

PRA1- Barcelona Rental Snapshot

1. Context

En els darrers dos anys, degut als efectes de la pandèmia del COVID s'han produït molts moviments en quan al mercat del lloguer. Els efectes negatius de la crisi econòmica han fet que moltíssimes persones no poguessin pagar els lloguers i per tant hi ha hagut molt de moviment en el mercat. Aquests moviments, han estat aprofitats molt cops per ens privats per especular amb el preu del habitatge. A fi de donar solució a aquest problema s'han establert unes limitacions ens els preus que pretenen controlar-ne l'evolució.

Les dades extretes a el nostre Dataset ens poden servir per detectar anomalies, possibles mals usos dels portals, també comprovar que es segueix la normativa de l'habitatge en quan a la regulació de preus i també poder calcular un preu a l'hora de crear un nou anunci.

Habitacalia és un dels principals portals webs per a l'anunci de oportunitats de comprar i lloguer. Actualment (en el moment de la redacció d'aquest document) hi ha 5 554 anuncis. Això ens podrà proporcionar una visió realista de la situació actual malgrat que hi ha altres portals que ens podrien també ser útils (ja que també tenen una molt alta popularitat)

El mateix portal ens proporciona informació bàsica i rellevant de tots els pisos en lloguer, com la zona, el nombre d'habitacions, banys i també la superfície i el preu a l'hora de realitzar les cerques. També ofereix plans Premium de posicionament dels anuncis.

2. Especificació de títol del Dataset

Un cop vist el motiu i la informació que pot ser extreta del nostre Dataset ens proposem establir el següent nom: "Barcelona Rental Snapshot"

Aquest nom es decideix ja que estem fent servir un portal web usat per molts usuaris que busquen pis per llogar i pretenem fer una foto de la situació actual.

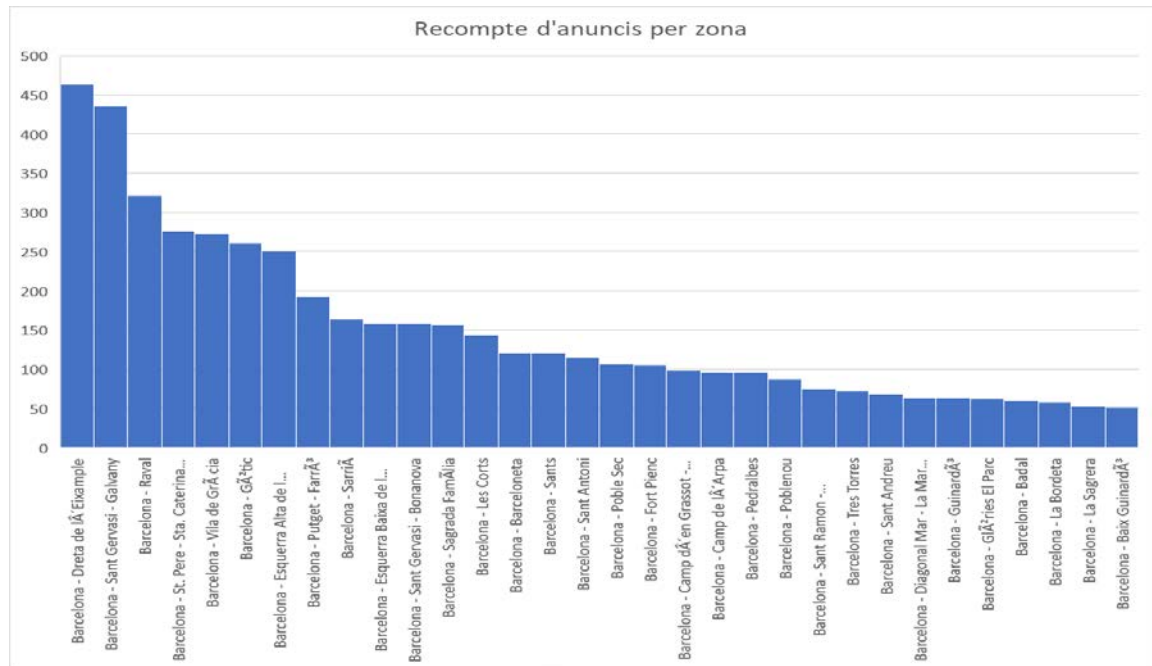
3. Descripció del Dataset

En el nostre Dataset tindrem a cada línia un anunci publicat a la web amb diferents atributs extrets directament des del web.

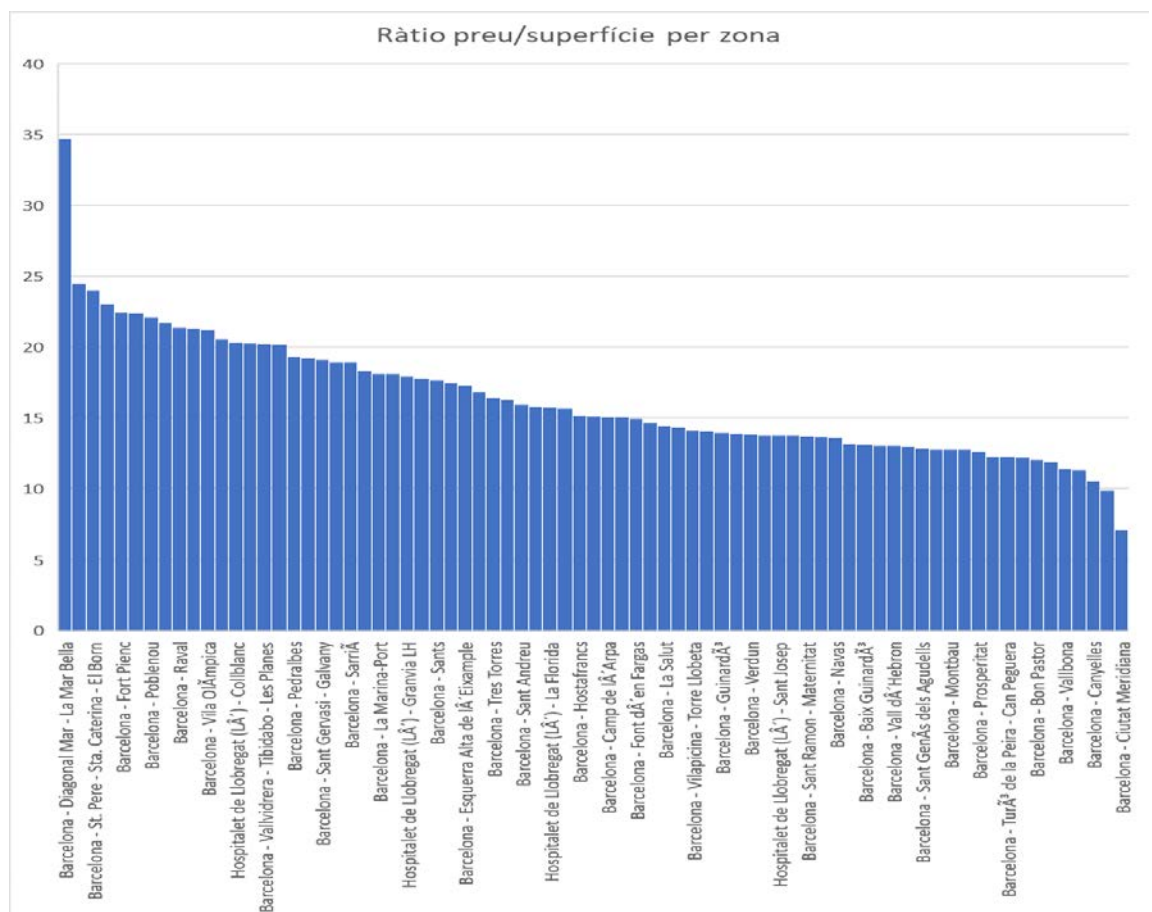
Aquestes dades s'han decidit extreure-les de les descripcions breus que ofereix el portal a l'hora de mostrar els resultat de la cerca. Atributs com la superfície, el nombre d'habitacions així com el preus seran de molta rellevància per poder extreure'n conclusions. En alguns anuncis algunes dades no són mostrades i per tant és possible que puguem tenir valors nuls o incorrectes.

4. Representació gràfica del Dataset

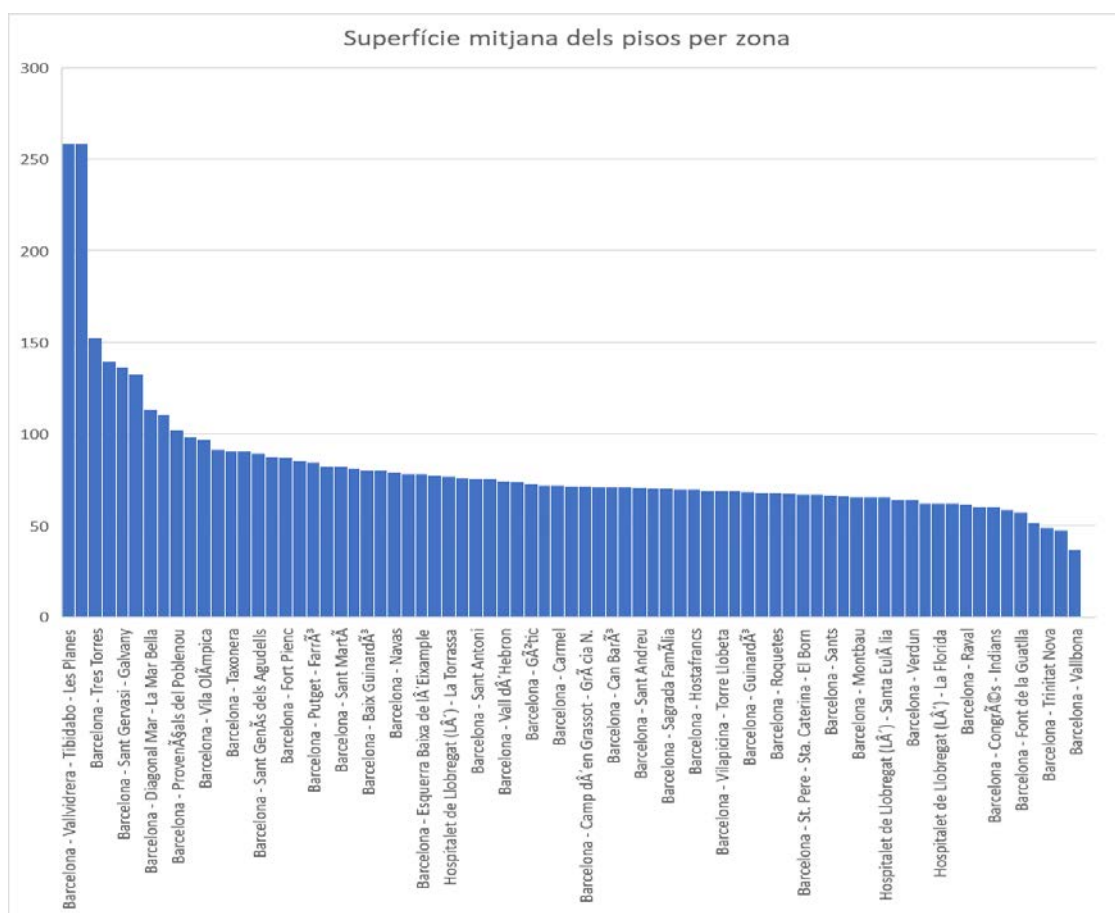
Amb unes quantes representacions gràfiques podem analitzar on es localitzen la majoria d'anuncis:



També podem veure el ràtio preu/superfície mig per cadascuna de les zones:



També podem veure la superfície mitjana dels immobles anunciats per zones.



5. Contingut del Dataset

A fi de poder descriure bé les propietats del nostre Dataset mostrem la següent taula:

- **barcelona_rents.csv**: Anuncis de pisos de lloguer a Barcelona presents en la web d'Habitacía a data de l'execució de l'script.

El fitxer creat té les següent característiques:

- Ve en format CSV (Comma separated values)
- Té headers.
- Fa servir “,” com a separadors de columnes.

Field name	Type	Example	Descripció	Observacions
location	Text	Barcelona - Esquerra Alta de l'Eixample	Nom de la zona de Barcelona on es troba el pis anunciat	Els noms de les zones són únics i per tant es poden agrupar de manera eficient i correcta. (és a dir no hi ha dos valors diferents que representin la mateixa zona)
surface	Num	105	Superfície en m2 del habitatge en lloguer.	Conté nombres enters
rooms	Num	3	Número d'habitacions del habitatge en lloguer	Conté nombres enters.
bathrooms	Num	1	Número d'habitacions del habitatge en lloguer	Conté nombres enters.
surface_price_rate	Num	16,22	Ràtio preu/superfície	És un valor numèric amb decimals. Pot contenir valors erronis, o degut a un error en la conversió o també degut a un valor incorrecte en origen.
premium?	Booelan	True	És l'anunciat Premium?	
days_from_last_update	Num	5	Dies transcorreguts des de la última actualització de l'anunci	Conté nombres enters.
price	Num	1150	Preus del immoble en lloguer	Conté nombres enters sense separador de milers

En total tenim 5656 registres.

6. Agraïments

En aquest projecte es presenta un recull de les dades ja publicades en un portal web en concret d'accés totalment obert i per tant sense haver incorregut en una infracció de revelació de dades confidencials o sensibles cap a la propietat del portal web en qüestió. Durant els darrers anys s'han realitzat diferents estudis de la situació dels lloguers a Barcelona, però moltes vegades les conclusions resultants no s'ajustaven a la realitat que la població podia percebre a l'hora de recórrer a portals generalistes com el analitzat en aquest estudi. Aquesta diferent percepció pot ser deguda a que una bona part dels acords de lloguer es fan fora d'aquests canals encara, i per tant no són públics. El codi emprat en aquest estudi és totalment propi, i les llibreries usades són de codi obert per la qual cosa tampoc s'ha vulnerat cap mena de propietat intel·lectual.

7. Inspiració

En aquest Dataset es pretén analitzar l'estat actual dels lloguers en un portal usat per molts usuaris.

Ens pot ser molt interessant a fi de poder analitzar la realitat de totes les persones que es troben en cerca de pis així com establir relacions entre les diferents zones, el preus i les característiques dels diferents habitatges anunciats.

Es pretén respondre preguntes com:

- S'està complint la normativa en quan al preu del lloguer
- A quina zona hi ha un major nombre d'anuncis, i per tant hi ha hagut més moviment?

8. Llicència

Per a aquest Dataset, donat que s'ha extret informació d'un portal públic s'ha decidit que la llicència serà:

- Released Under CC0: Public Domain License

9. Codi

El Codi es pot trobar al repositori:

https://github.com/Toni94bdn/UOC_TCVD_PRAC1

A fi d'executar el codi només cal assignar una nova URL segons la cerca a Habitacalia d'on es vulgui extreure les dades. En aquest cas en concret la URL d'entrada ja està escrita en el propi codi però pot ser canviada conveniència del usuari:

```
114 main("https://cataia.habitacalia.com/lloguer-barcelona.htm")
```

10.Dataset

El Dataset resultant de l'extracció s'ha publicat al portal Zenodo. Per accedir-hi es pot fer servir el següent enllaç:

<https://doi.org/10.5281/zenodo.5651993>

Contribucions	Signatura
Investigació prèvia	Toni Forcada
Redacció de les respostes	Toni Forcada
Desenvolupament del codi	Toni Forcada