

Data Description

csv file notes

The data.zip contains following a .csv and a folder with photos:

(The train dataset contain 100 houses with booking data of year 2017 and 2018. The test dataset contains the same houses' data of year 2019 but is not included in the data.zip)

1. `train.csv`:

- "house_pk" int (categorical)

Should not be used for training but is there to match images. Each image is named by its house_pk

- price

this is the target to be predicted based on the other features (including images). you should try to make the model generalizable

so that it works for houses not in the dataset.

categorical:

- agency_id

time (date)

- date_in

continuous:

- dis_water_real: Real distance measured between the accommodation and sea/river/lake

- dis_shopping

- no_bedrooms

- max_persons

- house_size

- land_size

- build_year

- renovation_year

ordinal:

- agency_rating

categorical (boolean)

- apartment

- indoor_pool

- spa

- internet

- pets_allowed

- water_view

- fire_stove