

Sílvia Benet Porta

Marc Garrido Casas

Pràctica 1 (35% nota final)

1. Context. Explicar en quin context s'ha recol·lectat la informació. Explicar per què el lloc web triat proporciona aquesta informació.

L'objectiu de la búsqueda era recopilar informació de preus de vivenda de tots els pobles d'espanya que s'actualitzessin de forma constant per poder veure com es comporta la seva evolució. Hem triat la pàgina web : <https://www.trovimap.com/> perquè ens proporciona unes taules ben detallades del preu de la vivenda per cadascuna de les províncies i poblacions d'espanya. A més del preu €/m² ens proporciona la variació mensual, dels últims tres mesos, anual i una taula comparativa segons el tamany de la vivenda.

En el procés inicial de recerca hem intentat altres webs com per exemple idealista però es feia imposible extreure informació de la web. Ens retornava un error de fallada de comunicació amb el servidor, d'alguna forma, la web estava preparada contra la petició de paquets tal com ho fem amb python. Segur que d'alguna manera es podria burlar aquesta seguretat, però vam decidir buscar una altre pàgina web menys protegida.

2. Definir un títol pel dataset. Triar un títol que sigui descriptiu.

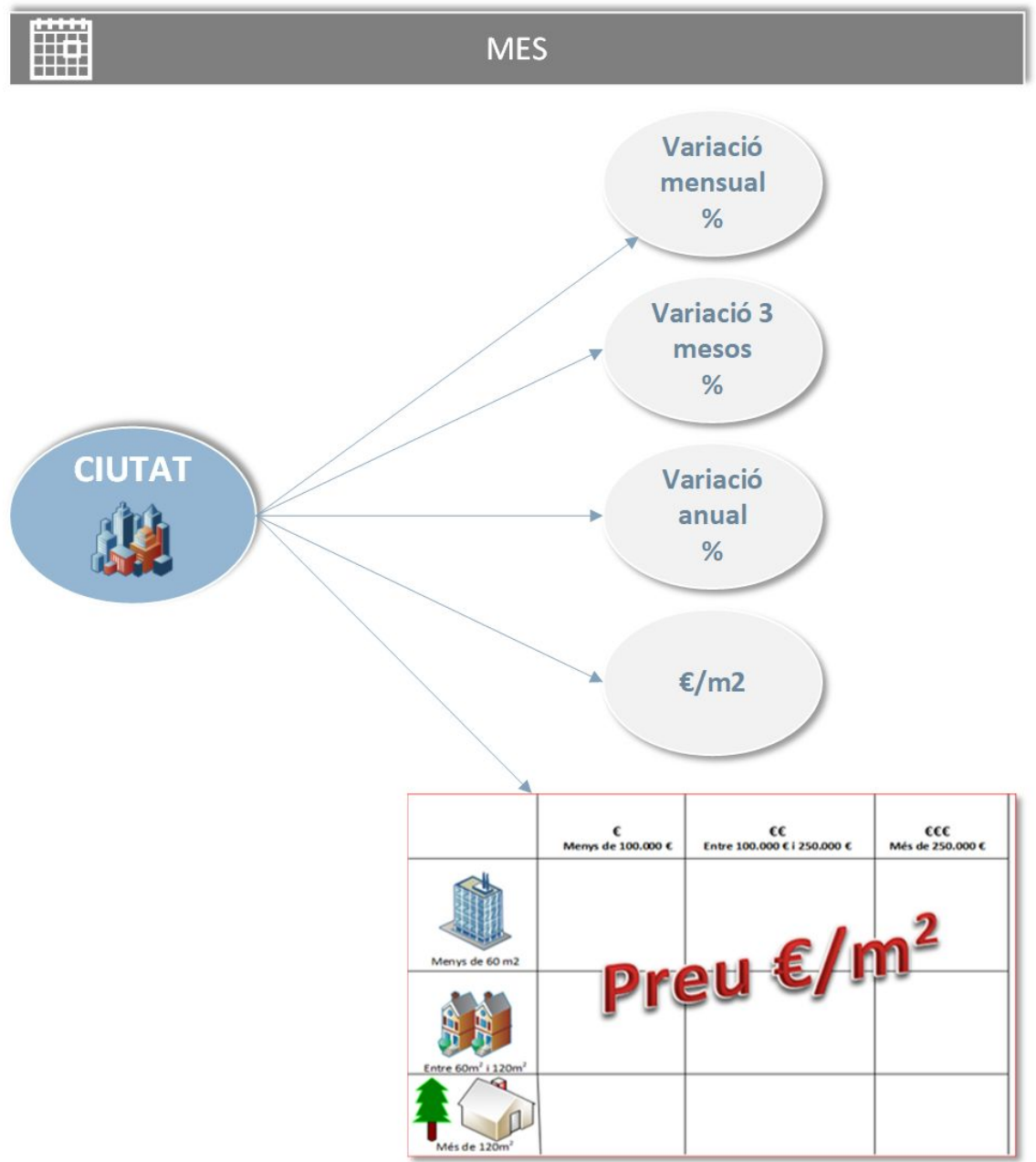
El títol del dataset és "PreusVivendaEspanya_mes_any"

3. Descripció del dataset. Desenvolupar una descripció breu del conjunt de dades que s'ha extret (és necessari que aquesta descripció tingui sentit amb el títol triat).

El dataset conté els preus de la vivenda de cada població d'espanya d'un mes, la variació amb el mes anterior, els 3 mesos i l'any anterior. També inclou preus agrupats segons tamany de vivenda i preu total.

4. Representació gràfica. Presentar una imatge o esquema que identifiqui el dataset visualment

PreusVivendaEspanya_mes_any.csv



5. Contingut. Explicar els camps que inclou el dataset, el període de temps de les dades i com s'ha recollit.

El dataset inclou els camps següents:

- ☐ **Ciutat:** Població d'espanya
- ☐ **Variació mensual:** % variació mensual respecte mes anterior.
- ☐ **Variació tres mesos:** % variació mensual respecte els tres mesos anteriors.
- ☐ **Variació Anual:** % variació anual.
- ☐ **€/m²:** Preu € / m²
- ☐ **Menys 60 m² / Menys 100.000 €:** Preu € / m² en vivendes de menys de 60 m² i de menys de 100.000 €
- ☐ **Menys 60 m² / Entre 100.000 € i 250.000 €:** Preu € / m² en vivendes de menys de 60 m² i entre 100.000 € i 250.000 €
- ☐ **Menys 60 m² / Més de 250.000 €:** Preu € / m² en vivendes de menys de 60 m² i de més de 250.000 €
- ☐ **Entre 60 m² i 120 m² / Menys 100.000 €:** Preu € / m² en vivendes entre 60 m² i 120 m² i de menys de 100.000 €
- ☐ **Entre 60 m² i 120 m² / Entre 100.000 € i 250.000 €:** Preu € / m² en vivendes entre 60 m² i 120 m² i entre 100.000 € i 250.000 €.
- ☐ **Entre 60 m² i 120 m² / Més de 250.000 €:** Preu € / m² en vivendes entre 60 m² i 120 m² i de més de 250.000 €
- ☐ **Més de 120 m² / Menys 100.000 €:** Preu € / m² en vivendes de més de 120 m² i de menys de 100.000 €
- ☐ **Més de 120 m² / Entre 100.000 € i 250.000 €:** Preu € / m² en vivendes de més de 120 m² i entre 100.000 € i 250.000 €.
- ☐ **Més de 120 m² / Més de 250.000 €:** Preu € / m² en vivendes de més de 120 m² i de més de 250.000 €

Les dades de preus són de l'últim mes actualitzat. En el nom del dataset s'inclou el mes i l'any al que pertanyen les dades i que s'ha extret de la pàgina web.

Les poblacions d'espanya i les seves variacions de preus s'han obtingut de les taules que ens proporciona la web mitjançant web scrapping. La web conté una taula amb totes les províncies i accedint a l'enllaç de cada província obtenim la taula de variacions per cada població. Mitjançant iteració aconseguim les dades de totes les poblacions d'espanya.

De la mateixa forma, accedint a l'enllaç de cada població obtenim la taula comparativa de valors dels immobles segons superfície i preu total.

6. Agraïments. Presentar el propietari del conjunt de dades. És necessari incloure cites de recerca o anàlisis anteriors (si n'hi ha).

Les dades l'han adquirit i ordenat els responsables de la immobiliària "Trovimap".

Nosaltres hem fet un web-scraping de la seva pàgina web.

7. Inspiració. Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre.

El conjunt de dades és interessant perquè d'una forma molt ràpida podem obtenir els preus de les vivendes de totes les poblacions d'Espanya de l'últim mes. Ho podem utilitzar per realitzar comparatives dels preus de les vivendes, per afegir-ho a la nostra web o aplicació i tindrem ràpidament tots els preus en un format molt simple i ràpid.

La pregunta que respon seria com evoluciona el preu de la vivenda a cada població espanyola, a cada província o al país en general. També com evolucionen els preus agrupant les vivendes per tamany i preu total.

8. Llicència. Seleccionar una d'aquestes llicències pel dataset resultant i explicar el motiu de la seva selecció:

Hem triat la llicència de creative commons attribution 4.0 ja que és la més nova i és totalment lliure de compartir.

9. Codi. Adjuntar el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

<https://github.com/Garriden/Tipologia-i-cicle-de-vida-de-les-dades>

10.Dataset. Publicar el dataset en format CSV a Zenodo amb una xicoteta descripció.

<https://zenodo.org/record/3749346#.XpNpTjdR2Hs>

11.Lliurar. Presentar el treball amb el DOI del dataset a Github

DOI: 10.5281/zenodo.3749346

Contribucions	Signa
Recerca prèvia	S.B., M.G.
Redacció de les respostes	S.B., M.G.
Desenvolupament codi	S.B., M.G.