

Capstone Project

Hotel Booking Analysis

by
Yukta Shinde

Exploratory Data Analysis



Hotel Booking sector is a customer oriented sector and highly competitive market so to be in a profitable & sustainable in market various factors help us to observe Market trend and analyze the data to discover important factors that govern the Hotel bookings

Libraries used

- 1.Numpy
- 2.Pandas
- 3.Matplotlib
- 4.Seaborn



matplotlib



seaborn

Data Preparation

- **Import Libraries**
- **Load Data & save copy**
- **Data Inspection**
- **Data Cleaning :**
 - i) **Handling Null values**
 - ii) **Removing Duplicates**
 - iii) **Outliers Detection**
- **Exploratory Data Analysis**
- **Conclusion**

Data Inspection:

- **Import dataset in Google colab from drive**
- **`data= pd.read_csv(path)`**
- **Saving copy of original dataset**
- **`df=data.cop()`**
- **Data Inspection**
- **`df.shape` : Rows & Columns in data : 119390, 32**
- **`df.size` : Total number of elements in dataset**
- **`df.head()` : Top 5 rows**
- **`df.tail()` : Bottom 5 rows**
- **`df.columns` : features in dataset**

	count	mean	std	min	25%	50%
is_canceled	119390.0	0.370	0.483	0.00	0.00	0.000
lead_time	119390.0	104.011	106.863	0.00	18.00	69.000
arrival_date_year	119390.0	2016.157	0.707	2015.00	2016.00	2016.000
arrival_date_week_number	119390.0	27.165	13.605	1.00	16.00	28.000
arrival_date_day_of_month	119390.0	15.798	8.781	1.00	8.00	16.000
stays_in_weekend_nights	119390.0	0.928	0.999	0.00	0.00	1.000
stays_in_week_nights	119390.0	2.500	1.908	0.00	1.00	2.000
adults	119390.0	1.856	0.579	0.00	2.00	2.000
children	119386.0	0.104	0.399	0.00	0.00	0.000
babies	119390.0	0.008	0.097	0.00	0.00	0.000
is_repeated_guest	119390.0	0.032	0.176	0.00	0.00	0.000
previous_cancellations	119390.0	0.087	0.844	0.00	0.00	0.000
previous_bookings_not_canceled	119390.0	0.137	1.497	0.00	0.00	0.000
booking_changes	119390.0	0.221	0.652	0.00	0.00	0.000

**Statistical Information
Of Numerical
features**

```
# statistical information about Object variables in data.  
df.describe(include='O').T
```

	count	unique	top	freq
hotel	119390	2	City Hotel	79330
arrival_date_month	119390	12	August	13877
meal	119390	5	BB	92310
country	118902	177	PRT	48590
market_segment	119390	8	Online TA	56477
distribution_channel	119390	5	TATO	97870
reserved_room_type	119390	10	A	85994
assigned_room_type	119390	12	A	74053
deposit_type	119390	3	No Deposit	104641
customer_type	119390	4	Transient	89613
reservation_status	119390	3	Check-Out	75166
reservation_status_date	119390	926	2015-10-21	1461

Statistical Information Of Categorical features

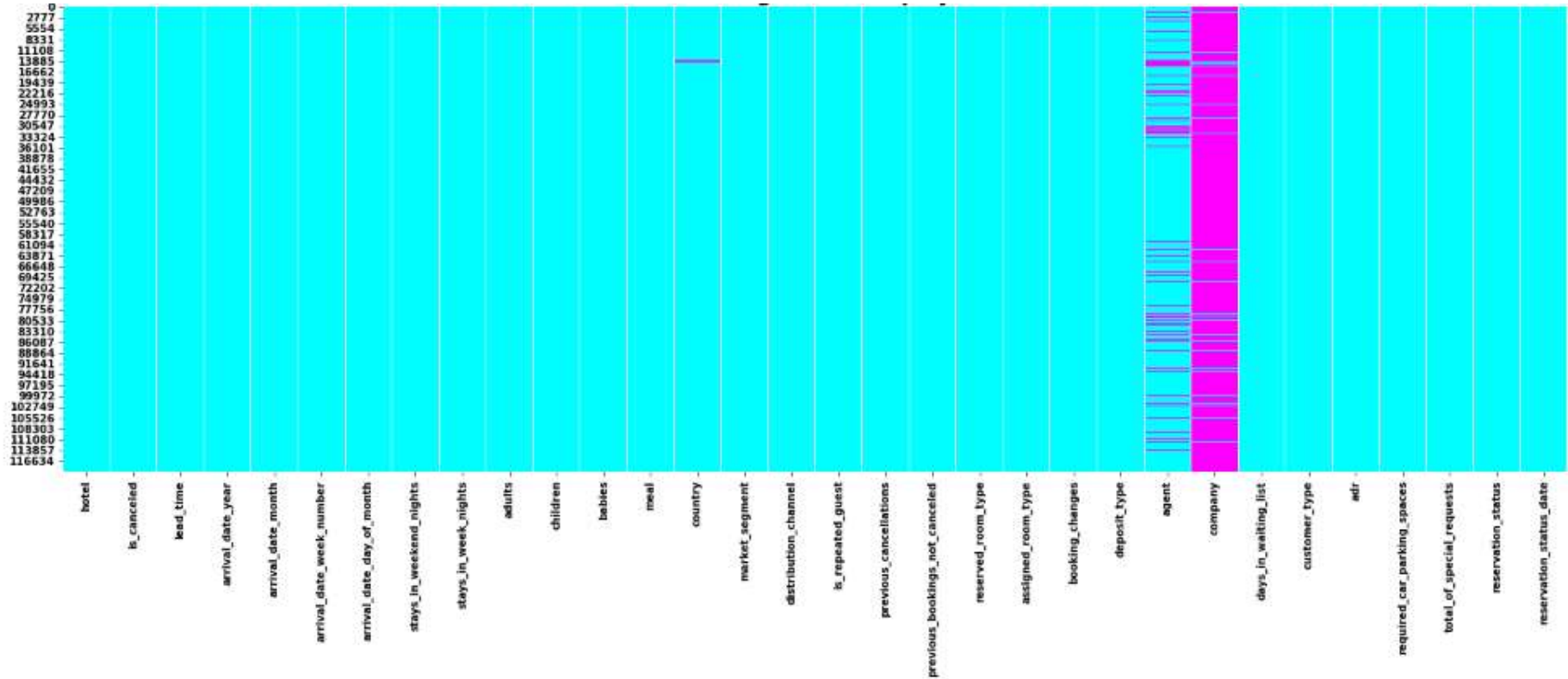
Steps involved in Data cleaning:

- 1) Handling Null values:
 - i) check if special characters in [+,@,#,%] in string datatype
 - ii) Percentage of Null values in columns & heatmap for visual

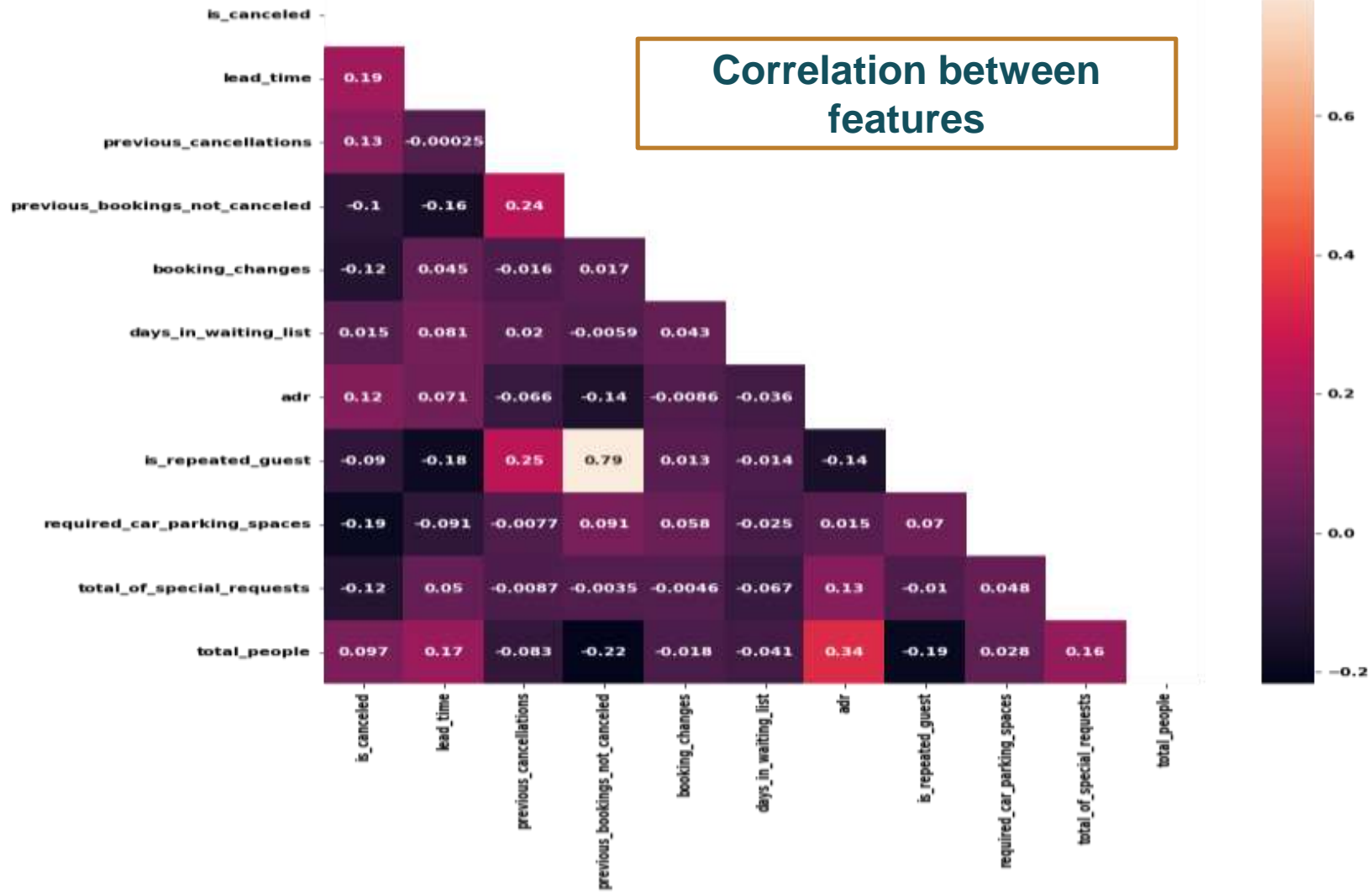
	null_val	Percentage
company	112593	94.306893
agent	16340	13.686238
country	488	0.408744
children	4	0.003350
reserved_room_type	0	0.000000

DATA TYPE	Replace by	Columns
Numerical	0	Children agent
categorical	'others'	country

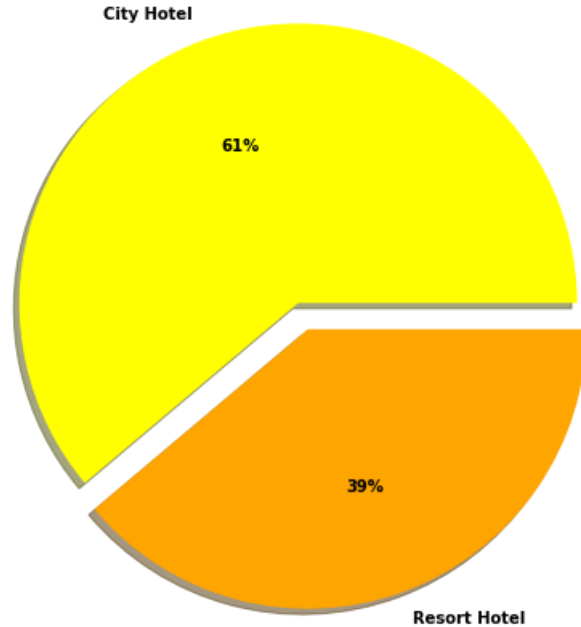
Null values-Heatmap



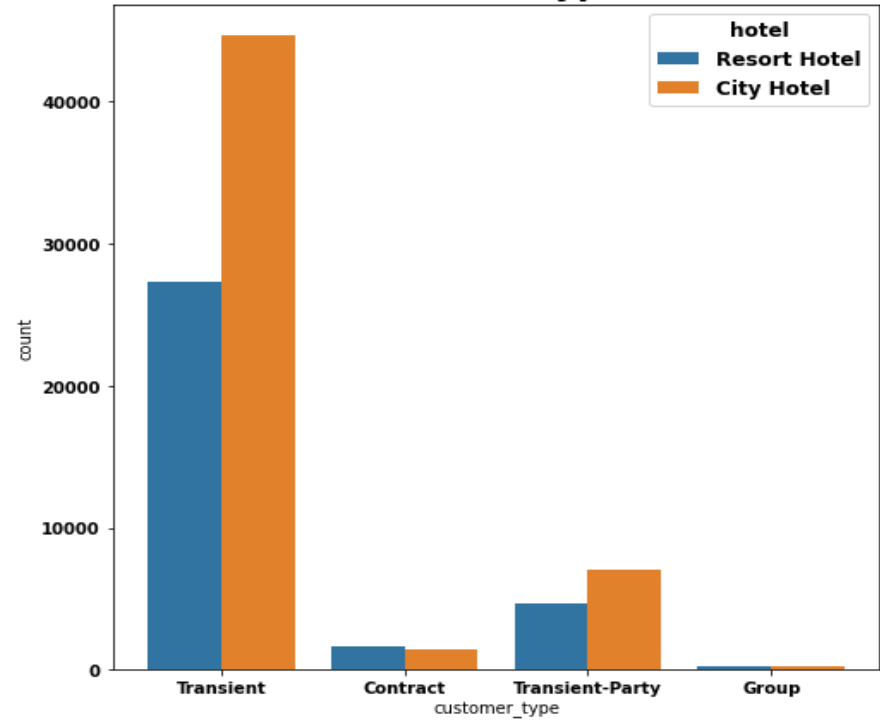
Correlation between features



Hotel preference distribution

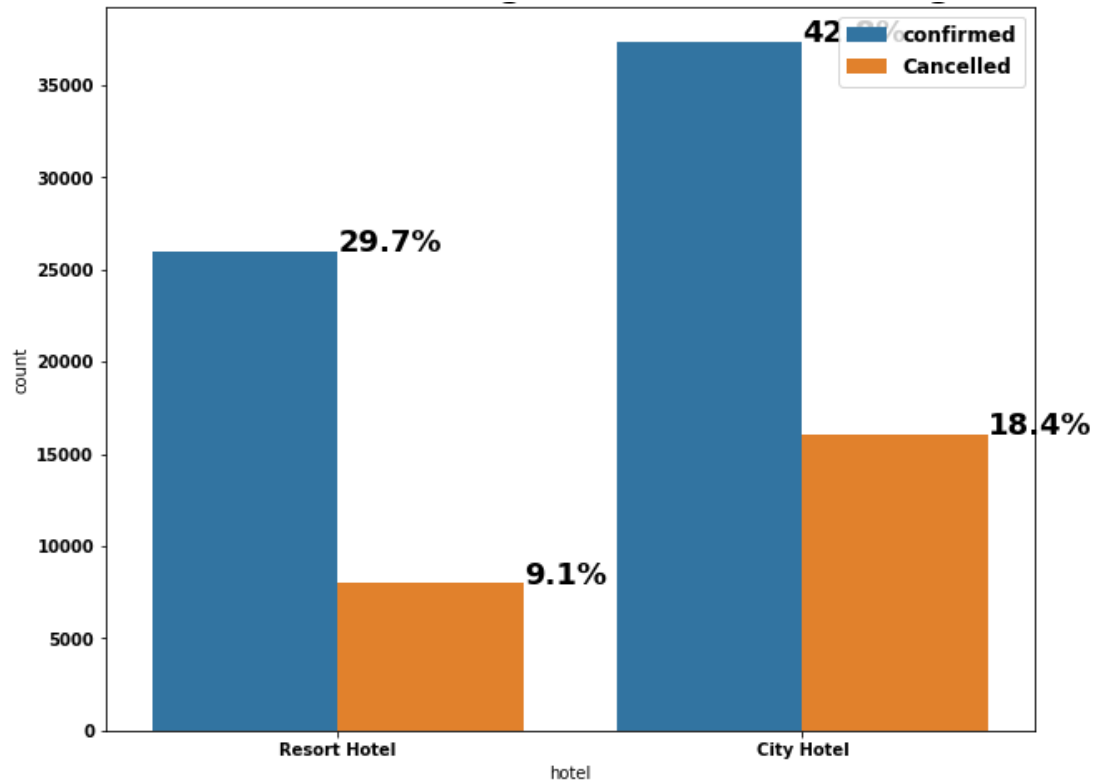


Customer types

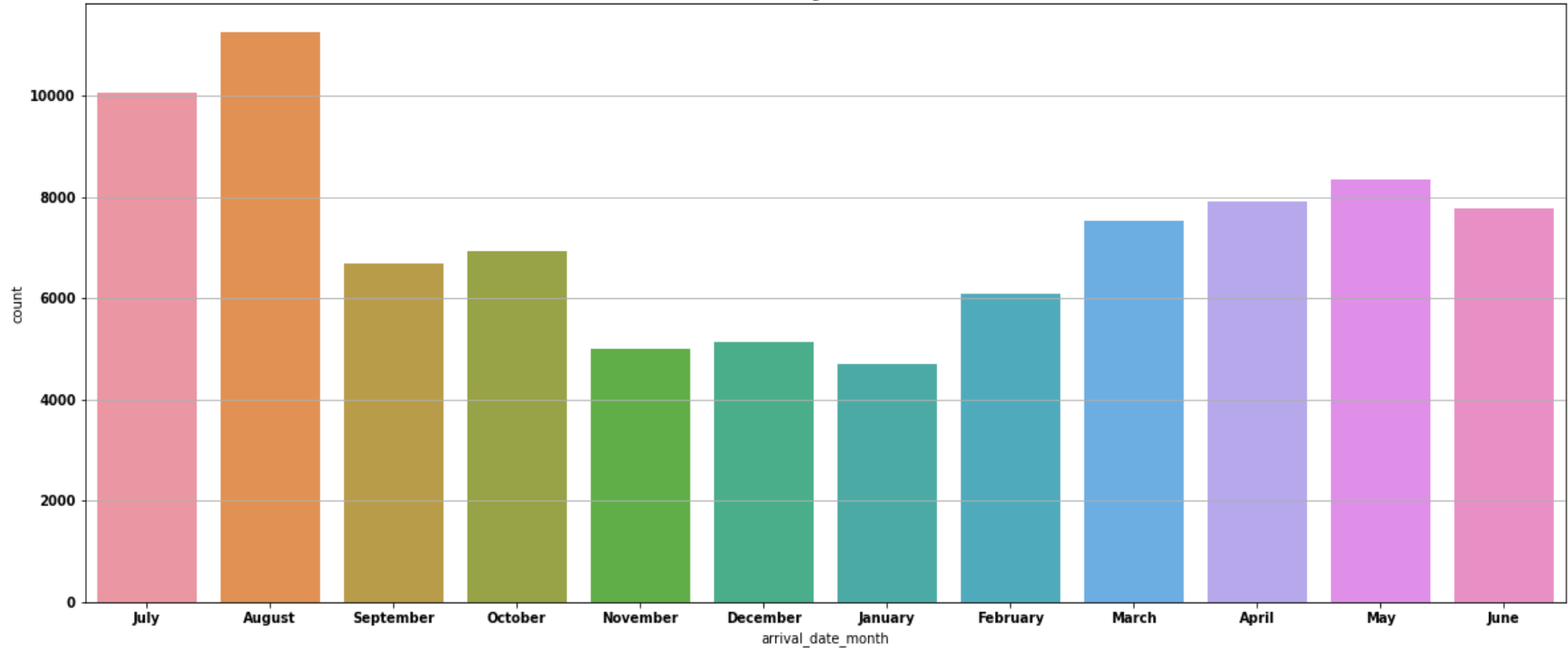


City hotel is more preferred over resort hotel
Most of the customers from Transient type

Cancelled & confirmed booking

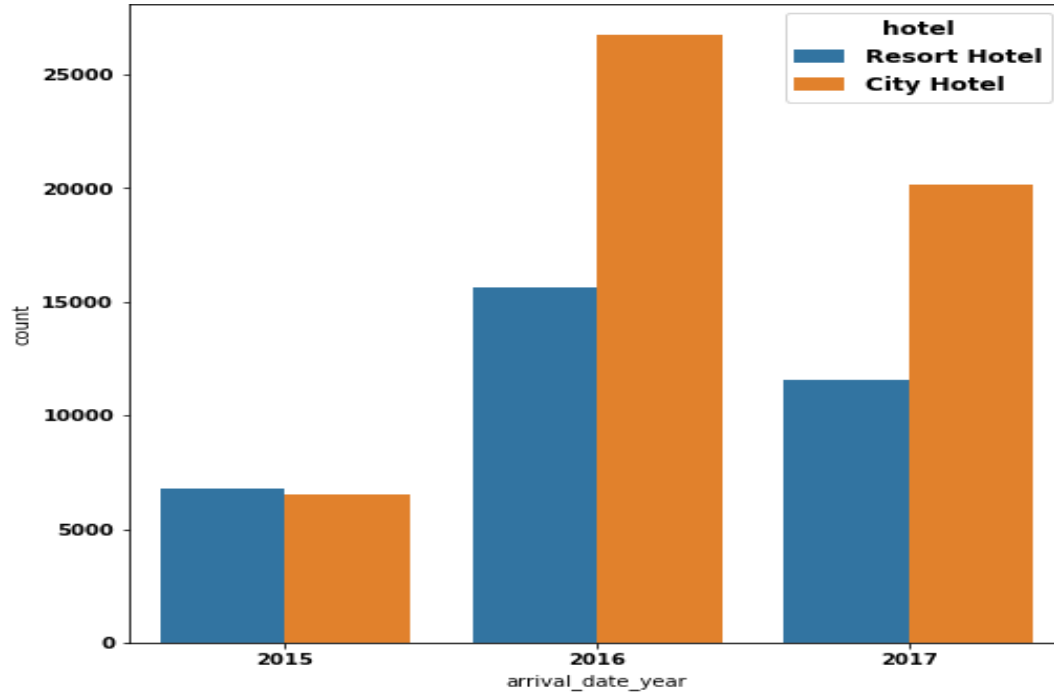


Monthly arrival



- Peak Season of Booking is August , July where maximum booking occurs.
- Minimum Booking seen in January , November.

Yearly Bookings



City Hotel in Demand for every year
Yearly Bookings doesn't increase proportionately

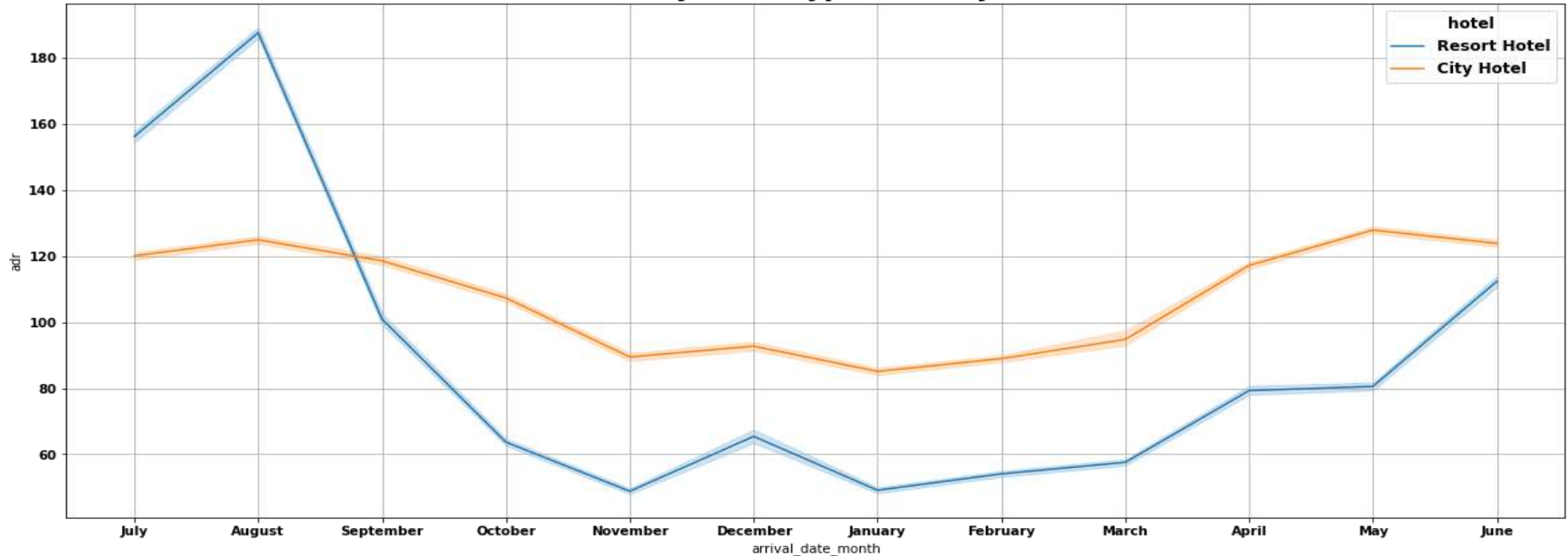
adr vs Month relation for every year



➤ ADR Highest in August, July (peak months) when demand increases. (high rush --> high booking --> high adr)

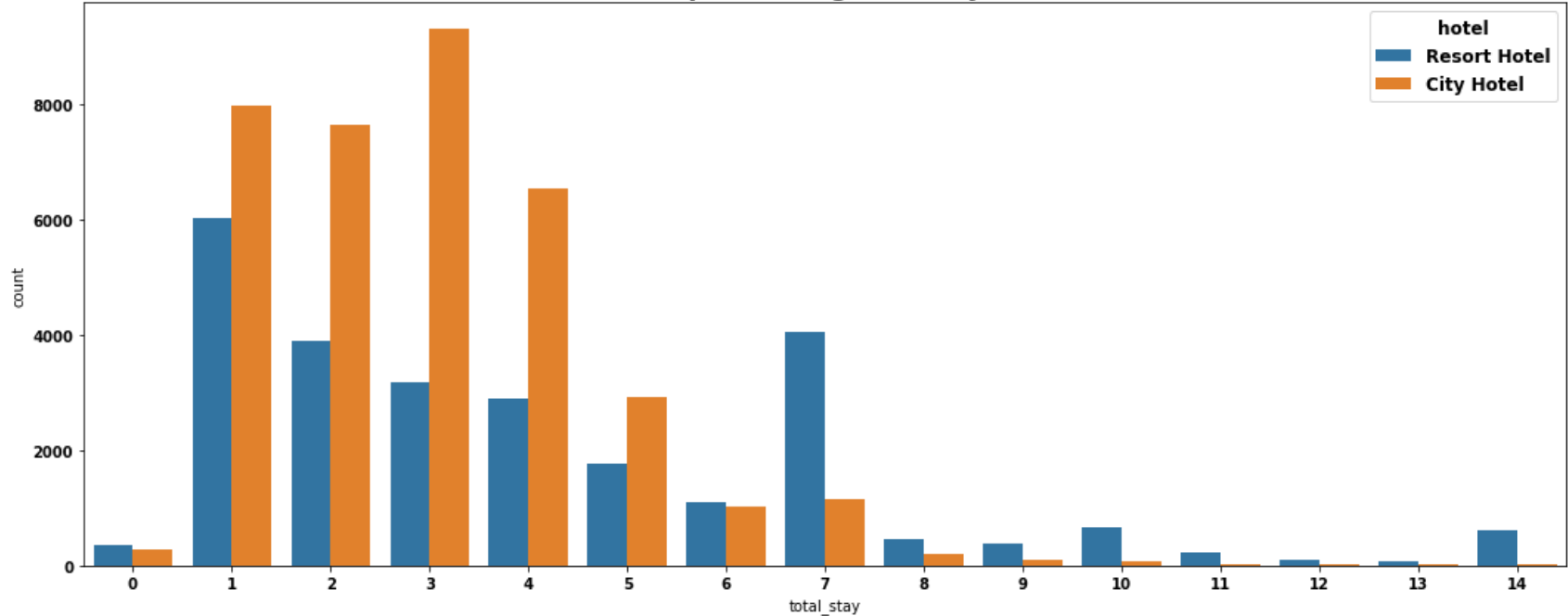
➤ adr is lowest in month of January, November hence this can be the best time for booking to get at best price.

ADR variation by hotel type Monthly distribution



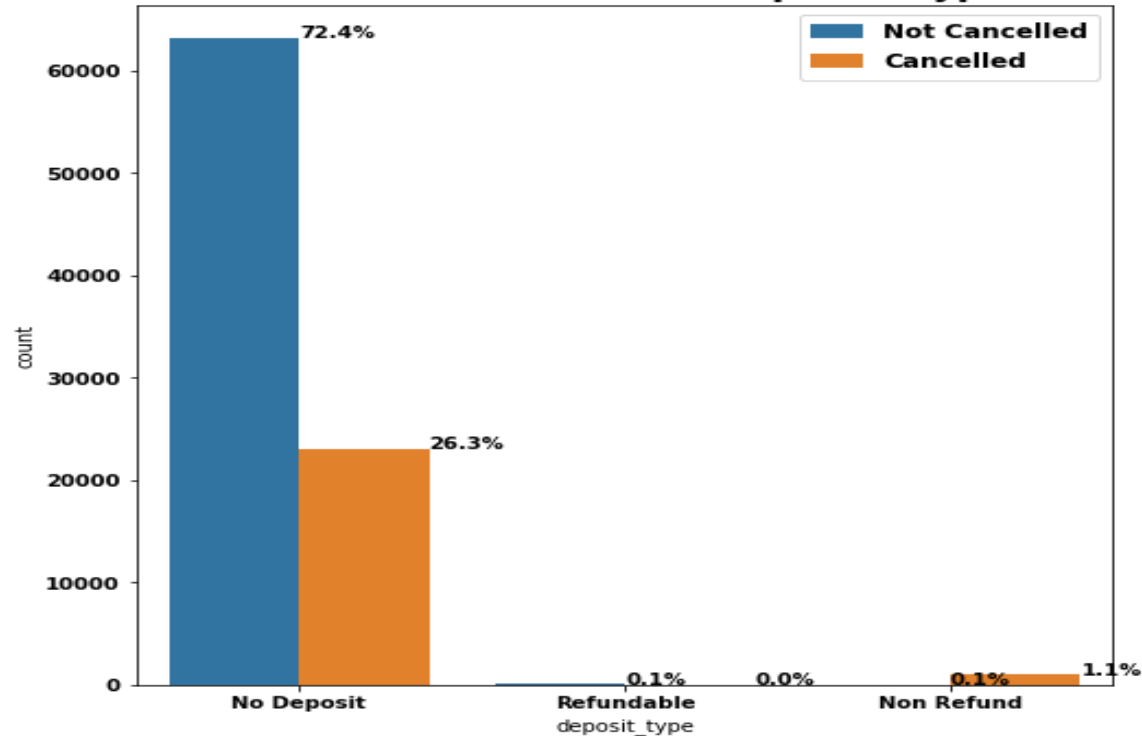
- It is seen that average daily rate for Resort hotel is fluctuating where as for city hotel it is quite stable.
- ADR resort hotel is higher then city hotel in month of July, august whereas it is lower in other months
- to book **resort in summer months**(July, aug, sept) at **best price** , booking can be done in advance when adr is less.

Optimal length of stay



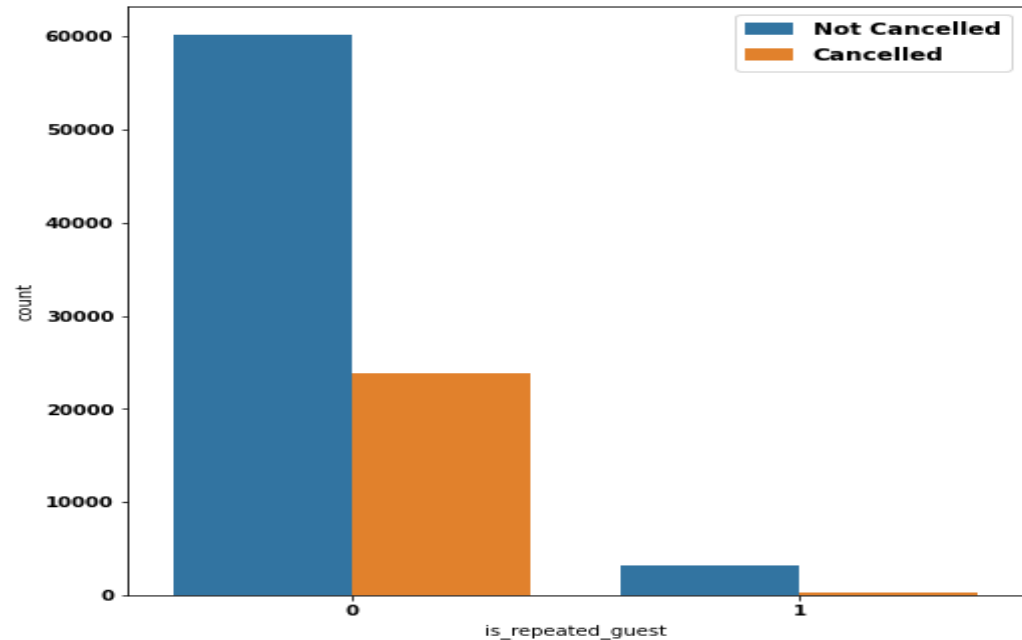
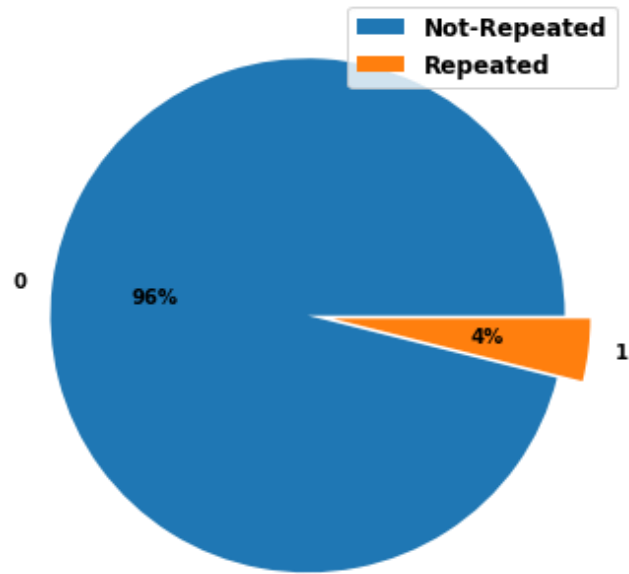
- usual stay length is about 4 days.
- Mostly for couple of days (less than week) city hotel is preferred over Resort Hotel
- for longer stay (more than week) Resort hotel is preferred.

Cancellation rate with Deposit type



No Deposit is highly preferred payment mode by customer but it comes with disadvantage for hotels as people with no deposit has highest chances of booking cancellation

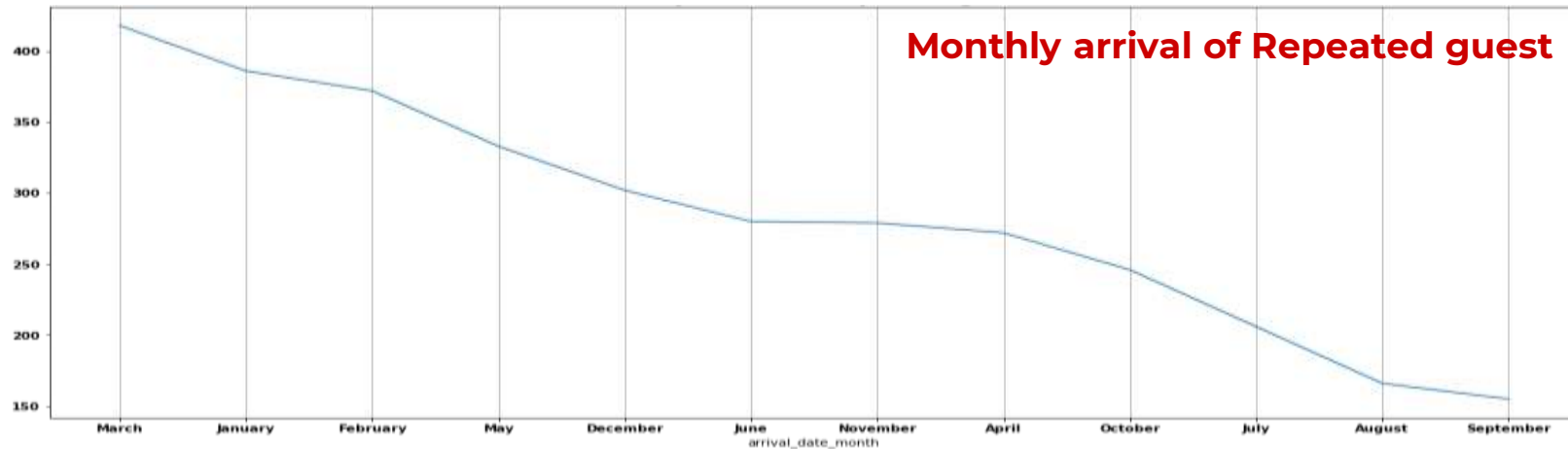
Cancellation rate by repeated & Non-Repeated guest



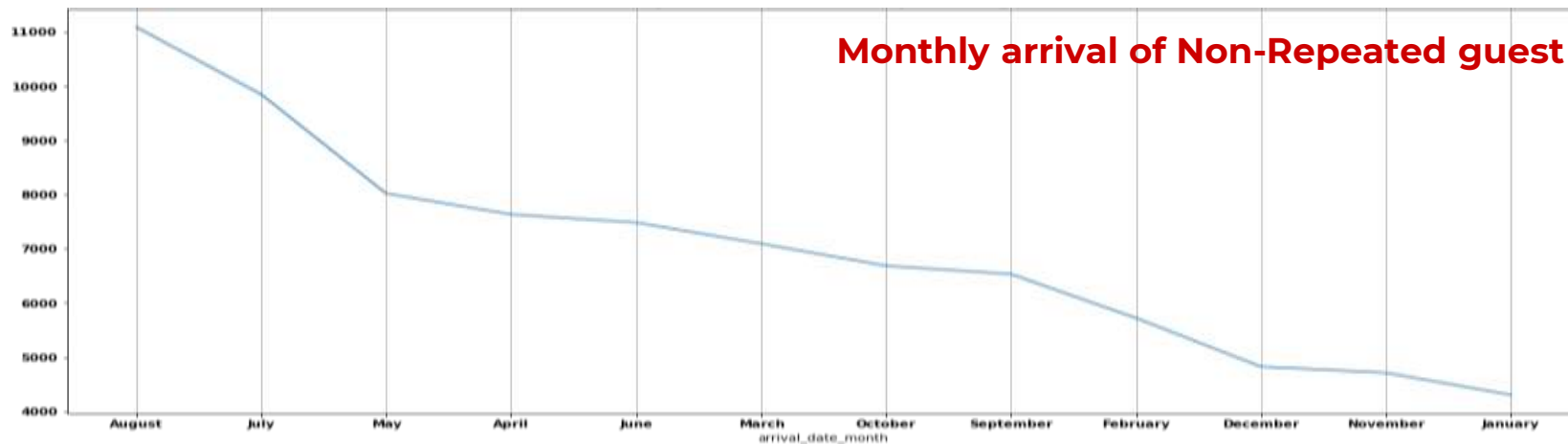
0 -Non repeated guest
1- Repeated Guest

Repeated guests are not likely to cancel their booking.

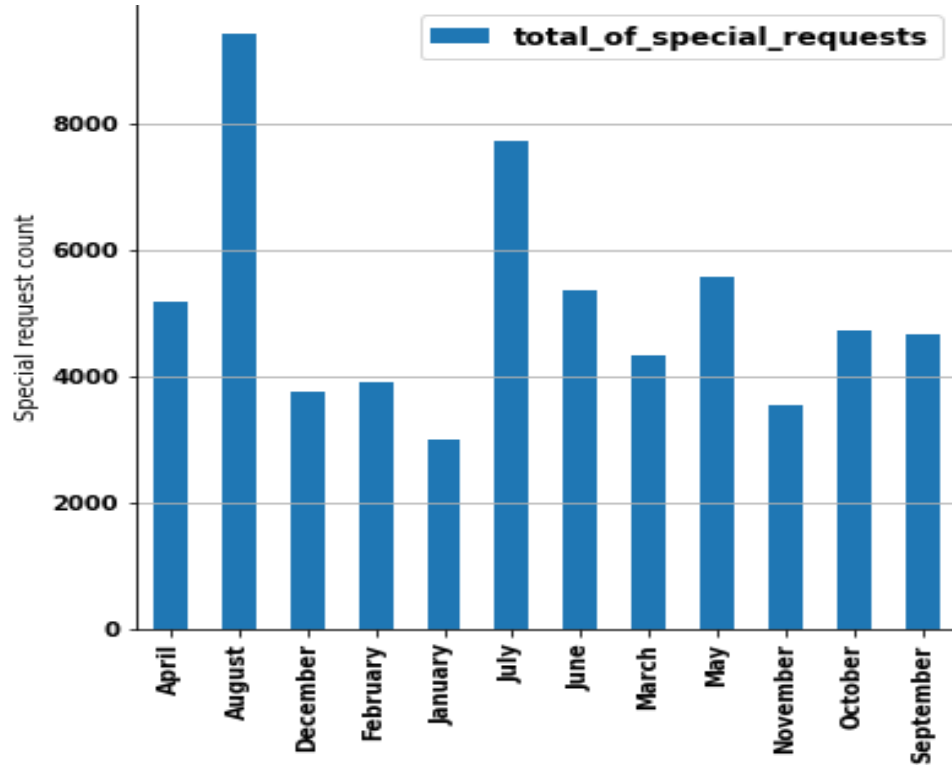
Monthly arrival of Repeated guest



Monthly arrival of Non-Repeated guest



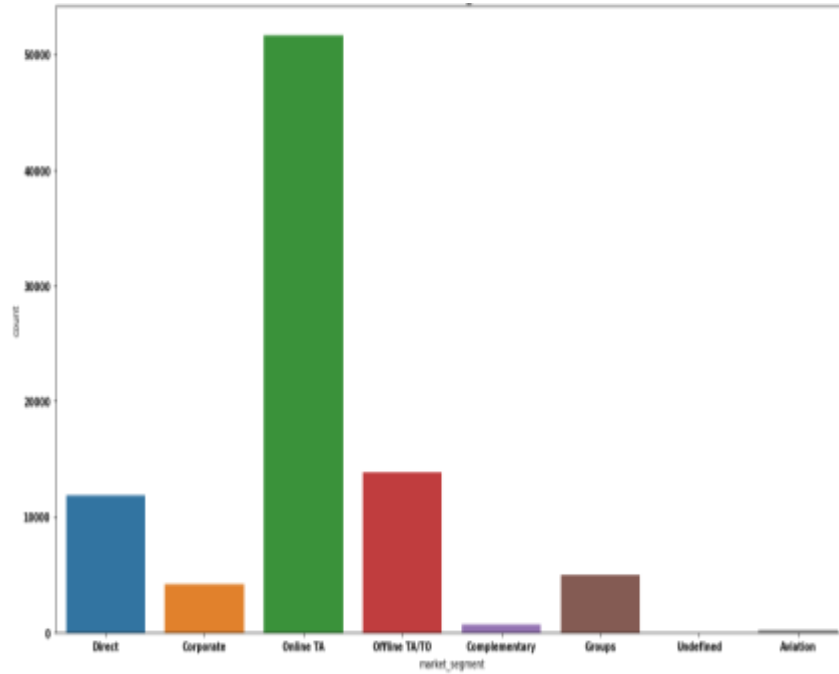
Special requests according to Month



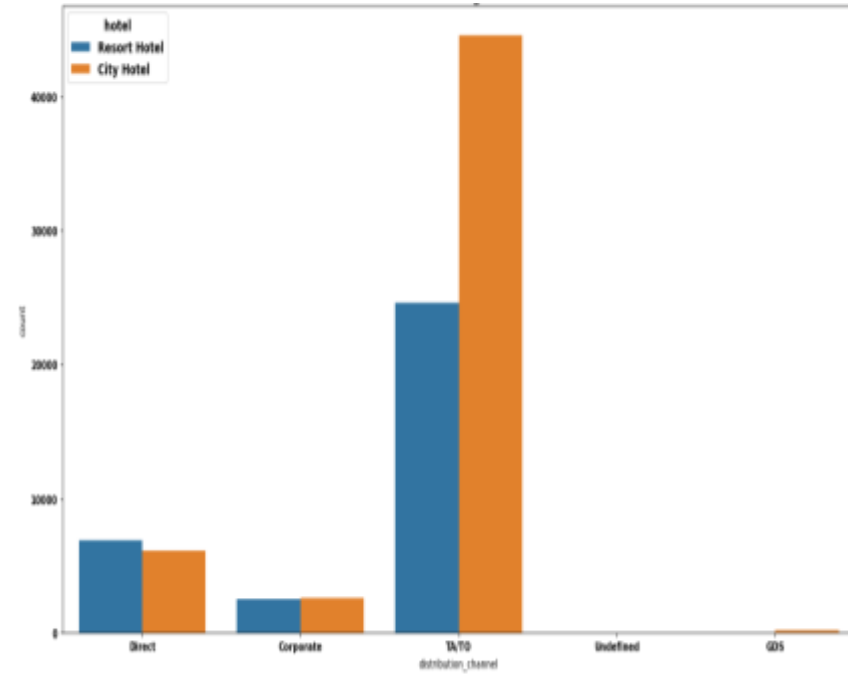
	arrival_date_month	total_of_special_requests
0	April	5173
1	August	9409
2	December	3756
3	February	3894
4	January	2985
5	July	7720
6	June	5345
7	March	4312
8	May	5557
9	November	3534
10	October	4706
11	September	4658

special request seen in peak Months that is august & July

Market segment

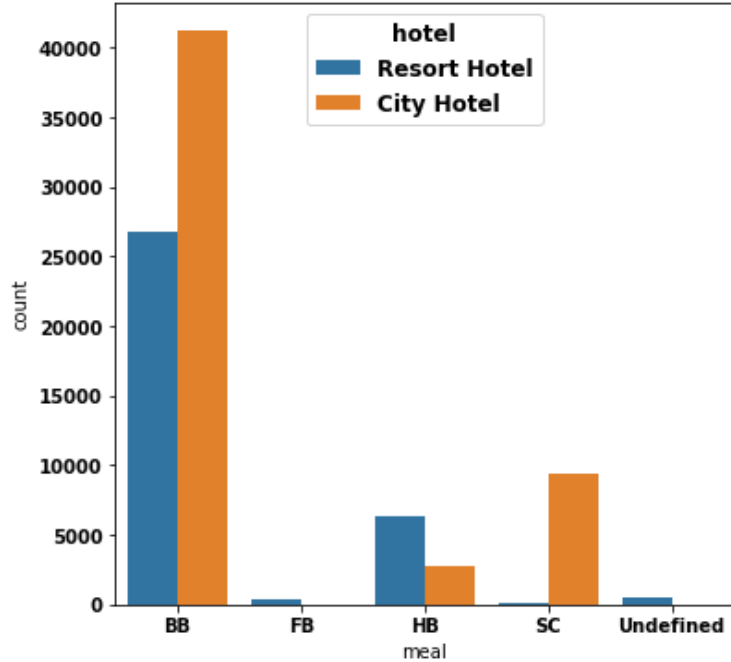


Distribution channel

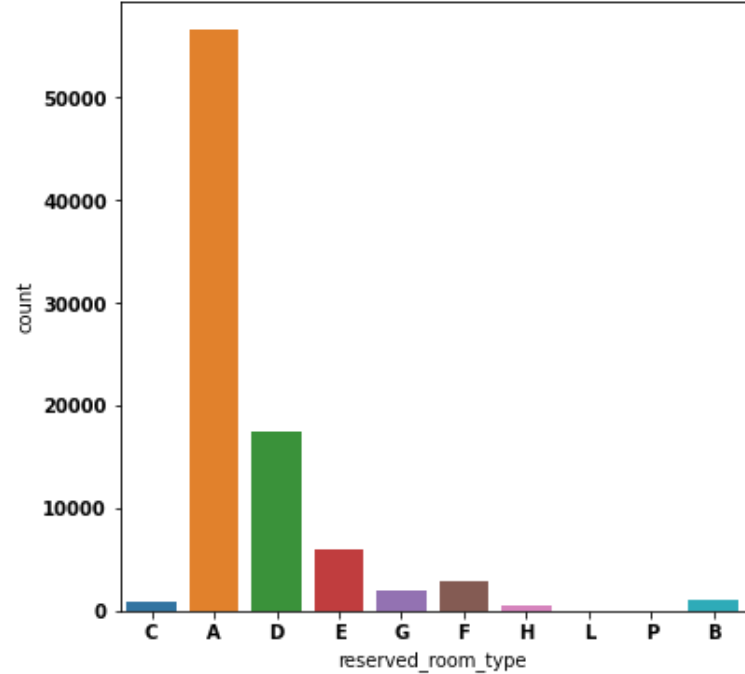


- Highest market captured by Online Tour operators in Market segment & by both Travel agents & Tour operators for distribution channel.
- in order increase revenue of company one should focus on TA/TO in marketing segment.

Popular Meal type in Resort & City hotel



Room Types



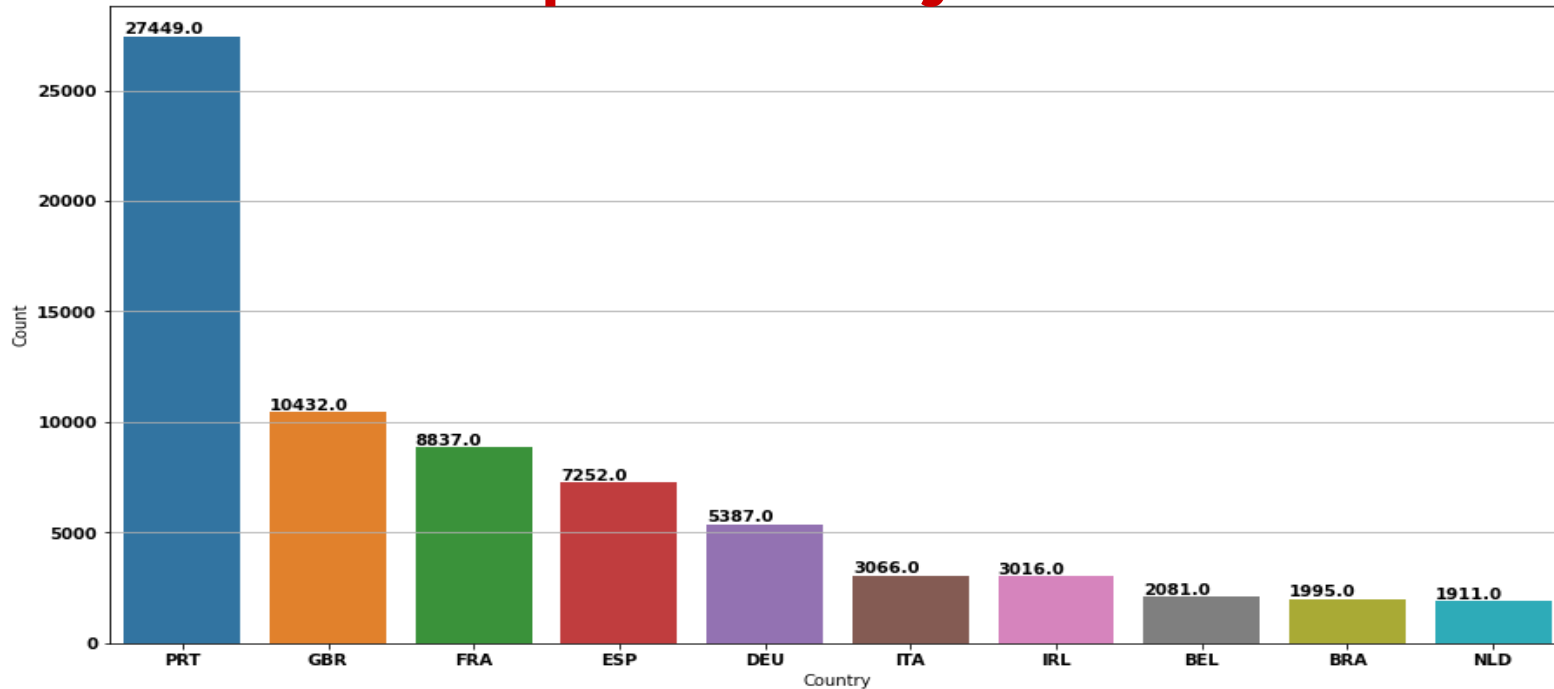
BB – Bed & Breakfast

HB – Half board (breakfast and one other meal – usually dinner)

FB – Full board (breakfast, lunch and dinner)

SC self-catering (no meals are included)

Top 10 Country visitors



Most of the visitors from western European countries in like Portugal has highest visitors 27449 followed by United Kingdom, France, Spain. more capital can be invested in this areas.

Conclusion

- City hotel is highly preferred over resort hotel & Most of the customers are from Transient type (booking is not part of a group or contract)
- **peak season - adr relation**
adr increases when demand increases & drops as demand decreases (high rush --> high booking --> high adr)
- **Best time to book hotel room**
Highly preferred arrival month is August , july but adr is high during this peak season , so one can do advance booking when prices are down
To get **resort hotels** at **least price** less than 60 adr one can book hotel in **November , january**
- **Revenue generation**
customers preferring resort-hotel more after june and crowd increases suddenly ,but **before peak season demand is very poor** in order to manage that Resort Hotels can increase revenue by attracting customers before peak season (june)
- **Market Capture**
Highest market captured by: **Online Tour operators** in Market segment & by **Travel agents & Tour operators** for distribution channel
thus focusing on TA/TO , maximizing tie-up with them can be escalating factor for Revenue generation
- **Special request**
Hotels are likely to receive high special requests in peak months like August , july
- **Optimal length of stay is :** about 3-4 days for which city hotel & resort hotel is preferred for longer stay (week)
- **Most of the visitors** from western European countries like Portugal, United Kingdom , france, spain. more capital can be invested in this areas.
- **Repeated Guest**
4% of Repeated guests , repeated guests are aware about peak months, as their arrival is high when adr drops
repeated customers are less likely to cancel bookings company should focus on increasing repeating visitors.

Thank you