Módulo 5: Prueba técnica A

Desarrollo en lenguaje Python

Año de realización: 2021

Índice

1. Introduction
2. Dataset
3. Exploratory Dataset Analysis
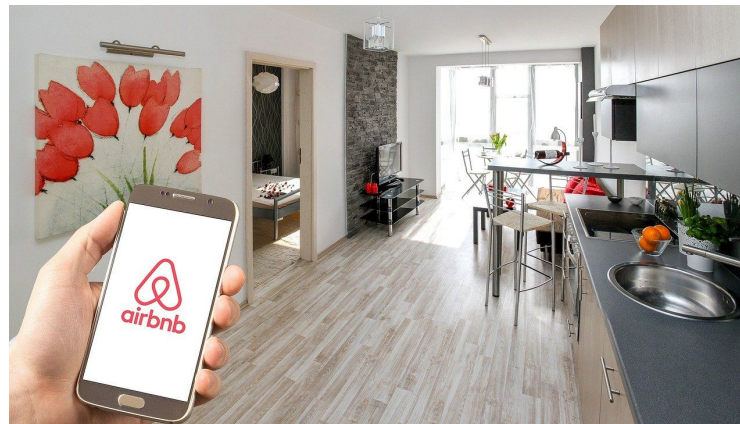4. Model Building
5. Submission

Introduction
Prueba técnica final Módulo 5 - A

Introduction

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

The purpose of the present exercise is to find some answers about activity generated by hosts and reach some kind of predictions about it.

Dataset

Prueba técnica final Módulo 5 - A

Dataset

You'll be using a dataset from Airbnb 2019 listing and metrics in New York City. You can download the dataset here:

New York City 2019 Airbnb

This dataset has around 49000 observations in it with 16 columns and it is a mix between categorical and numeric values

This public dataset is part of Airbnb, and the original source can be found on this website.

Columns

The fields in this dataset are:

- id: lodging id in listing
- name: lodging name in listing
- host_id: id of the host
- host_name: name of the host
- neighbourhood_group: location
- neighbourhood: area
- latitude
- longitude
- room_type: listing space type
- price: price in dollars
- minimun_nights: amount of nights minimun
- number_of_reviews

- last_review
- reviews_per_month
- calculated_host_listings_count: amount of listing (lodgings) per host
- availability_365: number of days when listing is available for booking

Exploratory Data Analysis

Prueba técnica final Módulo 5 - A

Tasks

1. Present the code and methods for acquiring the data. Loading the data into appropriate format for analysis. Explain the process and results

2. Explore the data by analyzing its statistics and visualizing the values of features and correlations between different features. Explain the process and the results

3. In case you use a scatterplot based on latitude and longitude coordinates, you may need this image to get a better understanding map plot:
   https://drive.google.com/file/d/1Fu6L8pDt8ujRhIJbubEiIag9PEd_2H_E/view?usp=sharing

Model Building
Prueba técnica final Módulo 5 - B

Task

The purpose of this task is to predict the price of NYC Airbnb rentals based on the data provided and any external dataset(s) with relevant information. Two main goals are suggested:

1. Users should submit a csv file with each listing from the data set and the model-predicted price:

    id, price
    2539, 149
    2595, 225
    3647, 150

    …. ,…

2. A solution with low root-mean-squared error (RMSE) based on cross-validation that can be reproduced and interpreted is ideal

Submission

Prueba técnica final Módulo 5 - B

Submission

Prepare a presentation with 4 or 5 slides explaining the process and conclusions.

Please, send files, graphs, images and everything you consider important to Carmen Bartolomé:

carmenbvg@gmail.com

If you prefer send the submission as a github link from your own repository, it's ok too.