

Dataset

El dataset està compost per propietats de l'estat d'Illinois (majoritàriament la Chicago metropolitan area), en l'últim any (Setembre 23 - Setembre 24).

El [train.csv](#) conté els 10 primers mesos (Setembre 23 - Juliol 24) i el [test.csv](#) conté els últims 2 (Agost 24 i Setembre 24).

Per més explicació sobre els fields del .csv podeu consultar el [Fields Description](#)

Us animem a complementar el dataset amb qualsevol informació open source que pugueu trobar, per exemple: Parcs, Census, Escoles, etc. Tot el que pugui ajudar-vos a enriquir el model és benvingut.

Recordeu que està completament prohibit pujar el dataset a qualsevol plataforma online.

Challenge

Volem que ens creeu un model de predicció de preus, conegut en la indústria com un AVM (automated valuation model).

És un problema recurrent que afronten els agents immobiliaris, donat un històric de dades, han d'aconseguir fer la millor estimació per la que es pot comprar / vendre el pis amb les característiques que els ha arribat.

Entrega

Volem que ens envieu un .csv basat en el test.csv (**no el train.csv**), compost per dos columnes:

- **Listing.ListingId:** Identificador únic de la propietat venduda.
- **Listing.Price.ClosePrice:** La vostra predicció del preu al que s'hauria de vendre l'immoble.

L'entrega final serà diumenge, però al llarg del dia d'avui us deixarem enviar alguns intents per que aneu evaluant com es comporta el model. No volem obrir-ho per intents infinits ja que no volem que us focalitzeu en intentar completament millorar l'accuracy, i més en la comprensió i anàlisis de les dades.

Per fer aquests intents, apropau-vos a l'stand de Restb.ai i allà us donarem el resultat i us intentarem guiar en el que puguem.

Tingueu en compte que el .csv ha de contenir tots els Listing.ListingId del test.csv i tenir una predicció per cadascun d'ells, d'altre manera l'entrega no serà vàlida.

Avaluació

La mètrica que farem servir per determinar el millor comportament del model serà: [Mean Squared Error](#) (MSE) i [Mean Absolute Error](#) (MAE). Tingueu en compte per això que aquest no serà l'únic factor a validar. Això ens servirà de llindar, però la millor mètrica no serà necessàriament la guanyadora,

Volem que us centreu en fer un anàlisis profund de les dades i que ens justifiqueu molt bé qualsevol decisió que prengueu (bones i dolentes). Tota decisió que prengueu per acabar aconseguint el model final ha d'estar justificada amb dades, **no volem** que prengueu una decisió basada exclusivament en la vostra intuïció.