Data analysis about hotel booking

In this project I analyze different information about the hotels guests like who cancels the booking, number of guests per country. For this project I did different steps to get the results as good data analysis I had to do first check the data source in this a CSV file, then I check the data using python, load the data in the software, then remove the duplicates, drop the irrelevant data and fix the data.

I also did a pivot table to check if there is any difference between assigned and reserved room types as follows:

```
In [57]: ns = data['assigned_room_type'], margins= True, normalize = 'index').round(2)*100
```

```
In [58]: pivot_normalize
```

Out[58]:

assigned_room_type reserved_room_type	A	B	C	D	E	F	G	H	I	K	L
A	81.0	2.0	2.0	11.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0
B	11.0	88.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
C	1.0	0.0	95.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0
D	2.0	0.0	0.0	92.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.0	90.0	6.0	2.0	0.0	1.0	0.0	0.0
F	0.0	0.0	0.0	0.0	1.0	93.0	4.0	0.0	0.0	0.0	0.0
G	0.0	0.0	0.0	0.0	0.0	1.0	98.0	0.0	1.0	0.0	0.0
H	0.0	0.0	0.0	0.0	0.0	0.0	2.0	97.0	1.0	0.0	0.0
L	17.0	17.0	17.0	0.0	0.0	17.0	0.0	17.0	0.0	0.0	17.0
All	53.0	2.0	2.0	26.0	8.0	4.0	3.0	1.0	0.0	0.0	0.0

I also did an analysis about market segment with the highest booking, I did a pie chart for visual representation.

```
In [58]: # show the pie chart
         fig.show()
```

```
In [54]: data['market_segment'].value_counts()
```

```
Out[54]: Online TA          51553
         Offline TA/TO      13855
         Direct             11780
         Groups              4922
         Corporate           4200
         Complementary        692
         Aviation             226
         Undefined              2
         Name: market_segment, dtype: int64
```
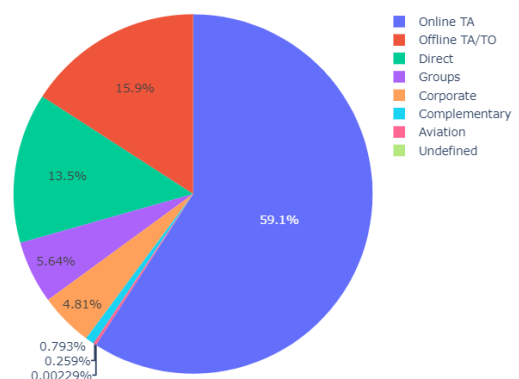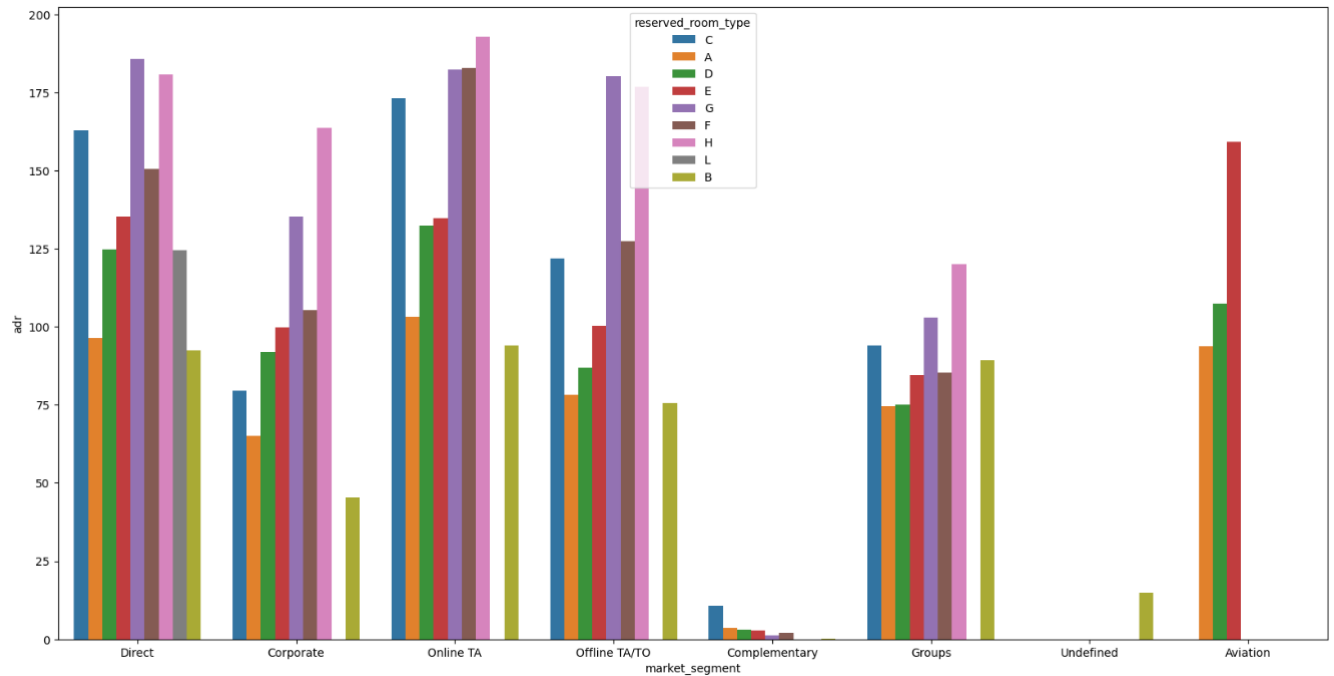
Then I did a analysis about average price per night (ADR) of various room-types for all the market segment, so I did a chart bar:

```
In [71]: plt.figure(figsize= (20,10))
         sns.barplot(x= 'market_segment', y = 'adr', hue= 'reserved_room_type', data=data, ci=None)
```

```
C:\Users\luist\AppData\Local\Temp\ipykernel_1520\3294906479.py:2: FutureWarning:


The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.
```
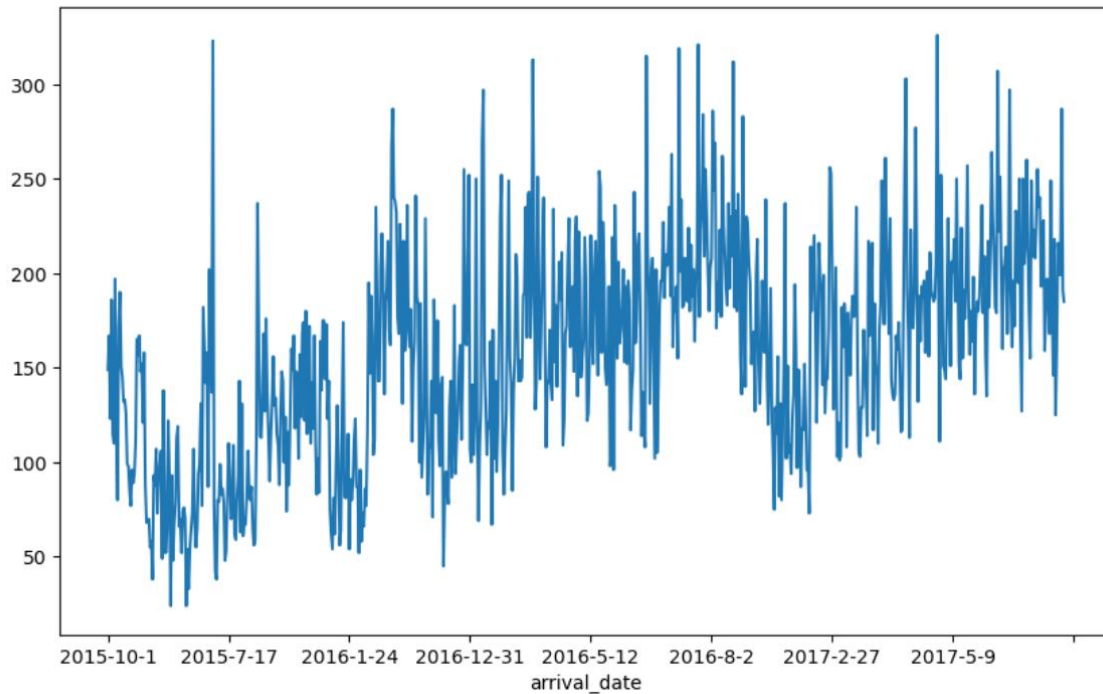
Out[71]: \<Axes: xlabel='market_segment', ylabel='adr'\>

Next I analyze the data about the total number of guests all included together: adults, children and babies combining the data of those columns in a single one. I also combine the columns of dates year, month, and day.
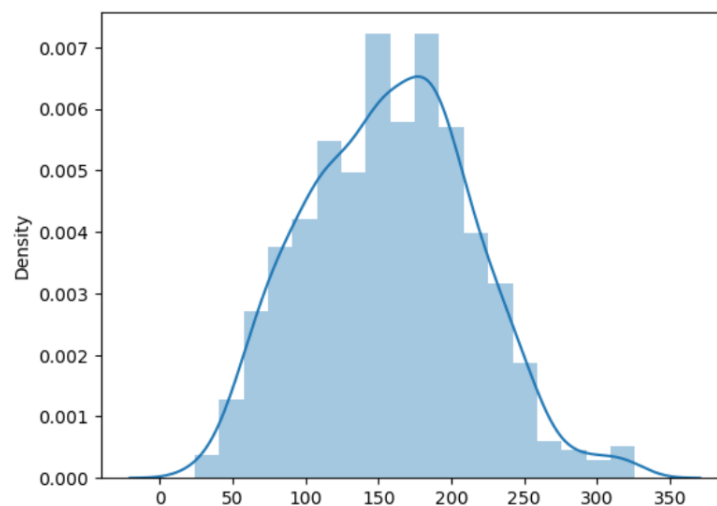
So I did a plot as follows with tha data arrival date and the total number of guests.

```
]: guest_arrival_series.plot(figsize=(10,6))
]: <Axes: xlabel='arrival_date'>
```


Following the analysis, I also created a histogram and density curve

```
7]: sns.distplot(guest_arrival_series.values)
7]: <Axes: ylabel='Density'>
```

As a conclusion, with this data I found interesting results about the different guests sectors as we can see I found that online TA is the biggest one in this case.

I also found that number of rooms reseved by the different sectors Varied and with different prices.

The results found about the different number and type of guests through time.