

**Realitzat per:** Gerard López Roig i Antonio Àngel Martínez Chamizo

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

Per a aquest projecte ens hem centrat en la venda de vehicles d'ocasió. La font de les nostres dades és el portal de compra-venda de la empresa Autocasión. La empresa es dedica a la venda de cotxes de primera i segona mà, així com a oferir notícies sobre el món del motor.

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

Audi Autocasion Espanya

**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 dades que fan referència a vehicles de segona mà de la marca Audi; vehicles que, en el moment de la recol·lecció de dades, hem trobat en venda a Espanya. La informació capturada al conjunt de dades reflexa les propietats quantitatives i qualitatives que fan referència al vehicle en venda, alguns exemples representatius serien: model, preu actual, localització del venedor i fitxa tècnica.

S'han ignorat aspectes més subjectius i de difícil anàlisi com l'equipament, les imatges del vehicle o la descripció no estructurada donada per l'anunciant en text lliure.

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



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

Descripcio: Model

Descripcio2: Descripció del model

Any: Any de matriculació del vehicle

Província: Localització del venedor del vehicle, província

Kms: Quilòmetres recorreguts pel vehicle, sencer en quilòmetres

Matriculacio: Mes i any matriculació del vehicle

Garantia: Nombre de mesos després de la compra on el vehicle està cobert per la garantia, sencer en mesos

Color: Color de la carroceria

Ambiental: Classificació del vehicle vinculada a la seva emissió de contaminants, codi de lletres de la DGT

Preu: Preu actual del vehicle en venda, nombre sencer en euros

Preu-nou: Preu original del vehicle (quan disponible), nombre sencer en euros

Llarg: Llargada del vehicle, sencer en mil·límetres

Ample: Amplada del vehicle, sencer en mil·límetres

Alt: Alçada del vehicle, sencer en mil·límetres

Batalla: Distància d'eix a eix del vehicle, sencer en mil·límetres

Maleter: Capacitat del maleter, sencer en litres

Peso-max: Pes del vehicle, sencer en kg

Carroceria: Tipus de carroceria utilitzada al vehicle

Portes: Número de portes del vehicle, sencer

Places: Número de places del vehicle, sencer

Combustible: Tipus de combustible

Cilindrada: Volum total dels cilindres del vehicle, real en cm<sup>3</sup>

Cilindres: Número de cilindres del vehicle, sencer

Sobrealimentacio: Tipus de sobrealimentació del motor

Traccio: Tipus de tracció del vehicle

Transmisio: Tipus de transmissió utilitzada per el vehicle (manual o automàtica)

Marxes: número de marxes del vehicle, sencer

Potencia-kw: Potencia del motor, sencer en KW

Potencia-cv: Potencia del motor, sencer en CV

Par: Moment de força del motor sobre la transmissió, sencer en N·m

Velocitat-max: Velocitat màxima del vehicle, sencer en km/h

Acceleracio: Temps transcorregut accelerant de 0 a 100km/h, real en segons

Consum-urba: Consum de combustible mig en entorn urbà, real en litres per 100km

Consum-carretera: Consum de combustible mig en carretera, real en litres per 100km

Consum-mitja: Consum de combustible mig, real en litres per 100km

Co2: Emissió de CO<sub>2</sub>, sencer en g/km

Diposit: volum del dipòsit, real en litres

Tipus-venedor: Tipus de venedor del vehicle (particular o professional)

Data-publicacio: Data de publicació de l'anunci

Les dades son una captura dels anuncis disponibles per audis d'ocasió el 07/11/2020.

La captura de les dades s'ha fet mitjançant un script de web scraping en python, utilitzant la llibreria Scrapy.

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

Volem agrair a Vocento, actuals propietaris del portal del motor AUTOCASION.COM, per la recopilació i allotjament de dades utilitzades com a font de informació per a aquest projecte. No s'han trobat recerques o anàlisis anteriors.

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

L'objectiu últim del dataset és intentar esbrinar quins són els principals factors físics darrera del preu i la depreciació d'un cotxe. A tal efecte, el dataset conté una descripció minuciosa dels vehicles audi de segona mà en venda a Espanya, acompanyats del seu preu actual.

Exemples de preguntes que es podrien respondre serien: Afecta el color en cotxes comparables? Hi ha models que perden valor més ràpidament que altres? Es poden trobar vehicles a millor preu en algunes localitzacions específiques? Quins elements de la fitxa tècnica són els millors predictors del preu?.

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

Els termes de ús de la pàgina s'especifica, en quant a la seva propietat intel·lectual, que l'usuari es compromet a utilitzar els continguts i / o elements als quals accedeixi a través dels Serveis de Autocasion.com per al seu propi ús i necessitats, i a no realitzar en cap cas una explotació comercial, directa o indirecta dels mateixos.

El nostre scraper no ha fet una acceptació activa dels termes d'ús en cap pas per a aconseguir les dades, no obstant, per a respectar la intenció moral de la clàusula, hem decidit fer servir la llicència "**Creative Commons Attribution-NonCommercial (CC-BY-NC)**" que protegeix el dataset contra l'explotació comercial de les dades obtingudes.

Contribucions	Signa
Recerca Previa	A.A.M.C, G.L.R
Redacció de les respostes	A.A.M.C, G.L.R
Desenvolupament del codi	A.A.M.C, G.L.R