

```
In [54]: import pandas as pd
```

```
In [55]: hotel_bookings=pd.read_csv(r"C:\Users\HP\Downloads\hotel_bookings.csv")  
hotel_bookings
```

Out[55]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
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

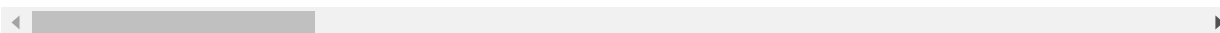
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
10	Resort Hotel	1	23	2015	July	27
11	Resort Hotel	0	35	2015	July	27
12	Resort Hotel	0	68	2015	July	27
13	Resort Hotel	0	18	2015	July	27
14	Resort Hotel	0	37	2015	July	27
15	Resort Hotel	0	68	2015	July	27
16	Resort Hotel	0	37	2015	July	27
17	Resort Hotel	0	12	2015	July	27
18	Resort Hotel	0	0	2015	July	27
19	Resort Hotel	0	7	2015	July	27
20	Resort Hotel	0	37	2015	July	27
21	Resort Hotel	0	72	2015	July	27
22	Resort Hotel	0	72	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
23	Resort Hotel	0	72	2015	July	27
24	Resort Hotel	0	127	2015	July	27
25	Resort Hotel	0	78	2015	July	27
26	Resort Hotel	0	48	2015	July	27
27	Resort Hotel	1	60	2015	July	27
28	Resort Hotel	0	77	2015	July	27
29	Resort Hotel	0	99	2015	July	27
...
119360	City Hotel	0	195	2017	August	35
119361	City Hotel	0	100	2017	August	35
119362	City Hotel	0	198	2017	August	35
119363	City Hotel	0	212	2017	August	35
119364	City Hotel	0	212	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
119365	City Hotel	0	54	2017	August	35
119366	City Hotel	0	210	2017	August	35
119367	City Hotel	0	212	2017	August	35
119368	City Hotel	0	261	2017	August	35
119369	City Hotel	0	207	2017	August	34
119370	City Hotel	0	201	2017	August	35
119371	City Hotel	0	175	2017	August	35
119372	City Hotel	0	175	2017	August	35
119373	City Hotel	0	175	2017	August	35
119374	City Hotel	0	201	2017	August	35
119375	City Hotel	0	200	2017	August	35
119376	City Hotel	0	165	2017	August	35
119377	City Hotel	0	185	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
119378	City Hotel	0	247	2017	August	35
119379	City Hotel	0	109	2017	August	35
119380	City Hotel	0	44	2017	August	35
119381	City Hotel	0	188	2017	August	35
119382	City Hotel	0	135	2017	August	35
119383	City Hotel	0	164	2017	August	35
119384	City Hotel	0	21	2017	August	35
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

119390 rows × 32 columns



```
In [56]: #1. Print the column names of the data frame.
hotel_bookings.columns
```

```
Out[56]: 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', 'agen
               t',
               'company', 'days_in_waiting_list', 'customer_type', 'adr',
               'required_car_parking_spaces', 'total_of_special_requests',
               'reservation_status', 'reservation_status_date'],
              dtype='object')
```

```
In [57]: #2. Find the unique hotel names.
hotel_bookings.hotel.unique()
```

```
Out[57]: array(['Resort Hotel', 'City Hotel'], dtype=object)
```

```
In [67]: grouped = hotel_bookings.groupby(hotel_bookings.hotel)
Resort_Hotel = grouped.get_group("Resort Hotel")
Resort_Hotel
```

```
Out[67]:
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
0	Resort Hotel	0	342	2015	July	27
1	Resort Hotel	0	737	2015	July	27
2	Resort Hotel	0	7	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
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
14	Resort Hotel	0	37	2015	July	27
15	Resort Hotel	0	68	2015	July	27

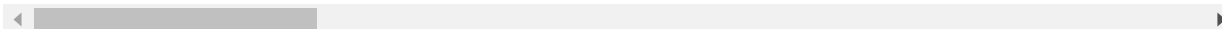
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
16	Resort Hotel	0	37	2015	July	27
17	Resort Hotel	0	12	2015	July	27
18	Resort Hotel	0	0	2015	July	27
19	Resort Hotel	0	7	2015	July	27
20	Resort Hotel	0	37	2015	July	27
21	Resort Hotel	0	72	2015	July	27
22	Resort Hotel	0	72	2015	July	27
23	Resort Hotel	0	72	2015	July	27
24	Resort Hotel	0	127	2015	July	27
25	Resort Hotel	0	78	2015	July	27
26	Resort Hotel	0	48	2015	July	27
27	Resort Hotel	1	60	2015	July	27
28	Resort Hotel	0	77	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
29	Resort Hotel	0	99	2015	July	27
...
40030	Resort Hotel	0	135	2017	August	35
40031	Resort Hotel	0	301	2017	August	35
40032	Resort Hotel	0	21	2017	August	35
40033	Resort Hotel	0	379	2017	August	34
40034	Resort Hotel	0	279	2017	August	35
40035	Resort Hotel	0	21	2017	August	35
40036	Resort Hotel	0	279	2017	August	35
40037	Resort Hotel	0	379	2017	August	34
40038	Resort Hotel	0	191	2017	August	35
40039	Resort Hotel	0	108	2017	August	35
40040	Resort Hotel	0	194	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40041	Resort Hotel	0	227	2017	August	35
40042	Resort Hotel	0	210	2017	August	35
40043	Resort Hotel	0	227	2017	August	35
40044	Resort Hotel	0	270	2017	August	34
40045	Resort Hotel	0	210	2017	August	34
40046	Resort Hotel	0	208	2017	August	34
40047	Resort Hotel	0	214	2017	August	35
40048	Resort Hotel	0	239	2017	August	35
40049	Resort Hotel	0	217	2017	August	35
40050	Resort Hotel	0	173	2017	August	35
40051	Resort Hotel	0	264	2017	August	34
40052	Resort Hotel	0	207	2017	August	34
40053	Resort Hotel	0	269	2017	August	34

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40054	Resort Hotel	0	169	2017	August	35
40055	Resort Hotel	0	212	2017	August	35
40056	Resort Hotel	0	169	2017	August	35
40057	Resort Hotel	0	204	2017	August	35
40058	Resort Hotel	0	211	2017	August	35
40059	Resort Hotel	0	161	2017	August	35

40060 rows × 32 columns



```
In [64]: grouped = hotel_bookings.groupby(hotel_bookings.hotel)
City_Hotel = grouped.get_group("City Hotel")
City_Hotel
```

Out[64]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40060	City Hotel	0	6	2015	July	27
40061	City Hotel	1	88	2015	July	27
40062	City Hotel	1	65	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40063	City Hotel	1	92	2015	July	27
40064	City Hotel	1	100	2015	July	27
40065	City Hotel	1	79	2015	July	27
40066	City Hotel	0	3	2015	July	27
40067	City Hotel	1	63	2015	July	27
40068	City Hotel	1	62	2015	July	27
40069	City Hotel	1	62	2015	July	27
40070	City Hotel	0	43	2015	July	27
40071	City Hotel	0	43	2015	July	27
40072	City Hotel	0	43	2015	July	27
40073	City Hotel	0	4	2015	July	27
40074	City Hotel	1	43	2015	July	27
40075	City Hotel	0	43	2015	July	27

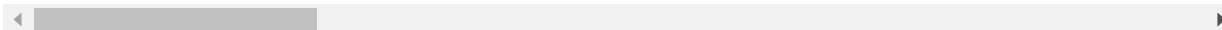
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40076	City Hotel	1	43	2015	July	27
40077	City Hotel	0	43	2015	July	27
40078	City Hotel	0	43	2015	July	27
40079	City Hotel	1	43	2015	July	27
40080	City Hotel	1	43	2015	July	27
40081	City Hotel	1	97	2015	July	27
40082	City Hotel	0	43	2015	July	27
40083	City Hotel	0	43	2015	July	27
40084	City Hotel	1	80	2015	July	27
40085	City Hotel	0	60	2015	July	27
40086	City Hotel	1	106	2015	July	27
40087	City Hotel	1	68	2015	July	27
40088	City Hotel	1	92	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40089	City Hotel	0	18	2015	July	27
...
119360	City Hotel	0	195	2017	August	35
119361	City Hotel	0	100	2017	August	35
119362	City Hotel	0	198	2017	August	35
119363	City Hotel	0	212	2017	August	35
119364	City Hotel	0	212	2017	August	35
119365	City Hotel	0	54	2017	August	35
119366	City Hotel	0	210	2017	August	35
119367	City Hotel	0	212	2017	August	35
119368	City Hotel	0	261	2017	August	35
119369	City Hotel	0	207	2017	August	34
119370	City Hotel	0	201	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
119371	City Hotel	0	175	2017	August	35
119372	City Hotel	0	175	2017	August	35
119373	City Hotel	0	175	2017	August	35
119374	City Hotel	0	201	2017	August	35
119375	City Hotel	0	200	2017	August	35
119376	City Hotel	0	165	2017	August	35
119377	City Hotel	0	185	2017	August	35
119378	City Hotel	0	247	2017	August	35
119379	City Hotel	0	109	2017	August	35
119380	City Hotel	0	44	2017	August	35
119381	City Hotel	0	188	2017	August	35
119382	City Hotel	0	135	2017	August	35
119383	City Hotel	0	164	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
119384	City Hotel	0	21	2017	August	35
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

79330 rows × 32 columns



```
In [75]: #3. Use shape command on the datasets and find the number of samples in
each dataset.
print('Resort_Hotel: ')
display(Resort_Hotel.shape)

print('\nCity_Hotel: ')
display(City_Hotel.shape)
```

Resort_Hotel:

(40060, 32)

City_Hotel:

(79330, 32)

```
In [77]: #4. Find the percentage of cancelled bookings for each hotel.
```



```
print('The resort hotel percentage is {} %'.format(round(Resort_Hotel['is_canceled'].sum()*100/Resort_Hotel['is_canceled'].count(),2)))
print('The city hotel percentage is {} %'.format(round(City_Hotel['is_canceled'].sum()*100/City_Hotel['is_canceled'].count(),2)))
```

The resort hotel percentage is 27.76 %
The city hotel percentage is 41.73 %

In [78]: *# 5. Take all the bookings where is_canceled == 0 from each dataset. Basically you need to remove all the rows where is_canceled == 1.*
display(Resort_Hotel[Resort_Hotel['is_canceled']==0])
display(City_Hotel[City_Hotel['is_canceled']==0])

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
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

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
11	Resort Hotel	0	35	2015	July	27
12	Resort Hotel	0	68	2015	July	27
13	Resort Hotel	0	18	2015	July	27
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 Hotel	0	12	2015	July	27
18	Resort Hotel	0	0	2015	July	27
19	Resort Hotel	0	7	2015	July	27
20	Resort Hotel	0	37	2015	July	27
21	Resort Hotel	0	72	2015	July	27
22	Resort Hotel	0	72	2015	July	27
23	Resort Hotel	0	72	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
24	Resort Hotel	0	127	2015	July	27
25	Resort Hotel	0	78	2015	July	27
26	Resort Hotel	0	48	2015	July	27
28	Resort Hotel	0	77	2015	July	27
29	Resort Hotel	0	99	2015	July	27
30	Resort Hotel	0	118	2015	July	27
31	Resort Hotel	0	95	2015	July	27
33	Resort Hotel	0	69	2015	July	27
36	Resort Hotel	0	15	2015	July	27
...
40030	Resort Hotel	0	135	2017	August	35
40031	Resort Hotel	0	301	2017	August	35
40032	Resort Hotel	0	21	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40033	Resort Hotel	0	379	2017	August	34
40034	Resort Hotel	0	279	2017	August	35
40035	Resort Hotel	0	21	2017	August	35
40036	Resort Hotel	0	279	2017	August	35
40037	Resort Hotel	0	379	2017	August	34
40038	Resort Hotel	0	191	2017	August	35
40039	Resort Hotel	0	108	2017	August	35
40040	Resort Hotel	0	194	2017	August	35
40041	Resort Hotel	0	227	2017	August	35
40042	Resort Hotel	0	210	2017	August	35
40043	Resort Hotel	0	227	2017	August	35
40044	Resort Hotel	0	270	2017	August	34
40045	Resort Hotel	0	210	2017	August	34

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40046	Resort Hotel	0	208	2017	August	34
40047	Resort Hotel	0	214	2017	August	35
40048	Resort Hotel	0	239	2017	August	35
40049	Resort Hotel	0	217	2017	August	35
40050	Resort Hotel	0	173	2017	August	35
40051	Resort Hotel	0	264	2017	August	34
40052	Resort Hotel	0	207	2017	August	34
40053	Resort Hotel	0	269	2017	August	34
40054	Resort Hotel	0	169	2017	August	35
40055	Resort Hotel	0	212	2017	August	35
40056	Resort Hotel	0	169	2017	August	35
40057	Resort Hotel	0	204	2017	August	35
40058	Resort Hotel	0	211	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40059	Resort Hotel	0	161	2017	August	35

28938 rows × 32 columns

◀		▶
---	--	---

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40060	City Hotel	0	6	2015	July	27
40066	City Hotel	0	3	2015	July	27
40070	City Hotel	0	43	2015	July	27
40071	City Hotel	0	43	2015	July	27
40072	City Hotel	0	43	2015	July	27
40073	City Hotel	0	4	2015	July	27
40075	City Hotel	0	43	2015	July	27
40077	City Hotel	0	43	2015	July	27
40078	City Hotel	0	43	2015	July	27
40082	City Hotel	0	43	2015	July	27

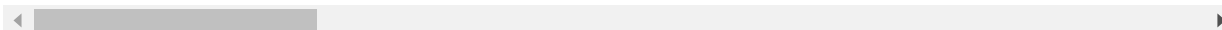
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40083	City Hotel	0	43	2015	July	27
40085	City Hotel	0	60	2015	July	27
40089	City Hotel	0	18	2015	July	27
40104	City Hotel	0	69	2015	July	28
40113	City Hotel	0	54	2015	July	28
40117	City Hotel	0	6	2015	July	28
40119	City Hotel	0	51	2015	July	28
40121	City Hotel	0	72	2015	July	28
40127	City Hotel	0	92	2015	July	28
40132	City Hotel	0	67	2015	July	28
40135	City Hotel	0	71	2015	July	28
40136	City Hotel	0	3	2015	July	28
40137	City Hotel	0	71	2015	July	28

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
40138	City Hotel	0	3	2015	July	28
40140	City Hotel	0	96	2015	July	28
40141	City Hotel	0	96	2015	July	28
40144	City Hotel	0	96	2015	July	28
40146	City Hotel	0	3	2015	July	28
40147	City Hotel	0	61	2015	July	28
40148	City Hotel	0	96	2015	July	28
...
119360	City Hotel	0	195	2017	August	35
119361	City Hotel	0	100	2017	August	35
119362	City Hotel	0	198	2017	August	35
119363	City Hotel	0	212	2017	August	35
119364	City Hotel	0	212	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
119365	City Hotel	0	54	2017	August	35
119366	City Hotel	0	210	2017	August	35
119367	City Hotel	0	212	2017	August	35
119368	City Hotel	0	261	2017	August	35
119369	City Hotel	0	207	2017	August	34
119370	City Hotel	0	201	2017	August	35
119371	City Hotel	0	175	2017	August	35
119372	City Hotel	0	175	2017	August	35
119373	City Hotel	0	175	2017	August	35
119374	City Hotel	0	201	2017	August	35
119375	City Hotel	0	200	2017	August	35
119376	City Hotel	0	165	2017	August	35
119377	City Hotel	0	185	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
119378	City Hotel	0	247	2017	August	35
119379	City Hotel	0	109	2017	August	35
119380	City Hotel	0	44	2017	August	35
119381	City Hotel	0	188	2017	August	35
119382	City Hotel	0	135	2017	August	35
119383	City Hotel	0	164	2017	August	35
119384	City Hotel	0	21	2017	August	35
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

46228 rows × 32 columns



```
In [85]: #6. Find the month which has the highest number of bookings using the mode command
from scipy import stats
```

```
In [86]: x = stats.mode(Resort_Hotel['arrival_date_month'])
y = stats.mode(City_Hotel['arrival_date_month'])
print('The month {} has the highest number of bookings in {} which is {}'.format(x.mode[0],Resort_Hotel.iloc[0,0],int(x.count)))
print('The month {} has the highest number of bookings in {} which is {}'.format(y.mode[0],City_Hotel.iloc[0,0],int(y.count)))
```

```
C:\Users\HP\AppData\Roaming\Python\Python36\site-packages\scipy\stats\stats.py:245: RuntimeWarning: The input array could not be properly checked for nan values. nan values will be ignored.
"values. nan values will be ignored.", RuntimeWarning)
```

The month August has the highest number of bookings in Resort Hotel which is 4894

The month August has the highest number of bookings in City Hotel which is 8983

```
In [87]: #7. Find the most frequent room type reserved
def frequent_room_type(hotel):
    x = stats.mode(hotel['reserved_room_type'])
    print('The type {} room has the highest number of bookings in {} which is {}'.format(x.mode[0],hotel.iloc[0,0],int(x.count)))
    return
```

```
In [88]: frequent_room_type(Resort_Hotel)
frequent_room_type(City_Hotel)
```

```
C:\Users\HP\AppData\Roaming\Python\Python36\site-packages\scipy\stats\stats.py:245: RuntimeWarning: The input array could not be properly checked for nan values. nan values will be ignored.
"values. nan values will be ignored.", RuntimeWarning)
```

The type A room has the highest number of bookings in Resort Hotel which

h is 23399

The type A room has the highest number of bookings in City Hotel which is 62595

```
In [90]: #8. Find the average lead time for the bookings
print('The average lead time of resort hotel is {}'.format(Resort_Hotel
['lead_time'].mean()))
print('The average lead time of city hotel is {:.2f}'.format(City_Hotel
['lead_time'].mean()))
```

The average lead time of resort hotel is 92.67568647029456

The average lead time of city hotel is 109.74

```
In [92]: #9. Find the average number of adults who stayed for each booking
print('The average adults of resort hotel is {:.0f}'.format(Resort_Hotel
['adults'].mean()))
print('The average adults of city hotel is {:.0f}'.format(City_Hotel['a
dults'].mean()))
```

The average adults of resort hotel is 2

The average adults of city hotel is 2

```
In [94]: #10. Find the unique distribution channels for the bookings made
print('The unique distribution channels for resort hotel is {}'.format(
list(Resort_Hotel['distribution_channel'].unique())))
print('The unique distribution channels for city hotel is {}'.format(li
st(City_Hotel['distribution_channel'].unique())))
```

The unique distribution channels for resort hotel is ['Direct', 'Corporate', 'TA/T0', 'Undefined']

The unique distribution channels for city hotel is ['TA/T0', 'Direct', 'Undefined', 'Corporate', 'GDS']

```
In [97]: #11. Using pivot table find the total number of adults who stayed in ea
ch hotel grouped by the distribution channels they used for the booking
import numpy as np
pivot_table_adults=pd.pivot_table(hotel_bookings, index='hotel',columns
='distribution_channel',values='adults', aggfunc=np.sum,fill_value=0)
pivot_table_adults
```

Out[97]:

distribution_channel	Corporate	Direct	GDS	TA/TO	Undefined
hotel					
City Hotel	4237	12246	210	130136	9
Resort Hotel	4579	14768	0	55449	2

```
In [99]: #12. Identify the unique deposit types. Find the most frequent deposit
type
frequent_deposit_type = stats.mode(hotel_bookings['deposit_type'])
print('Unique deposit types are: {} \nand the most frequent deposit typ
e is {} with total counts {}'.format(hotel_bookings['deposit_type'].uni
que(),frequent_deposit_type.mode[0],int(frequent_deposit_type.count)))
```

Unique deposit types are: ['No Deposit' 'Refundable' 'Non Refund']
and the most frequent deposit type is No Deposit with total counts 104641

C:\Users\HP\AppData\Roaming\Python\Python36\site-packages\scipy\stats\s
tats.py:245: RuntimeWarning: The input array could not be properly chec
ked for nan values. nan values will be ignored.
"values. nan values will be ignored.", RuntimeWarning)

```
In [101]: #13. Find the unique country of origin of the people who made bookings
to the hotel
unique_country = hotel_bookings[hotel_bookings['booking_changes'] != 0]
print('The unique country of origin of the people who made bookings to
the hotel are \n\n{}'.format(unique_country['country'].unique()))
```

The unique country of origin of the people who made bookings to the hotel are

```
['PRT' 'ESP' nan 'GBR' 'IRL' 'CHE' 'FRA' 'NLD' 'DNK' 'RUS' 'CZE' 'DEU'
 'ITA' 'FIN' 'POL' 'SWE' 'USA' 'MOZ' 'BWA' 'LUX' 'ROU' 'IND' 'MAR' 'BR
A'
 'LVA' 'SRB' 'BEL' 'AGO' 'AUT' 'CPV' 'ZAF' 'KOR' 'AUS' 'HUN' 'CN' 'IRN'
 'JEY' 'EST' 'CAF' 'URY' 'ISR' 'CHN' 'NOR' 'HRV' 'PHL' 'COL' 'GEO' 'GI
B'
```

```
'BHR' 'CYP' 'SVN' 'ARG' 'DOM' 'MKD' 'NZL' 'LTU' 'MEX' 'UKR' 'CUB' 'GR
C'
'THA' 'CIV' 'OMN' 'PRI' 'LBN' 'AZE' 'ARE' 'DZA' 'BLR' 'SUR' 'SVK' 'BG
R'
'MYS' 'IDN' 'TUR' 'JPN' 'JAM' 'CHL' 'KAZ' 'NGA' 'MAC' 'MDV' 'TWN' 'PE
R'
'SGP' 'KNA' 'JOR' 'KWT' 'LKA' 'IRQ' 'EGY' 'SEN' 'TUN' 'ISL' 'NIC' 'ML
T'
'SAU' 'TZA' 'HKG' 'AND' 'ZWE' 'CRI' 'CMR' 'MUS' 'KEN' 'ARM' 'LBY' 'PA
N'
'GNB' 'VEN' 'BOL' 'TMP' 'BIH' 'PYF' 'GUY' 'MNE' 'GTM' 'ALB' 'SYR' 'QA
T'
'GHA' 'ECU' 'STP' 'PAK' 'TJK' 'VNM' 'FRO']
```

```
In [103]: #14. Using pivot table find the aggregate number of people who have mad
e bookings from each unique country of origin
import numpy as np
pivot_table_aggregate=pd.pivot_table(unique_country, index='country',co
lums='hotel',values='booking_changes', aggfunc=np.sum,fill_value=0)
pivot_table_aggregate
```

Out[103]:

hotel	City Hotel	Resort Hotel
country		
AGO	128	9
ALB	3	0
AND	1	0
ARE	3	3
ARG	33	9
ARM	1	0

hotel	City Hotel	Resort Hotel
Country		
AUS	83	28
AUT	326	158
AZE	4	3
BEL	341	163
BGR	24	2
BHR	0	4
BIH	2	0
BLR	5	1
BOL	2	0
BRA	452	137
BWA	0	3
CAF	0	1
CHE	274	124
CHL	23	7
CHN	181	22
CIV	2	3
CMR	4	0
CN	102	164
COL	18	2
CPV	5	2
CRI	8	0

hotel	City Hotel	Resort Hotel
Country	3	1
CYP	22	4
CZE	37	27
...
PRI	0	2
PRT	3943	4309
PYF	1	0
QAT	2	0
ROU	67	64
RUS	97	69
SAU	8	0
SEN	3	0
SGP	12	0
SRB	30	2
STP	2	0
SUR	0	1
SVK	7	1
SVN	12	4
SWE	172	111
SYR	3	0
THA	10	1

hotel	City Hotel	Resort Hotel
ADN	2	0
TMP	1	0
TUN	8	0
TUR	55	3
TWN	10	2
TZA	1	0
UKR	15	9
URY	2	5
USA	553	157
VEN	1	0
VNM	1	0
ZAF	10	12
ZWE	4	0

126 rows × 2 columns

```
In [105]: #15. At how many occasions reserved room type was not the same as the assigned room type
occasions_nonsimilar = hotel_bookings['reserved_room_type'] != hotel_bookings['assigned_room_type']
reserverd_assigned_not_same = hotel_bookings[occasions_nonsimilar]
display(reserverd_assigned_not_same)
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_day_of_month
2	Resort Hotel	0	7	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date
12	Resort Hotel	0	68	2015	July	27
15	Resort Hotel	0	68	2015	July	27
17	Resort Hotel	0	12	2015	July	27
18	Resort Hotel	0	0	2015	July	27
24	Resort Hotel	0	127	2015	July	27
33	Resort Hotel	0	69	2015	July	27
36	Resort Hotel	0	15	2015	July	27
48	Resort Hotel	0	90	2015	July	27
49	Resort Hotel	0	50	2015	July	27
53	Resort Hotel	0	3	2015	July	27
57	Resort Hotel	0	0	2015	July	27
59	Resort Hotel	0	14	2015	July	27
63	Resort Hotel	0	93	2015	July	27

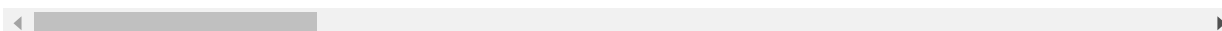
	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
65	Resort Hotel	0	10	2015	July	27
66	Resort Hotel	0	3	2015	July	27
88	Resort Hotel	0	12	2015	July	27
89	Resort Hotel	0	9	2015	July	27
90	Resort Hotel	0	1	2015	July	27
100	Resort Hotel	0	4	2015	July	27
103	Resort Hotel	0	95	2015	July	27
111	Resort Hotel	0	78	2015	July	27
119	Resort Hotel	0	2	2015	July	27
120	Resort Hotel	0	30	2015	July	27
122	Resort Hotel	1	36	2015	July	27
127	Resort Hotel	0	100	2015	July	27
130	Resort Hotel	0	100	2015	July	27

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
131	Resort Hotel	0	9	2015	July	28
133	Resort Hotel	0	0	2015	July	28
135	Resort Hotel	0	2	2015	July	28
...
118796	City Hotel	0	205	2017	August	34
118810	City Hotel	0	2	2017	August	34
118826	City Hotel	0	179	2017	August	34
118833	City Hotel	0	6	2017	August	34
118856	City Hotel	0	143	2017	August	34
118885	City Hotel	0	457	2017	August	34
118931	City Hotel	0	224	2017	August	34
118946	City Hotel	0	31	2017	August	34
118952	City Hotel	0	457	2017	August	34

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
118979	City Hotel	0	16	2017	August	34
119042	City Hotel	0	6	2017	August	34
119044	City Hotel	0	3	2017	August	34
119121	City Hotel	0	240	2017	August	35
119143	City Hotel	0	323	2017	August	35
119146	City Hotel	0	39	2017	August	35
119153	City Hotel	0	22	2017	August	35
119155	City Hotel	0	214	2017	August	35
119173	City Hotel	0	95	2017	August	35
119183	City Hotel	0	209	2017	August	35
119234	City Hotel	0	189	2017	August	35
119235	City Hotel	0	189	2017	August	35
119256	City Hotel	0	279	2017	August	35

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_dat
119258	City Hotel	0	163	2017	August	35
119269	City Hotel	0	178	2017	August	35
119270	City Hotel	0	0	2017	August	35
119273	City Hotel	0	213	2017	August	35
119274	City Hotel	0	213	2017	August	35
119289	City Hotel	0	25	2017	August	35
119297	City Hotel	0	332	2017	August	35
119357	City Hotel	0	47	2017	August	35

14917 rows × 32 columns



```
In [106]: #16. Find the average length of the stay. For this you need to sum the
           weekend and weekday stay columns and then find the mean
           print('The average length of the stay is {}'.format(np.mean(hotel_booki
           ngs['stays_in_weekend_nights'] + hotel_bookings['stays_in_week_nights'
           ])))
```

The average length of the stay is 3.4279001591423066