

1. Context

Davant de l'escalada del índex de preu al consum (IPC), el qual afecte de manera generalitzada als preus de béns finals, i la repoblació de les zones més rurals degut a les noves tendències i a la normalització del teletreball, es pretén extreure un data set que contingui informació sobre el preu dels lloguers de habitatges de la web Habitacalia, així com les característiques que determinen aquest. Concretament, per tal de acotar l'estudi, s'analitzarà solament les ofertes de lloguer existents en la comarca del Vallès Occidental.

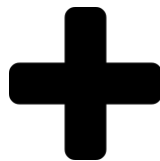
2. Títol

Habitatges de lloguer del Vallès Occidental

3. Descripció del data set

El data set 'Habitatges de lloguer del Vallès Occidental' pretén recollir el màxim d'informació sobre l'oferta d'habitatges de lloguer del Vallès Occidental que es mostren a la pàgina web https://www.habitacalia.com/alquiler-en-valles_occidental.htm. Aquest set de dades està compost de diverses variables tant de tipus 'string' com de tipus numèric per tal de descriure de la forma més completa possible les característiques principals de cada habitatge.

4. Representació gràfica



5. Contingut

El data set està compost per 9 variables les quals són:

- Títol: Títol o descripció del habitatge en qüestió que fa la funció de tag per cada registre.
- Ciutat: Ciutat on es troba l'habitatge
- Barri: Barri on es troba l'habitatge
- M2: metres quadrats de l'habitatge
- Habitacions: Nombre d'habitacions que conformen l'habitatge
- Lavabos : Nombre de lavabos que hi ha en l'habitatge
- Descripció breu: Petita descripció del immoble
- Preu m2: Preu del metre quadrat de l'habitatge
- Preu: preu de lloguer de l'habitatge (Variable target)

Aquestes dades han estat publicades en diversos períodes de temps difícils de recopilar i han estat extretes el 08/04/2022 mitjançant tècniques de web scrapping

6. Agraïments

El propietari de les dades presentades en aquest data set és HABITACLIA S.L., una empresa especialitzada en intermediar i posar en contacte propietaris/agències immobiliàries i compradors/llogaters a través de la seva pàgina web.

En aquest projecte, no es recullen ni s'exploten dades d'àmbit individual ni personal i en cap cas es vulnera el Reglament General de Protecció de Dades (RGPD) ni la Llei Orgànica 3/2018, de 5 de desembre, de protecció de Dades de Caràcter Personal y Garantia dels Drets Digitals.

Les dades publicades en aquest data set ja són d'àmbit públic i en cap cas es revelen dades sensibles com, per exemple, la geolocalització del habitatge, fotografies de l'interior, dades del propietari ni de l'agència immobiliària, etc.

Finalment, les dades capturades prenen com objectiu formar en conjunt, un data set destinat al estudi general i d'interès públic del preu de l'habitatge mitjançant l'exploració d'aquest i la construcció d'algorismes de Machine Learning, entre d'altres, però, en cap cas, l'objecte d'aquest estudi és l'explotació per obtenir rèdits econòmics.

7. Inspiració

La motivació principal d'aquest estudi és la de realitzar un informe on s'analitzin de forma exhaustiva i s'intentin respondre les següents preguntes:

- Quines són les variables que més influeixen al preu de l'habitatge?
- Quines són les ciutats i els barris on més alt és el preu del lloguer?
- De quina manera varia el preu per metre quadrat?

- Influeix el nombre d'habitacions i lavabos d'una forma determinant en el preu?

Posteriorment, també es pretén realitzar la construcció de algorismes de Machine Learning que proporcionin classificacions i regressions per tal de predir el preu de lloguer d'un immoble introduint les característiques desitjades.

8. Llicència

La llicència triada per aquest data set és la de Released Under CC BY-NC-SA 4.0 License ja que aquesta permet copiar, distribuir, adaptar i transformar-lo però no permet el ús d'aquestes dades per finalitats comercials.