Two Sigma Connect: Rental Listing Inquiries

RentHop makes apartment search smarter by using data to sort rental listings by quality. But while looking for the perfect apartment is difficult enough, structuring and making sense of all available real estate data programmatically is even harder.

The task is to predict the number of inquiries a new listing receives based on the listing’s creation date and other features. Doing so will help RentHop better handle fraud control, identify potential listing quality issues, and allow owners and agents to better understand renters’ needs and preferences.

We will be able to predict how popular an apartment rental listing is based on the listing content like text description, photos, number of bedrooms, price, etc. The data comes from [renthop.com](https://www.renthop.com/), an apartment listing website. These apartments are located in New York City.

The target variable, **Interest\_Level**, is defined by the number of inquiries a listing has in the duration that the listing was live on the site.

Data Fields

* bathrooms: number of bathrooms
* bedrooms: number of bathrooms
* building\_id
* created
* description
* display\_address
* features: a list of features about this apartment
* latitude
* listing\_id
* longitude
* manager\_id
* Photos: a list of photo links. You are welcome to download the pictures yourselves from renthop's site, but they are the same as imgs.zip.
* price: in USD
* street\_address
* interest\_level: this is the target variable. It has 3 categories: 'high', 'medium', 'low'

The data contains numerical, text and categorical variables. We perform exploratory data analysis to describe the different variables via different plots and diagrams and find the distributions of the data.

After performing the exploratory data analysis, we perform logistic regression on the data.