

Open data: Wat kunnen en willen we ermee?

Published on April 17, 2018



BAG

De Nederlandse overheid stimuleert al jaren het gebruik van opendata. Immers, waarom zou data die met geld van de burger is ingewonnen niet terugvloeien naar die burger? Klinkt logisch toch? Eén van de datasets die ik interessant vind is de BAG. Dat staat voor Basisregistratie Adressen en Gebouwen en bevat naast informatie over adressen en gebouwen ook informatie over o.a. ligplaatsen (woonboten), standplaatsen (sta caravan). En verder o.a. over het functioneel gebruik van "adressen" (wonen, gezondheidszorg, onderwijs etc.) oppervlaktes van "woningen", woonplaatsen en gemeentes. En je kunt ook gegevens afleiden zoals locatie, waardoor je het kan weergeven op een landkaart.

Een wens van velen is om een "adressentabel" van Nederland te maken zodat je bij het intikken van een postcode en huisnummer meteen kunt zien welke straat, woonplaats het betreft. De vrienden van NLExtract leveren al sinds jaar en dag een maandelijks bijwerkte adressentabel van NL. Een prima tabel maar ik wilde van een adres meer zien dan hetgeen NLExtract levert. Ik wilde weten wat de oppervlakte van een pand is of wat voor type woning (vrijstaand, hoekwoning etc.) het betreft. Bovendien wilde ik ook adressen zonder postcode zien. (ja .. die zijn er ook, bv van los staande garages) Gelukkig levert NLExtract ook maandelijks een postgresql/postgis database dump met alle BAG data. Met die dumps ben ik aan de slag gegaan om een uitgebreide adressentabel te maken.

De adressentabel, wat is het en wat kunnen we ermee?

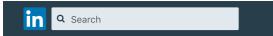
Met de omschrijving van deze tabel (met een voorbeeldadres) krijg je een beeld van de gegevens die er in staan maar je kunt ook de data zelf bekijken. Als je een csv-bestand met alle adressen van NL met 9.147.695 rijen en 50 kolommen wilt downloaden (situatie april 2018) dan kan dat hier maar bedenk je wel dat niet alle programma's dit soort aantallen kunnen verwerken. Voor de beeldvorming kun je ook een csv-tabel met alleen de adressen van de woonplaats Utrecht hier downloaden. Heb je die liever in Excel formaat neem dan deze.

Er zijn vele toepassingen van deze tabel te bedenken. Ik noem er een aantal.

Koppelen aan je eigen (klanten) adressen:









Door je eigen adressengegevens te koppelen aan deze tabel kun je exact op ee zien waar dat adres is en vervolgens dat gegeven verder gebruiken.

3. Aggregeren

Je kunt nu aggregeren naar het niveau woonplaats, gemeente of provincie.

(In welke provincie zitten de meesten van jouw klanten?)

4. Koppelen externe data

Koppelen aan by CBS data waar veel gegevens vastliggen op gemeente- of pr

(Je hebt misschien wel de meeste klanten in bv Zuid-Holland maar hoe dat tot het aantal inwoners van deze provincie?)

5. Risico analyse

(Welke type woning (hoek, tussen, vrijstaand) heeft de grootste kans op inbraak? Is er een verband tussen de afstand tot een brandweerkazerne en de mate van brandschade? En tussen bouwjaar van het pand en de brandschade? Is er een verband tussen de afstand tot een snelweg of drukke weg en longaandoeningen)

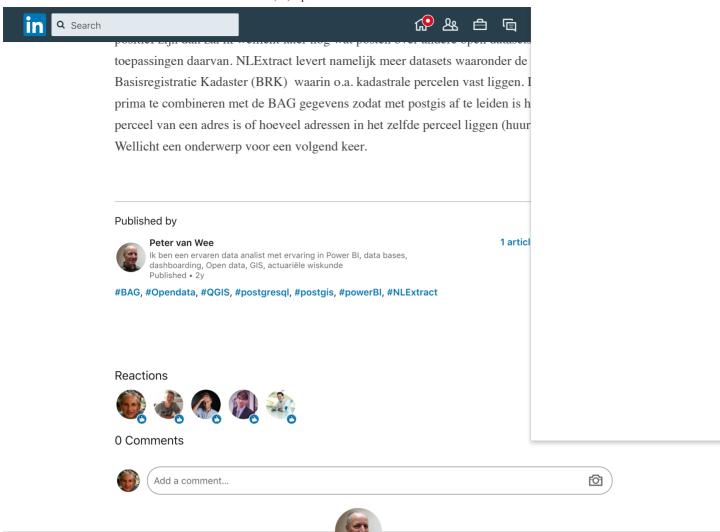
Er zijn uiteraard veel meer toepassingen te bedenken. Het mooie van het vak van data analist is dat je door met open-, eigen-, of externe data aan de slag te gaan inzichten opdoet die je weer kunt gebruiken voor het toetsen van hypothesen. Business intelligence (BI) tools zoals by PowerBI desktop kunnen daar nuttige diensten bewijzen.

Het leuke van de workflow die ik gebruik voor deze tabel is dat het gedaan is met open data en alleen opensource tooling. Dus naast de benodigde hardware kun je de kosten binnen de perken houden. Het staat je vrij om met de tabel aan de slag te gaan. Je zou de tabel bijvoorbeeld kunnen laden in PowerBI desktop (welke provincie heeft de meeste adressen met vrijstaande huizen?) of kunnen gebruiken om met Qgis aan de slag te gaan om mooie kaartjes te maken. Beiden zijn gratis te gebruiken.

Mocht je fouten ontdekken in de data kijk dan eerst even in de BAG Viewer of de fout wellicht in de BAG zelf zit. Zo niet geef dan s.v.p. een seintje zodat ik kan kijken of er een fout in mijn script zit. Dat moet je nooit uitsluiten;)

Vervolg





Peter van Wee

Ik ben een ervaren data analist met ervaring in Power BI, data bases, dashboarding, Open data, GIS, actuariële wiskunde



