Investigate Business Hotel Using Data Visualization

CREATED BY

ARIF BUDIMAN
Data Scientist

Let's Connect!

✉ abudiman.work@gmail.com

in https://www.linkedin.com/in/arfbudiman/

○ https://github.com/abudiman357

Data scientist with expertise in data wrangling, preprocessing, visualization, interpreted complex data to drive business decisions and generated machine learning as predictive engine. Skilled in using Python, SQL, Tableau, multiple Python libraries. Experienced in collaborating and working as part of a team on data analytics projects. Individual with a growth mindset and strong interpersonal skills who consistently delivers high-quality work in professional.

Overview

Analyzing business performance is a crucial task for any company. In this instance, I will delve into the hospitality sector, specifically focusing on understanding the behavior of hotel customers when making reservations. Additionally, the project aims to investigate the correlation between customer behavior and the cancellation rate of hotel bookings. Presenting the findings through visualized data is essential, as it enhances comprehension and persuasiveness.

Environment
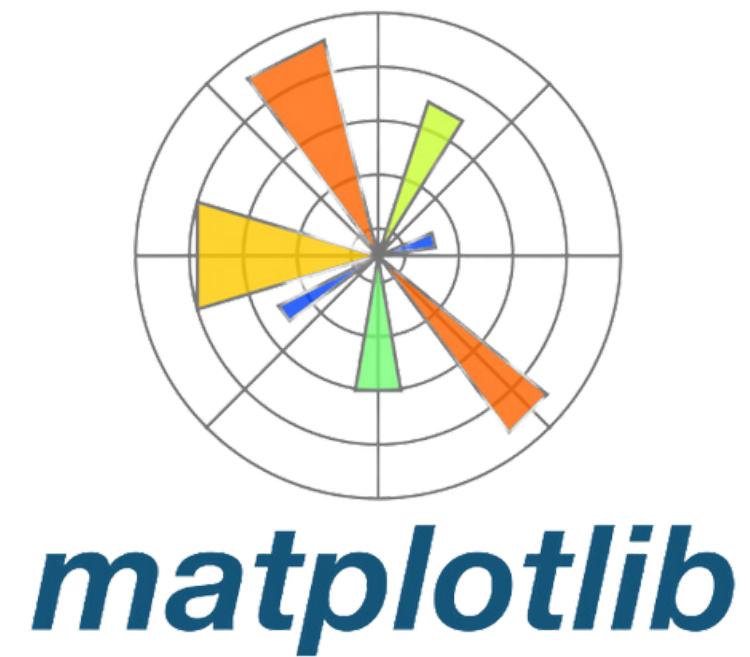
Dataset

	Dtype	Missing value	Missing value (%)	N-Unique	Unique value
company	float64	112593	94.306893	352	[nan, 110.0, 113.0, 270.0, 178.0, 240.0, 154.0...
agent	float64	16340	13.686238	333	[nan, 304.0, 240.0, 303.0, 15.0, 241.0, 8.0, 2...
city	object	488	0.408744	177	[Kota Denpasar, Kabupaten Bangka, Kabupaten Sl...
children	float64	4	0.003350	5	[0.0, 1.0, 2.0, 10.0, 3.0, nan]
adr	float64	0	0.000000	8879	[0.0, 75.0, 98.0, 107.0, 103.0, 82.0, 105.5, 1...
is_canceled	int64	0	0.000000	2	[0, 1]
stays_in_weekend_nights	int64	0	0.000000	17	[0, 1, 2, 4, 3, 6, 13, 8, 5, 7, 12, 9, 16, 18,...
stays_in_weekdays_nights	int64	0	0.000000	35	[0, 1, 2, 3, 4, 5, 10, 11, 8, 6, 7, 15, 9, 12,...
reservation_status	object	0	0.000000	3	[Check-Out, Canceled, No-Show]
required_car_parking_spaces	int64	0	0.000000	5	[0, 1, 2, 8, 3]
previous_cancellations	int64	0	0.000000	15	[0, 1, 2, 3, 26, 25, 14, 4, 24, 19, 5, 21, 6, ...
previous_bookings_not_canceled	int64	0	0.000000	73	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,...
meal	object	0	0.000000	5	[Breakfast, Full Board, Dinner, No Meal, Undef...
market_segment	object	0	0.000000	8	[Direct, Corporate, Online TA, Offline TA/TO, ...
lead_time	int64	0	0.000000	479	[342, 737, 7, 13, 14, 0, 9, 85, 75, 23, 35, 68...
is_repeated_guest	int64	0	0.000000	2	[0, 1]
deposit_type	object	0	0.000000	3	[No Deposit, Refundable, Non Refund]
hotel	object	0	0.000000	2	[Resort Hotel, City Hotel]
distribution_channel	object	0	0.000000	5	[Direct, Corporate, TA/TO, Undefined, GDS]
adults	int64	0	0.000000	14	[2, 1, 3, 4, 40, 26, 50, 27, 55, 0, 20, 6, 5, 10]
days_in_waiting_list	int64	0	0.000000	128	[0, 50, 47, 65, 122, 75, 101, 150, 125, 14, 60...
customer_type	object	0	0.000000	4	[Personal, Contract, Family, Bussiness]
booking_changes	int64	0	0.000000	21	[3, 4, 0, 1, 2, 5, 17, 6, 8, 7, 10, 16, 9, 13,...
babies	int64	0	0.000000	5	[0, 1, 2, 10, 9]
arrival_date_year	int64	0	0.000000	3	[2017, 2018, 2019]
arrival_date_week_number	int64	0	0.000000	53	[27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 3...
arrival_date_month	object	0	0.000000	12	[September, October, November, December, Janua...
arrival_date_day_of_month	int64	0	0.000000	31	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...

Content

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Shape

119.390 data rows, 29 features

Dtype

Float64 (4 features), int64 (16 features), object (9 features)

Missing Value

company: 94%, agent: 13%, city: 0,4%, children: 0,003%

Detail Codes

Data Preprocessing

Handling Missing Values

1. **Company (94.30% Missing):** Fill with 0 since NaN indicates bookings that didn't come from corporate.
2. **Agent (13.68% Missing):** Fill with 0 because NaN indicates bookings that didn't use any agent channel.
3. **City (0.40% Missing):** Fill with 'unknown' because the city of origin is unknown.
4. **Children (0.003% Missing):** Fill with 0 because its median is 0, indicating that the customer didn't have any children with them.

Correcting Value

Change 'undefined' to be 'No Meal' in meal feature

Drop Unnecesarry Data

Drop data rows that has 0 guests and 0 stay_durations

Detail Codes

Monthly Booking Analysis By Hotel Type

Average Number of Hotel Bookings per Month by Hotel Type

The peak season for hotel bookings is June-Augustus, coinciding with summer holidays. City hotel highest in July (11.18%), resort hotel in June (9.82%).
There is another surge in December (10.32% city, 9.61% resort), during Christmas and New Year holidays.
Both hotels have less customer during not holiday season (Jan-Mar)
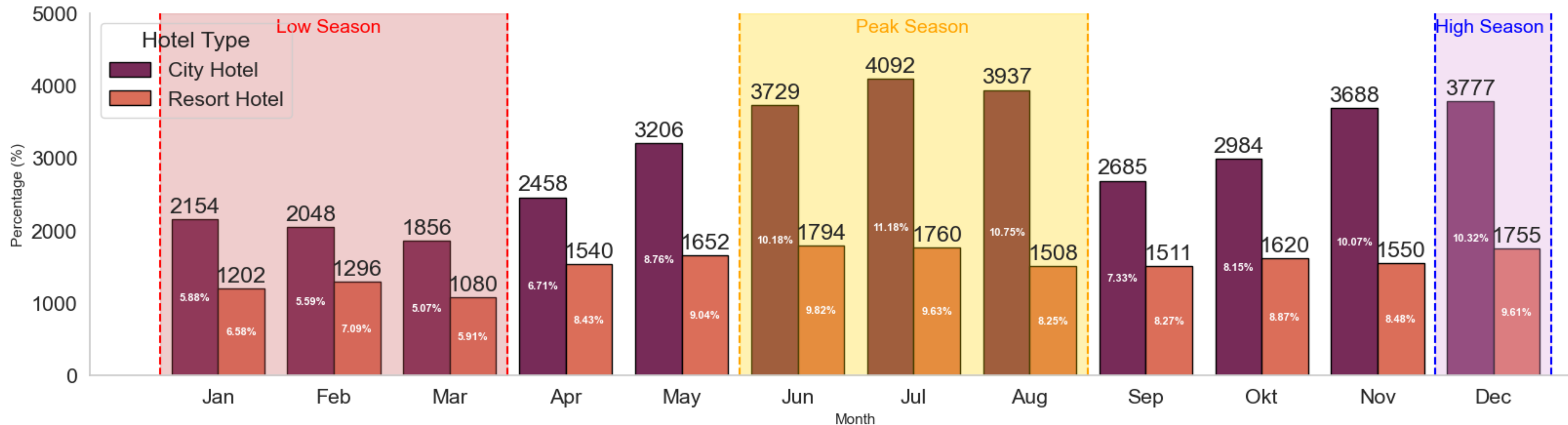
- There are three seasons: Low, Peak, and High.
- City hotels experience more volatile demand, varying significantly across different seasons.
- In contrast, resort hotels exhibit a comparatively more stable demand pattern

Detail Codes

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

Positive Trend on Total Nights and Percentage of Booking Canceled

The longer the total nights booked, the higher the percentage of booking being canceled, also City Hotel has more steeper trend compared to Resort Hotel
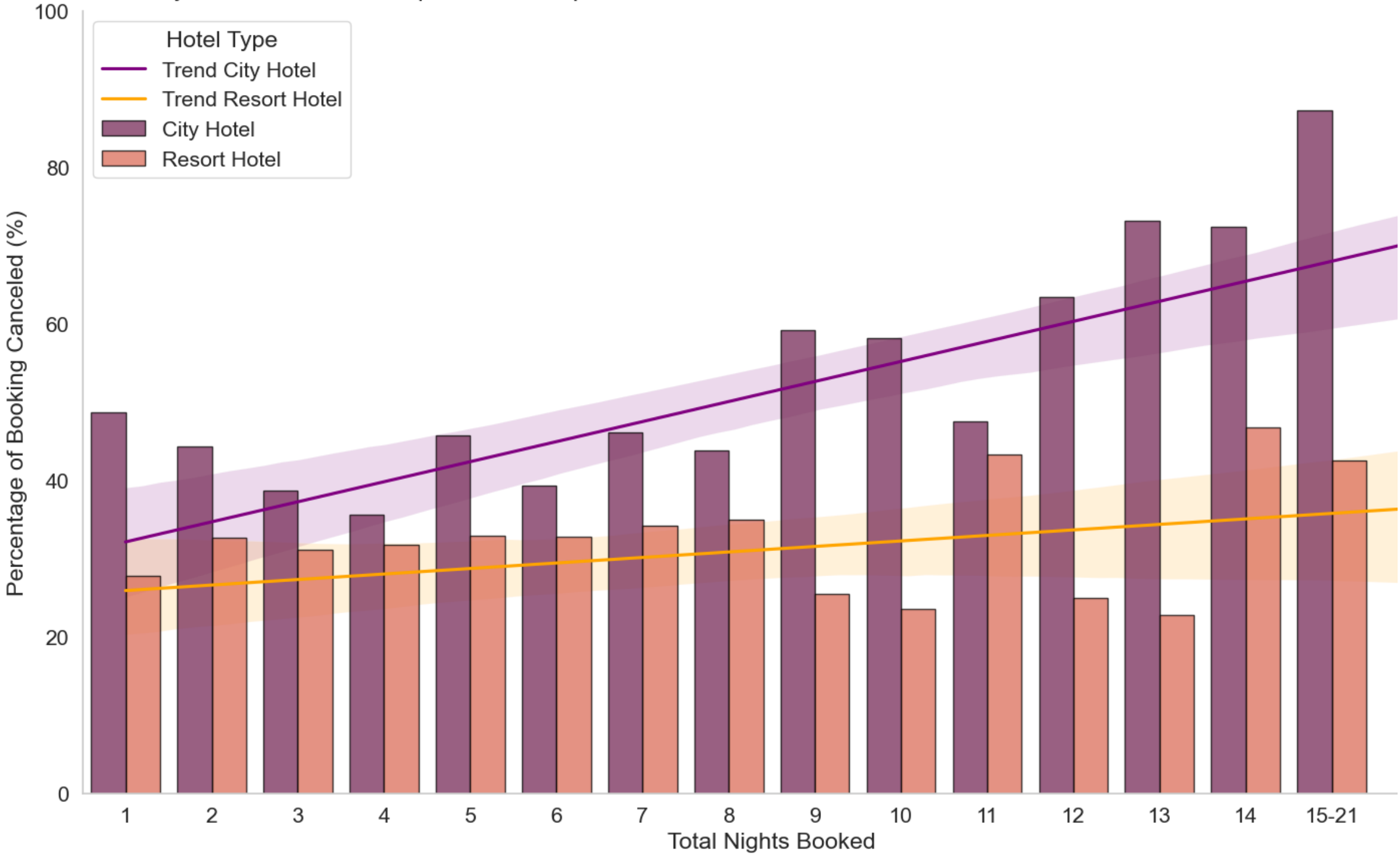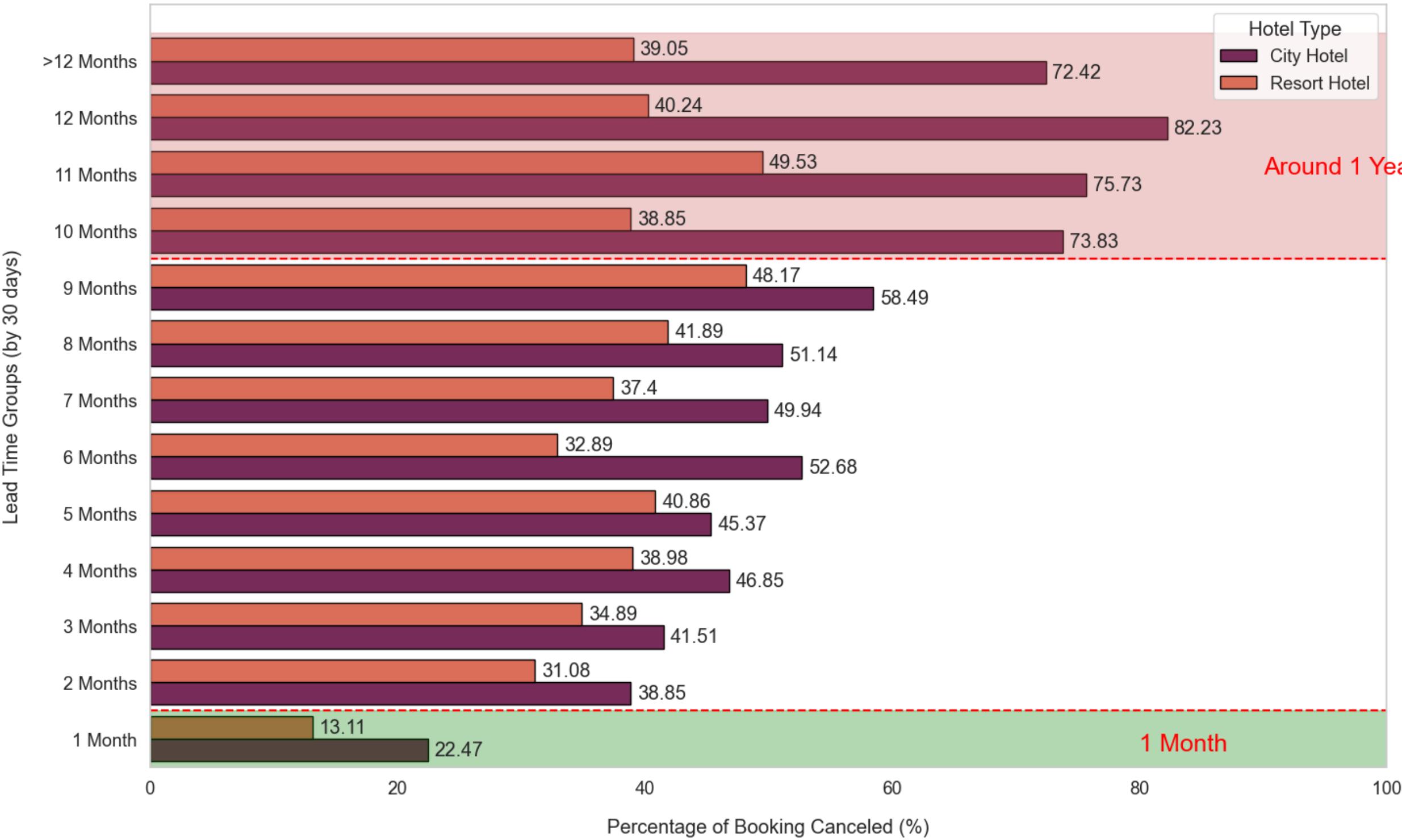
- **There is a positive trend in the cancellation rate of hotel bookings based on stay duration and hotel types.**
- **For city hotels, cancellation rates increase significantly to above 80% for stay longer than 2 weeks**
- **For resort hotels, cancellation rates increase by more than 40% for stays longer than 2 weeks.**
- **The observed trend may be attributed to human errors during hotel reservations.**
- **Long vacation plans might be canceled due to urgent reasons that customers couldn't decline, contributing to the higher cancellation rates.**

Detail Codes

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

Both Hotels Have the Lowest Canceled Ratio within 1 Month of Leads Time

Resort Hotel has quite stagnant (in around 40%) canceled ratio, while City Hotel has quite high canceled ratio (above 60%) with leads time around 1 year

Lead Time Groups (by 30 days):

Lead Time	Resort Hotel	City Hotel
>12 Months	39.05	72.42
12 Months	40.24	82.23
11 Months	49.53	75.73
10 Months	38.85	73.83
9 Months	48.17	58.49
8 Months	41.89	51.14
7 Months	37.4	49.94
6 Months	32.89	52.68
5 Months	40.86	45.37
4 Months	38.98	46.85
3 Months	34.89	41.51
2 Months	31.08	38.85
1 Month	13.11	22.47

Hotel Type
- City Hotel
- Resort Hotel

Around 1 Year

1 Month

Percentage of Booking Canceled (%)

- The lowest booking cancellation rate is for bookings that have a waiting time of less than 30 days and applies to both types of hotels
- Significant growth of cancellation rate for city hotels each month from around 20% to around 80%, also it happens for resort hotels each month from around 10% to around 40%.

Detail Codes

Thank You

Contact:

abudiman.work@gmail.com   https://www.linkedin.com/in/arfbudiman/   https://github.com/abudiman357