carta

v2.2

Data Scientist Take Home - Product

INTRODUCTION

Airbnb, Inc. is a company headquartered in San Francisco that operates an online marketplace and hospitality service. Members can use the service to arrange or offer lodging. The company does not own any of its real estate listings; as a broker, it receives commissions from every booking.

In this exercise, we'd like you to answer some business questions with SQL and Python/R to showcase your technical and analytical skills. You'll be working with a modified version of the Bayarea Dataset from Inside Airbnb.

Data

calendar.csv

- listing_id: unique identifier for a listing
- date: calendar date for a listing
- available: if the listing is available on that calendar date
- price: listing price on that calendar date

listings.csv

- id: unique identifier for a listing
- host_id: id of listing's host
- host_name: name of host
- host_since: host's starting date
- host_response_rate: message response rate of host
- host_acceptance_rate: rate of booking requests accepted by host
- host_is_super_host: if the host is a super host (superhosts are experienced hosts who provide a shining example for other hosts, and extraordinary experiences for their guests.)
- host_verifications: how host being verified
- host_identity_verified: if host's identity is verified

- neighbourhood: listing's neighbourhood
- room_type: listing's room type
- accommodates: listing's space
- bedrooms: # of bedrooms
- beds: # of beds
- amenities: listing's amenities that guest can access
- room_type: listing's room type
- minimum_nights: min nights of booking
- maximum_nights: max nights of booking
- has_availability: if the listing is available in 365 days
- availability_365: how many days of the listing is available in 365 days
- number_of_reviews: # of reviews of the listing
- review_scores_rating: listing's review score
- instant_bookable: if the listing can be instantly booked
- reviews_per_month: # of reviews of the listing per month

Grading

Your submission will be evaluated on three axes:

1. Are the answers **correct**?
2. Is the code **clear** and **readable**?
3. The **methodology** on how you approach the problems and the **analytical thinking** behind the answers.

Question

Please provide the **SQL statement** for these questions:

1. How many different listings were there on 2021-01-10? By how many different hosts?

2. What are the top 10 most expensive (price-wise) listings?

3. Which listing has the lowest **Calendar** vacancy rate?

4. What 5 listings have had the most frequent day-over-day price increases?

Please answer the following questions with **Python**:

1. Please review the integrity of the data. Do you notice any data anomalies? If so, please describe them.

2. Calculate the average listing price by calendar day. Based on the results, do you notice any performance trends of listing price over time? If so, please describe them and give short explanations on why you think this could happen?

3. Suppose Airbnb charges 10% commission fee on each booked listing. Please use the data to give business operation recommendations on how to increase the commission revenue.

SUBMISSION

Please submit your answers to the unique link provided in the email, splitting into two parts.

1. **PDF** Format for SQL Statement Questions
2. **Python notebook** for Python Questions

The naming convention on the title should be your Initials- Carta Science (i.e. XY-CartaScience), and please refrain from using your full name in the assignment if possible.

Once you've done so we'll get back to you in a few days with next steps. Thank you for your time!