

Dataset:

Dades d'immobles en lloguer a la ciutat de Barcelona al web Fotocasa.es al octubre del 2020

Descripció

El dataset generat recull dades dels pisos que es lloguen a la ciutat de Barcelona a una data determinada, extretes de la web de fotocasa.es. A hores d'ara, webs com Fotocasa ofereixen [visualització de dades obertes](#), però es centren bàsicament en mostrar els preus mitjans dels lloguers de cada comunitat, província, districte o barri de l'Estat Espanyol.

Per tal d'accedir a un conjunt de dades més ric, on podem trobar diverses característiques de cada lloguer més enllà del preu i localització, hem portat a terme un procés de scraping a la web de Fotocasa. Posteriorment, hem inclòs les dades de preus mitjans de lloguers per barris i districtes oferts per l'Agència de l'Habitatge de la Generalitat de Catalunya a cada ítem del dataset. Això ens permetrà contrastar els preus dels lloguers a Fotocasa amb una altra font d'informació de preus, en aquest cas provinent de l'administració pública.

Imatge identificativa <Pol>

Context

El conjunt de dades extret correspon als pisos que es lloguen a la ciutat de Barcelona al mes d'octubre de 2020 a la web fotocasa.es. Per cada lloguer, s'indica també el preu mitjà del barri i districte segons els preus mitjans de lloguers per barris i districtes de la ciutat de Barcelona que posa a disposició l'Agència de l'Habitatge de la Generalitat de Catalunya a la seva [web](#). Tot plegat ens permet comparar el preu ofertat a Fotocasa contra els preus mitjans dels que disposa l'administració pública, a més d'altres característiques de la oferta de lloguer.

Contingut

El dataset inclou els següents camps:

- **address:** l'adreça de l'habitatge, on es pot especificar carrer, barriada i districte
- **bathrooms:** nombre de banys de l'habitatge
- **building_type:** tipus d'habitatge, que habitualment és *Flat*.
- **building_subtype:** subtipus d'habitatge (*Flat*, *Duplex*, *Apartment*, ...)
- **conservation_state:** valor numèric que indica l'estat de conservació de l'habitatge.
- **date:** data de recollida de la informació
- **discount:** import de descompte sobre el preu original
- **floor_elevator:** camp booleà que indica si l'habitatge disposa o no d'ascensor
- **is_new_construction:** camp booleà que indica si l'habitatge és de nova construcció o no
- **link:** enllaç a la web de Fotocasa amb el detall de l'habitatge
- **price:** preu de l'habitatge en euros
- **realestate:** nom de l'agència que publica la oferta de lloguer
- **realestate_id:** identificador a Fotocasa de l'agència que publica la oferta de lloguer
- **rooms:** nombre d'habitacions de l'habitatge
- **sqft_m2:** superfície en metres quadrats de l'habitatge
- **neighb_meanprice:** preu mitjà al 2019 dels habitatges de la barriada on s'oferta el lloguer, segons la informació de preus mitjans de lloguers que posa a disposició l'Agència de l'Habitatge de la Generalitat de Catalunya
- **district_meanprice:** preu mitjà al 2019 dels habitatges del districte on s'oferta el lloguer, segons la informació de preus mitjans de lloguers que posa a disposició l'Agència de l'Habitatge de la Generalitat de Catalunya

Agraïments

Els nostres agraïments van dirigits, en primera instància, al nostre professor, que ens ha orientat i aconsellat en la elaboració d'aquesta pràctica. També agraïm a Fotocasa que no ens ho hagi posat especialment difícil. Per últim, agraïm a la comunitat del software lliure per crear llibreries tan útils com Scrapy i a totes aquelles persones anònimes que comparteixen exemples i responen dubtes a la web.

Inspiració

El principal focus del *dataset* generat és recollir informació de la web d'ofertes de lloguer de Fotocasa, i de preus mitjans de lloguer per barri i districte de l'Agència de l'Habitatge de Catalunya, amb la finalitat de poder comparar els preus de cadascuna de les ofertes amb el suposat preu mitjà del barri i districte de l'habitatge.

Aquest conjunt de dades també habilita tota una sèrie de d'anàlisis d'estadística descriptiva que no es limiten als preus de lloguer d'habitatges a la web de Fotocasa, si no que també els podem portar a terme sobre les característiques d'aquests habitatges (nombre d'habitacions, nombre de banys, metres quadrats, etc.).

Per últim, aquesta informació també podria ser útil per tal de generar un model capaç de predir el preu d'un habitatge a la ciutat de Barcelona donades les seves característiques més rellevants.

Llicència

Aquest *dataset* es publica amb la llicència CC BY-NC-SA (Reconeixement-No comercial-Compartir igual). S'ha escollit per què es tracta d'una llicència oberta que exigeix reconeixement però que no permet un ús comercial, ni tan sols en les obres derivades, ja que aquestes s'han de compartir amb la mateixa llicència. Creiem que és el més adient donat que fer un ús comercial d'aquestes dades no és ètic ni probablement legal, ja que Fotocasa ho prohibeix explícitament:

El Usuario deberá utilizar los materiales, elementos e información a la que acceda a través de la utilización del Portal y de cada uno de los correspondientes Servicios únicamente para sus propias necesidades, obligándose a no realizar ni directa ni indirectamente una explotación

comercial ni de los servicios ni de los materiales, elementos e información obtenidos a través de los mismos. ¹

¹ Fotocasa. “Aviso Legal”. Consultat el 28/10/2020. [En línia <https://www.fotocasa.es/es/aviso-legal/in>]

Contribucions	Signa
Recerca prèvia	PBL, JMOB
Redacció de les respostes	PBL, JMOB
Desenvolupament codi	PBL, JMOB