

```
In [1]: import pandas as pd

In [2]: bookings = pd.read_csv('hotel_booking_data.csv', parse_dates = True)

In [3]: bookings

Out[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 | ... | customer_type | |
|--|--------|--------------|-----------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|-----|---------------|----------|
| | 0 | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| | 1 | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| | 2 | Resort Hotel | 0 | 7 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| | 3 | Resort Hotel | 0 | 13 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| | 4 | Resort Hotel | 0 | 14 | 2015 | July | 27 | 1 | 0 | 2 | 2 | ... | Transier |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | 119385 | City Hotel | 0 | 23 | 2017 | August | 35 | 30 | 2 | 5 | 2 | ... | Transier |
| | 119386 | City Hotel | 0 | 102 | 2017 | August | 35 | 31 | 2 | 5 | 3 | ... | Transier |
| | 119387 | City Hotel | 0 | 34 | 2017 | August | 35 | 31 | 2 | 5 | 2 | ... | Transier |
| | 119388 | City Hotel | 0 | 109 | 2017 | August | 35 | 31 | 2 | 5 | 2 | ... | Transier |
| | 119389 | City Hotel | 0 | 205 | 2017 | August | 35 | 29 | 2 | 7 | 2 | ... | Transier |

119390 rows × 36 columns

```
In [4]: bookings.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 36 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
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
32  name                                119390 non-null object
33  email                                119390 non-null object
34  phone-number                        119390 non-null object
35  credit_card                         119390 non-null object
dtypes: float64(4), int64(16), object(16)
memory usage: 32.8+ MB

In [5]: bookings.head()

Out[5]:
```

| | 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 | ... | customer_type |
|---|--------------|-------------|-----------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|-----|---------------|
| 0 | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| 1 | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| 2 | Resort Hotel | 0 | 7 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| 3 | Resort Hotel | 0 | 13 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| 4 | Resort Hotel | 0 | 14 | 2015 | July | 27 | 1 | 0 | 2 | 2 | ... | Transier |

5 rows × 36 columns

```
In [6]: city = bookings[bookings["hotel"] == "City Hotel"]
resort = bookings[bookings["hotel"] == "Resort Hotel"]

In [11]: city.head()

Out[11]:
```

| | 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 | ... | customer_type |
|-------|------------|-------------|-----------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|-----|---------------|
| 40060 | City Hotel | 0 | 6 | 2015 | July | 27 | 1 | 0 | 2 | 1 | ... | Transier |
| 40061 | City Hotel | 1 | 88 | 2015 | July | 27 | 1 | 0 | 4 | 2 | ... | Transier |
| 40062 | City Hotel | 1 | 65 | 2015 | July | 27 | 1 | 0 | 4 | 1 | ... | Transier |
| 40063 | City Hotel | 1 | 92 | 2015 | July | 27 | 1 | 2 | 4 | 2 | ... | Transier |
| 40064 | City Hotel | 1 | 100 | 2015 | July | 27 | 2 | 0 | 2 | 2 | ... | Transier |

5 rows × 36 columns

```
In [12]: resort.head()

Out[12]:
```

| | 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 | ... | customer_type |
|---|--------------|-------------|-----------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|-----|---------------|
| 0 | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| 1 | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transier |
| 2 | Resort Hotel | 0 | 7 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| 3 | Resort Hotel | 0 | 13 | 2015 | July | 27 | 1 | 0 | 1 | 1 | ... | Transier |
| 4 | Resort Hotel | 0 | 14 | 2015 | July | 27 | 1 | 0 | 2 | 2 | ... | Transier |

5 rows × 36 columns

```
In [16]: bookings['total_stay'] = bookings['stays_in_week_nights']+bookings['stays_in_weekend_nights']

In [17]: # Total number of guests
bookings['guests'] = bookings['adults']+bookings['children'] +bookings['babies']

In [19]: bookings = bookings[bookings.deposit_type=="No Deposit"]
bookings.groupby(['deposit_type','is_canceled']).count()

Out[19]:
```

| | hotel | 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 | ... | customer_type |
|--|--------------|-------------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|----------|-------|---------------|
| | deposit_type | is_canceled | | | | | | | | | | |
| | No Deposit | 0 | 74947 | 74947 | 74947 | 74947 | 74947 | 74947 | 74947 | 74947 | 74947 | 74947 |
| | | 1 | 29694 | 29694 | 29694 | 29694 | 29694 | 29694 | 29694 | 29694 | 29694 | 29694 |

2 rows × 36 columns

```
In [21]: bookings["lead_time"].max()

Out[21]: 737

In [23]: bookings.shape

Out[23]: (104641, 38)

In [24]: bookings.dropna()

Out[24]:
```

| | 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 | ... | required_room_type |
|--------|--------------|-------------|-----------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|-----|--------------------|
| 2392 | Resort Hotel | 0 | 6 | 2015 | October | 42 | 11 | 2 | 0 | 2 | ... | ... |
| 2697 | Resort Hotel | 0 | 24 | 2015 | October | 44 | 26 | 7 | 15 | 1 | ... | ... |
| 2867 | Resort Hotel | 0 | 24 | 2015 | November | 45 | 3 | 0 | 3 | 2 | ... | ... |
| 2877 | Resort Hotel | 0 | 24 | 2015 | November | 45 | 3 | 2 | 10 | 1 | ... | ... |
| 2878 | Resort Hotel | 0 | 24 | 2015 | November | 45 | 3 | 3 | 10 | 2 | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 112499 | City Hotel | 0 | 13 | 2017 | May | 21 | 24 | 0 | 1 | 2 | ... | ... |
| 113046 | City Hotel | 0 | 13 | 2017 | May | 22 | 29 | 1 | 3 | 1 | ... | ... |
| 113082 | City Hotel | 0 | 13 | 2017 | May | 22 | 29 | 1 | 3 | 2 | ... | ... |
| 113627 | City Hotel | 0 | 210 | 2017 | June | 23 | 9 | 0 | 1 | 2 | ... | ... |
| 116451 | City Hotel | 0 | 191 | 2017 | July | 29 | 16 | 2 | 2 | 2 | ... | ... |

213 rows × 38 columns

```
In [25]: bookings.drop("lead_time",axis=1)

Out[25]:
```

| | hotel | is_canceled | 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 | ... | required_room_type |
|--------|--------------|-------------|-------------------|--------------------|--------------------------|---------------------------|-------------------------|----------------------|--------|----------|-----|--------------------|
| 0 | Resort Hotel | 0 | 2015 | July | 27 | 1 | 0 | 0 | 2 | 0.0 | ... | ... |
| 1 | Resort Hotel | 0 | 2015 | July | 27 | 1 | 0 | 0 | 2 | 0.0 | ... | ... |
| 2 | Resort Hotel | 0 | 2015 | July | 27 | 1 | 0 | 1 | 1 | 0.0 | ... | ... |
| 3 | Resort Hotel | 0 | 2015 | July | 27 | 1 | 0 | 1 | 1 | 0.0 | ... | ... |
| 4 | Resort Hotel | 0 | 2015 | July | 27 | 1 | 0 | 2 | 2 | 0.0 | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 119385 | City Hotel | 0 | 2017 | August | 35 | 30 | 2 | 5 | 2 | 0.0 | ... | ... |
| 119386 | City Hotel | 0 | 2017 | August | 35 | 31 | 2 | 5 | 3 | 0.0 | ... | ... |
| 119387 | City Hotel | 0 | 2017 | August | 35 | 31 | 2 | 5 | 2 | 0.0 | ... | ... |
| 119388 | City Hotel | 0 | 2017 | August | 35 | 31 | 2 | 5 | 2 | 0.0 | ... | ... |
| 119389 | City Hotel | 0 | 2017 | August | 35 | 29 | 2 | 7 | 2 | 0.0 | ... | ... |

104641 rows × 37 columns

```
In [ ]:
```