

Dataset: Preus PVP empresa ACO Iberia

Xavier Parramon Boada

Abril 2020

Context

Una de les principals empreses de drenatge del sector vol fer un estudi de mercat per veure com es posicionen els seus productes respecte al principal competidor *ACO Iberia*. Per això, utilitza la seva pròpia web ACO Iberia per extreure tots els productes que tenen en el catàleg online amb el preu de venda al públic (PVP). Així posteriorment podrà creuar aquesta informació amb la seva pròpia i fer una comparativa de preus per saber en quin tipus de producte pot competir millor i en quins no.

Descripció

El conjunt de dades generat com a part d'aquesta activitat, està format per tots els productes del catàleg online de l'empresa ACO Iberia. Hi ha diferents tipus de productes com: canals, boneres, separadors, vàlvules, reixes, etc. Algunes de les variables que s'han recollit en el conjunt de dades són: la data, el codi del producte, el pvp, la descripció curta del producte i una classificació de 2 nivells.

Representació gràfica.

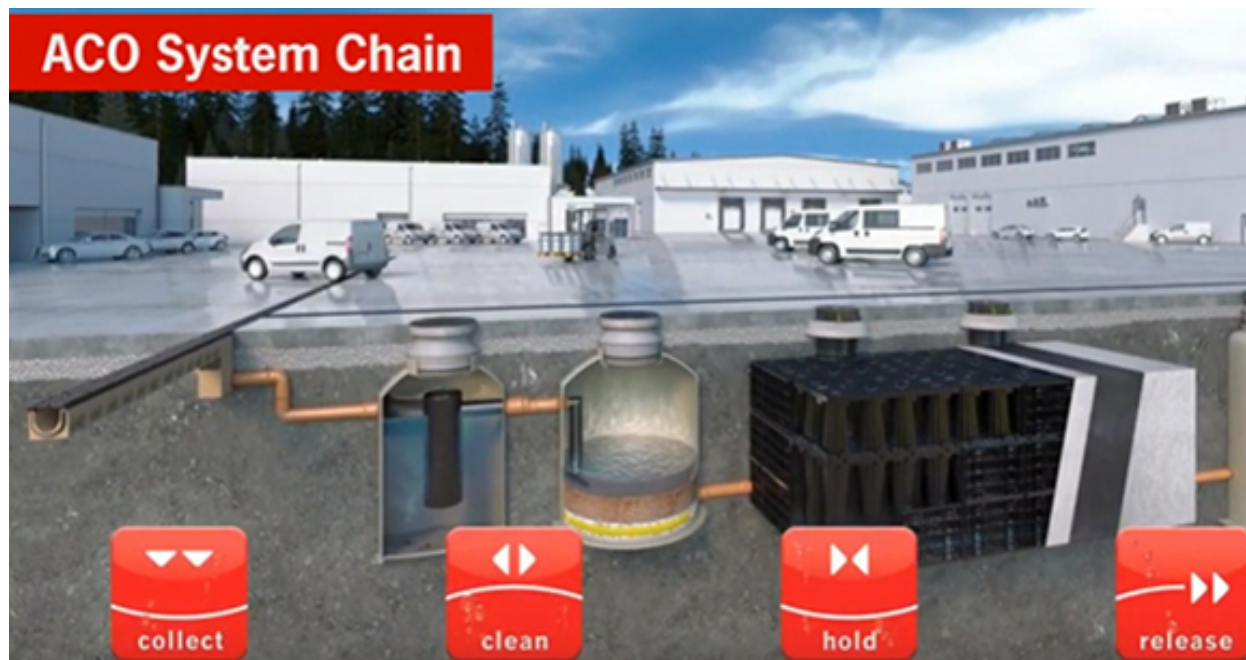


Figure 1: Instal·lació en cadena dels productes del sistema de drenatge ACO

Contingut

Per a cada producte, el qual es correspon amb un registre en el conjunt de dades, es recullen les següents característiques:

- **Date:** Dia en el que s'ha recollit la informació
- **Code:** Codi del producte
- **PVP:** Preu de venda al públic en €
- **Description1:** Descripció curta del producte
- **Description2:** Especificació del tipus de producte que és i de quina gamma forma part
- **Tipe:** Classificació dins el rang de productes

L'empresa ACO Iberia va actualitzant periòdicament la seva web amb els productes que tenen disponibles per a vendre en el mercat Espanyol i Portuguès. Els preus varien aproximadament cada any (no es pot obtenir històric, però el podem anar enregistrant amb el dataset), i també van introduint productes nous i en van eliminant d'obsolets.

Agraïments.

Les dades s'han extret del catàleg online de la pàgina web ACO Iberia. Per a fer-ho, s'ha utilitzat el llenguatge de programació python i tècniques de *web scraping* per extreure la informació allotjada en les pàgines HTML.

Inspiració

El conjunt de dades pot ser interessant per a les empreses del sector de drenatge d'aigües. Aquest dataset els permet, juntament amb les seves dades internes, fer una comparativa de posicionament en el mercat. Per saber, a nivell de preu, com estan posicionats respecte de la competència. Amb el temps també es pot anar creant un historic de preus per a poder veure com va evolucionant l'empresa i el mercat, i si s'extengues aquest procés a més empreses em podria fer un anàlisis molt més complet i precís.

Llicència

La llicència seleccionada per la publicació del conjunt de dades ha sigut **CC BY-SA 4.0 License**. Els motius per els que s'ha seleccionat aquesta llicència són:

- *S'ha de donar crèdit al creador del conjunt de dades, indicant els canvis que s'han realitzat.* Així és reconeix la feina que s'ha dut a terme i com això ha ajudat a altres treballs.
- *Es permet l'ús comercial.* L'objectiu principal de la base de dades és per poder dur a terme un estudi de mercat i de posicionament respecte la competència. Tot i que per no perjudicar l'empresa objectiu d'aquesta pràctica també hi havia la possibilitat d'utilitzar una llicència: **CC BY-NC-SA 4.0 License**.
- *Els treballs que es basin en el conjunt de dades creat, també tindran que distribuir-se sota aquesta mateixa llicència.*

Codi i dataset

Tant el codi font escrit per a l'extracció de les dades, com el dataset creat, és pot accedir a través de l'enllaç. El conjunt de dades en format csv també es pot trobar a la web Zenodo

Recursos

1. Lawson, R. (2015). *Web Scraping with Python*. Packt Publishing Ltd. Chapter 1 Introduction to web Scraping and 2. Scraping the Data.
2. subirats Maté, L. i Calvo González, M. *Web scraping*.
3. Mitchel, R. (2015). *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
4. web: *Beautiful Soup Documentation*. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/#>