INVESTIGATE

# HOTEL BUSINESS

USING DATA VISUALIZATION

# About Dataset

**11930** rows of data

**29** Columns

| Column | Data Type |
|---|---|
| hotel | object |
| is_canceled | int64 |
| lead_time | int64 |
| arrival_date_year | int64 |
| arrival_date_month | object |
| arrival_date_week_number | int64 |
| arrival_date_day_of_month | int64 |
| stays_in_weekend_nights | int64 |
| stays_in_weekdays_nights | int64 |
| adults | int64 |

| Column | Data Type |
|---|---|
| children | float64 |
| babies | int46 |
| meal | object |
| city | object |
| market_segment | object |
| distribution_channel | object |
| is_repeated_guest | int64 |
| previous_cancellations | int64 |
| previous_bookings_not_canceled | int64 |
| booking_changes | int64 |

| Column | Data Type |
|---|---|
| deposit_type | object |
| agent | float64 |
| company | float64 |
| days_in_waiting_list | int64 |
| customer_type | object |
| adr | float64 |
| required_car_parking_spaces | int64 |
| total_of_special_requests | int64 |
| reservation_status | object |

See more here

# Pre Processing

## Checks for null values

| | |
|---|---|
| hotel | 0 |
| is_canceled | 0 |
| lead_time | 0 |
| arrival_date_year | 0 |
| arrival_date_month | 0 |
| arrival_date_week_number | 0 |
| arrival_date_day_of_month | 0 |
| stays_in_weekend_nights | 0 |
| stays_in_weekdays_nights | 0 |
| adults | 0 |
| children | 4 |
| babies | 0 |
| meal | 0 |
| city | 488 |
| market_segment | 0 |
| distribution_channel | 0 |
| is_repeated_guest | 0 |
| previous_cancellations | 0 |
| previous_bookings_not_canceled | 0 |
| booking_changes | 0 |
| deposit_type | 0 |
| agent | 16340 |
| company | 112593 |
| days_in_waiting_list | 0 |
| customer_type | 0 |
| adr | 0 |
| required_car_parking_spaces | 0 |
| total_of_special_requests | 0 |
| reservation_status | 0 |

## Fill in null values

1. Replace the null value in the city column with the city that appears most often
2. Fill in the null values children, agent and company with 0
3. Replace 'undefined with 'no meal'

See more here

# Comparison of the Number of Hotel Bookings

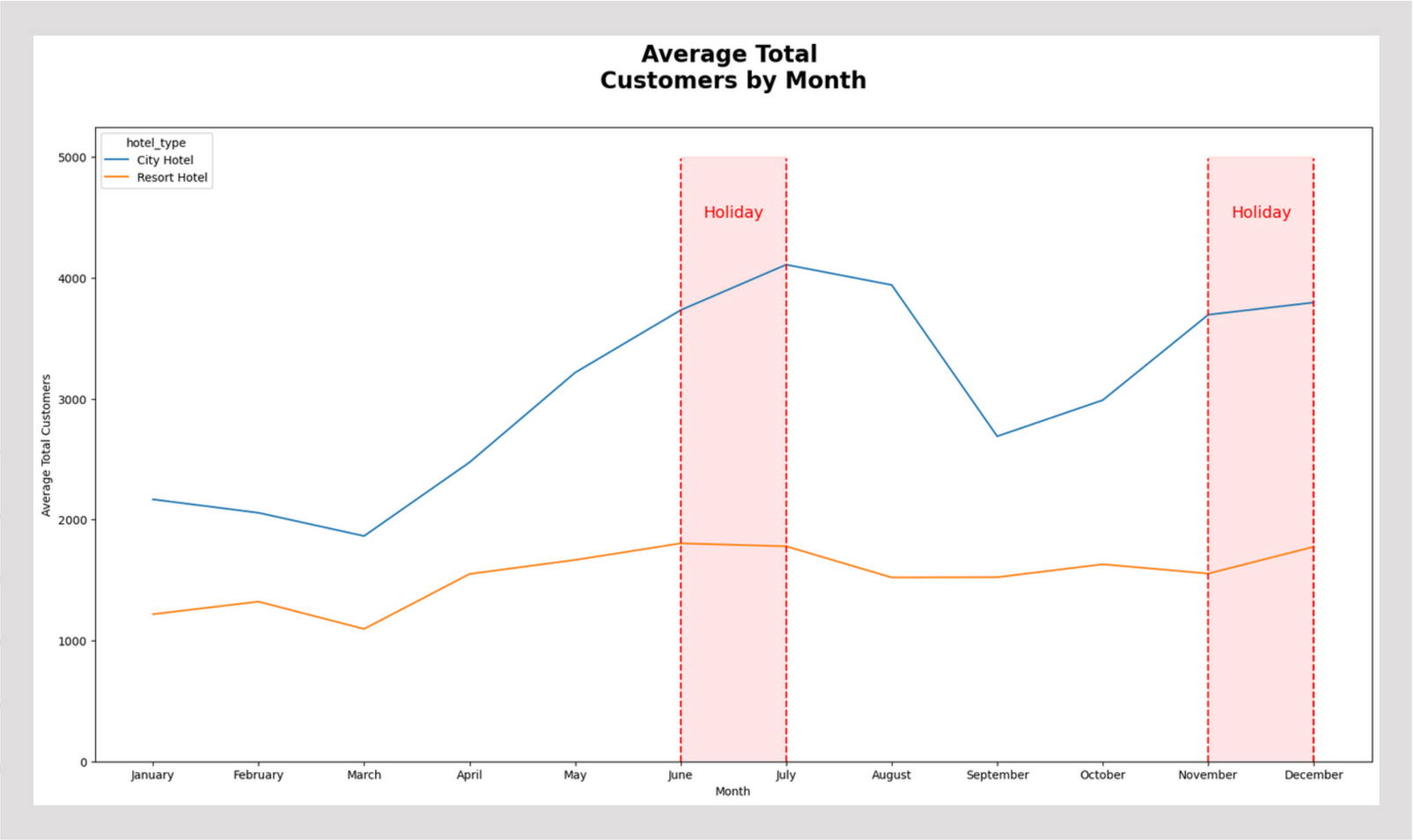
```
# Membuat tabel agregat yang menunjukan perbandingan jumlah
pemesanan hotel setiap bulannya berdasarkan jenis hotel
df_tabel = df_prep.groupby(['hotel','arrival_date_month'])
['arrival_date_year'].agg(['nunique','count']).reset_index()
df_tabel.columns = ['hotel_type', 'arrival_month',
'unique_year', 'count_guest']

# Mengurutkan data berdasarkan bulan
df_tabel["arrival_month"] =
pd.Categorical(df_tabel["arrival_month"],
               categories=["January", "February", "March",
"April", "May", "June", "July",
                   "August", "September", "October",
"November", "December"],
               ordered=True)
df_tabel = df_tabel.sort_values(["arrival_month"],
ignore_index=True)

df_tabel['average_booking'] =
round(df_tabel['count_guest']/df_tabel['unique_year'])
df_tabel.head()
```

| hotel_type | arrival_month | unique_year | count_guest | average_booking |
|---|---|---|---|---|
| City Hotel | January | 2 | 4336 | 2168.0 |
| Resort Hotel | January | 2 | 2435 | 1218.0 |
| City Hotel | February | 2 | 4114 | 2057.0 |
| Resort Hotel | February | 2 | 2645 | 1322.0 |
| City Hotel | March | 2 | 3730 | 1865.0 |

See more here



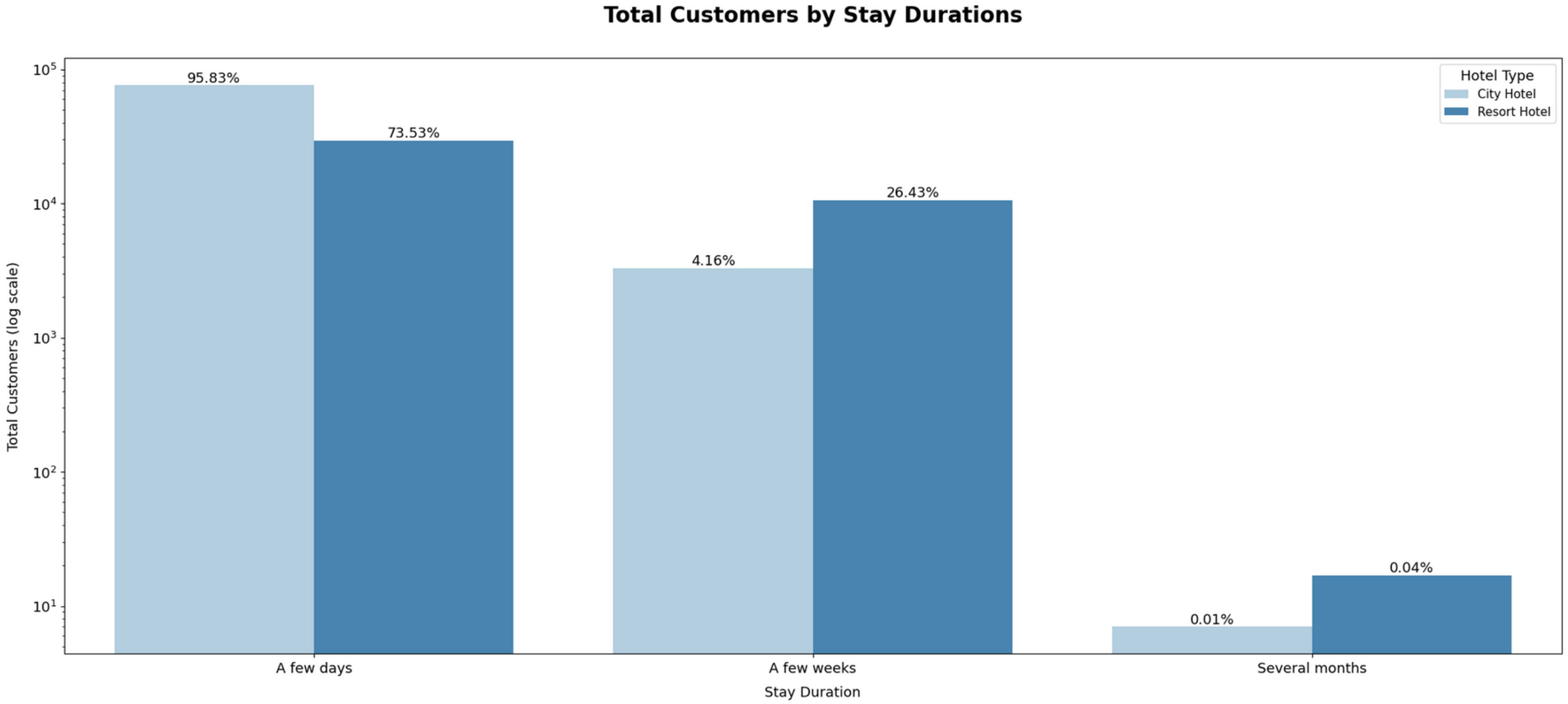

# Insights

- City Hotel bookings are always higher than Resort Hotelsl
- The highest increase in hotel bookings occurred in June-July and November-December, this was because those months were school holidays.
- The lowest bookings occurred in March, both for City Hotels and Resort Hotels

See more here

# Duration of Customer Stay

Group by is carried out based on hotel type and duration of stay by aggregating the total count of customers who stay overnight and grouping the duration of stay into 3 categories

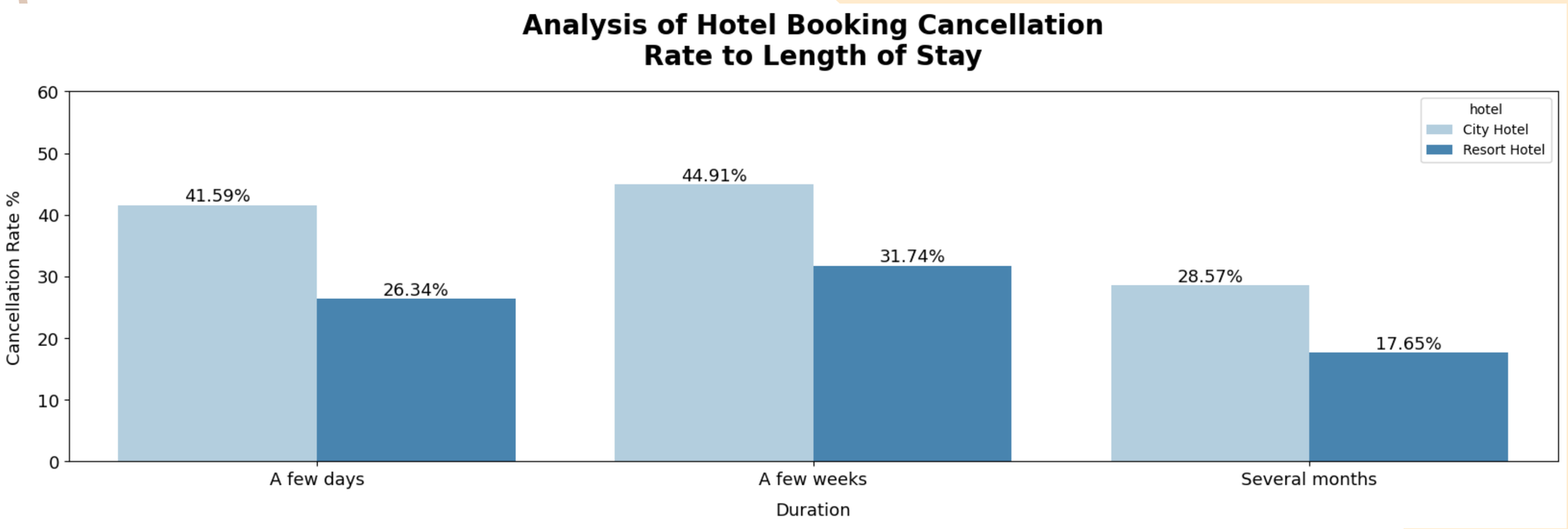



See more here

# Ratio of Hotel Cancellations

| hotel | duration | not canceled | canceled | ratio |
|---|---|---|---|---|
| City Hotel | A few days | 44405 | 31618 | 41.590045 |
| City Hotel | A few weeks | 1818 | 1482 | 44.909091 |
| City Hotel | Several months | 5 | 2 | 28.571429 |
| Resort Hotel | A few days | 21697 | 7758 | 26.338482 |
| Resort Hotel | A few weeks | 7227 | 3361 | 31.743483 |
| Resort Hotel | Several months | 14 | 3 | 17.647059 |

See more here

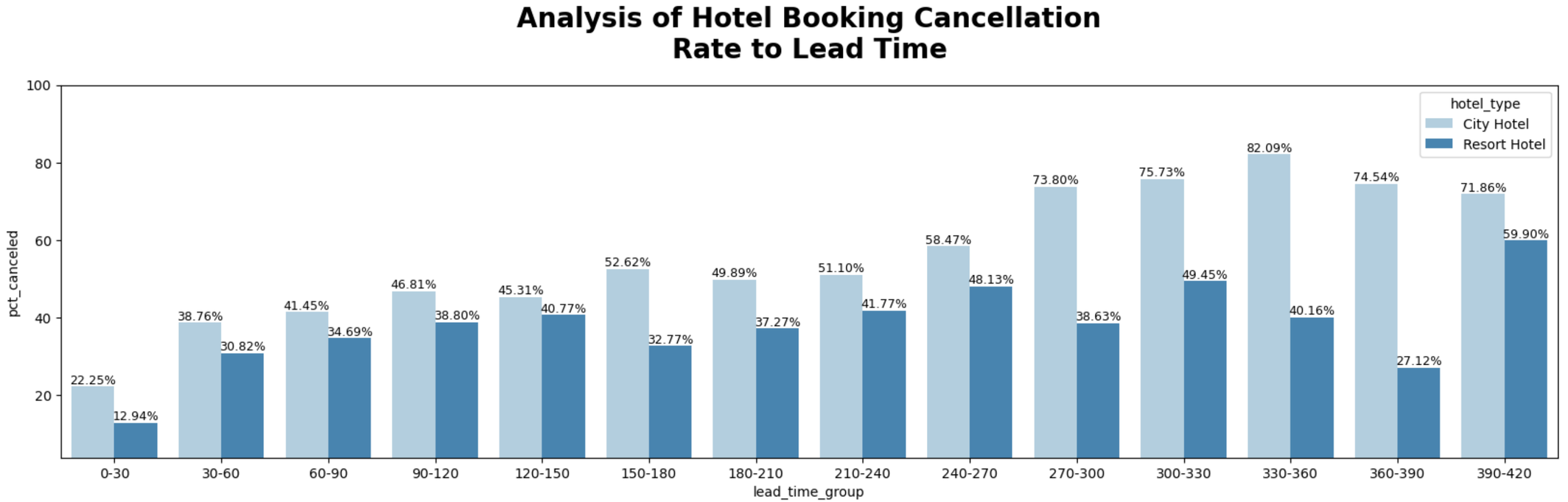



# Insights

- The duration of a customer's stay is usually 2-3 days
- The highest cancellation ratio for City Hotels and Resort hotels is in the 'A few weeks' category (Less than 31 days)
- Even though the most cancellations were made in the 'A few days' category, this was offset by the number of customers who did not cancel orders
- It is necessary to analyze more deeply why the 'A few weeks' category has the highest ratio of cancels

See more here

# ‘Lead Time’ Column Categorization




```
# Categorize the lead_time column
df3['lead_times'] = np.where(df3['lead_time'] > 400, 401, df3['lead_time'])
bins = list(np.arange(0,max(df3['lead_times'])+30, 30))
labels = ["{0}-{1}".format(str(bins[x]), str(bins[x+1])) for x in range(len(bins[:-1]))]
df3['lead_time_group'] = pd.cut(df3['lead_times'],bins=bins, include_lowest=True, labels=labels)

# perform data aggregation to calculate the number of bookings
df3_book = df3.groupby(['hotel','lead_time_group', 'is_canceled'])['hotel'].count().reset_index(name='num_bookings')
df3_pivot = df3_book.pivot_table(index=['hotel','lead_time_group'], columns='is_canceled', values='num_bookings').reset_index()
df3_pivot.columns = ['hotel_type','lead_time_group', 'not_canceled', 'canceled']
df3_pivot['all_bookings'] = df3_pivot['not_canceled']+df3_pivot['canceled']
df3_pivot['pct_canceled'] = df3_pivot['canceled']*100.0/(df3_pivot['all_bookings'])
```

## Insights

- Cancel for City Hotels is always bigger than for Resort Hotels
- Cancels above 50% for City Hotels occur when the lead time is above 210 days
- Cancel for Resort Hotels tends to be stable and does not show a significant graph
- The lowest cancellations for City Hotels and Resort hotels are in the lead time of 0-30, this could be because the time to book is close so there is no time to look for other accommodation

See more here

# Business
## Recommendation

Berdasarkan durasi menginap, diketahui bahwa tamu resort hotel cenderung menginap lebih lama. Informasi ini dapat dimanfaatkan untuk strategi pemasaran, seperti program diskon untuk sewa kamar dalam beberapa hari dan penyelenggaraan acara jangka panjang seperti festival tahun baru atau program liburan lebaran. Hal ini bertujuan untuk menarik perhatian calon pelanggan yang berencana berlibur. Kerjasama dengan tempat wisata juga dapat menjadi strategi yang efektif. Dengan menawarkan paket penawaran menarik bagi tamu yang menginap. Dengan memanfaatkan tamu yang menginap dalam jangka waktu yang lebih lama, hotel dapat menawarkan fasilitas berbayar, seperti layanan sewa kendaraan.

Dengan karakteristik tamu city hotel yang cenderung menginap singkat karena keperluan pekerjaan atau perjalanan bisnis, disarankan untuk mengimplementasikan program fee tambahan yang memungkinkan fleksibilitas waktu check-in dan check-out. Program ini dapat menarik perhatian pelanggan yang memiliki keterbatasan waktu akibat pekerjaan namun menghargai fleksibilitas dalam jadwal menginap. Selain memberikan kenyamanan, program ini juga dapat menjadi sumber tambahan pemasukan bagi hotel.

Berdasarkan tingkat pembatalan reservasi yang signifikan dalam satu bulan menuju check-in, disarankan untuk melibatkan analisis lebih lanjut dengan menambahkan fitur pemilihan alasan pembatalan oleh pelanggan. Hal ini akan memungkinkan untuk memberikan rekomendasi yang lebih tepat guna mengurangi tingkat pembatalan.

Selain itu, direkomendasikan pula adanya pengingat kepada pelanggan, seperti seminggu sebelum dan sehari sebelum tanggal check-in. Tujuan dari pengingat ini adalah untuk meningkatkan keterlibatan pelanggan serta memastikan tidak adanya kendala yang mungkin dihadapi oleh pelanggan dalam proses check-in. Implementasi pengingat ini diharapkan dapat meningkatkan retensi pelanggan dan mengurangi tingkat pembatalan reservasi.