

Project “Big Data, Business Intelligence and NoSQL”

House prices



Je werkt op de “**Ames housing dataset**” (samengesteld door *Dean De cock*). In feite is deze dataset een modernere, uitgebreide versie van de Boston Housing dataset (waar je heel wat kant en klaar onderzoeksmateriaal voor terugvindt).

Het uiteindelijke doel van de analyse op deze dataset is het voorspellen van de “sales price” voor elk huis. M.a.w. voor elk huis in de “test set” moet je de `SalePrice` variabele kunnen voorspellen.

Om je toch wat op weg te helpen geven wij hier een minimaal stappenplan die moet uitgevoerd worden (extra stappen of zij-onderzoeken zijn uiteraard toegelaten).

- 1) Beschrijvend data-onderzoek m.b.v. Tableau (hoe zijn de variabelen verdeeld, vind je grafische verbanden tussen features?, ...). Maak een overzichtelijk dashboard met nuttige grafieken die een meerwaarde zou betekenen voor beleidvoerders die op basis hiervan beslissingen kunnen nemen.
- 2) Data-onderzoek m.b.v. Python (ga op zoek naar ontbrekende data, kies een gepaste manier om met die ontbrekende data om te gaan, maak hiervoor gebruik van de technieken die je in de datacamp modules geleerd hebt, ...)
- 3) Stel een correlatiematrix op waarbij je de features die het meest correleren zoekt (neem als maatstaf een correlatie, $|r|$ van $> 0,5$). Doe een grafische analyse van de meest correlerende features.
- 4) Onderzoek lineaire regressie met behulp van scikit-learn en de ingebouwde methodes (ridge, LASSO, ElasticNet, ...)
- 5) Bouw een lineair model op basis van je resultaten in (4) en test dit uit.

Deadline: ten laatste 15/05/2019 tijdens de les wordt een onderzoeksrapport (.PDF) ingediend op Leho. In dit onderzoeksrapport verwerk je duidelijk de stappen die hierboven beschreven staan.

De presentaties gaan door op **15 mei of 22 mei 2018** in de voormiddag. Zie *Leho* voor de planning.

Tijdens de presentatie illustreer je de stappen uit je onderzoeksrapport. Zorg ervoor dat elk groepslid op de hoogte is van zowel het Tableaudeel als het Pythongedeelte!

Veel succes!

Thijs Martens & Brian Baert