

Capstone Project

Hotel Booking Analysis BY

Dinesh Anna Wagh

Points to Discuss

- **Agenda**
- **Data Summary**
- **Hotel wise analysis**
- **Best arrival period**
- **Meal Wise analysis**
- **country wise hotel bookings**
- **Room type wise bookings**
- **Distribution Channel wise analysis**
- **Booking cancellation analysis**
- **Market segments wise booking**
- **Conclusion**

Agenda

To discuss the analysis of given hotel bookings data set.

We'll be doing analysis of given data set in following ways

- Hotel wise analysis
- Distribution Channel wise analysis
- Booking cancellation analysis
- country wise hotel bookings
- Market segments wise booking

By doing this we'll try to find out key factors driving the hotel bookings trends.

Data Summary

Given data set has different columns of variables crucial for hotel bookings.

Some of them are:

Hotel: The category of hotels, which are two resort hotel and city hotel.

Is_cancelled : The value of column show the cancellation type. If the booking was cancelled or not. Values[0,1], where 0 indicates not cancelled.

Lead_time : The time between reservation and actual arrival.

Stayed_in_weekend_nights: The number of weekend nights stay per reservation

Stayed_in_weekday_nights: The number of weekday nights stay per reservation.

Meal: Meal preferences per reservation.[BB , FB ,HB ,SC ,Undefined.]

Data Summary

Country: The origin country of guest.

Market_segment: This column show how reservation was made and what is the purpose of reservation. E.g. corporate means corporate trip, TA for travel agency.

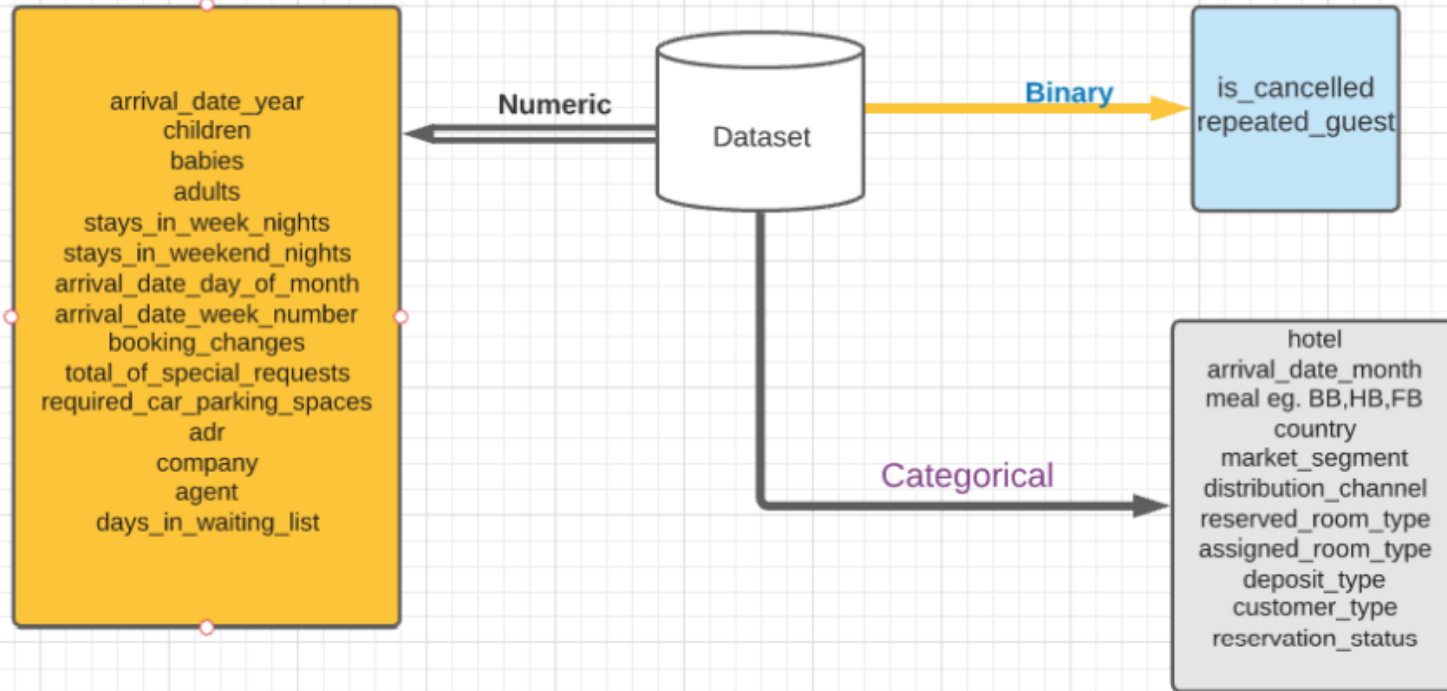
distribution_channel: The medium through booking was made.[Direct,Corporate,TA/TO, undefined, GDS.]

reserved_room_type: The type of rooms available for bookings (A , B , C , D , E etc.)

days_in_waiting_list: Number of days between actual booking and transact.

customer_type: Type of customers(Transient, group, etc.)

Data Summary

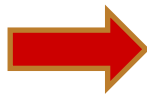


Data Cleaning & Manipulation

- There were 4 columns Company , Agent , Country and Children with missing values.

```
booking_data.isnull().sum().sort_values(ascending=False)[0:8]
```

```
company      112593
agent        16340
country       488
children       4
reserved_room_type  0
assigned_room_type  0
booking_changes  0
deposit_type  0
dtype: int64
```



```
booking_data['company'].fillna(0,inplace=True)
booking_data['agent'].fillna(0,inplace=True)
booking_data['country'].fillna('other',inplace=True)
booking_data['children'].fillna(0,inplace=True)
```

```
booking_data.isnull().sum().sort_values(ascending =False)[0:5]
```

```
hotel      0
is_canceled 0
reservation_status 0
total_of_special_requests 0
required_car_parking_spaces 0
dtype: int64
```

- Handling Duplicates – Data had duplicates value so we dropped it from data.



```
booking_data.duplicated().value_counts()
```

```
False      87396
True       31994
dtype: int64
```

```
[ ] booking_data.drop_duplicates(inplace=True)
```

```
[ ] booking_data.shape
```

```
(87396, 32)
```

Hotel wise analysis

- Question 1 - Which hotel is mostly preferred for bookings?

```
[22] ob1
```

```
City Hotel      79330  
Resort Hotel    40060  
Name: hotel, dtype: int64
```



- As we can see from above chart city hotels are mostly preferred by peoples while booking hotels.
- We can say city hotel is the busiest hotel.

Best Arrival Period

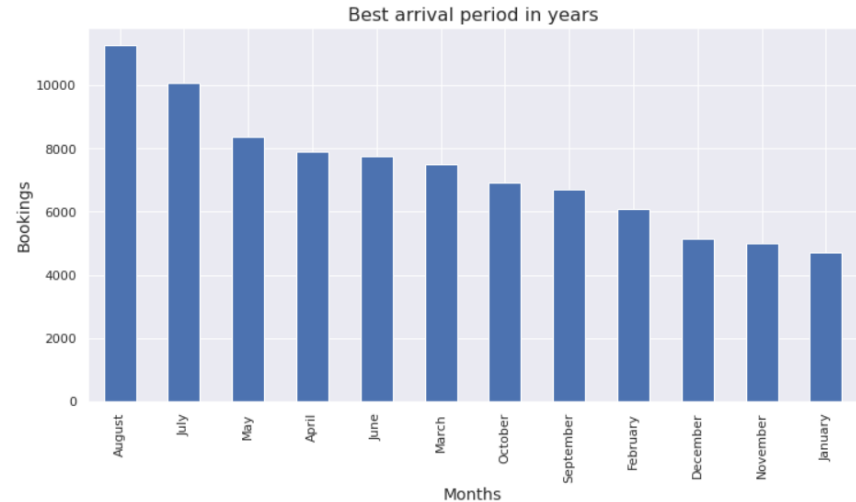
- Question 2 - Which is the best arrival period in the year when maximum bookings are done by guest.



ob2

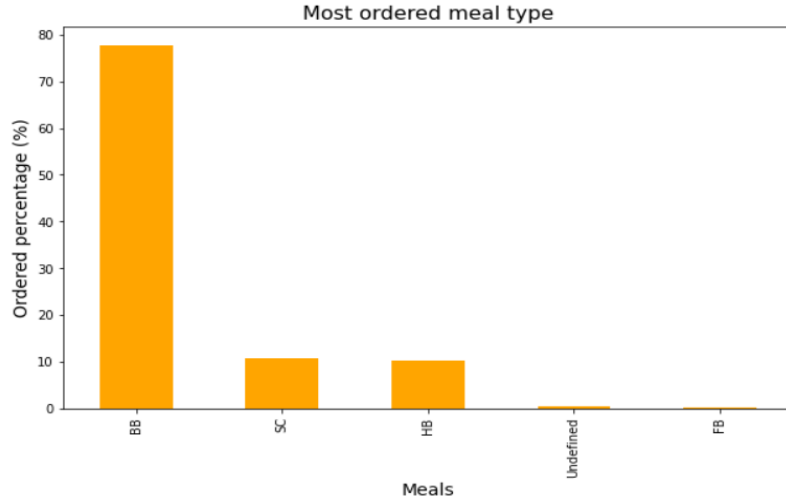
| | |
|-----------|-------|
| August | 13877 |
| July | 12661 |
| May | 11791 |
| October | 11160 |
| April | 11089 |
| June | 10939 |
| September | 10508 |
| March | 9794 |
| February | 8068 |
| November | 6794 |
| December | 6780 |
| January | 5929 |

Name: arrival_date_month, dtype: int64

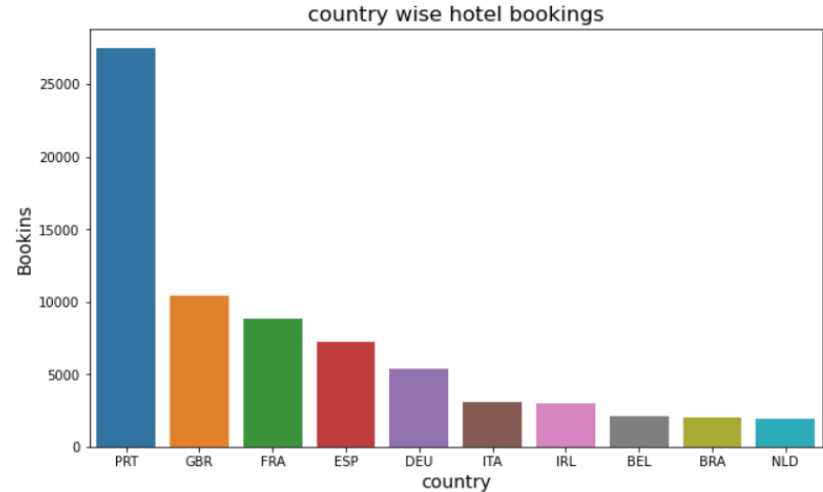


- Month of August is mostly arrival period followed by months July and May.

Meal & Country wise Analysis



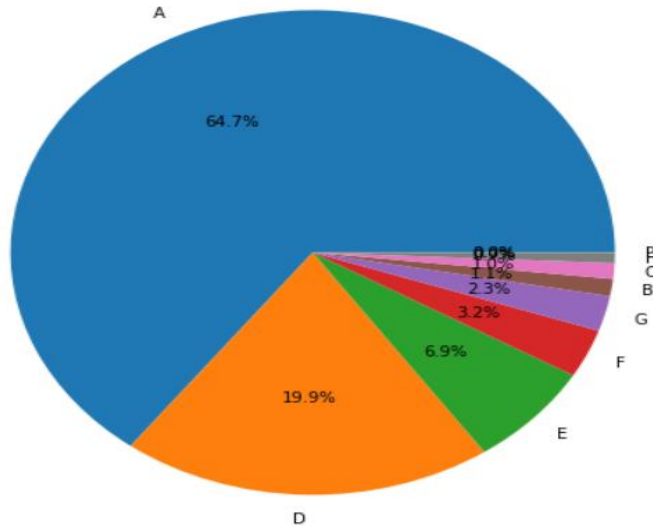
- Most preferred meal type is BB (Bed and breakfast) by customers.
- Almost 75-80% Guest like BB type Meal followed by SC & HB type



- Most of the customers from European countries like Portugal, Great Britain, France and Spain.
- Portugal is top most country for booking hotels.

Reserved room type Analysis

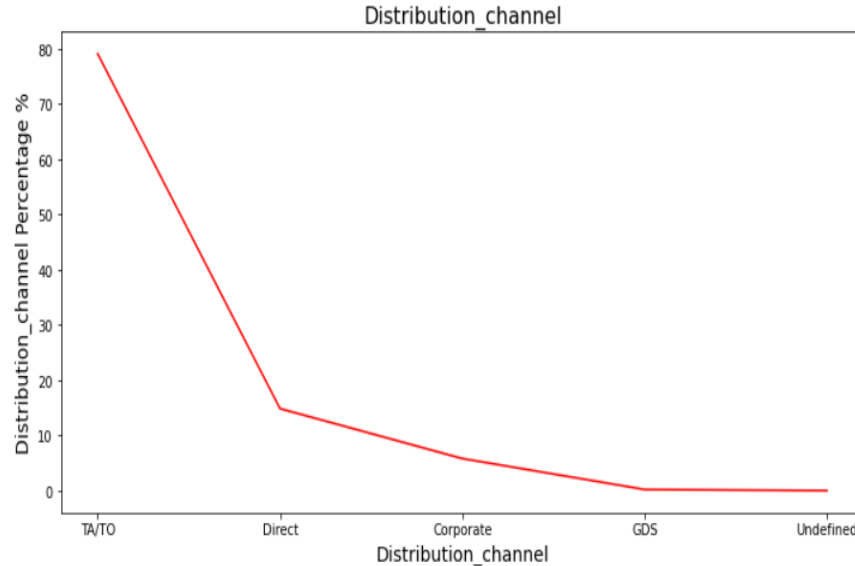
- Which room type mostly preferred by guests while booking.



- Type A room is most demanded by customers.
- Almost 65% Booking for Room type A & 20% bookings for room type D

Distribution Channel wise Analysis

- Question 6 = Mostly used distribution channel for hotel booking?



```
TA/TO      69141
Direct     12988
Corporate   5081
GDS        181
Undefined   5
Name: distribution_channel, dtype: int64
```

- Most of guests are making reservations through TA/TO channel which is travel agency and tour operator.
- Direct Booking also done by almost 15% Guests.

Booking cancellation Analysis

- Question 7 = What is percentage of booking cancellation by hotel types.

```
[40] ob7 = booking_data.groupby('hotel')['is_canceled'].describe()
ob7
```

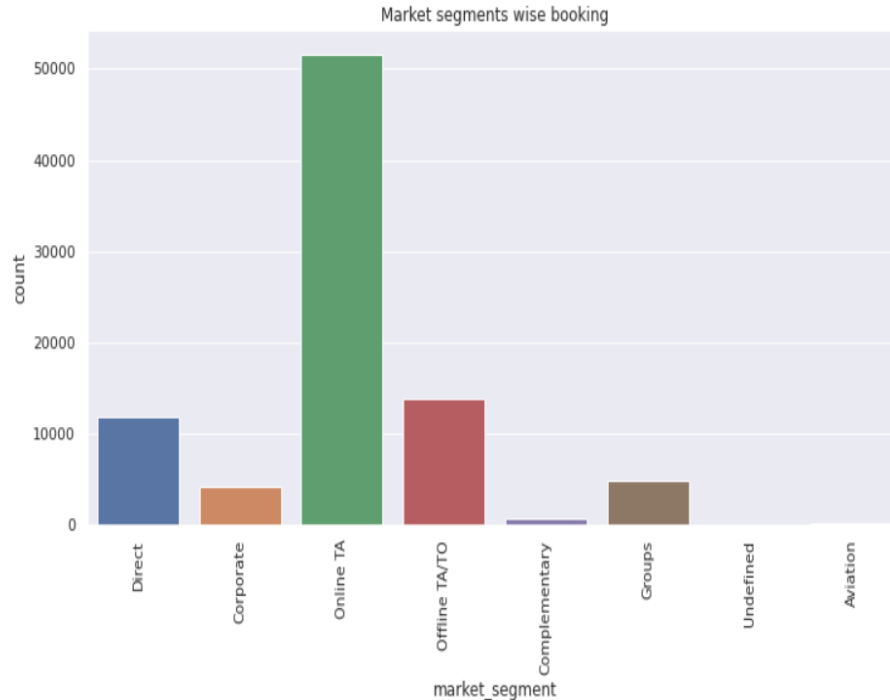
| | count | mean | std | min | 25% | 50% | 75% | max |
|--------------|---------|----------|----------|-----|-----|-----|-----|-----|
| hotel | | | | | | | | |
| City Hotel | 79330.0 | 0.417270 | 0.493111 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |
| Resort Hotel | 40060.0 | 0.277634 | 0.447837 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |



- Around 40% bookings were cancelled in city hotels and 30% in resort hotels.
- Rate of booking cancellation is lower in resort hotel compared to city hotels.

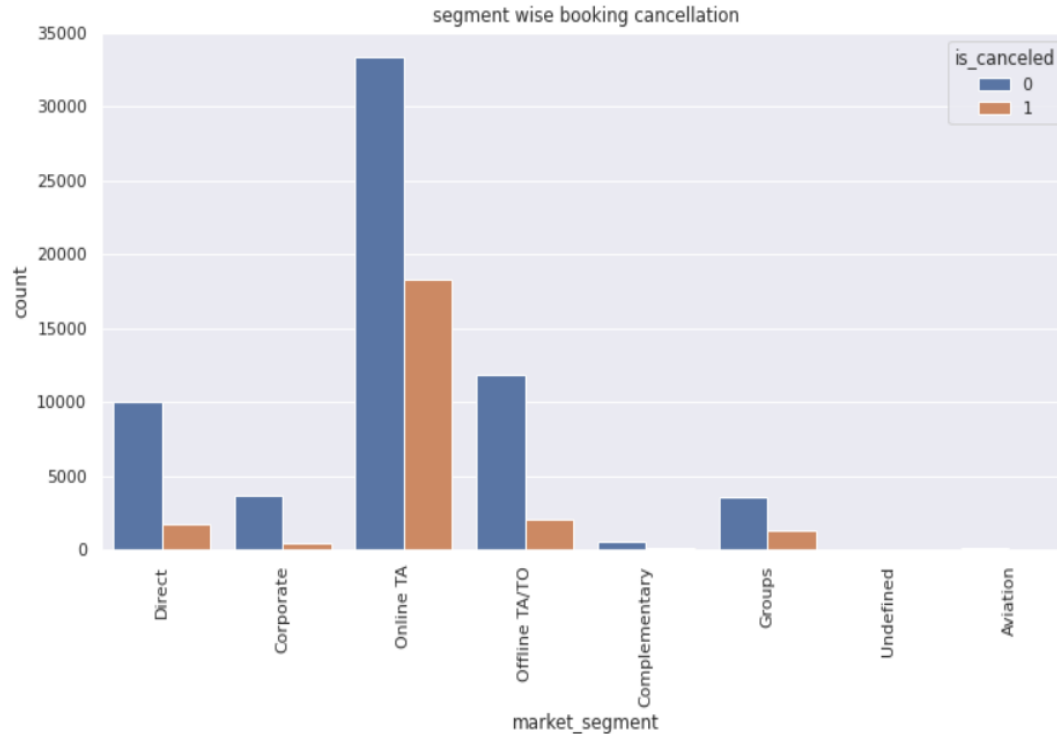
Market Segment wise Analysis

- Question 8 = What is the relationship between market segments and booking cancellation?



- Bookings through online TA And offline TA/TO are higher followed by direct bookings and groups booking.
- More than 50000 Guests used Online TA for bookings.

- Relationship between market segments and booking cancellation.



- Most of bookings were canceled in Online TA followed by Offline TA/TO , Direct bookings and Groups bookings.
- Cancellations are high when done through Online TA compared to direct bookings. Hotels need to do marketing and give special incentives for direct bookings.

Conclusion

- Around 60% bookings are for city hotel & 40% bookings for resort hotel, therefore city hotel is busier than resort hotel.
- July & August are the most busier months in years for both hotels. April to August is the peak season of bookings.
- BB type meal mostly preferred for ordering food followed by SC type meal & FB is least preferred meal by guests.
- Maximum number of guests came from PRT i.e., more than 25000. After PRT GBR and FRA are countries from where most of the guests came.
- Room type A is selected by almost 64% guests & room type D & E selected by 20% and 7% guests respectively. Room type A generating most revenue for hotels.
- Guests use different distribution channels for making bookings out of which most preferred way is TA/TO.
- Both hotels have significantly higher booking cancellation , Around 40% of bookings were cancelled in city hotels & 30% bookings were cancelled in resort hotels.
- Cancellations are high when done through Online TA compared to direct bookings. Hotels need to do marketing and give special incentives for direct bookings.