Data Engineer challenge

We'd like you to submit your exercise as a Git repository so that we can see how you elaborated on your solution. Preferably, share a private **GitHub** repository with the user **s4l-techrec**. You have a week to complete it so please take your time. There are no extra points for doing it quickly. We would prefer that you take your time and deliver a solution that you're proud of.

Problem definition

We have a data set with booking information, and we'd like you to help us better understand our business. You can download the data set here:
https://d3ony5rmo3oa45.cloudfront.net/careers/data_take_home_bookings.zip

This is what we want to know about the data:
1. What are the most popular guest origin countries?
2. What is the room price per night?
3. How does the previous price vary throughout the year?
4. What is the distribution of guests per month?
5. For how long do guests book their stays?

The solution should provide different dimensions, when necessary, to help the business better understand the data. Take into account that cancellations and accommodation types are important business concepts.

Requirements

- Include a README.md file in your repository explaining *at least* how to run your solution.

Nice to have

- Besides the numerical information in the answers, a visual representation would be great.
- Try to commit as often as necessary while providing some value or functionality in each commit message.

Notes

- Use the technology stack you feel more comfortable with.
- Assume Docker will be available in the machine running your solution.

Data set information

Here you can find a definition of all the columns included in the provided CSV file.

- hotel: Defines the accommodation type, either a Hotel or an Apartment
- is_canceled: Indicates whether the booking was canceled (1) or not (0)
- booking_to_arrival_time: The difference in days between the booking date and the arrival date
- arrival_date_year: Year of the arrival date
- arrival_date_month: Month of the arrival date
- arrival_date_week_number: Week of the year of the arrival date
- arrival_date_day_of_month: Month of the arrival date
- stays_in_weekend_nights: The number of nights the guest stayed during the weekend
- stays_in_week_nights: The number of weekday nights the guest stayed
- adults: Total number of adults
- children: Total number of children
- babies: Total number of babies
- board: Represents the meal type. Undefined/SC: no meal package; BB: Bed & Breakfast; HB: Half-board; FB: Full board
- country: Guest country of origin in ISO alpha-3 format
- market_segment: TA: Travel Agents; TO: Tour Operators
- acquisition_channel: TA: Travel Agents; TO: Tour Operators
- is_repeated_guest: Whether the guest is a returning customer
- previous_cancellations: Number of cancellations before the current booking
- previous_bookings_not_canceled: Number of bookings before the current
- reserved_room_type: Booked room type
- assigned_room_type: Assigned room type
- booking_changes: Total number of changes to the booking since its creation
- deposit_type: Whether the customer has guaranteed the booking with a deposit
- agent: Related travel agency that managed the booking
- Company: Related company that managed the booking
- days_in_waiting_list: Days between booking request and confirmation
- adr: Booking cost divided by the number of nights. (Average Daily Rate)
- parking_lot: Car parking spaces.
- total_of_special_requests: Total number of special requests
- reservation_status: Booking status. Canceled, Check-Out, or No-Show
- reservation_status_date: Date at which reservation_status last changed.