

Seazone Challenge

Junior Data Scientist

This challenge goal is to test your coding skills, logical thinking and analytical capabilities. This is based on the work done in Seazone, with real data. You will be judged both on code structure (variable names, git commits, function abstraction, etc.) and code efficiency (How long does it take to run? How scalable is it? How much memory will your solution require?). All the code you produce will be tested by our technology team, so you must be clear in your README about all the steps needed to produce the intended results.

We also require a report in english covering the problems tackled, the solutions found, and some brief feedback about the challenge. Feel free to expose other possible solutions you considered and possible future improvements to your code. Both the report (in PDF) and your code must be published in a publicly accessible git repository.

We look forward to seeing your results!

1. Data description

- table listings
 - Código - The listing code
 - Localização - The listing location code
 - Comissão - Seazone's commission on a listing's revenue
 - Cama Casal - Amount of double beds in the listing
 - Cama Solteiro - Amount of single beds in the listing
 - Cama Queen - Amount of queen sized beds in the listing
 - Cama King - Amount of king sized beds in the listing
 - Sofá Cama solteiro - Amount of sofa-beds in the listing
 - Travesseiros - Amount of pillows in the listing
 - Banheiros - Amount of bathrooms in the listing
 - Taxa de limpeza - Listing's cleaning fee
 - Capacidade - Listing's maximum guest capacity
 - Hotel - 'Sim' means its a listing inside a hotel, 'Não' means a listing outside a hotel
 - Data inicial do contrato - Listing's start date with Seazone.
 - Status - current listing status. "Ativo" means it is still available for renting.
 - Tipo - Determines if it's inside a hotel, Apartment or house.

- table daily revenue
 - listing - Listing's code
 - date - Booking date
 - last_offered_price - Listing last offered price for this date
 - Occupancy - 1 for booked or blocked, 0 for still available.
 - Revenue - The listing's revenue for this date
 - blocked - 1 for a blocked date (no revenue and no reservation made, despite not being available)
 - creation_date - the creation date for a night's booking.

In Seazone, we sell nightly stays in several apartments, houses and hotels. We use OTAs (Online Travel Agencies) like Airbnb, Booking and Expedia, as well as our own website, to find guests interested in booking a short stay - usually lasting less than a month per reservation. Each property is advertised as a "Listing". For this analysis, you can use the words "listing" and "property" interchangeably.

Guests usually book rooms a few days before they intend to check-in. We call this lead-time the *reservation advance*, and it is one key metric of our business.

On the **table "Listings"** you can find information about our apartments, houses and hotels, such as number of rooms, maximum guest capacity, and location.

We classify our listings into 5 tiers of quality - SIM<JR<SUP<TOP<MASTER, in ascending order. This means that the listings classified as SIM will be the cheapest, while the ones classified as MASTER will be the most expensive. This, along with the number of rooms of each listing, is noted in the "*Categoria*" column. On houses, we also add a "HOU" prefix to the category. A listing tagged with MASTER3Q is, for example, a very luxurious apartment with 3 bedrooms, while one classified as HOUSIM2Q is a simple house with 2 bedrooms.

We also assign a code for each location. JUR represents Jurerê, and ING Ingleses, both are neighborhoods in Florianópolis. Some codes can represent entire cities, such as ITA for Itajaí, or a specific hotel - JBV and ILC are hotels in Jurerê.

On the **table "Daily Revenue"**, you can find information about the revenue of each listing, on a day to day basis. Note that, in this table, we have two columns formatted as date, the one labeled as "*date*" refers to the reservation date, that is, for instance, the check in date. The other, labeled as "*creation_date*" refers to the date the reservation was created/sold. Therefore, the "*creation_date*" column of a given row should be always lesser or equal to the "*date*" column of the same row.

Sometimes, a given listing may be unavailable for a few dates - usually for maintenance. We say that those dates are "blocked", and we use a boolean on the **"Daily Revenue" table** to indicate when this happens. A blocked listing date always has 0 revenue.

Some of our past data for prices were not stored in this database. For this reason, we can see that the column "*last_offered_price*" is often 0 for unsold nights. Also, bear in mind that a single reservation can be composed of many consecutive nights.

2. Data analysis

We can use this dataset to answer a few questions.

- What is the expected price and revenue for a listing tagged as JUR MASTER 2Q in march?
- What is Seazone's expected revenue for 2022? Why?
- How many reservations should we expect to sell per day? Why?
- At what time of the year should we expect to have sold 10% of our new year's nights? And 50%? And 80%?
 - How can this information be useful for pricing our listings?

Optional (solve if you finished the other questions with spare time):

- On the impact of the COVID-19 pandemic:
 - Can we estimate Seazone's revenue loss due to the pandemic? How?
 - Has the industry recovered?
 - If Yes, when can we state that we came back from pre-pandemic levels of sales/revenue?
 - If No, when do you expect this recovery to happen?

To answer these questions, feel free to use Excel, Google Sheets, Python, R, Power BI, Datastudio and/or Metabase. Remember to justify your stack choices in your PDF report.
Do not use Jupyter Notebook.

We recommend the use of the following python libraries: Pandas/Dask, Dash, Matplotlib, seaborn and Scikit-learn.

Remember: Deliver all your work in a git repository! Make sure to make it public, and to add both your code **and** the PDF report to it. Then, you can send the link to the repository via email.