StreetEasy Dataset

StreetEasy Dataset

Machine Learning Fundamentals & Data Science Path

StreetEasy is New York City's leading real estate marketplace — from studios to high-rises, Brooklyn Heights to Harlem.

We have partnered with the StreetEasy Research team for the Multiple Linear Regression (MLR) lesson, and you will be working with a **.csv** file that contains a sample of 5,000 rentals listings in `Manhattan`, `Brooklyn`, and `Queens`. You'll find the correlations between several features and the rent, build/evaluate a MLR model, and use the model to present interesting findings:

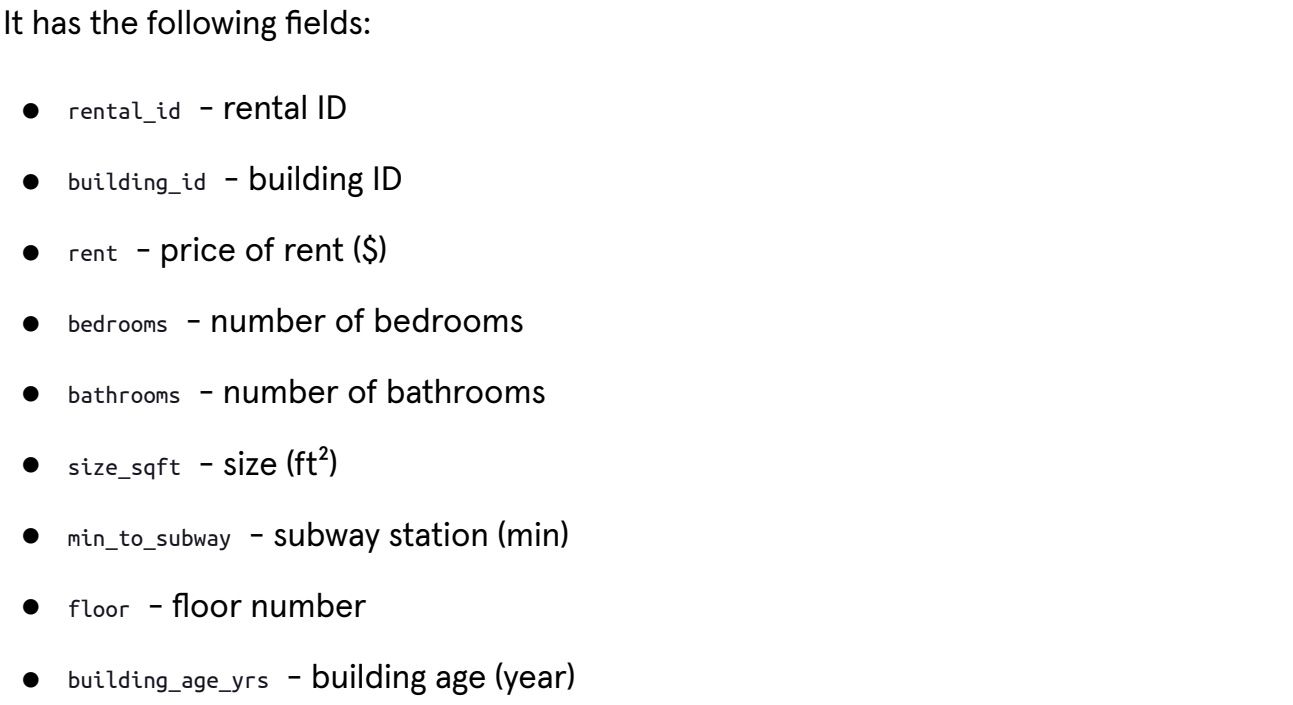
- "Does having a washer/dryer in unit increase the price of rent?"
- "How costly is living by a subway station in Brooklyn/Queens?"
- And most importantly, "Is a tenant over or underpaying?"

Samples Total	**5000**
Dimensionality	20
Features	text & real, positive

It has the following fields:

- `rental_id` - rental ID
- `building_id` - building ID
- `rent` - price of rent ($)
- `bedrooms` - number of bedrooms
- `bathrooms` - number of bathrooms
- `size_sqft` - size (ft^2)
- `min_to_subway` - subway station (min)
- `floor` - floor number
- `building_age_yrs` - building age (year)
- `no_fee` - has no broker fee (0/1)
- `has_roofdeck` - has roof deck (0/1)
- `has_washer_dryer` - has in-unit washer/dryer (0/1)
- `has_doorman` - has doorman (0/1)
- `has_elevator` - has elevator (0/1)
- `has_dishwasher` - has dishwasher (0/1)
- `has_patio` - has patio (0/1)
- `has_gym` - has gym (0/1)
- `neighborhood` - neighborhood (ex: Greenpoint)
- `submarket` - submarket (ex: North Brooklyn)
- `borough` - borough (ex: Brooklyn)

To understand the data better, take a look at the apartments on StreetEasy: www.streeteasy.com/rentals.



Thank you StreetEasy for this partnership and especially:

- Grant Long, Sr. Economist
- Lauren Riefflin, Sr. Marketing Manager

If you would like to follow along this lesson off-platform (locally on your computer), you can download the **.csv** file from our GitHub [download].

Next Get Unstuck