



MEMO

Onderwerp Meten van drukte en populaire tijden van locaties en voorzieningen
Aan (10)(2e)
Van Geo4COVID team
Datum 11 april 2020
Status Concept

Google verzamelt locatiegegevens van gebruikers die dat aan hebben staan op hun smartphone, tablet of laptop (*consent*). Google anonimiseert en aggregeert deze locatiegegevens en leidt hieruit af hoe druk het op een locatie of een voorziening - bijvoorbeeld een winkel, park, of weg - was, en maakt met het analyseren van de locatiegeschiedenis¹ een inschatting hoe druk het *wordt*.

Wij stellen nadrukkelijk vast dat het hier niet om persoonsgegevens, zoals gedefinieerd in de GDPR, maar enkel om maten die de drukte op locaties uit drukken.

Google publiceert de 'populaire tijden' van een locatie of voorziening in de browser bij een zoekresultaat. Ook geeft Google, indien beschikbaar², de *live*³ populariteit van deze locatie of voorziening weer. Met deze informatie kan een gebruiker beoordelen of het op deze locatie nu drukker of rustiger is dan *gepeild en voorspeld* door Google.



Twee voorbeelden van populaire tijden bij de gegevens van een voorziening in een Google zoekresultaat: Praxis (links) en Intratuin (rechts) in Breda op Goede vrijdag (10 april 2020) gepeild 17:30uur.

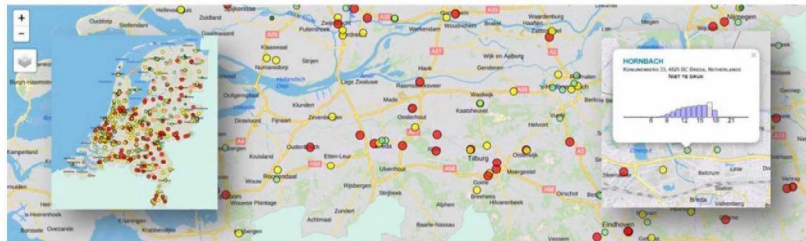
Google stelt deze gegevens niet via de reguliere API⁴ beschikbaar, maar de gegevens zijn wel met behulp van de Google API en een aanvullend script⁵ op te halen. Het respons uit dit script is onder meer een waarde voor de huidige drukte (*current_popularity*) en waarden voor de verwachte drukte over de dag (*populartimes*) voor een hotspot.

Om de mogelijkheden te verkennen, hebben we dit script geïnstalleerd op een testserver en de locatiegegevens van 600 bouwmarkten en tuincentra in Nederland gepeild op Goede vrijdag (10 april 2020). Vervolgens hebben we deze gegevens op de kaart gepresenteerd, hieronder een snapshot van het resultaat rond 17:00uur: op veel locaties is het 'drukker dan normaal' (geel) of 'redelijk druk' (rood) en op een aantal plaatsen is het 'niet te druk' (groen)⁶.

T (10)(2e)
E info@geonovum.nl
I www.geonovum.nl

bezoekadres
 Borchman Wuytterslaan 10
 3818 LH Amersfoort

postadres
 Postbus 506
 3800 AM Amersfoort



Wat kunnen we met deze gegevens?

- 1) *Community Mobility Reports*⁷ samenstellen van (type) hotspots in bijvoorbeeld een gemeente.
- 2) Real-time hotspots traceren waar het drukker is dan verwacht.

Deze informatie is relatief en indicatief⁶, rekening houdend met de volgende kanttekeningen en onzekerheden:

- *Consistentie*: We hebben steekproefsgewijs de grafieken van de testomgeving gecontroleerd met de grafieken in Google: de grafieken lieten dezelfde trend zien. De classificatie van 'drukte' tussen Google en onze presentatie verschilde in sommige gevallen.⁹
- *Actualiteit*: We hebben over de dag elk uur de gegevens van 600 locaties gepeild¹⁰. Het is niet bekend met welke updatefrequentie Google de gegevens van populaire tijden en actuele drukte bijwerkt.
- *Nauwkeurigheid*: Er wordt een waarde voor actuele drukte verkregen, maar het is niet precies bekend wat die waarde betekent¹¹. Google acht de gegevens over actuele drukte zelf te onbetrouwbaar omdat de locatiebepaling van gebruikers niet altijd even nauwkeurig en betrouwbaar is¹².
- *Compleetheid*: Voor sommige locaties en voorzieningen zijn geen gegevens over populaire tijden en actuele drukte beschikbaar. We verwachten dit verband houdt met betrouwbaarheid: alleen gegevens van locaties waar er genoeg devices ooit zijn of nu worden gepeild.
- *Juridisch*: volgens onze informatie¹³ is het niet toegestaan om bulk-gegevens over locaties of voorzieningen geautomatiseerd via Google Maps te verzamelen (*scrapen*).

Hoe zien we een vervolg?

- *Behoeft*: moet worden verkent welke meerwaarde deze informatie biedt voor het meten van drukte en monitoring van beleid rond 'social distancing' voor COVID-19
- *Juridisch*: moet nader worden onderzocht in hoeverre het toepassen van deze techniek voor het meten van drukte voor bestrijding COVID-19 door Google wordt toegestaan (grondige check op de licentievoorwaarden) of gedoogd.
- *Implementatie*: moet worden onderzocht hoe de kanttekeningen en onzekerheden een productie versie in de weg staan.

¹ Voor de inschatting worden verschillende gegevens gecombineerd, bijvoorbeeld weersverwachting, normale of bijzondere dag (Goede vrijdag/ tweede paasdag) etc.

² Deze gegevens zijn naar alle waarschijnlijkheid alleen beschikbaar als er voldoende telefoons (privacy) met voldoende locatienauwkeurigheid (betrouwbaarheid) gepeild worden.



³ Live wordt verondersteld in de zin van real-time gepeilde devices.

⁴ Gegevens over nuttige plaatsen worden via Google Places API beschikbaar gesteld. Google stelt dat de gegevens over 'populaire tijden' niet nauwkeurig/goed genoeg zijn om ze beschikbaar te stellen via deze API. Wel worden de gegevens bij zoekresultaten van locaties of voorzieningen via de browser gepresenteerd.

⁵ Er is een open source Python script toegepast om de gegevens te scrapen, zie <https://github.com/m-wrzt/populartimes>. Het script werkt op basis van de originele Google API waarvoor een API-key moet worden meegestuurd.

⁶ De classificatie van drukte is gebaseerd op de peiling (*current_popularity*) afgezet tegen verwachting (*populartimes*) voor deze geselecteerde hotspots.

⁷ Google publiceert in het kader van COVID-19 Community Mobility Reports door locatiegeschiedenis van type hotspots per provincie over een bepaalde periode te aggregeren, zie <https://www.google.com/covid19/mobility/>.

⁸ Deze gegevens kunnen *indicatief* worden toegepast om te analyseren op welke locaties het *relatief* meer of minder druk is dan verwacht berekend op basis van Google's real-time waarnemingen versus Google's voorspelling uit locatiegeschiedenis.

⁹ De classificatie van 'Redelijk druk', 'Niet te druk', 'Drukker dan normaal' moet nog worden verfijnd om dezelfde aanduiding als Google te presenteren. De exacte formule achter 'bij welke actuele waarde' en 'welk verschil in gepeilde en verwachte waarde' een drukte classificatie hoort is ons (nog) niet bekend.

¹⁰ Het ophalen van locatiegegevens van 600 locaties met de Google API en het script duurt ca. 8 tot 10 minuten; de Google API rekent kosten per x aantal requests.

¹¹ Het is niet bekend of dit getal een aggregatie is van het aantal gepeilde Google devices die locatiedelen hebben aanstaan is, dus een absoluut aantal, maar naar alle waarschijnlijkheid wel. De waarde fluctueerde ergens tussen de 20 en soms ver boven de 100, dus het is dus geen percentage. Let wel: gebruikers die de functie 'locatiedelen' hebben uitstaan en andere devices of personen zonder devices worden niet gepeild.

¹² Het zou dus kunnen zijn dat het device niet in deze voorziening, maar in de voorziening ernaast zich bevindt. In ieder geval wordt op basis van locatiegeschiedenis en actuele locatiegegevens van het device, het device aan deze locatie of voorziening gerelateerd. Google valideert de eigen waarnemingen door gebruikers achteraf een bericht te sturen met oproepen in de trant van 'Hoe was het hier?' of 'Geef jouw beoordeling over deze plaats...'.
¹³ Zie <https://github.com/m-wrzt/populartimes/issues/88> en <https://cloud.google.com/maps-platform/terms#3.-license->.