

OPDRACHTEN DATA SCIENCE FUNDAMENTALS

De opdrachten kunnen worden gemaakt met Microsoft Excel, PowerBI, R en Python. Kies een omgeving waarmee je enigszins vertrouwd bent.

Bij de opdrachten worden enkele datasets gebruikt:

- IRIS: een dataset met kwantitatieve gegevens van drie soorten iris-sen.
- Ames Housing: een dataset met kwalitatieve en kwantitatieve gegevens over huizen en de verkoop daarvan
- Mtcars: een dataset met kenmerken en karakteristieken van enkele automodellen.
- GM_LF: een dataset met de data over de ontwikkeling van de levensverwachting in de verschillende landen

1. Maak een grafiek die de ontwikkeling van de levensverwachting toont voor de landen: Afghanistan, Nederland en Brazilië.
2. Maak, op basis van het bestand AmesHousing, een overzicht waarin het aantal verkopen en de gemiddelde prijs zijn geaggregeerd op het jaar van de verkoop en de buurt waarin het huis staat.
3. Breng, op basis van het bestand mtcars, het verband in kaart tussen het aantal mijlen dat per gallon kan worden gereden (= target) en de features PK, cylinderinhoud, gewicht. Welke factor heeft de grootste invloed op het aantal mijlen dat per gallon kan worden gereden?
4. Breng, op basis van het bestand iris, het verband in kaart tussen de lengte en de breedte van de kroonbladeren, en tussen de lengte en de breedte van de kelkbladeren.
5. Breng, op basis van het bestand AmesHousing, het verband in kaart tussen de verkoopprijs (= target) en de features bouwjaar, verkoopjaar, omvang perceel en algemene kwaliteit. Welke factor heeft de grootste invloed op de prijs van AmesHousing?

Tk, 24 februari 2024