

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.

Model: Nom del model del vehicle

Preu: Preu actual del vehicle en venda, nombre sencer en euros
Preu de nova venda: Preu original del vehicle (quan disponible), nombre sencer en euros
Data de matrícula: Data original de matriculació del vehicle, format datetime
Quilometratge: Quilòmetres recorreguts pel vehicle, sencer en quilòmetres
Combustible: Tipus de combustible
Color: Color de la carroceria
Distintiu ambiental: Classificació del vehicle vinculada a la seva emissió de contaminants, codi de lletres de la DGT
Garantia: Nombre de mesos després de la compra on el vehicle està cobert per la garantia, sencer en mesos
Localització: Localització del venedor del vehicle, provincia
Data-publicacio: data de publicació de la referència, datetime
Carroceria-capacitat: Capacitat del maleter, sencer en litres
Carroceria-carroceria: Tipus de carroceria utilitzada al vehicle
Carroceria-places: Número de places al vehicle, sencer
Carroceria-batalla: Distància d'eix a eix del vehicle, sencer en milimetres
Carroceria-pes: Pes del vehicle, sencer en kg
Carroceria-portes: Número de portes del vehicle, sencer
Motor-combustible: Tipus de combustible utilitzat pel motor (diesel o gasolina)
Motor-cilindres: Número de cilindres del vehicle, sencer
Motor-cilindrada: Volum unitari de cada un dels cilindres del vehicle, real en cm3
Motor-sobrealimentacio: Tipus de sobrealimentació del motor
Transmissio-traccio: Tipus de tracció del vehicle
Transmissio-marches: número de marches del vehicle, sencer
Transmissio-tipus: Tipus de transmissió utilitzada per el vehicle (manual o automàtica)
Prestacions-potencia: potencia del motor, sencer en KW
Prestacions-par: moment de força del motor sobre la transmissió, sencer en N·m
Prestacions-acceleracio: temps transcorregut accelerant de 0 a 100km/h, real en segons
Prestacions-potencia: potencia del vehicle, sencer en cavalls de vapor
Prestacions-vmx: velocitat màxima del vehicle, sencer en km/h
Consum-urba: consum de combustible mig en entorn urbà, real en litres per 100km
Consum-mig: consum de combustible mig, real en litres per 100km
Consum-deposit: volum del dipòsit, real en litres
Consum-carretera: consum de combustible mig en carretera, real en litres per 100km
Consum-co2: emissió de CO2, sencer en g/km

Les dades son una captura dels anuncis disponibles per audis d'ocasió el 05/11/2020.

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.

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