

HOTEL BOOKING ANALYSIS

GASER ZAGHLOUL

INTRODUCTION

- Overview: Analysis of hotel booking data to understand guest behavior, booking trends, pricing, and cancellation patterns.
- Objective: Provide insights into factors affecting bookings and cancellations to optimize hotel operations

EDA PROCESS

- **Data Loading and Cleaning**
 - Libraries used (e.g., Pandas, Matplotlib, Seaborn)
 - Data types, missing values, and duplicates
- **Descriptive Analysis**
 - Summary statistics of numerical columns
 - Range of week & weekend nights
 - Min and max number of adults and children
- **Visualizations and Insights**
 - Distribution and correlation of numerical data
 - Booking trends by day of the week and month
 - Pricing analysis across room and meal types

- Read data & check data types

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36285 entries, 0 to 36284
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Booking_ID                            36285 non-null  object
1   number of adults                      36285 non-null  int64
2   number of children                    36285 non-null  int64
3   number of weekend nights              36285 non-null  int64
4   number of week nights                 36285 non-null  int64
5   type of meal                          36285 non-null  object
6   car parking space                     36285 non-null  int64
7   room type                            36285 non-null  object
8   lead time                            36285 non-null  int64
9   market segment type                   36285 non-null  object
10  repeated                              36285 non-null  int64
11  P-C                                   36285 non-null  int64
12  P-not-C                              36285 non-null  int64
13  average price                         36285 non-null  float64
14  special requests                      36285 non-null  int64
15  date of reservation                  36285 non-null  object
16  booking status                       36285 non-null  object
dtypes: float64(1), int64(10), object(6)
memory usage: 4.7+ MB
```

Booking_ID		number of adults	number of children	number of weekend nights	number of week nights	type of meal	car parking space	room type	lead time	market segment type	repeated	P-C	P-not-C	average price	special requests	date of reservation	booking status
0	INN00001	1	1	2	5	Meal Plan 1	0	Room_Type 1	224	Offline	0	0	0	88.00	0	10/2/2015	Not_Canceled
1	INN00002	1	0	1	3	Not Selected	0	Room_Type 1	5	Online	0	0	0	106.68	1	11/6/2018	Not_Canceled
2	INN00003	2	1	1	3	Meal Plan 1	0	Room_Type 1	1	Online	0	0	0	50.00	0	2/28/2018	Canceled
3	INN00004	1	0	0	2	Meal Plan 1	0	Room_Type 1	211	Online	0	0	0	100.00	1	5/20/2017	Canceled
4	INN00005	1	0	1	2	Not Selected	0	Room_Type 1	48	Online	0	0	0	77.00	0	4/11/2018	Canceled
...
36280	INN36282	2	0	0	2	Meal Plan 2	0	Room_Type 1	346	Online	0	0	0	115.00	1	9/13/2018	Canceled
36281	INN36283	2	0	1	3	Meal Plan 1	0	Room_Type 1	34	Online	0	0	0	107.55	1	10/15/2017	Not_Canceled
36282	INN36284	2	0	1	3	Meal Plan 1	0	Room_Type 4	83	Online	0	0	0	105.61	1	12/26/2018	Not_Canceled
36283	INN36285	3	0	0	4	Meal Plan 1	0	Room_Type 1	121	Offline	0	0	0	96.90	1	7/6/2018	Not_Canceled
36284	INN36286	2	0	0	5	Meal Plan 1	0	Room_Type 4	44	Online	0	0	0	133.44	3	10/18/2018	Not_Canceled
36285 rows x 17 columns																	

MISSING VALUES & DUPLICATES

```
df.isnull().sum()
```

✓ 0.0s

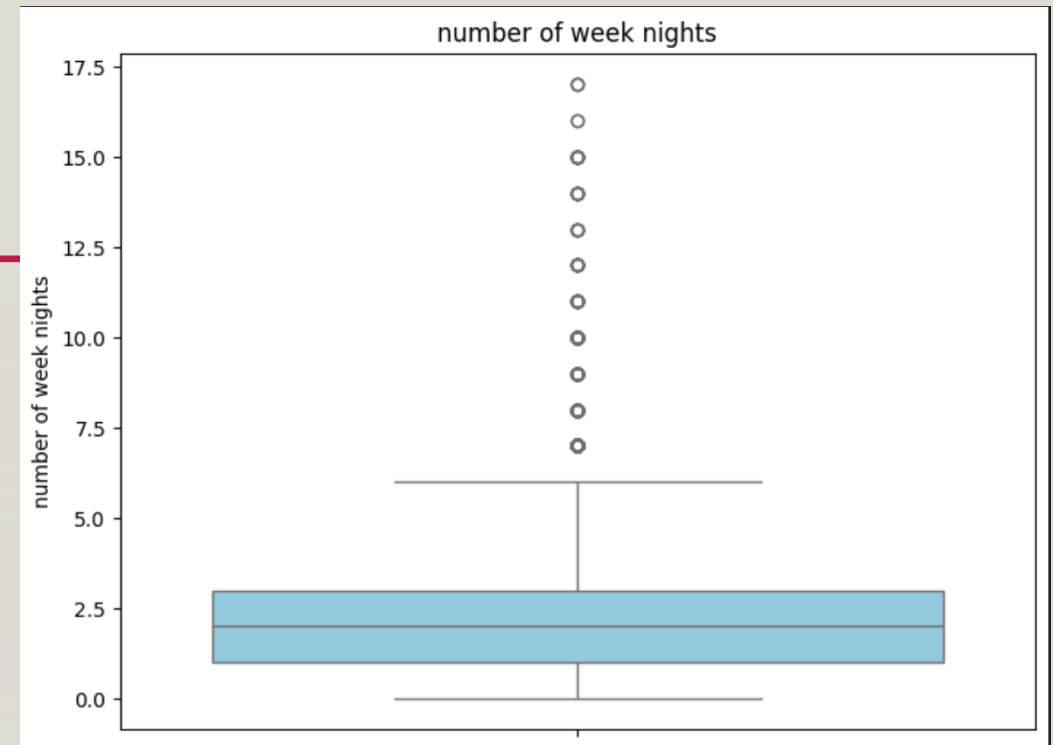
Booking_ID	0
number of adults	0
number of children	0
number of weekend nights	0
number of week nights	0
type of meal	0
car parking space	0
room type	0
lead time	0
market segment type	0
repeated	0
P-C	0
P-not-C	0
average price	0
special requests	0
date of reservation	0
booking status	0
dtype: int64	

check any duplicates

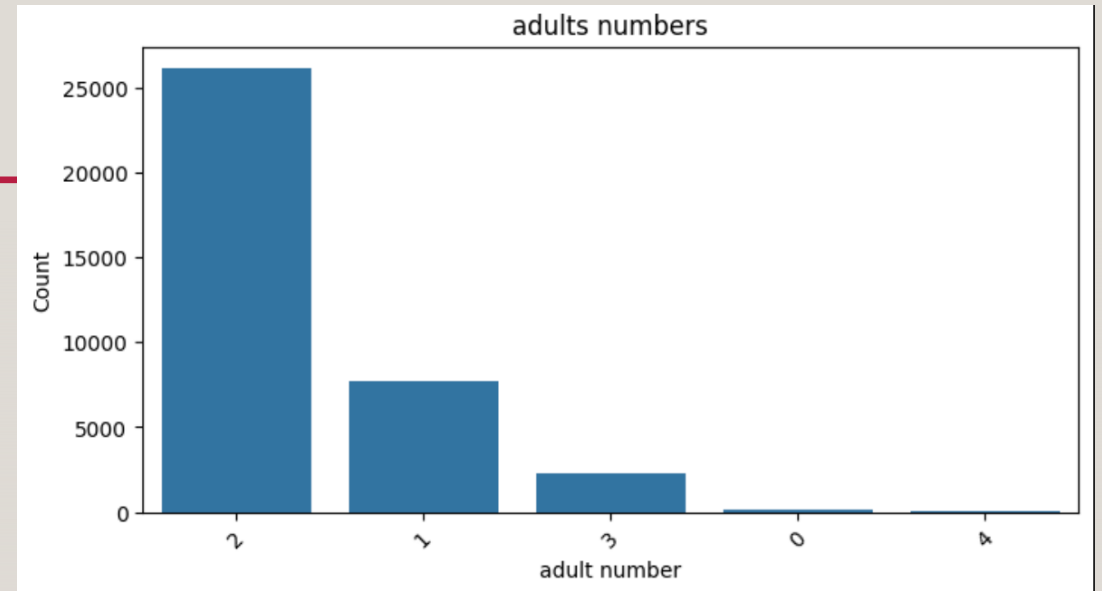
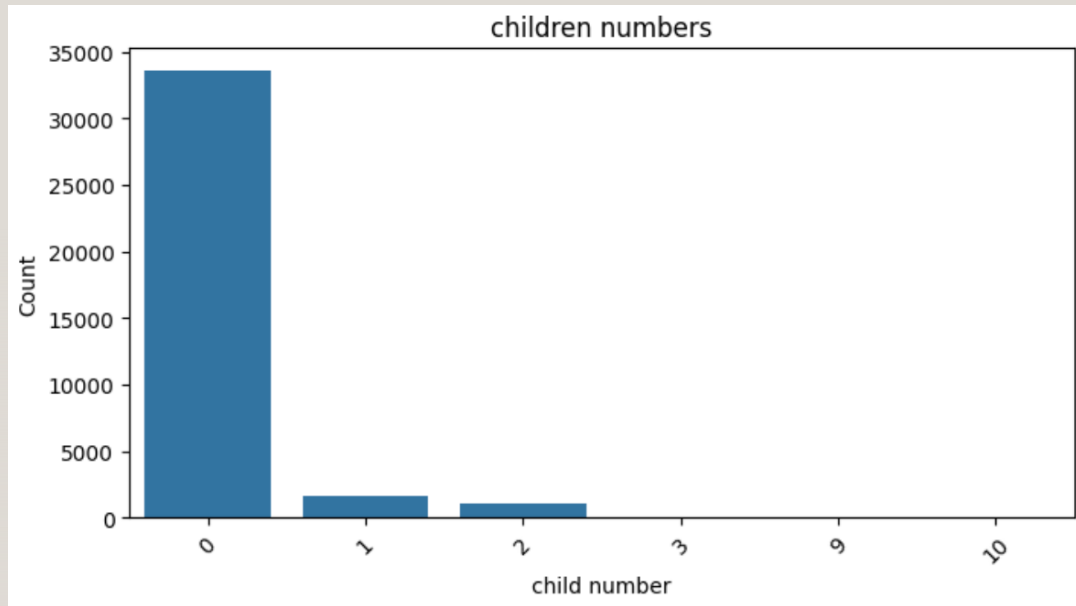
```
df.duplicated().sum()
```

✓ 0.0s

0



WEEK & WEEKEND NIGHTS RANGE



NUMBER OF ADULT & CHILDREN REQUEST

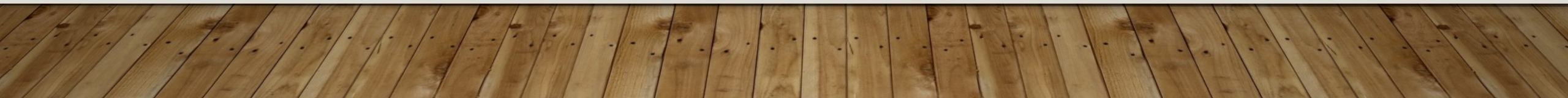
df.describe()

✓ 0.0s

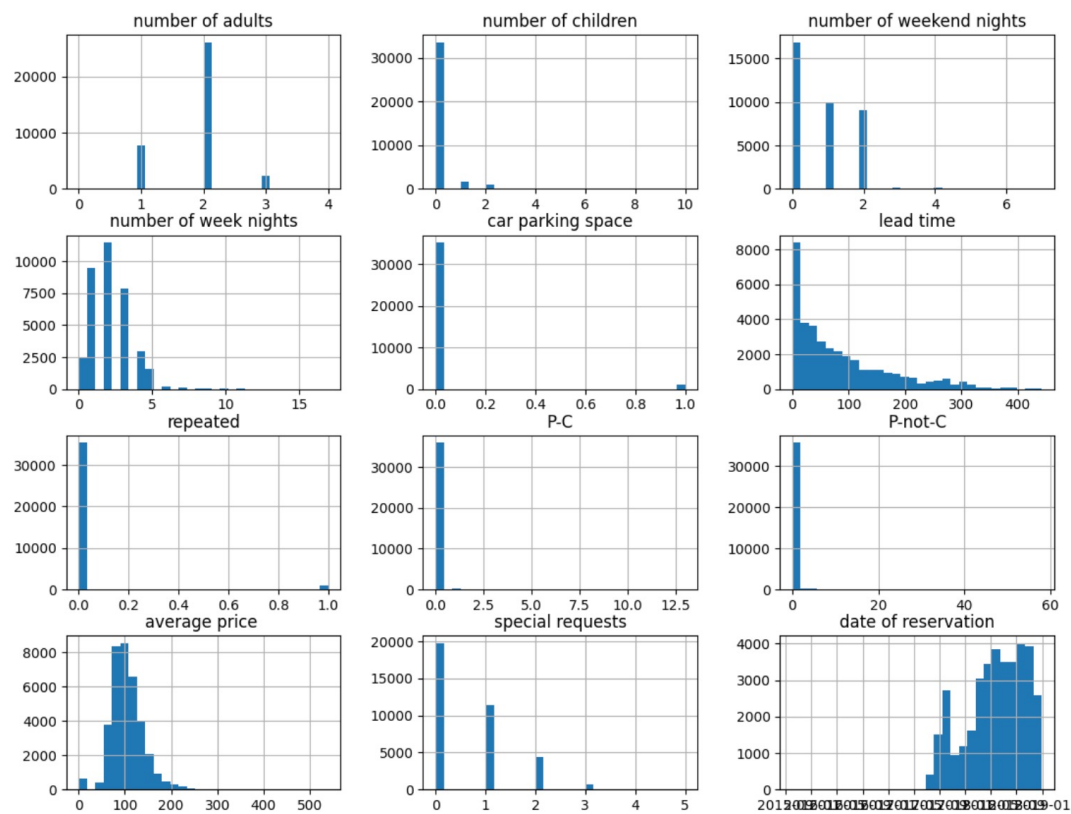
Python

	number of adults	number of children	number of weekend nights	number of week nights	car parking space	lead time	repeated	P-C	P-not-C	average price	special requests
count	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000	36285.000000
mean	1.844839	0.105360	0.810693	2.204602	0.030977	85.239851	0.025630	0.023343	0.153369	103.421636	0.619733
std	0.518813	0.402704	0.870590	1.410946	0.173258	85.938796	0.158032	0.368281	1.753931	35.086469	0.786262
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.000000	0.000000	0.000000	1.000000	0.000000	17.000000	0.000000	0.000000	0.000000	80.300000	0.000000
50%	2.000000	0.000000	1.000000	2.000000	0.000000	57.000000	0.000000	0.000000	0.000000	99.450000	0.000000
75%	2.000000	0.000000	2.000000	3.000000	0.000000	126.000000	0.000000	0.000000	0.000000	120.000000	1.000000
max	4.000000	10.000000	7.000000	17.000000	1.000000	443.000000	1.000000	13.000000	58.000000	540.000000	5.000000

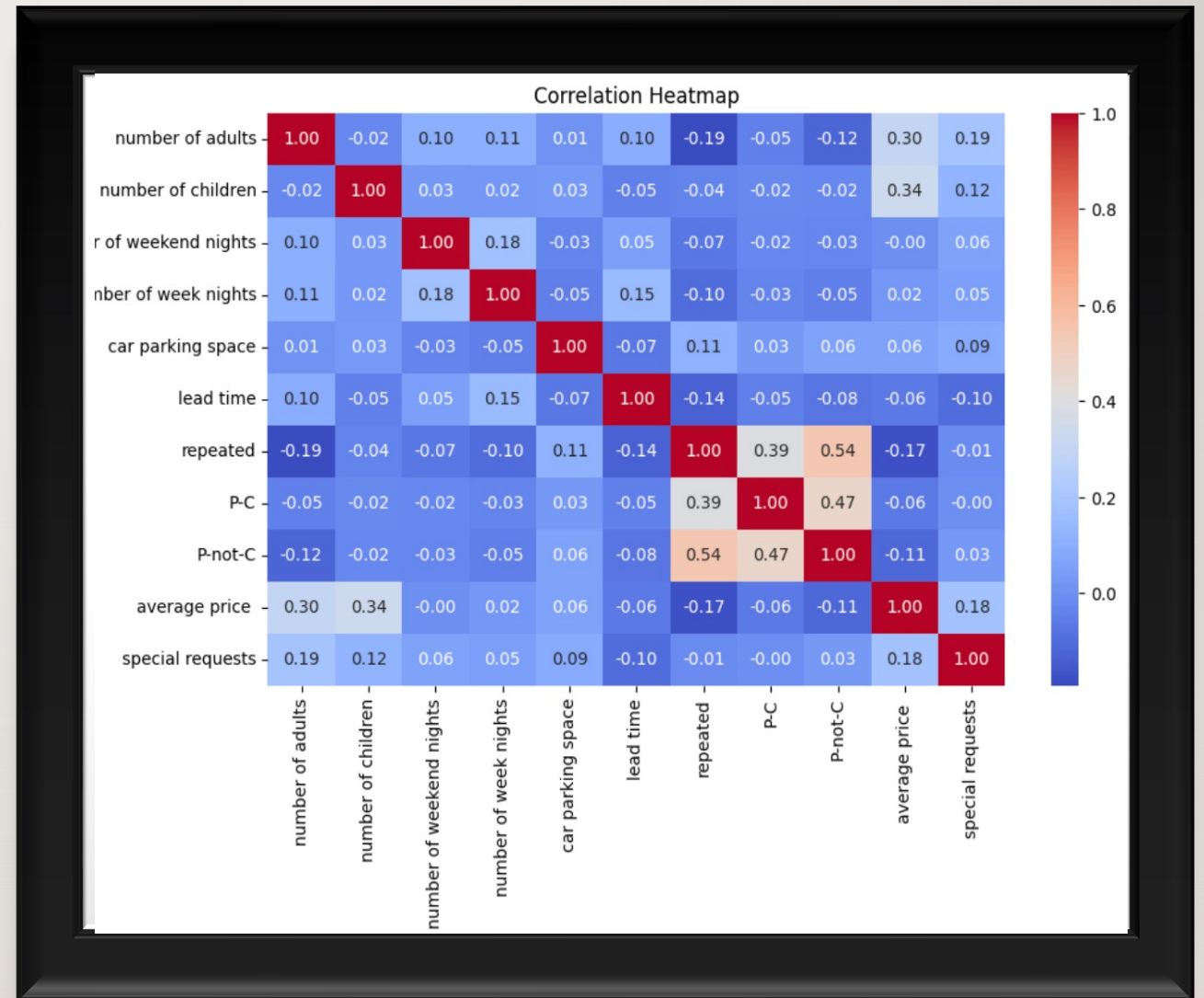
SUMMARY STATISTICS OF NUMERICAL COLUMNS



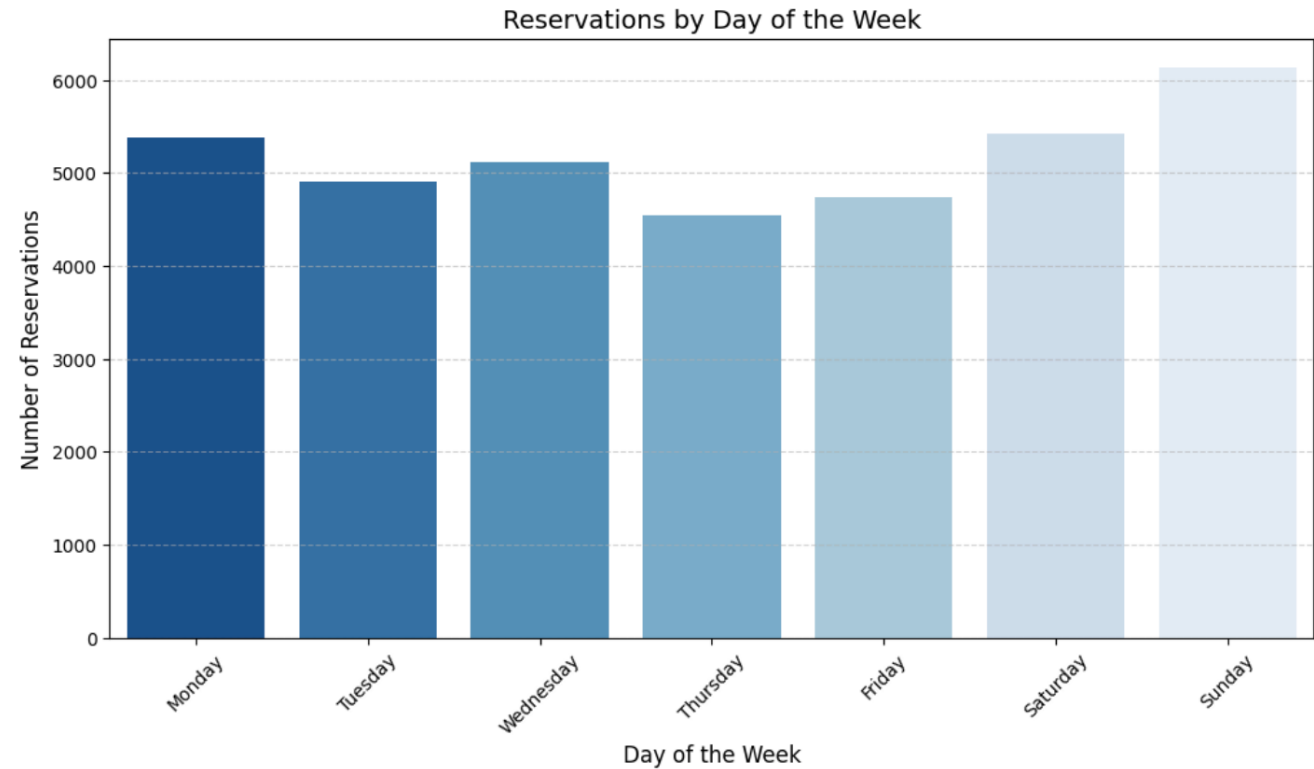
DISTRIBUTION OF NUMERICAL COLUMNS



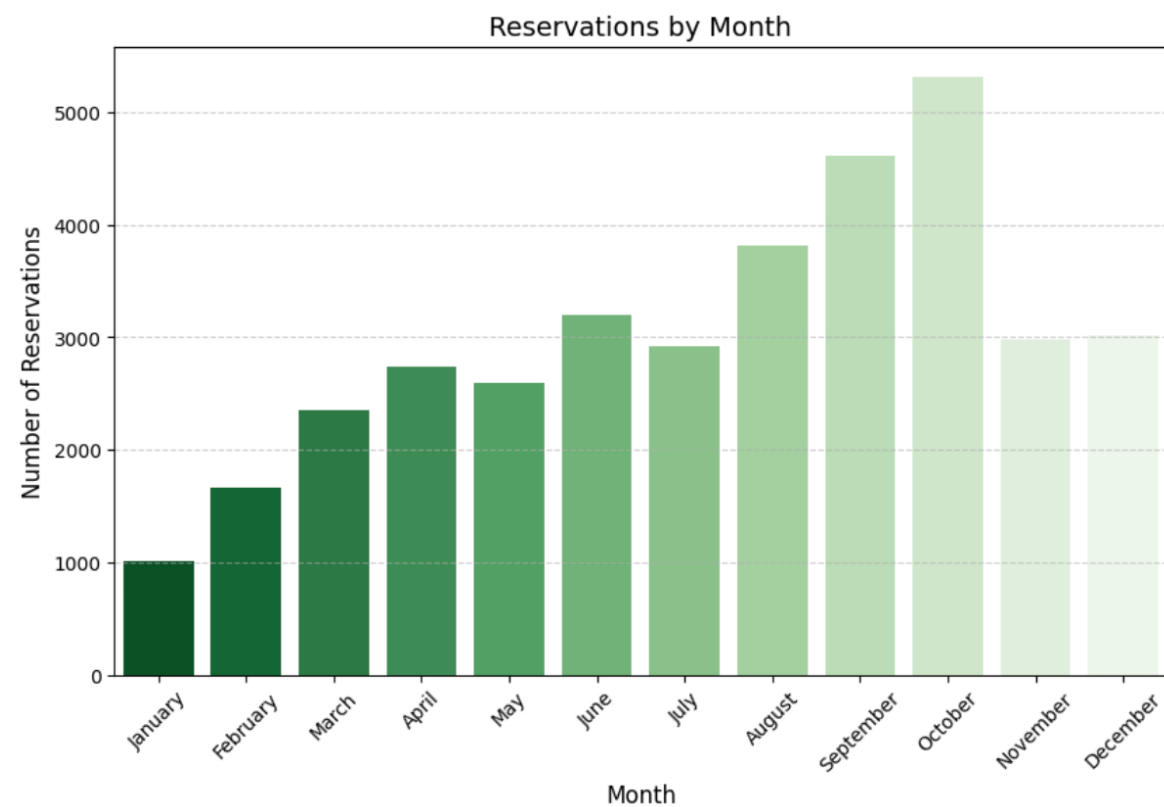
CORRELATION BETWEEN NUMERICAL COLUMN



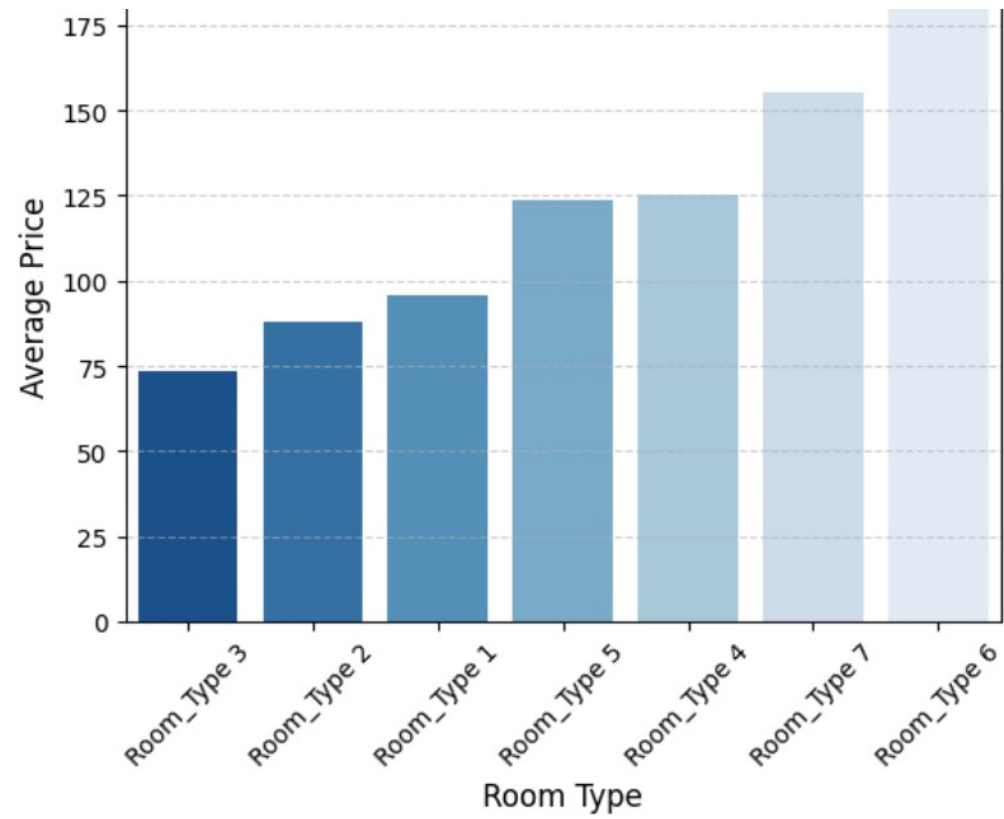
HIGHEST RESERVATIONS BY DAY OF WEEK



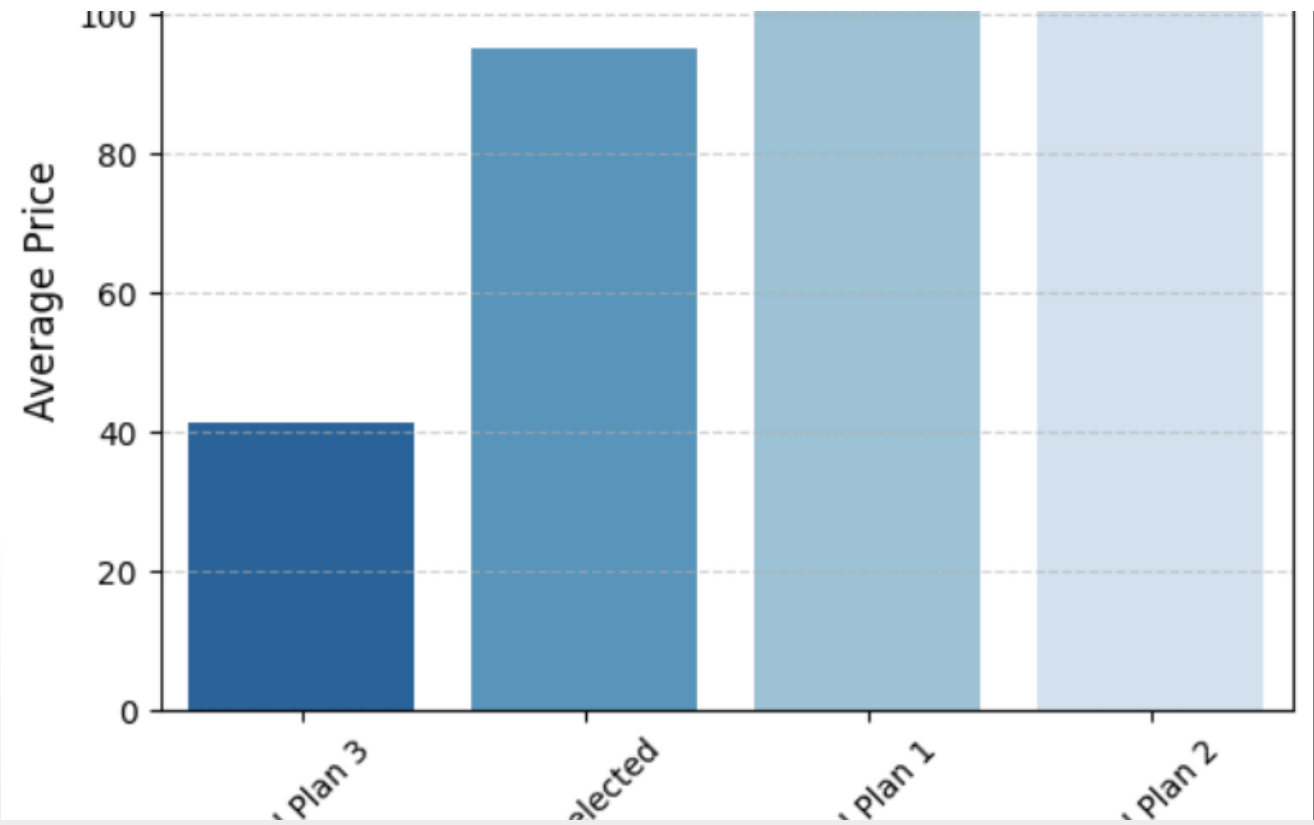
HIGHEST RESERVATIONS BY MONTH



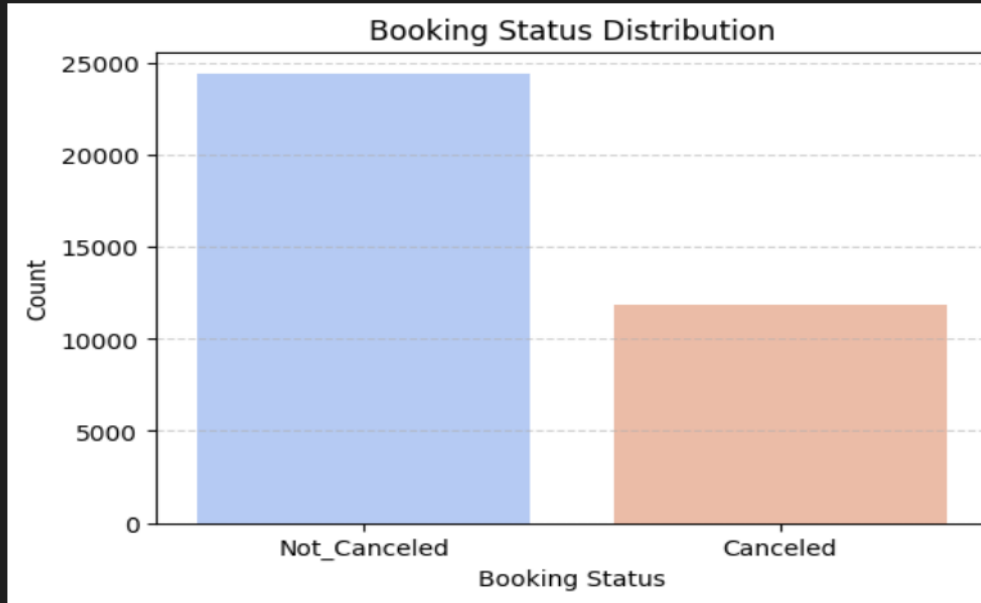
AVERAGE PRICE BY ROOM TYPE



AVERAGE PRICE BY MEAL TYPE



CANCELED BOOKINGS & PERCENTAGE

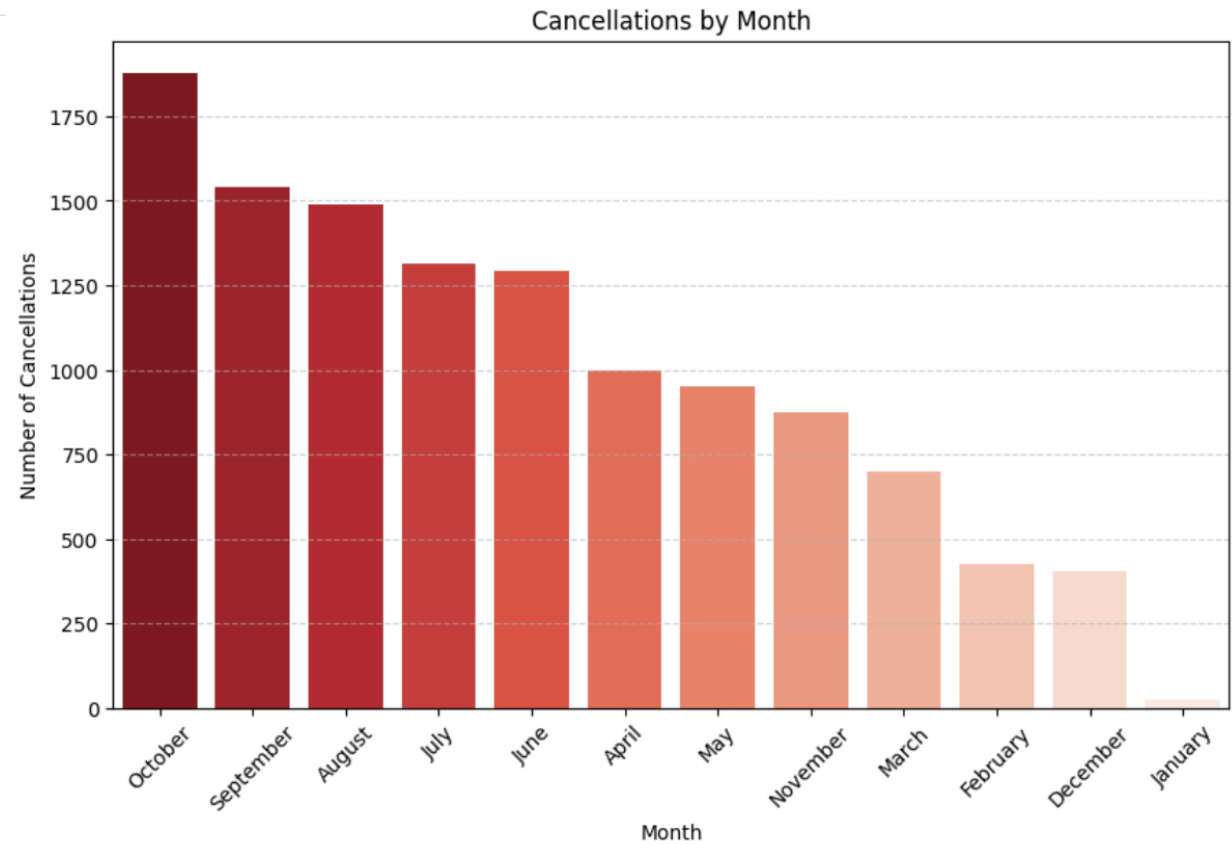


```
cancellation_rate = (df['booking status'].value_counts(normalize=True) * 100)
print(cancellation_rate)
```

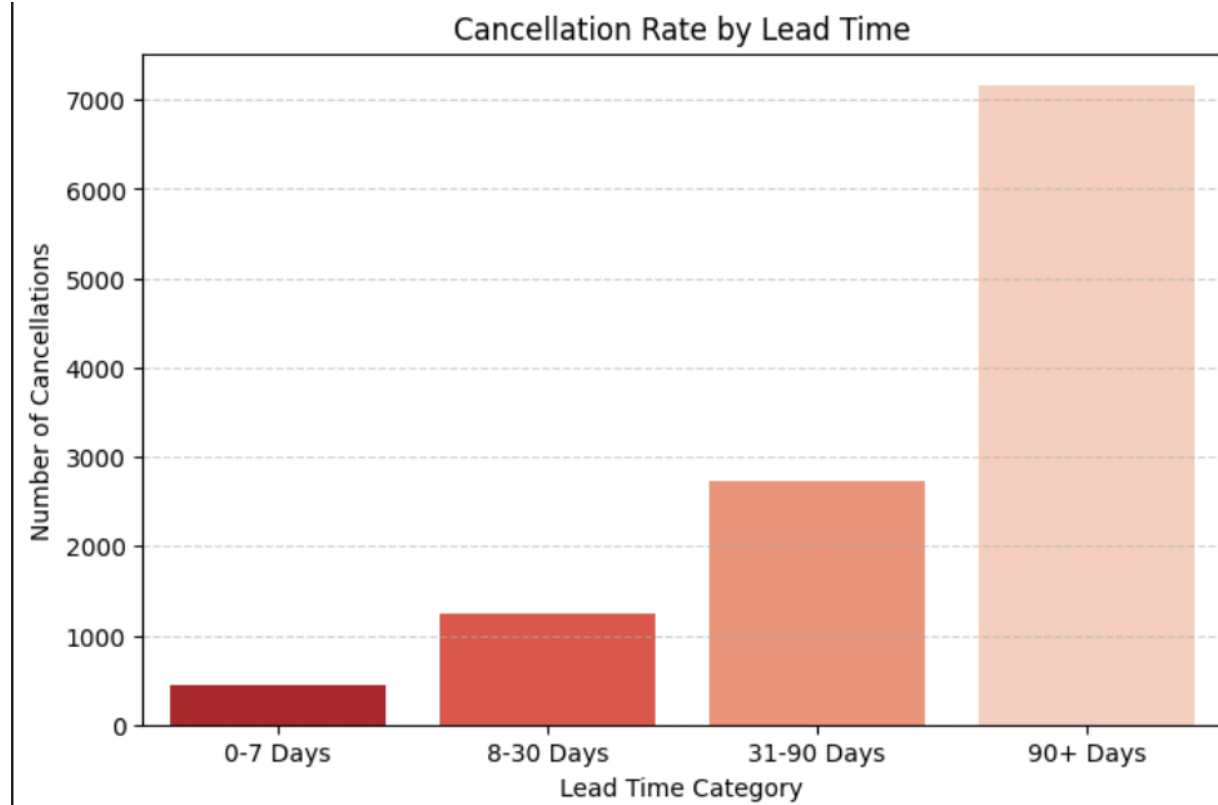
✓ 0.0s

```
booking status
Not_Canceled    67.234394
Canceled        32.765606
Name: proportion, dtype: float64
```

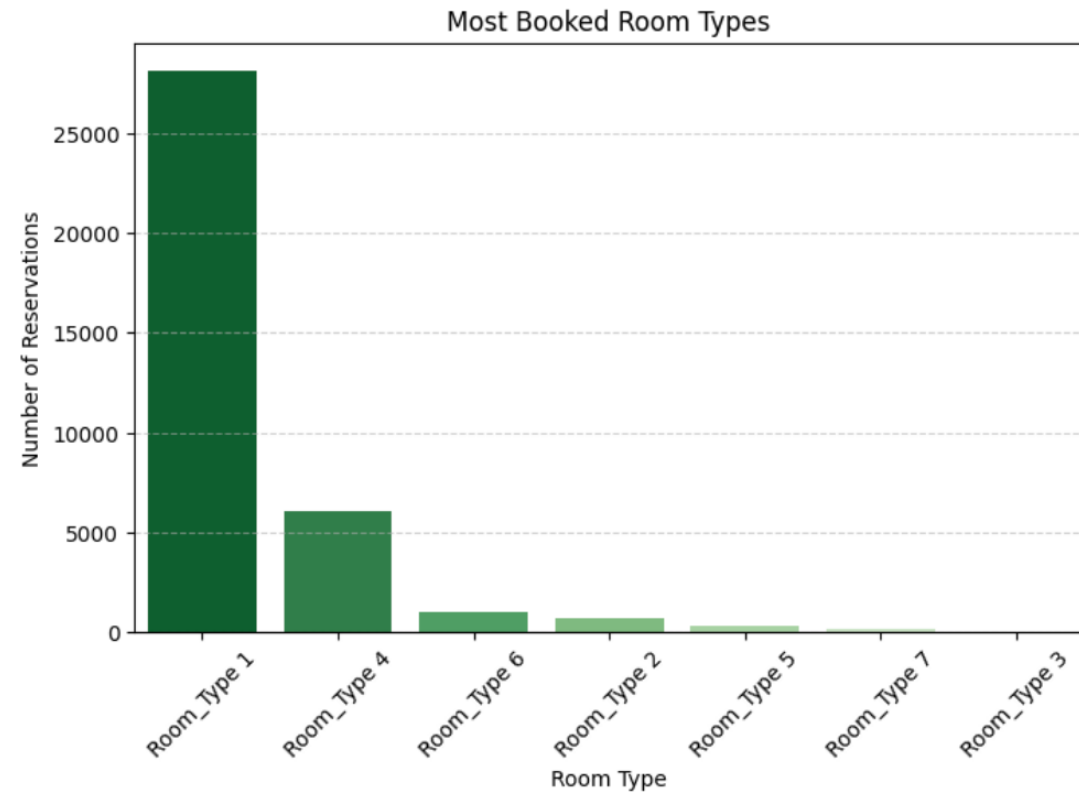
HIGHEST CANCELLATIONS BY MONTH



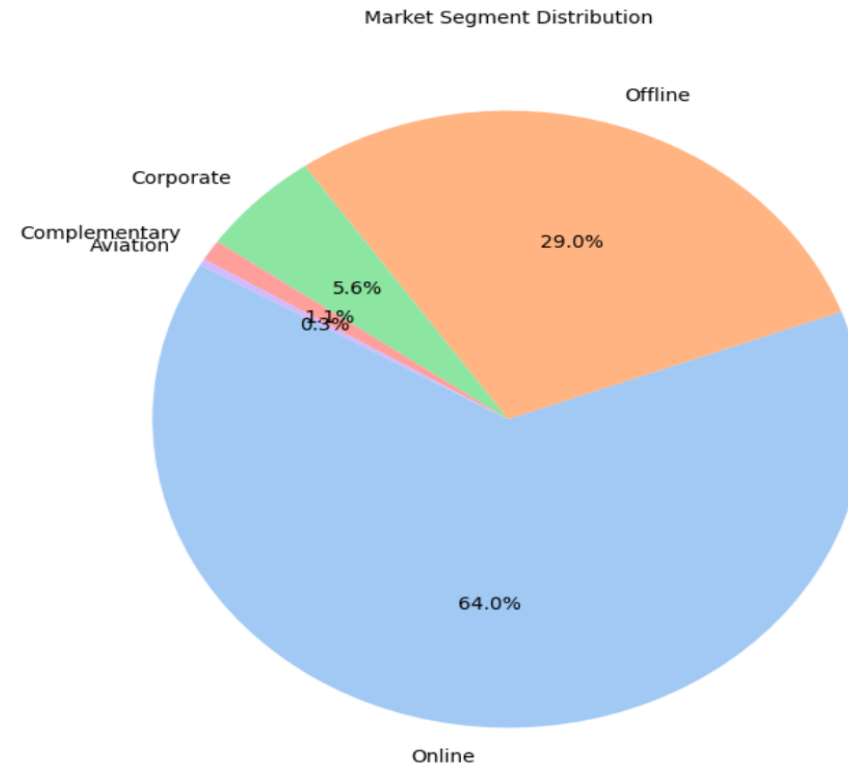
LAST-MINUTE BOOKING CANCELLATIONS



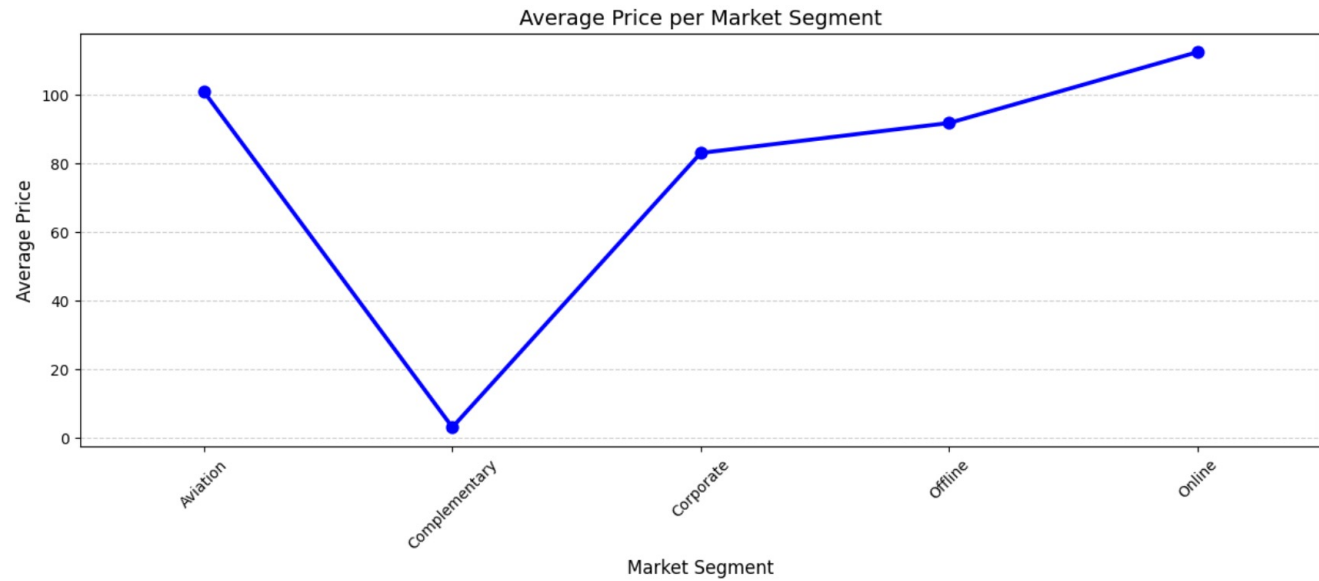
MOST BOOKED ROOM TYPE



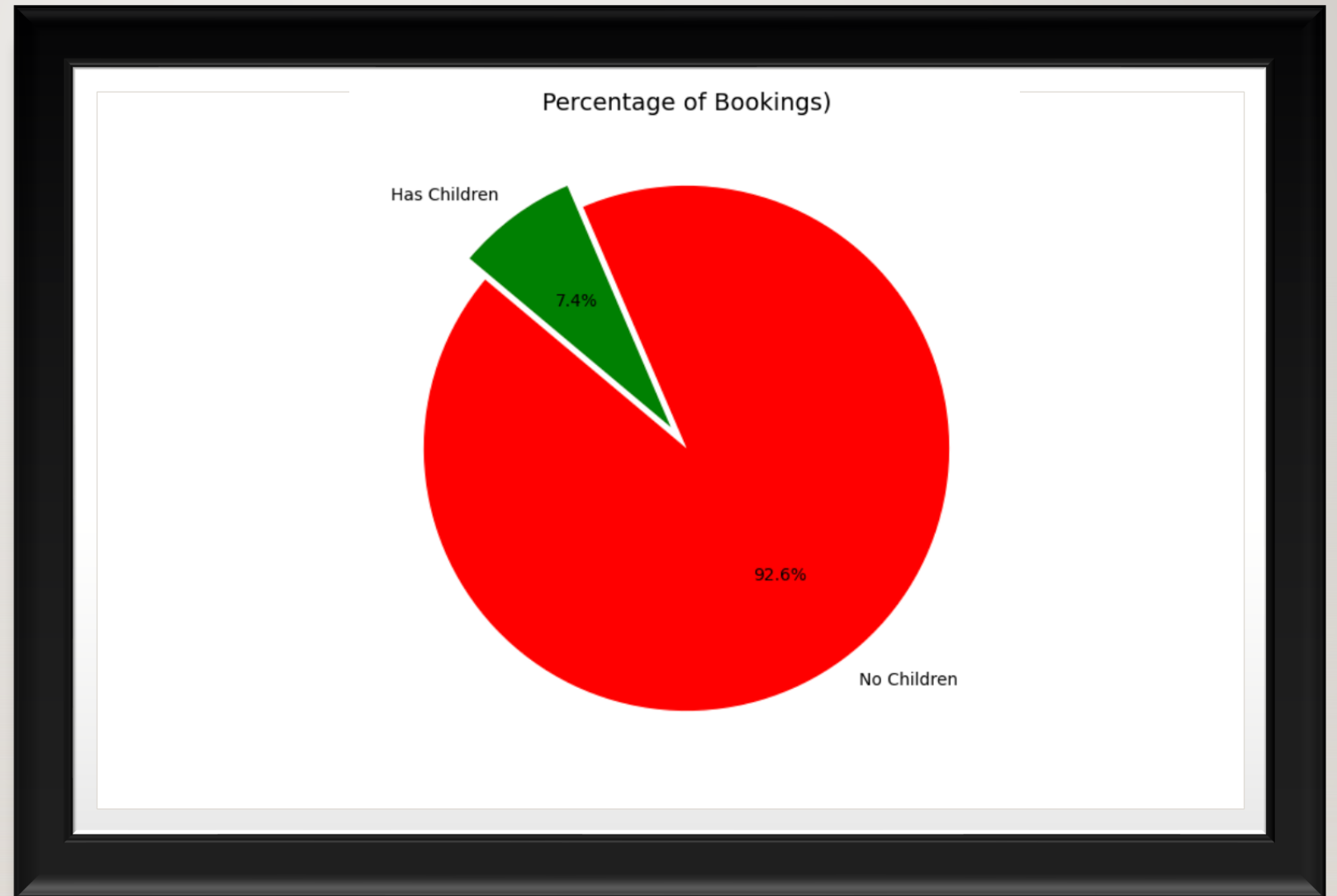
MOST COMMON MARKET SEGMENT



PRICING BY MARKET TYPE



BOOKING FOR FAMILY VS NOT FAMILY



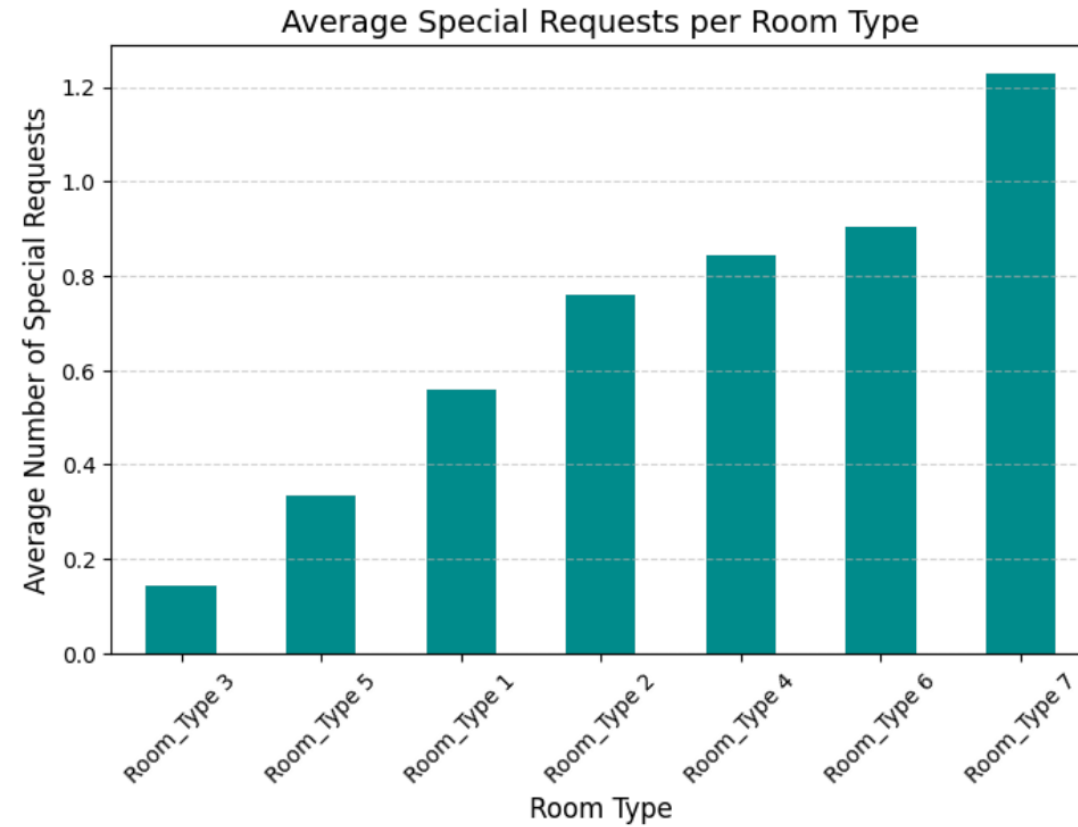
IMPACT OF SPECIAL REQUESTS ON CANCELLATIONS



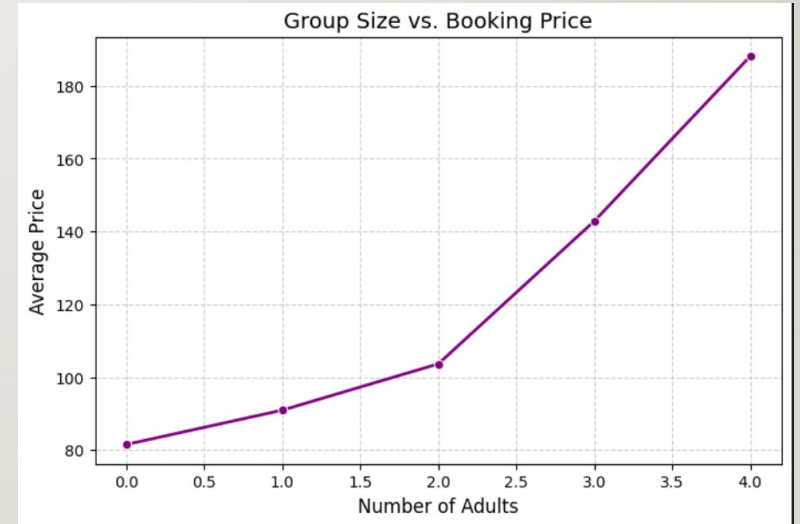
AVERAGE PRICE BY SPECIAL REQUESTS



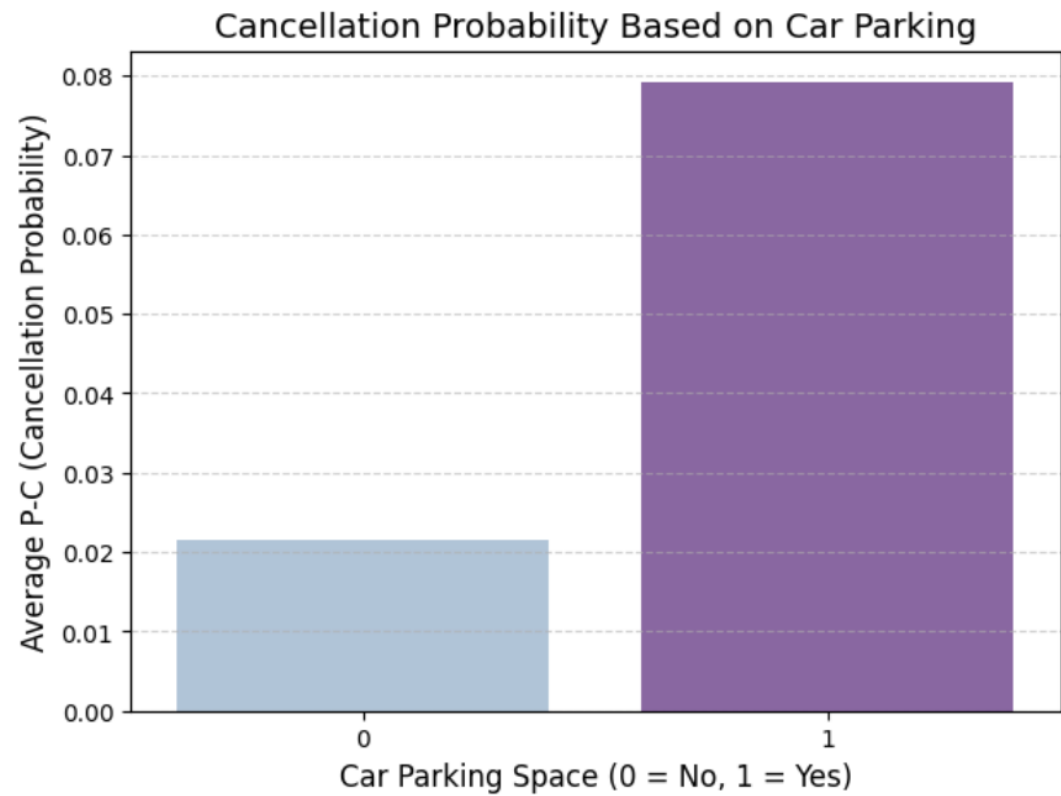
ROOM TYPE WITH MOST SPECIAL REQUESTS



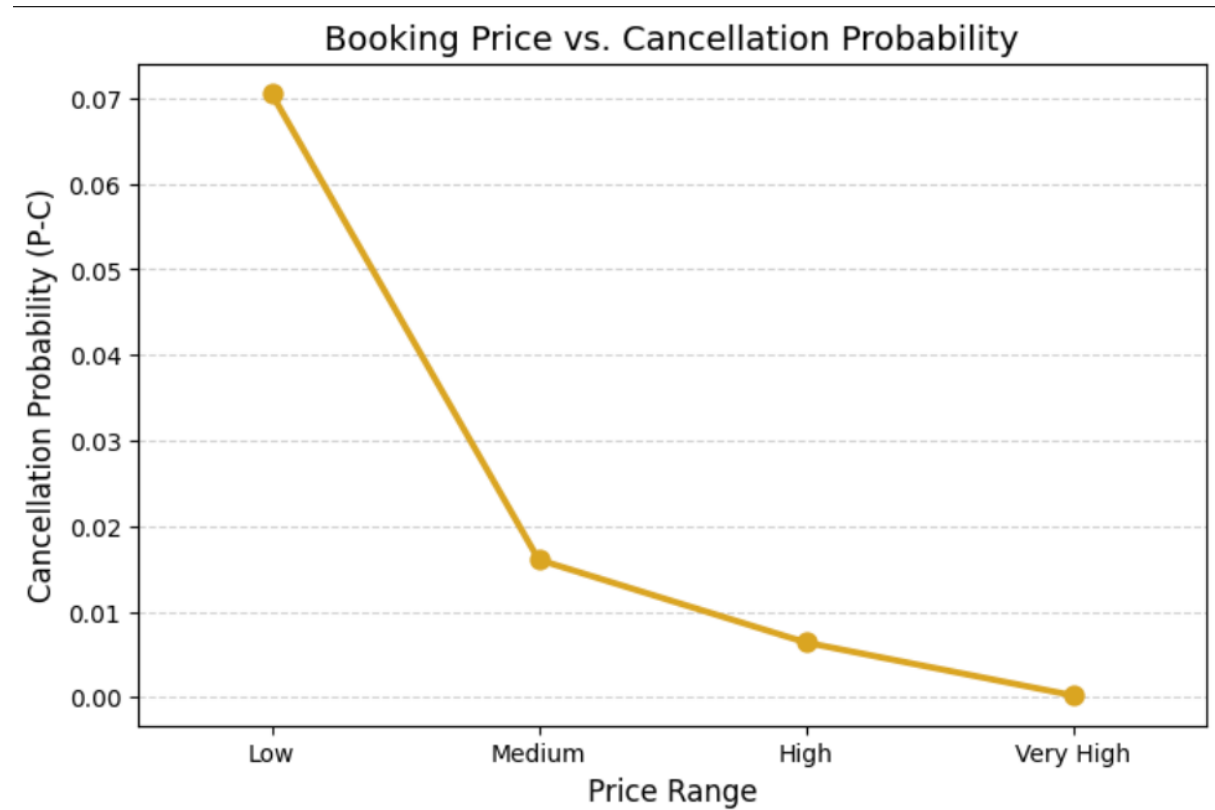
FAMILY IMPACT ON BOOKING PRICE



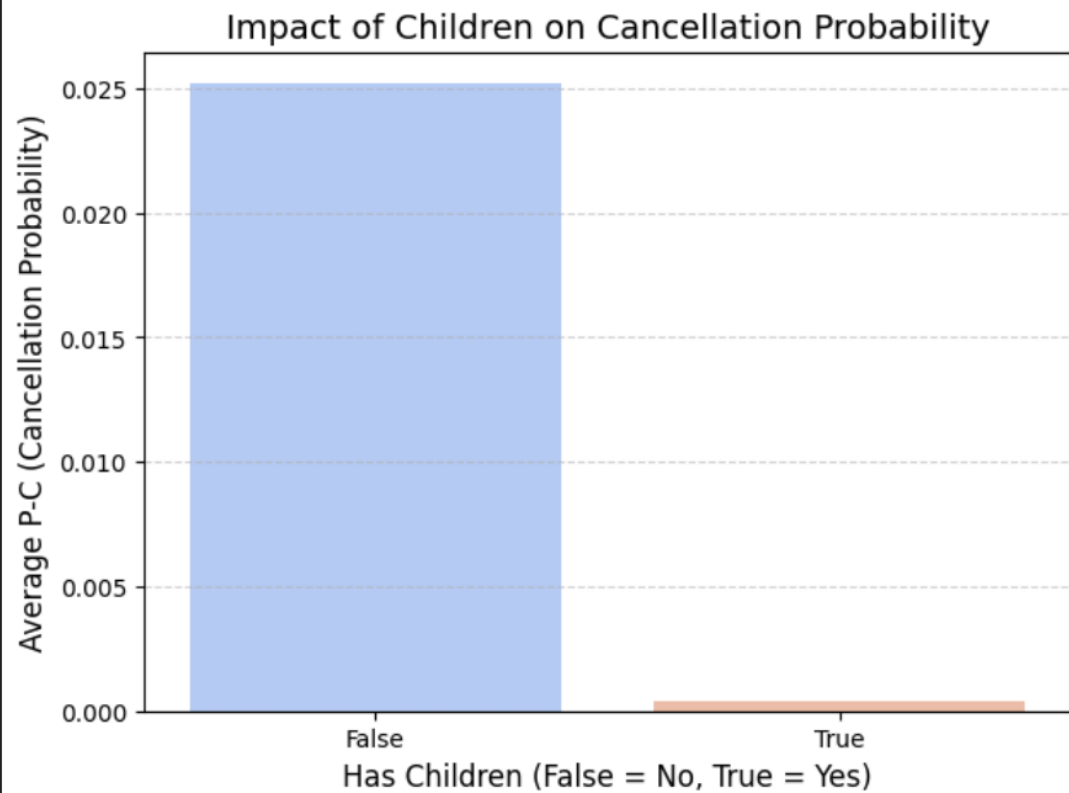
CAR PARKING & CANCELLATION PROBABILITY



HIGH PRICE & CANCELLATION RATE



CHILDREN & CANCELLATION PROBABILITY



WEEKEND NIGHTS & CANCELLATIONS

