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

El nostre objectiu és crear un dataset que contingui informació dels pisos de lloguer de les 4 capitals de província de Catalunya. El lloc a cercar aquesta informació doncs, seran els diferents portals de compra/lloguers de pisos de Catalunya i Espanya.

Hem fet una recerca sobre 8 portals de lloguers de pisos per escollir el que més fàcilment ens proporcionarà les dades que volem. Aquests portals són: fotocasa, idealista, habitaclia, API.cat, buscocasa, pisos.com, yaencontre i tucasa.

A continuació mostrem les conclusions d’aquesta recerca i el portat escollit per recol·lectar la informació:

- **Fotocasa** i **Idealista:**  són els primers en descartar ja que retornen un codi 405 (mètode no permès) i 403 (prohibit) respectivament. Ja que tenim alternatives decidim descartar aquests portals. Els següents tots retornen un codi 200 (Ok).

**- Buscocasa:** s’ha descartat perquè hi ha poques ofertes i totes de fa més d’un mes. Tenim alternatives amb un volum de dades més gran i més actualitzat.

**- API.cat:** s’ha descartat perquè aporta poca informació dels pisos de cada anunci. No aporta informació de número d’habitacions o número de lavabos cosa que altres portals sí que fan.

- **Habitaclia:** ara ja comencem a trobar portals molt complets en informació i en volum de pisos. Segurament a partir d’aquí tots ens servirien per crear un bon dataset però, ja que podem escollir, triarem el portal que presenti la informació de la manera més còmode de recol·lectar. En aquest cas hem descartat aquest portal perquè la informació del pis (metres quadrats, habitacions, lavabos...) no està diferenciada per etiquetes sinó que és una sola cadena de caràcters en la mateixa etiqueta. Com hem dit no impossibilita la recol·lecció de informació però tenim alternatives que ens deixaran un codi més clar i net.

**- Pisos.com:** també té la informació molt completa i en aquest cas sí que la diferent informació està separada en diferents etiquetes. S’ha descartat, però, perquè totes les classes d’aquestes etiquetes es diuen igual (class=item) i tenim alternatives que ho tenen més ben classificat.

**- Yaencontre i Tucasa:** finalment arribem als dos últims portals. Ens presenten informació completa i un número de pisos molt elevat. A més la informació que volem està molt ben classificada per etiquetes úniques. Així doncs tenim dues opcions molt vàlides. Decidm quedar-nos amb **Tucasa** ja que en aquest portal els títols dels anuncis són els noms dels barris d’una forma més estandarditzada que Yaencontre i permet fer cerques a nivell de província.