

Project “Big Data, Business Intelligence and NoSQL”

Je werkt in grote lijnen de challenge uit van Datatonic (zie het document [CHALLENGE.pdf](#) op Leho).

Wij hebben echter een aantal minimale vereisten!

1) **Preprocessing van de dataset** die terug te vinden is via onderstaande links. Denk aan de tips die gegeven worden in de cursussen van DataCamp. Schoon de data op, bekijk of er “verkeerde” of “ontbrekende” data in zit en kies een gepaste oplossing hiervoor.

- https://storage.googleapis.com/datatonic-joinus-challenge/flights_2010_000000000000.zip

- http://datatonic-joinus-challenge.storage.googleapis.com/flights_2010_000000000001.zip

- https://storage.googleapis.com/datatonic-joinus-challenge/flights_2010_000000000002.zip

- https://storage.googleapis.com/datatonic-joinus-challenge/flights_2010_000000000003.zip

- https://storage.googleapis.com/datatonic-joinus-challenge/flights_2010_000000000004.zip

- http://datatonic-joinus-challenge.storage.googleapis.com/flights_2010_000000000005.zip

2) **Beschrijvende statistiek met Tableau**, je maakt een dashboard waarin je de voor jullie relevante informatie weergeven. Zorg ervoor dat dit dashboard nuttig is en een meerwaarde levert voor beleidsvoerders die op basis van je samengevatte info beslissingen kunnen nemen.

3) **Machine learning m.b.v. Scikit learn**

Ga op zoek in de dataset naar verbanden die je kunt uitwerken m.b.v. Scikit learn. Minimaal vereisen wij dat je minstens één machine learning techniek implementeert. Dit kan een clustering zijn, een regressiemodel, ...

4) **Maak gebruik van de NetworkX library** (Python) om één of meerdere grafische weergaven te maken van (een gedeelte van) de data.

5) **Wees creatief** met de data.

Deadline: 14 mei 2018 dien je een verslag in (PDF) waarin je duidelijk documenteert wat je hebt ondernomen en wat het uiteindelijke resultaat is van elke stap.

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

Veel succes en laat je niet afschrikken door een nieuwe technologie, want iedereen moet ooit wel eens met een nieuwe technologie werken. Deel van dit project is dan ook dat je zelf in groep nieuwe technologieën onderzoekt en verwerft.

Brian Baert, Thijs Martens en team Datatonic