CAPSTONE PROJECT:

Hotel Booking Analysis

SAMEER THETE

Content

- Importing and loading data of Hotel booking analysis
- Data cleaning
- Data preparation
- Data visualization
- Challenges
- Conclusion
- Suggestions

Data summary

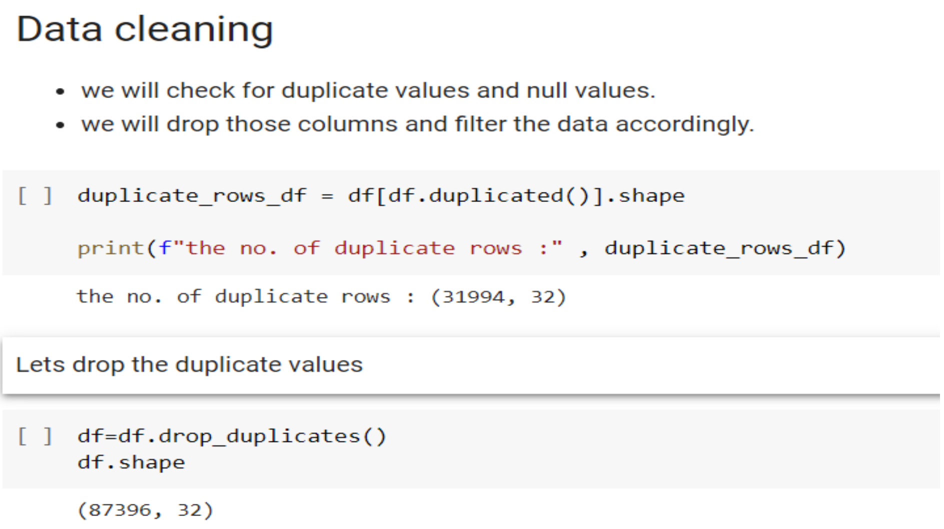
Data set – Hotel booking analysis database includes information about the hotels booked between the year 2015 to year 2017

Shape:

Rows – 119390

Columns – 32

Important columns- lead time, arrival date, no. of persons, repeated guest, no. of kids



Data cleaning

- we will check for duplicate values and null values.
- we will drop those columns and filter the data accordingly.

```
[ ]  duplicate_rows_df = df[df.duplicated()].shape

     print(f"the no. of duplicate rows :" , duplicate_rows_df)
```

```
     the no. of duplicate rows : (31994, 32)
```

Lets drop the duplicate values

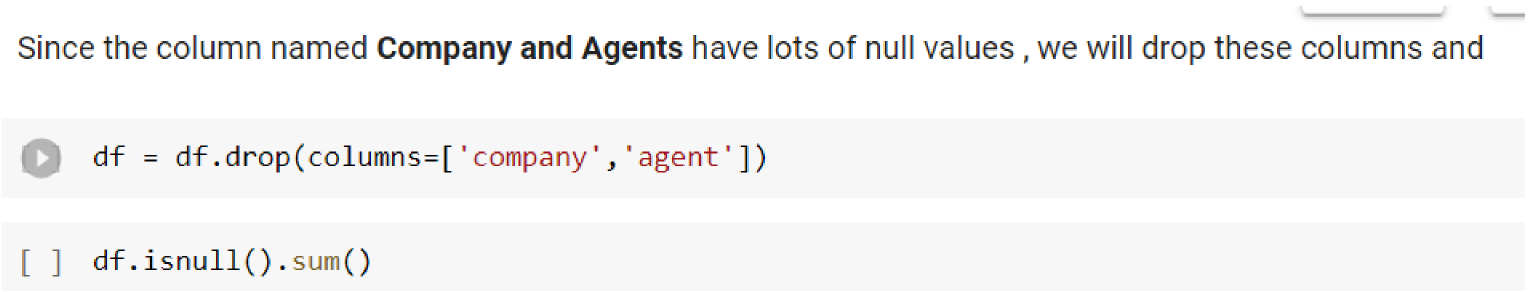
```
[ ]  df=df.drop_duplicates()
     df.shape
```

```
     (87396, 32)
```

Data Cleaning:

Checking for Null Values and duplicate values.

Dropping the necessary columns accordingly.

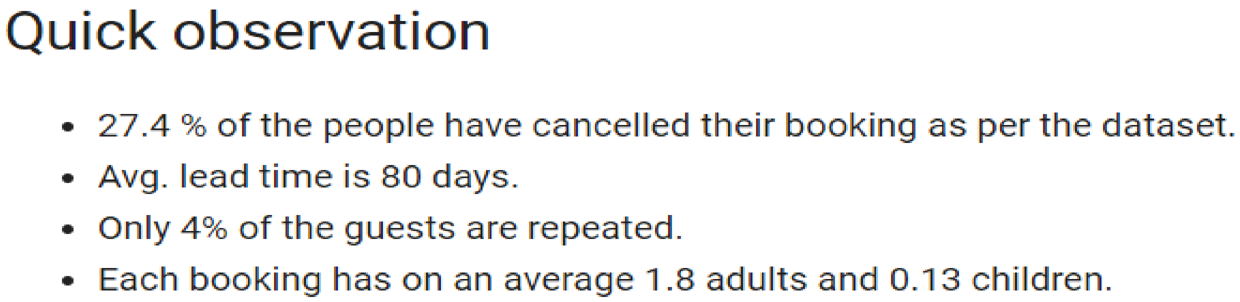


Since the column named **Company and Agents** have lots of null values , we will drop these columns and

```
df = df.drop(columns=['company','agent'])
```

```
[ ]  df.isnull().sum()
```

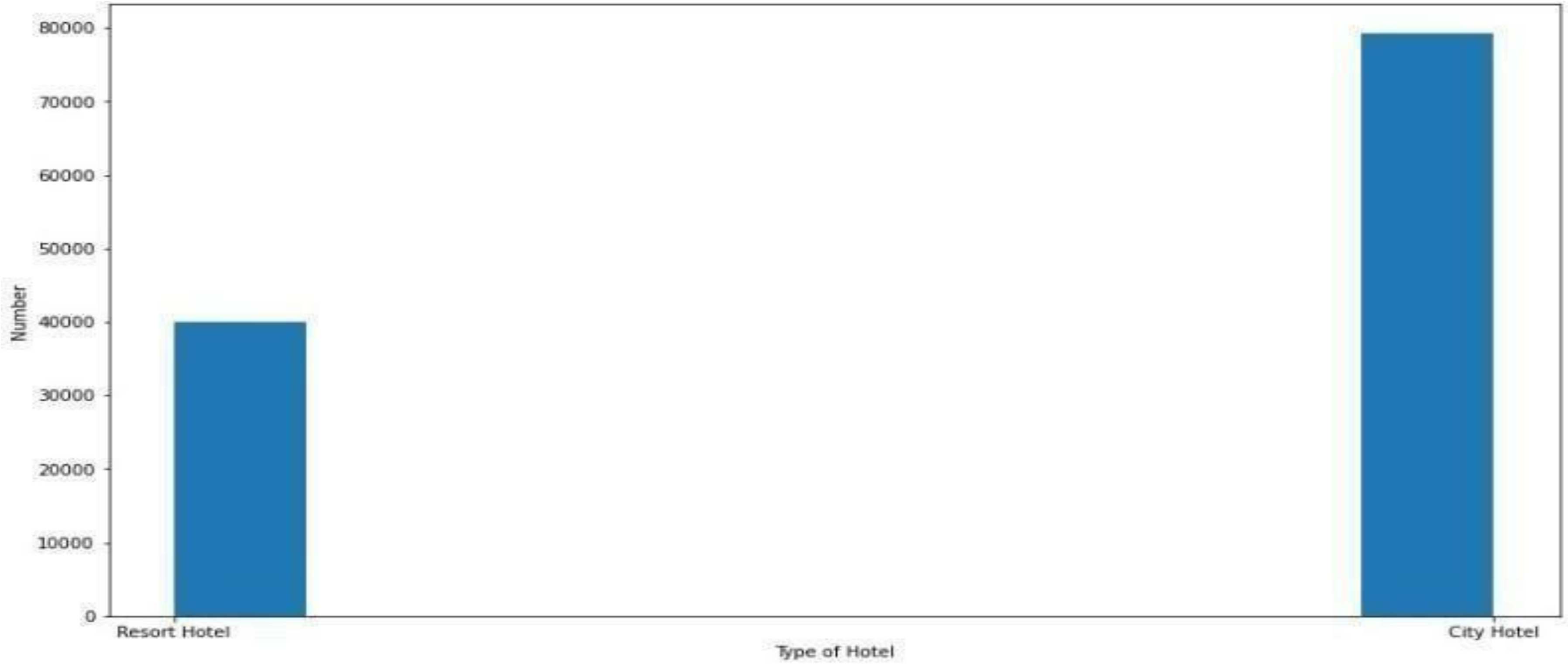
After Dealing with nulls ,these are some of the quick observations:



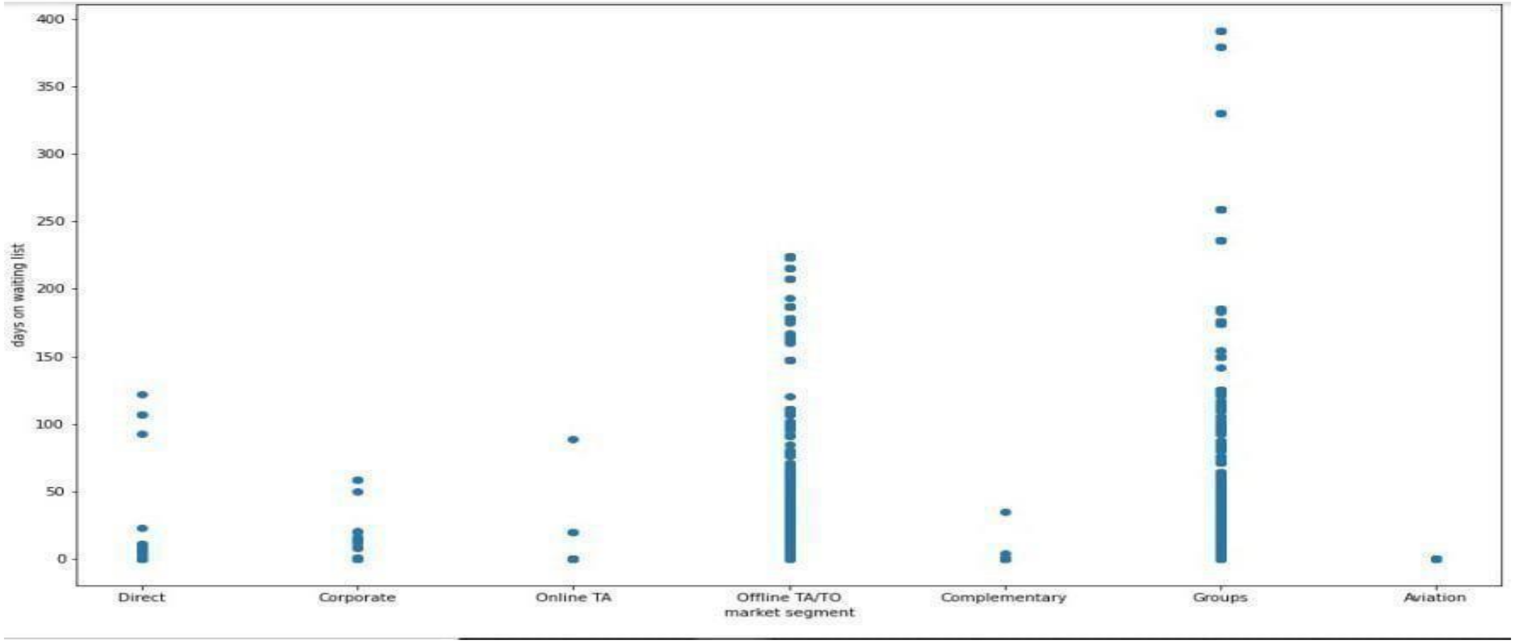
Quick observation

- 27.4 % of the people have cancelled their booking as per the dataset.
- Avg. lead time is 80 days.
- Only 4% of the guests are repeated.
- Each booking has on an average 1.8 adults and 0.13 children.

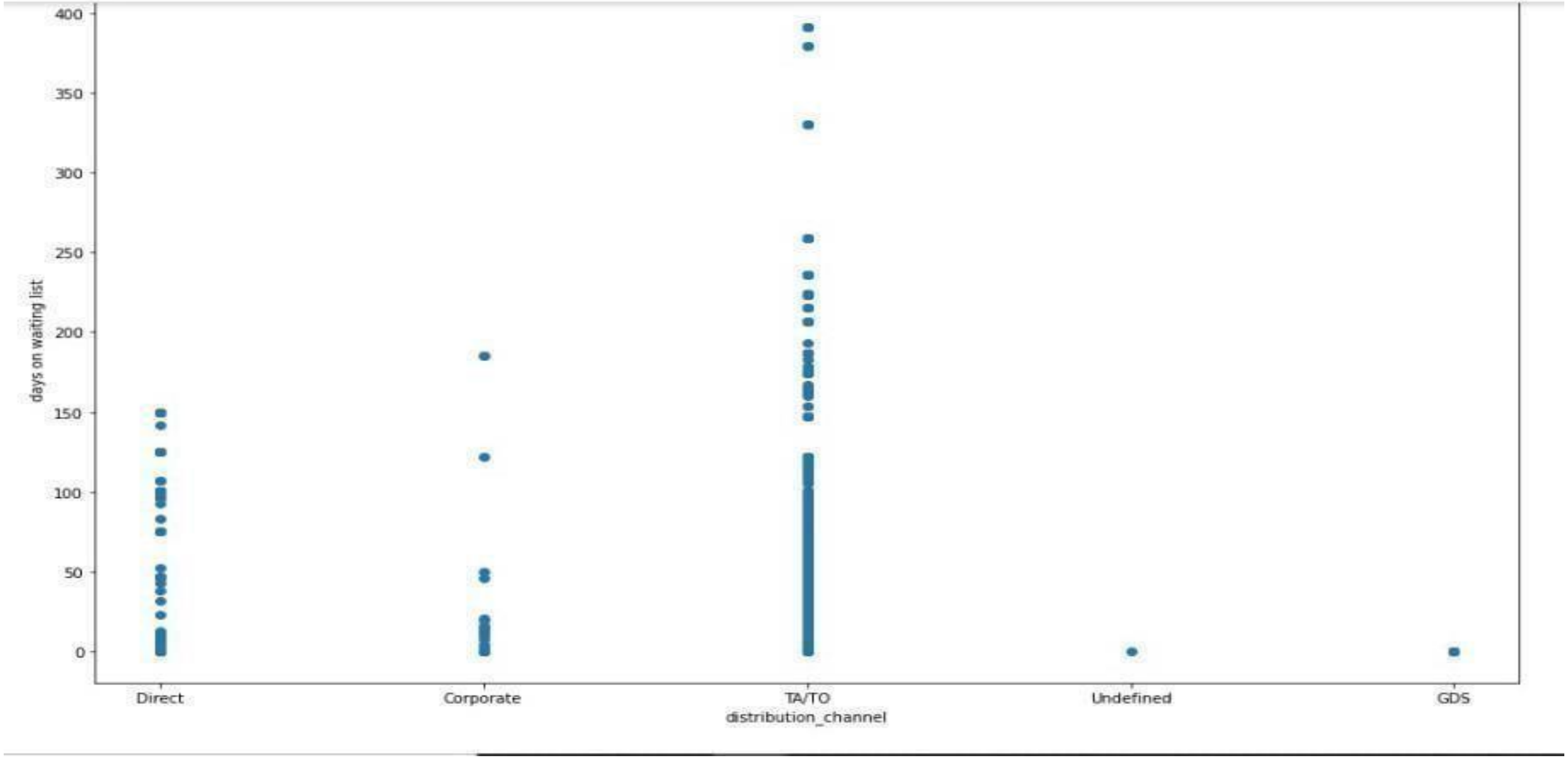
Number of Bookings for various types of hotels (Histogram)



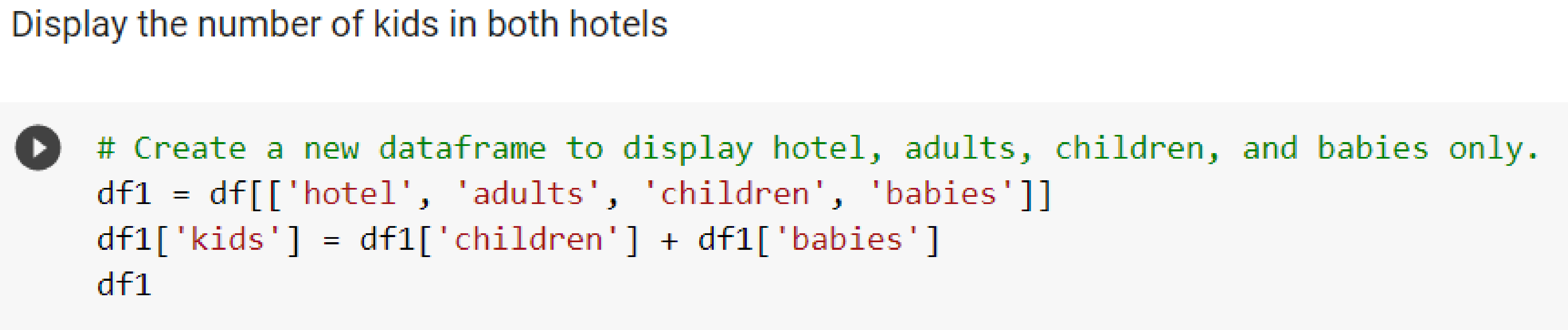
Scatter plot between Type of market segment and waiting list for the booking



Scatter Plot between Distributing Channel and Days on the waiting



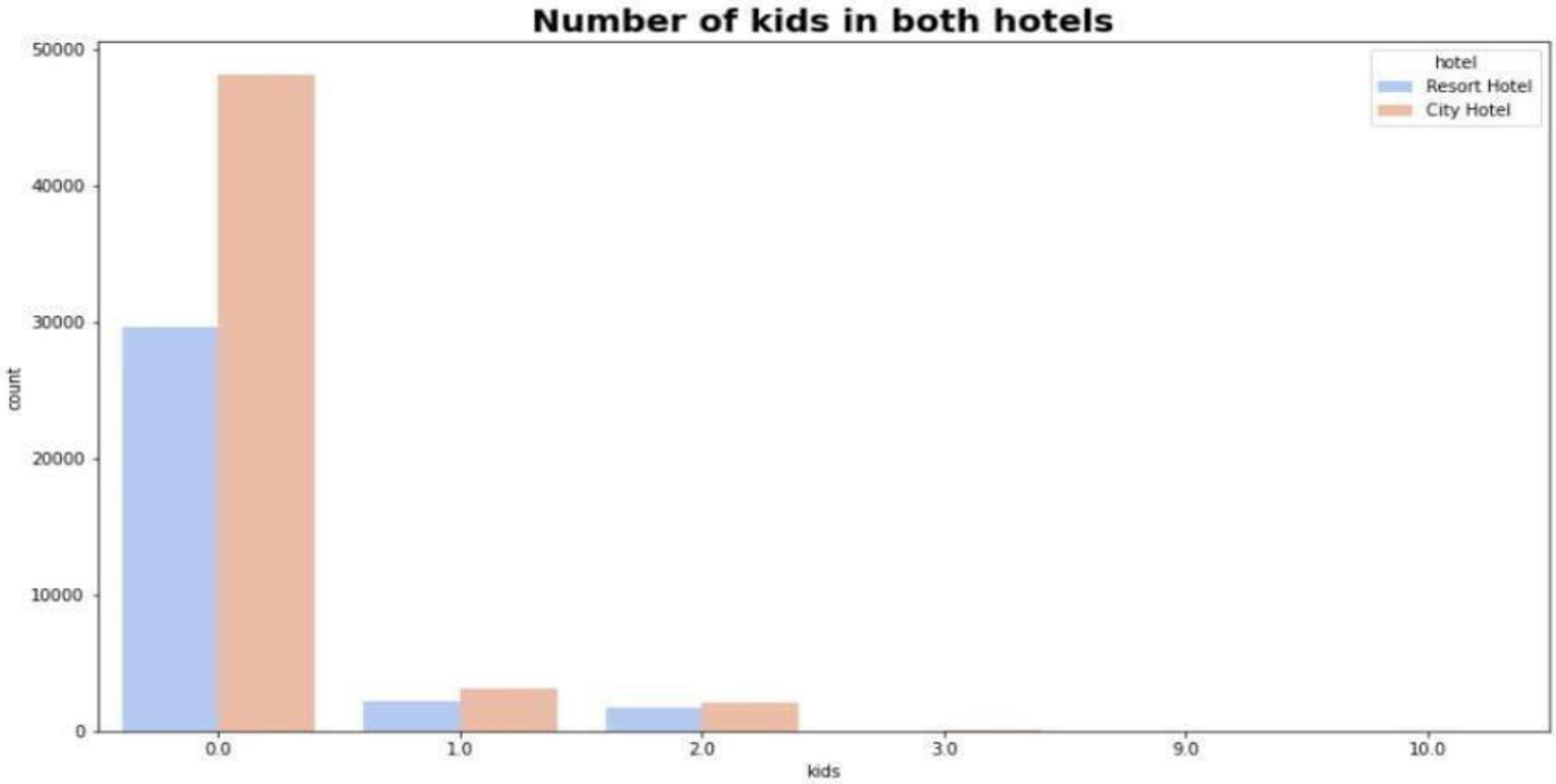
For Displaying the number of kids in both hotels the new data frame is created :



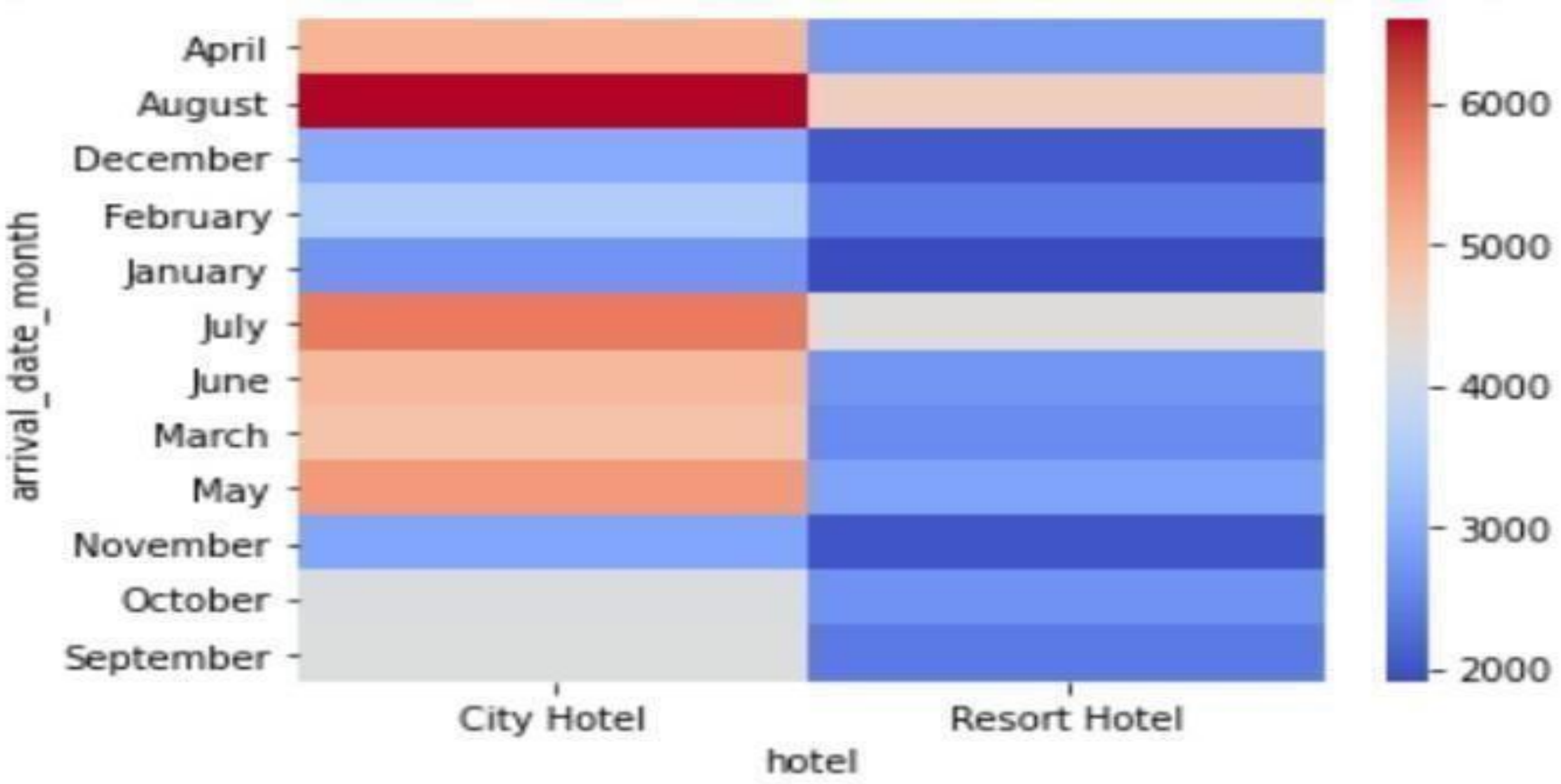
Display the number of kids in both hotels

```
# Create a new dataframe to display hotel, adults, children, and babies only.
df1 = df[['hotel', 'adults', 'children', 'babies']]
df1['kids'] = df1['children'] + df1['babies']
df1
```

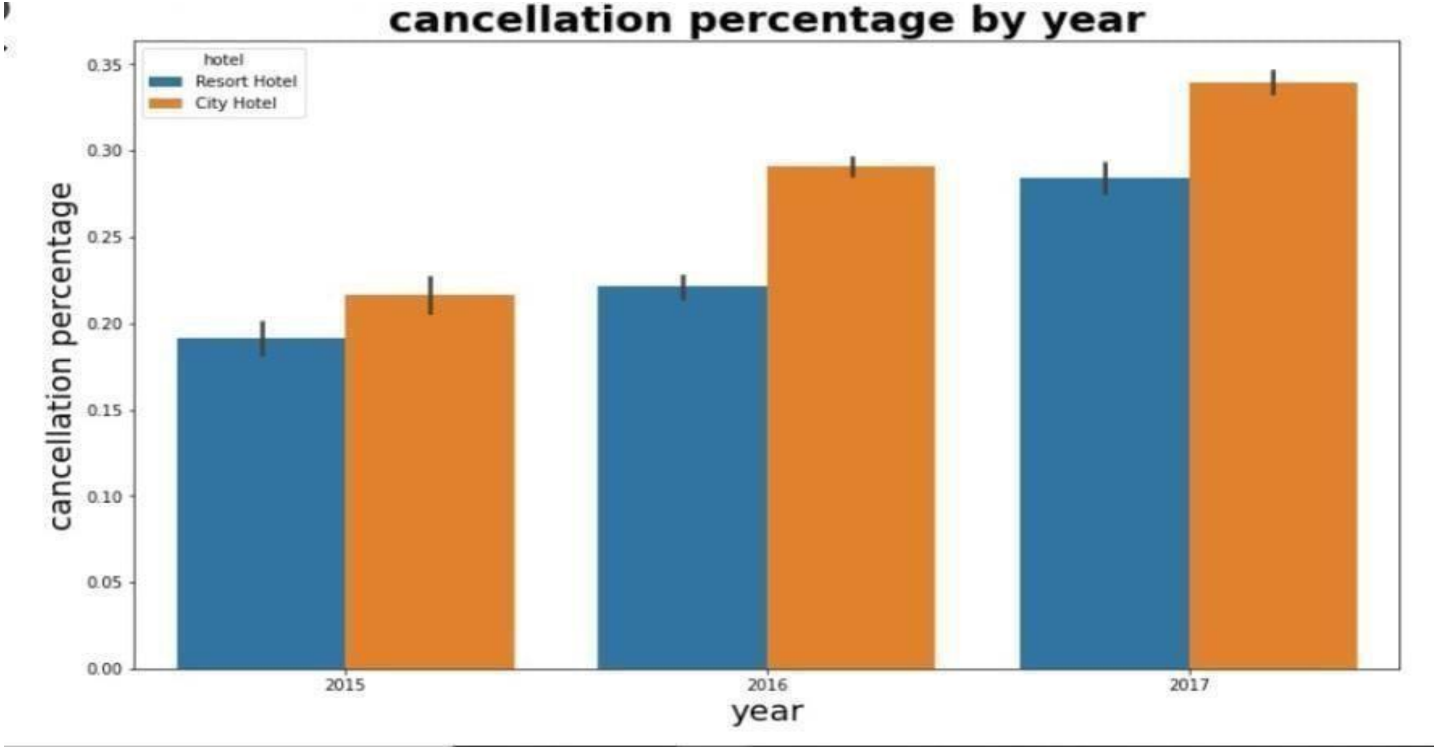
Number of kids in both hotels by countplot



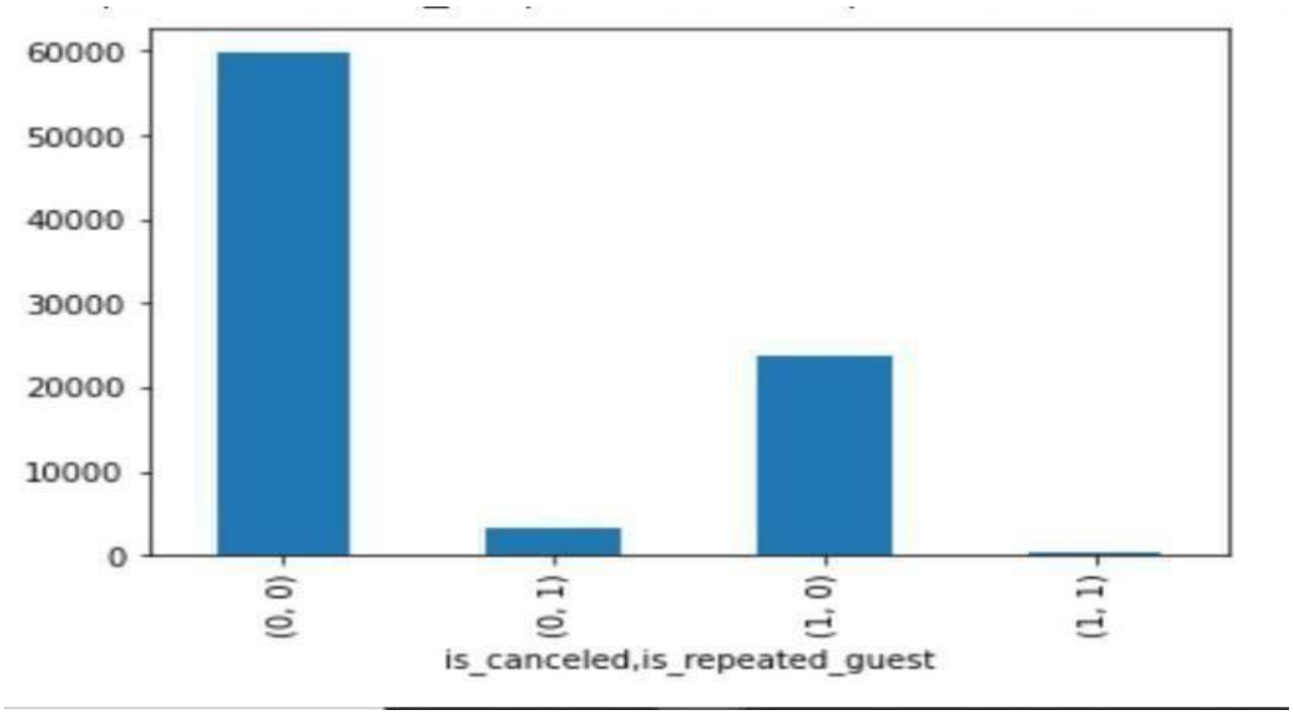
Heatmap between type of hotel and arrival month



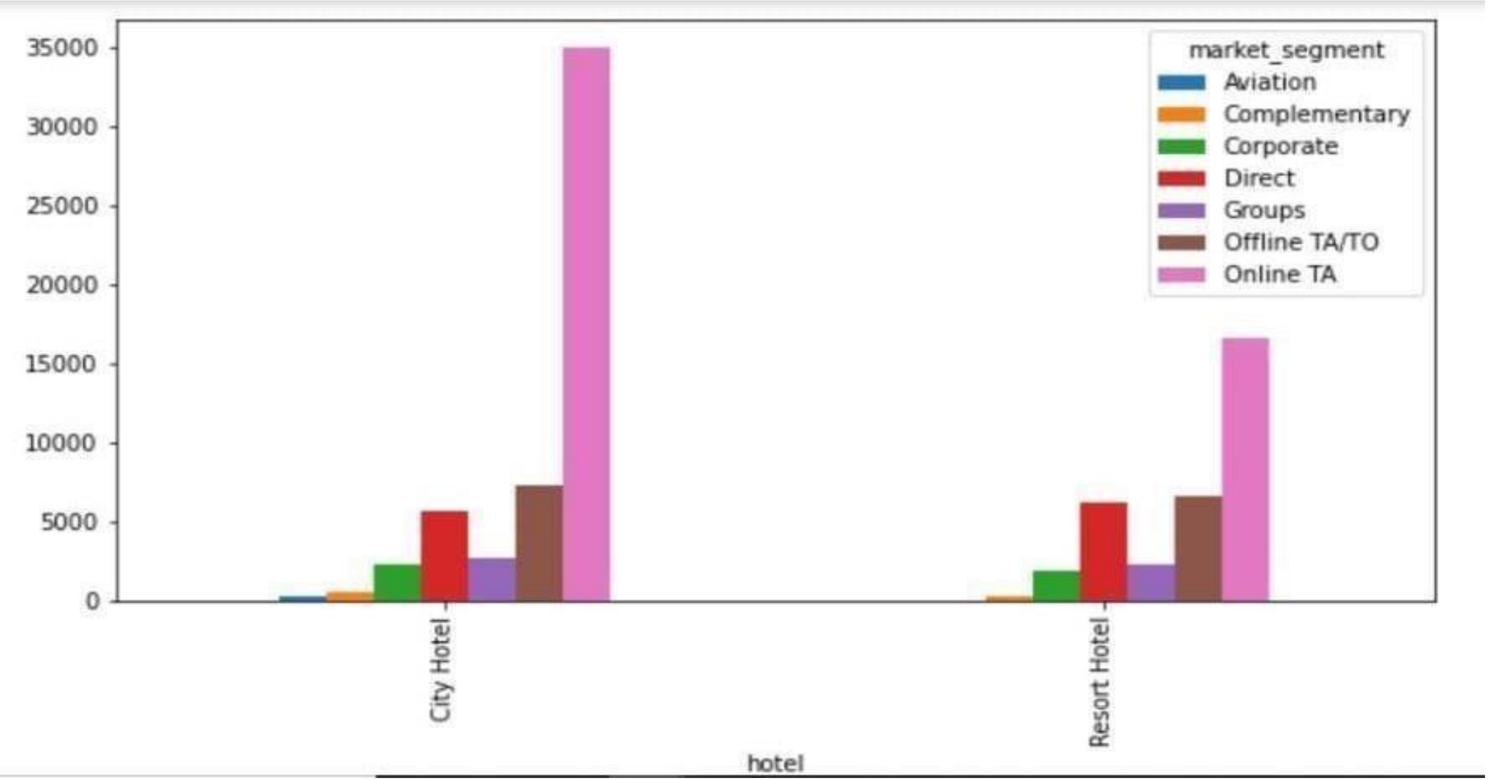
Cancellation percentage by year



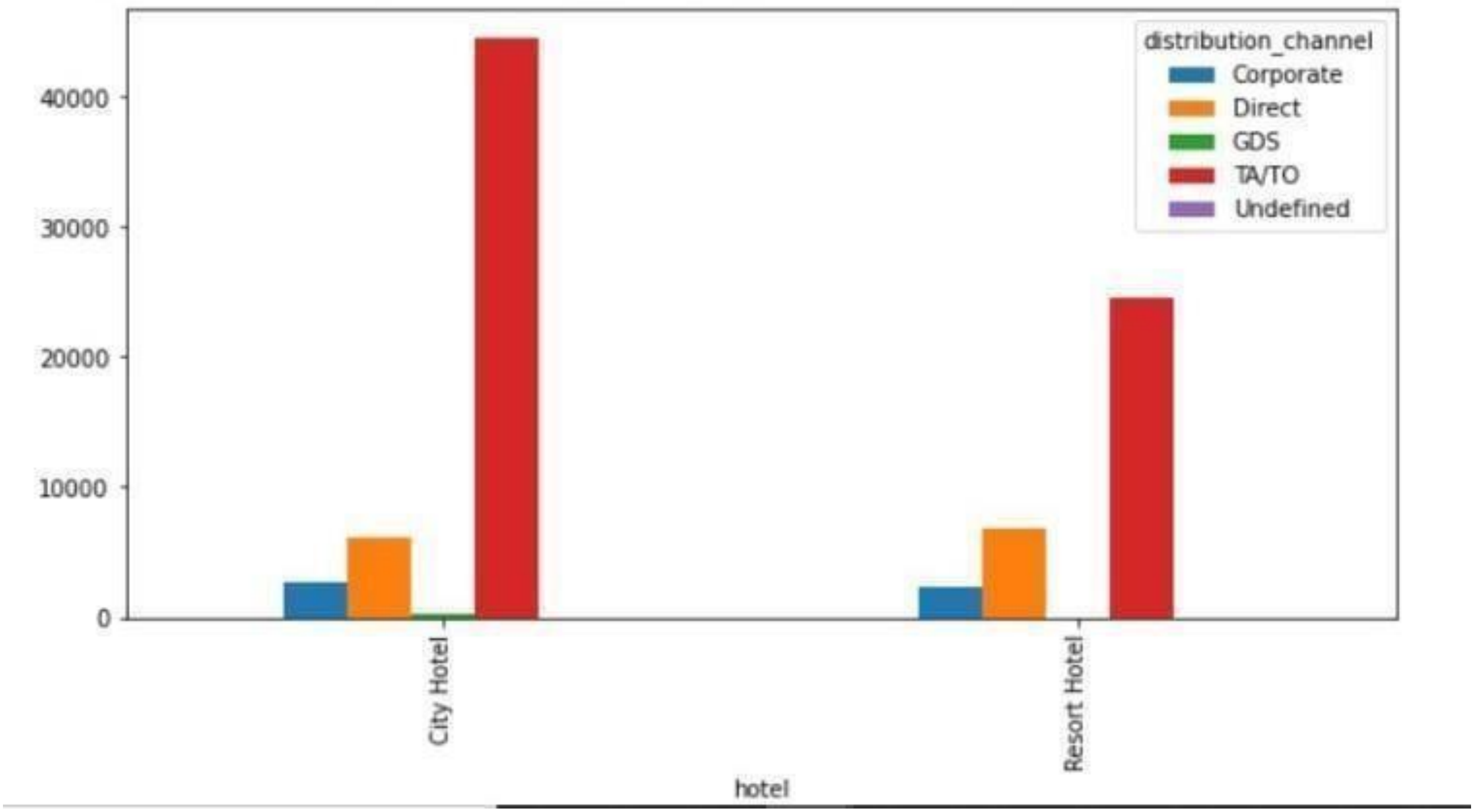
Plot between cancellation type & repeated guest



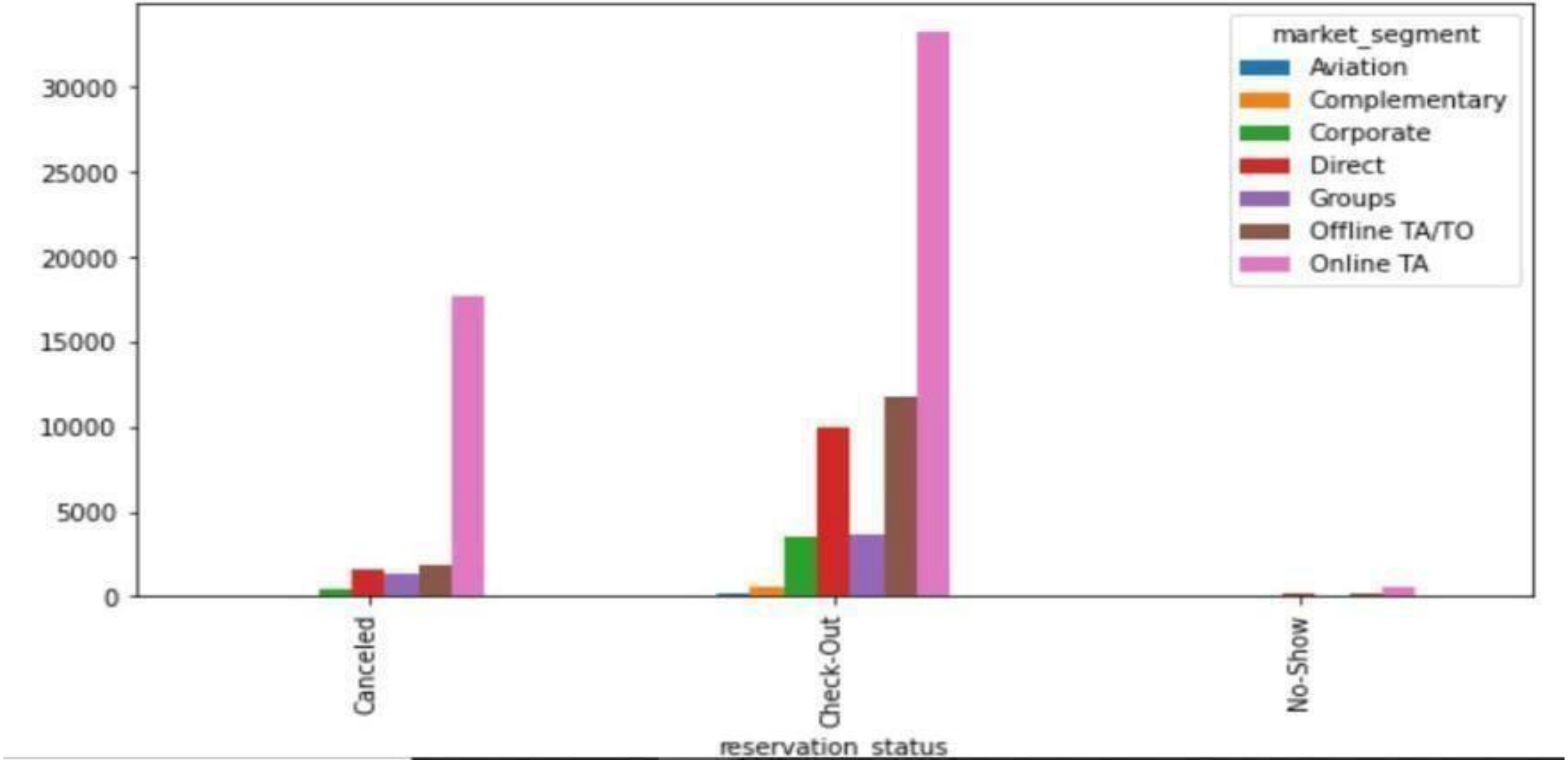
Plot between hotel and market segment



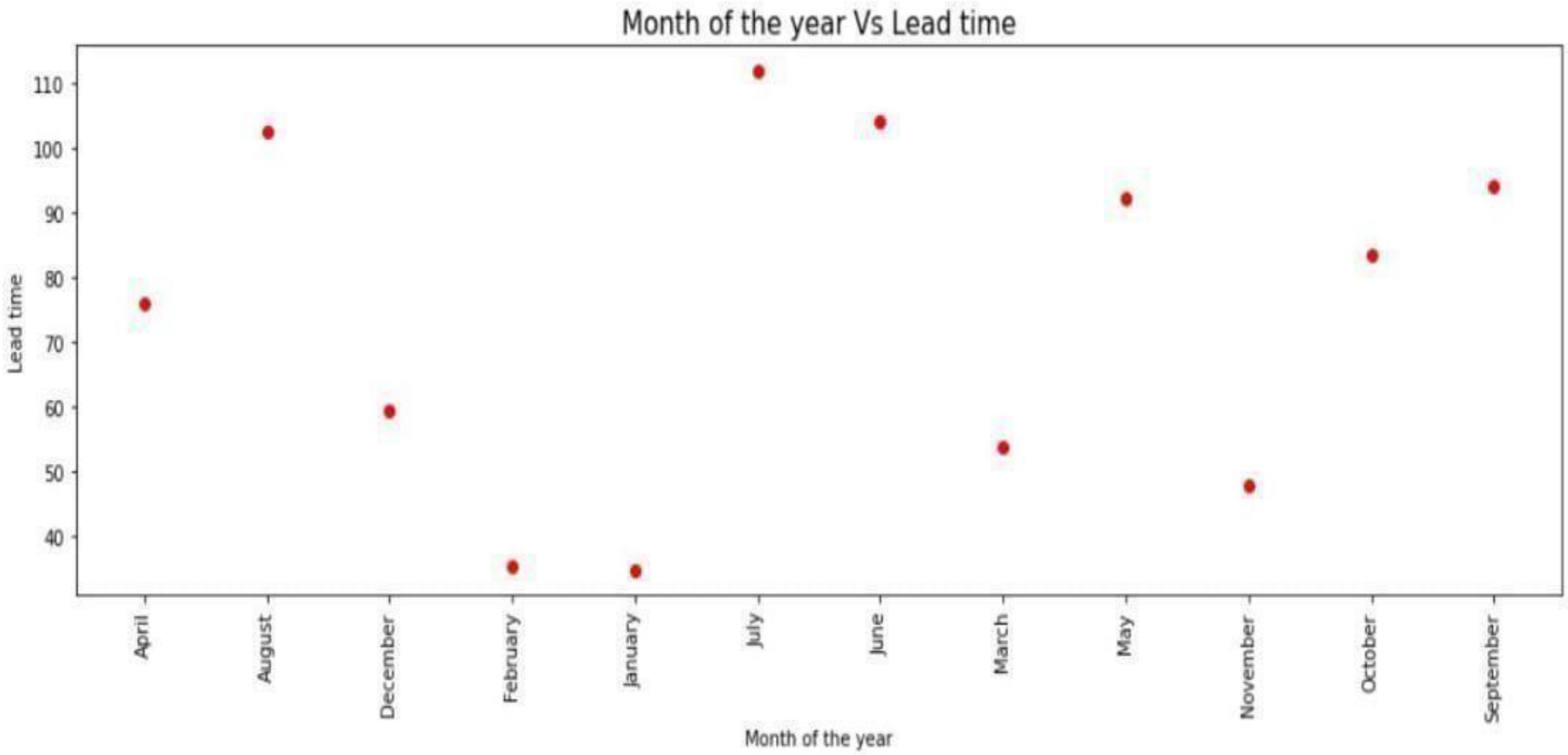
Plot between hotel and distribution channel



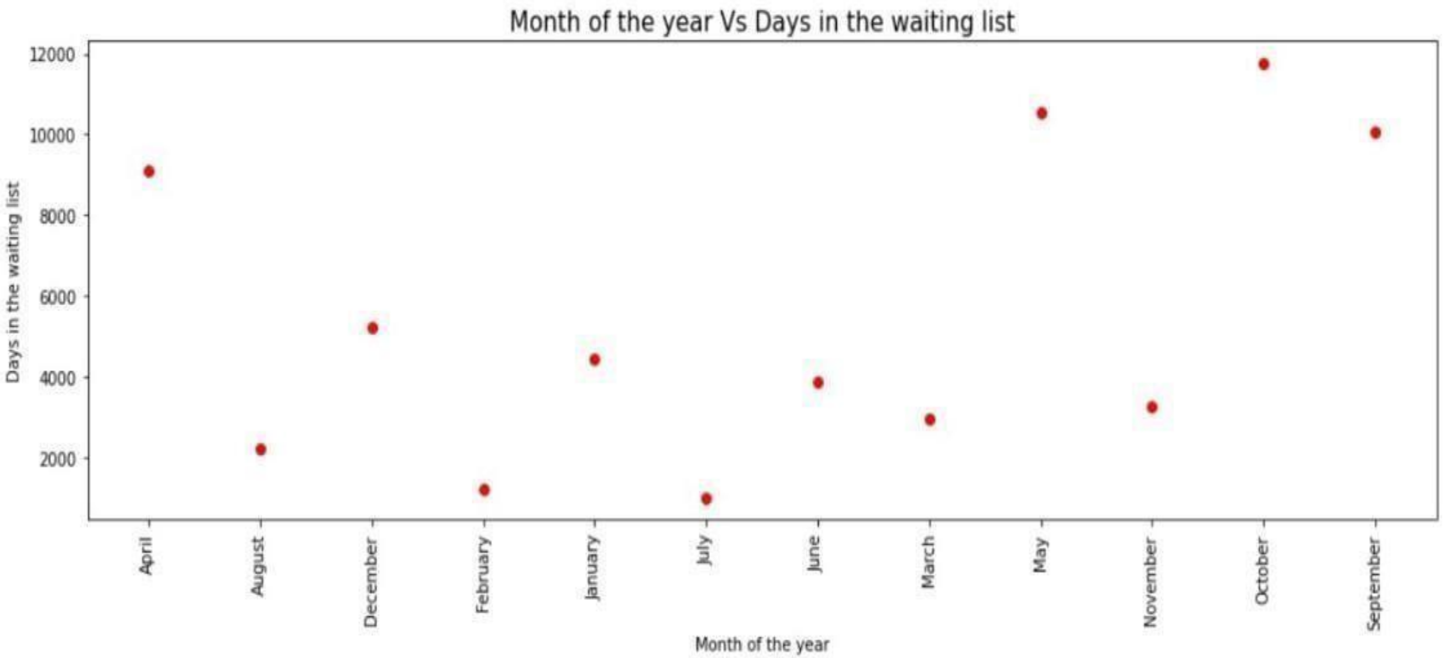
Plot between reservation status and market segment



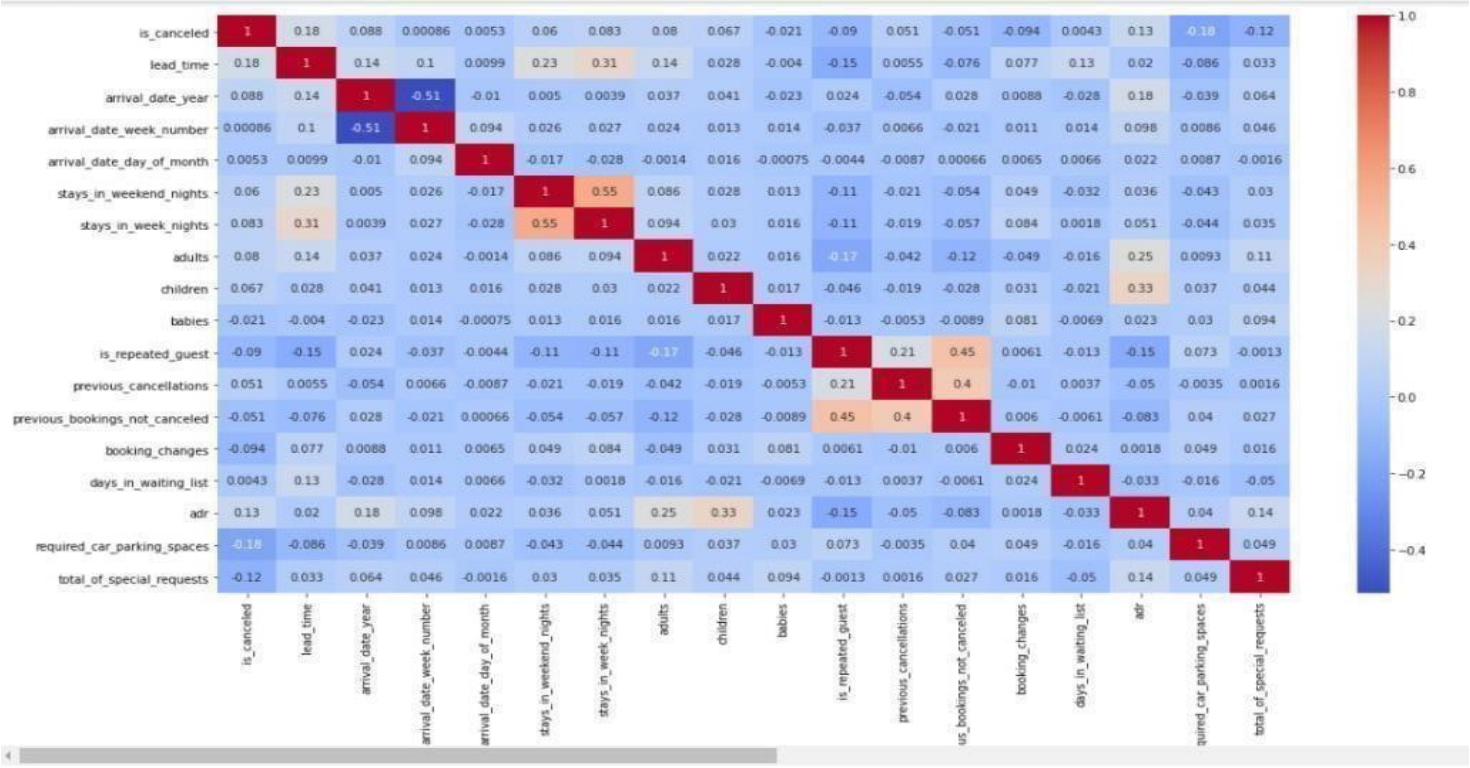
Plot between Month of year & Lead time



Plot between Month of year & Days in the waiting list



Heatmap to show the relation between variables



	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeated_guest	previous_cancellations	us_bookings_not_canceled	booking_changes	days_in_waiting_list	adr	quired_car_parking_spaces	total_of_special_requests
is_canceled	1	0.18	0.088	0.00086	0.0053	0.06	0.083	0.08	0.067	-0.021	-0.09	0.051	-0.051	-0.094	0.0043	0.13	-0.18	-0.12
lead_time	0.18	1	0.14	0.1	0.0099	0.23	0.31	0.14	0.028	-0.004	-0.15	0.0055	-0.076	0.077	0.13	0.02	-0.086	0.033
arrival_date_year	0.088	0.14	1	-0.51	-0.01	0.005	0.0039	0.037	0.041	-0.023	0.024	-0.054	0.028	0.0088	-0.028	0.18	-0.039	0.064
arrival_date_week_number	0.00086	0.1	-0.51	1	0.094	0.026	0.027	0.024	0.013	0.014	-0.037	0.0066	-0.021	0.011	0.014	0.098	0.0086	0.046
arrival_date_day_of_month	0.0053	0.0099	-0.01	0.094	1	-0.017	-0.028	-0.0014	0.016	-0.00075	-0.0044	-0.0087	0.00066	0.0065	0.0066	0.022	0.0087	-0.0016
stays_in_weekend_nights	0.06	0.23	0.005	0.026	-0.017	1	0.55	0.086	0.028	0.013	-0.11	-0.021	-0.054	0.049	-0.032	0.036	-0.043	0.03
stays_in_week_nights	0.083	0.31	0.0039	0.027	-0.028	0.55	1	0.094	0.03	0.016	-0.11	-0.019	-0.057	0.084	0.0018	0.051	-0.044	0.035
adults	0.08	0.14	0.037	0.024	-0.0014	0.086	0.094	1	0.022	0.016	-0.17	-0.042	-0.12	-0.049	-0.016	0.25	0.0093	0.11
children	0.067	0.028	0.041	0.013	0.016	0.028	0.03	0.022	1	0.017	-0.046	-0.019	-0.028	0.031	-0.021	0.33	0.037	0.044
babies	-0.021	-0.004	-0.023	0.014	-0.00075	0.013	0.016	0.016	0.017	1	-0.013	-0.0053	-0.0089	0.081	-0.0069	0.023	0.03	0.094
is_repeated_guest	-0.09	-0.15	0.024	-0.037	-0.0044	-0.11	-0.11	-0.17	-0.046	-0.013	1	0.21	0.45	0.0061	-0.013	-0.15	0.073	-0.0013
previous_cancellations	0.051	0.0055	-0.054	0.0066	-0.0087	-0.021	-0.019	-0.042	-0.019	-0.0053	0.21	1	0.4	-0.01	0.0037	-0.05	-0.0035	0.0016
previous_bookings_not_canceled	-0.051	-0.076	0.028	-0.021	0.00066	-0.054	-0.057	-0.12	-0.028	-0.0089	0.45	0.4	1	0.006	-0.0061	-0.083	0.04	0.027
booking_changes	-0.094	0.077	0.0088	0.011	0.0065	0.049	0.084	-0.049	0.031	0.081	0.0061	-0.01	0.006	1	0.024	0.0018	0.049	0.016
days_in_waiting_list	0.0043	0.13	-0.028	0.014	0.0066	-0.032	0.0018	-0.016	-0.021	-0.0069	-0.013	0.0037	-0.0061	0.024	1	-0.033	-0.016	-0.05
adr	0.13	0.02	0.18	0.098	0.022	0.036	0.051	0.25	0.33	0.023	-0.15	-0.05	-0.083	0.0018	-0.033	1	0.04	0.14
required_car_parking_spaces	-0.18	-0.086	-0.039	0.0086	0.0087	-0.043	-0.044	0.0093	0.037	0.03	0.073	-0.0035	0.04	0.049	-0.016	0.04	1	0.049
total_of_special_requests	-0.12	0.033	0.064	0.046	-0.0016	0.03	0.035	0.11	0.044	0.094	-0.0013	0.0016	0.027	0.016	-0.05	0.14	0.049	1

Challenges

- The name of the countries was not in the proper format, because of which weare not able to plot the geomap plot
- Company and agent column has lots of duplicate value
- There were many rows with almost similar data
- Lots of null values in the dataset

CONCLUSION

- Month of August and July receives most no. of booking.
- Booking for city hotels is twice as for resort hotels.
- Repeated costumers cancel their hotel in very rare cases.
- Customers coming from aviation industry has very less time i.e. they book urgently
- People with no kid prefer to choose city hotel over resort hotel

Strategies to counter high cancellations at Hotel

- Since we see, our repetitive customers are most loyal customers
- To maintain them we can provide them with some bonus points, which can be redeem in the next booking
- Month of January and December receives less no. of booking, hotels can offer discounted packages for these months.
- Family with kids prefer resorts, we can provide with holiday family packages.
- Great number of the bookings are coming from travel agents, so we can provide them some commission.

Thank You