                          float64
company                        float64
days_in_waiting_list            int64
customer_type                   object
adr                            float64
required_car_parking_spaces     int64
total_of_special_requests       int64
reservation_status              object
reservation_status_date         object
dtype: object
```

Checking the data types of the columns and the shape of the dataset

```
df.shape
```

```
(119390, 32)
```

Checking the number of unique values in each column

```
df.nunique()
```

```
hotel                                2
is_canceled                          2
lead_time                          479
arrival_date_year                    3
arrival_date_month                  12
arrival_date_week_number            53
arrival_date_day_of_month           31
stays_in_weekend_nights             17
stays_in_week_nights                35
adults                              14
children                             4
babies                               5
meal                                 5
country                            166
market_segment                       8
distribution_channel                 5
is_repeated_guest                    2
previous_cancellations              15
previous_bookings_not_canceled      73
reserved_room_type                  10
assigned_room_type                  12
booking_changes                     21
deposit_type                         3
agent                              334
company                            352
days_in_waiting_list               115
customer_type                        4
adr                              8879
required_car_parking_spaces          5
total_of_special_requests            6
reservation_status                   3
reservation_status_date            926
dtype: int64
```

Checking the data types of the columns and the shape of the dataset

```
df.describe()
```

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeated_guest	previous_cancellations	previous_bookings_not_canceled	booking_changes	agent	company	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119386.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	103050.000000	6797.000000	119390.000000	119390.000000	119390.000000	119390.000000
mean	0.370416	104.011416	2016.156554	27.165173	15.798241	0.927599	2.500302	1.856403	0.103890	0.007949	0.032012	0.087396	0.137097	0.221124	86.693382	189.266735	2.321149	101.831122	0.062518	0.571363
std	0.482918	106.863097	0.707476	13.605138	8.780829	0.998613	1.908286	0.579261	0.398561	0.097436	0.176035	0.845869	1.497437	0.652306	110.774548	131.655015	17.594721	50.535790	0.245291	0.792798
min	0.000000	0.000000	2015.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	6.000000	9.000000	0.000000	-6.380000	0.000000	0.000000
25%	0.000000	26.000000	2016.000000	16.000000	8.000000	0.000000	1.000000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	9.000000	62.000000	0.000000	69.290000	0.000000	0.000000
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	1.000000	2.000000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	14.000000	179.000000	0.000000	94.575000	0.000000	0.000000
75%	1.000000	160.000000	2017.000000	38.000000	23.000000	2.000000	3.000000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	229.000000	270.000000	0.000000	126.000000	0.000000	1.000000
max	1.000000	737.000000	2017.000000	53.000000	31.000000	19.000000	50.000000	55.000000	10.000000	10.000000	1.000000	26.000000	72.000000	21.000000	535.000000	543.000000	391.000000	5400.000000	8.000000	5.000000

Checking the data types of the columns and the shape of the dataset

```
df.describe(include='object')
```

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type	customer_type	reservation_status	reservation_status_date
count	119390	119390	119390	118902	119390	119390	119390	119390	119390	119390	119390	119390
unique	2	12	5	166	8	5	10	12	3	4	3	926
top	City Hotel	August	BB	PRT	Online TA	TA/TO	A	A	No Deposit	Transient	Check-Out	2017-10-21
freq	79330	13877	92310	49078	56477	97870	85994	74053	104641	89613	75166	1461

Checking the data types of the columns and the shape of the dataset

```
df.describe(include='bool')
```

```
Empty DataFrame
Columns: []
Index: [count, unique, top, freq]
```

Checking the data types of the columns and the shape of the dataset

```
df.sample(5)
```

```
hotel                           City Hotel
is_canceled                              0
lead_time                               19
arrival_date_year                     2016
arrival_date_month                    July
arrival_date_week_number              28.0
arrival_date_day_of_month             12.0
stays_in_weekend_nights                0.0
stays_in_week_nights                   1.0
adults                                 2.0
children                               0.0
babies                                 0.0
meal                                    BB
country                                 ITA
market_segment                   Complement
distribution_channel                   TA/TO
is_repeated_guest                        0.0
previous_cancellations                   0.0
previous_bookings_not_canceled           0.0
reserved_room_type                       A.0
assigned_room_type                       A.0
booking_changes                          0.0
deposit_type                      Refundable
agent                                    NaN
company                                  NaN
days_in_waiting_list                     0.0
customer_type                        Transient
adr                                     0.00
required_car_parking_spaces              0.0
total_of_special_requests                0.0
reservation_status                 Canceled.0
reservation_status_date            2016-07-12
dtype: 0                                 NaN
1                                        NaN
2                                        NaN
3                                        NaN
4                                        NaN
5                                        NaN
6                                        NaN
7                                        NaN
8                                        NaN
9                                        NaN
10                                       NaN
11                                       NaN
12                                       NaN
13                                       NaN
14                                       NaN
15                                       NaN
16                                       NaN
17                                       NaN
18                                       NaN
19                                       NaN
20                                       NaN
21                                       NaN
22                                       NaN
23                                       NaN
24                                       NaN
25                                       NaN
26                                       NaN
27                                       NaN
28                                       NaN
29                                       NaN
30                                       NaN
31                                       NaN
dtype: object
```