



Bedrijven QuickScan Tool (BQST)

Functionele specificaties bij BQST versie 4

januari 2025



januari 2025

Steller: M. van Loon
Rol: Product Owner
Organisatie: RIEC Oost-Brabant en RIEC Zeeland West-Brabant
(voorheen Taskforce-RIEC Brabant-Zeeland)
Contact: Michiel.van.Loon@taskforce-riec.nl

Versie document v1.0
Versie BQST tool v4.0

Inhoudsopgave

INLEIDING	3
<i>Functionele Specificaties.....</i>	<i>3</i>
1 OPHALEN ALLE ADRESSEN VAN INGEVOERDE POSTCODE GEBIED(EN)	4
2 PER ADRES OPHALEN INGESCHREVEN ONDERNEMINGEN BIJ DE KAMER VAN KOOPHANDEL	4
3 PER ONDERNEMING DE WEBSITE VASTSTELLEN.....	4
3.1 WEBSITE VASTSTELLEN VIA KvK-INSCHRIJVING.....	5
3.2 WEBSITE VASTSTELLEN VIA GOOGLE PLACES API.....	5
3.3 WEBSITE VASTSTELLEN VIA GOOGLE SEARCH API	5
4. HANDMATIG AANPASSEN VAN DEZE GEGEVENS (VIA CONTROLEBESTAND).	5
<i>Controle en correctie van websites</i>	<i>6</i>
5. HET LLM STELT BEDRIJFSACTIVITEIT VAST EN VERGELIJKT DEZE	6
5.1 WEBSITE ONDERNEMING OPENEN EN VASTSTELLEN BEDRIJFSACTIVITEIT	6
5.2 EXTRACTIE EN SAMENVATTING VAN BEDRIJFSACTIVITEIT	7
5.3 VERGELIJKING MET SBI-CODES.....	7
5.4 KEUZE VAN SEMANTISCHE MATCHES.....	7
5.5 TOEWIJZING VAN SBI-CODES.....	7
DE GESELECTEERDE SBI-CODES WORDEN GEKOPPELD AAN DE ONDERNEMING EN GEBRUIKT VOOR VERDERE ANALYSE. HIERBIJ WORDT BEOORDEELD OF DE VASTGESTELDE BEDRIJFSACTIVITEIT OVEREENKOMT MET DE BIJ DE KvK GEREgistREERDE SBI-CODE EN OF ER SPRAKE IS VAN MOGELIJKE RISICO-INDICATOREN.	7
5.6 OVER HET GEBRUIK VAN HET LARGE LANGUAGE MODEL	8
6. WEERGAVE VAN ADRESSEN IN KAART	8
KAARTLAAG 1: ONTBREKEN WEBSITE ONDERNEMING	8
KAARTLAAG 2: GEREgistREERDE BEDRIJFSACTIVITEIT IN KvK KWETSBAAR VOOR ONDERMIJNING.....	9
KAARTLAAG 3: BEDRIJFSACTIVITEIT VOLGENS DE WEBSITE KWETSBAAR VOOR ONDERMIJNING	9
KAARTLAAG 4: BEDRIJFSACTIVITEIT GEPRESENTEERD OP WEBSITE KWETSBAAR VOOR ONDERMIJNING EN KOMT NIET OVEREEN MET KvK.....	10
KAARTLAAG 5: ADRES IN PAND MET KENMERKEN DOMICILIE	10
KAARTLAAG 6: VERHOOGDE KANS OP LEEGSTAND.....	10
KAARTLAAG 7: AFWIJKEND GEBRUIKSDOEL ADRES	10
7. WEERGAVE IN EXCEL BESTAND	11
INFORMATIE OVER ANALYSE	11
ONDERNEMING OUTPUT	11
8. BEPERKINGEN EN BEKENDE ONNAUWKEURIGHEDEN IN DE ANALYSE.....	12

Inleiding

Dit document biedt inzicht in de functionele werking van de Bedrijven QuickScan Tool (BQST). De tool is ontwikkeld door Taskforce-RIEC Brabant-Zeeland.

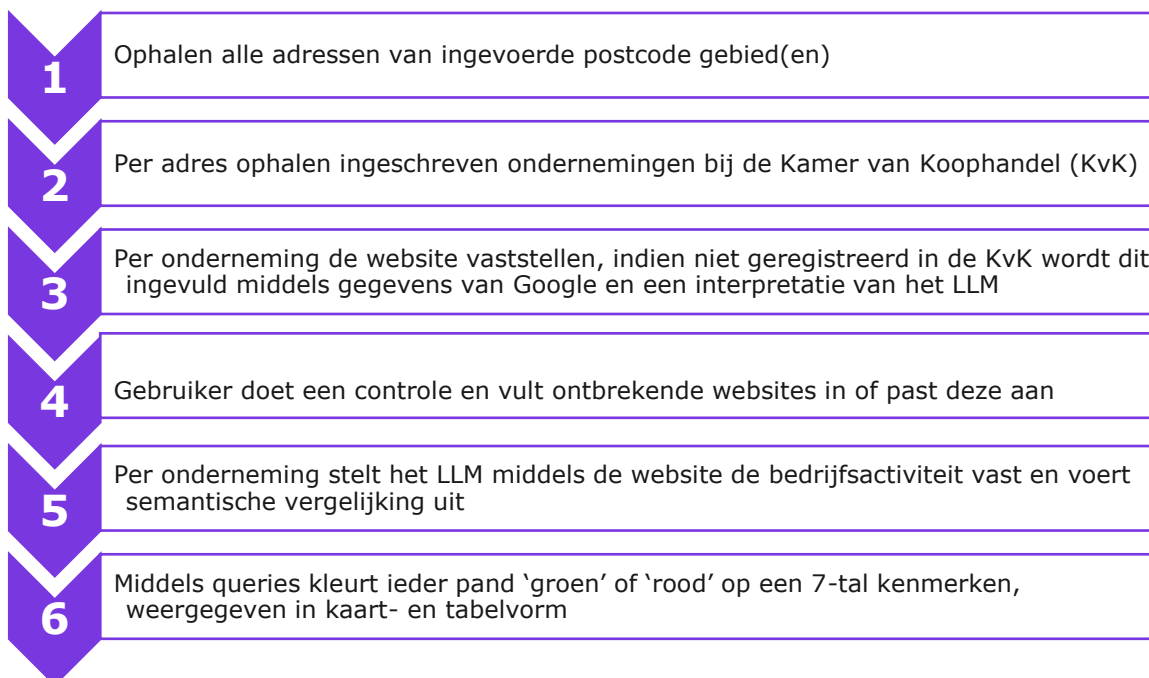
De tool gebruikt een postcodegebied als input. Vervolgens brengt de tool gegevens van verschillende bronnen samen, analyseert deze door onder andere cross-checks uit te voeren, en presenteert de resultaten in de vorm van een kaart (abc-MAP.html) in de webbrowser en een corresponderende Excel-tabel (abc-DATA.xlsx).

De kaart bestaat uit zeven kaartlagen met individuele panden en adressen, die met behulp van kleuren (groen-rood) aangeven of een pand bepaalde kenmerken al dan niet bezit. De tabel bevat per pand aanvullende informatie.

Dit document is bedoeld voor degenen die een beter begrip willen krijgen van de werking van de tool. Dit document is geen werkinstructie, maar kan wel helpen om de uitkomsten beter te begrijpen en op waarde te schatten. De tool doorloopt enkele stappen. De eerste stappen zijn gebaseerd op gestandaardiseerde overheidsgegevens met voorspelbare output. De latere stappen maken gebruik van een Large Language Model (LLM), een AI-model, waarbij de uitkomsten minder voorspelbaar zijn en mogelijk onverwachte uitkomsten kunnen ontstaan, zogenaamd fantoomgedrag van een AI.

Functionele Specificaties

De BQST doorloopt de volgende stappen en analyses.



1 Ophalen alle adressen van ingevoerde postcode gebied(en)

Nadat de gebruiker een postcode of een serie van postcodes heeft ingevoerd, worden van deze postcodes de adressen opgehaald uit een data dump van de BAG.

De datadump van de BAG is afkomstig van <https://www.kadaster.nl/zakelijk/producten/adressen-en-gebouwen/bag-2.0-extract>, deze dump is omgezet naar een relationele datastructuur met behulp van https://github.com/digitaldutch/BAG_parser. De datadump die ten tijde van het opstellen van dit document is gebruikt, dateert van 17 juni 2024.

2 Per adres ophalen ingeschreven ondernemingen bij de Kamer van Koophandel

Per adres wordt via de KvK API opgehaald welke bedrijven daar op dat adres zijn ingeschreven. Er worden per adres nul, één of meer inschrijvingen geretourneerd.

Indien er nul inschrijvingen zijn op een adres, worden er geen gegevens geretourneerd. Op deze manier worden alleen adressen meegenomen in de analyse waar volgens de KvK een bedrijf is gevestigd.

De specificaties van de KvK API zijn beschreven in de documentatie van de Zoeken API, te vinden op <https://developers.kvk.nl/nl/documentation/zoeken-api>. Voor het vestigingsnummer dat terugkomt uit de Zoeken API worden aanvullende gegevens ontsloten via de documentatie van de Vestigingsprofiel API, beschikbaar op <https://developers.kvk.nl/nl/documentation/vestigingsprofiel-api>.

3 Per onderneming de website vaststellen

Per ingeschreven onderneming wordt vastgesteld wat de website is. Allereerst wordt de website gebruikt die is opgegeven bij inschrijving in de KvK. Indien daar geen website bekend is, wordt Google Places api gebruikt met een specifieke zoekopdracht. Indien daar geen **geschikte match is**, wordt Google Search api in combinatie met het LLM gebruikt om de website vast te stellen. Deze stappen worden hier in meer detail beschreven.

3.1 Website vaststellen via KvK-inschrijving

Als een website aanwezig is binnen de KvK-registratie, wordt deze gebruikt.

3.2 Website vaststellen via Google Places API

Via de Google Places API wordt een zoekopdracht uitgevoerd, dit is vergelijkbaar met zoeken via de website Google Maps (<https://www.google.nl/maps/>). Specifiek wordt gebruik gemaakt van het coördinaat (lat lon) in combinatie met de naam of namen van de onderneming. Het coördinaat is afkomstig van de BAG. De namen zijn terug te vinden zijn in kolom 'namen' van de Excel-output.

Indien Google Places bedrijfsregistraties vindt met een overeenkomstige naam en locatie, dan vindt hier een check en keuze plaats. Hiervoor wordt een algoritme gebruikt, specifiek 'levenshtein' met een vooraf ingestelde marge (40%). De website van de onderneming die het beste past wordt gebruikt. Indien er geen website is gevonden die valt binnen de marge, wordt overgegaan naar de volgende stap.

3.3 Website vaststellen via Google Search API

Met de Google Search API wordt een zoekopdracht uitgevoerd naar de combinatie van de bedrijfsnaam en de plaats waar het bedrijf is ingeschreven. Deze zoekopdracht is vergelijkbaar met zoeken via Google (<https://www.google.nl/>). De eerste 10 zoekresultaten worden geanalyseerd met behulp van het LLM, specifiek het model GPT-4-o van OpenAI. Het model kiest het meest relevante resultaat.

De prompt die hiervoor gebruikt wordt, is als volgt:

*System prompt: Jij bent een behulpzame assistent die alleen antwoordt in JSON formaat.
User prompt: Hier volgt tussen haakjes de zoekresultaten van Google voor '{name}':
[{results}]. Het is aan jou om op basis van deze resultaten vast te stellen
of er één van deze websites de officiële website is van '{name}'.
Als je de officiële site vindt geef dan de bijbehorende link terug.
Als je de officiële site niet vindt, geef je een lege string terug.
In ieder geval dien je een JSON terug te geven, met één key: 'link' *
Voorbeeld: {"link": "https://www.example.nl"}.

Indien er geen website is gevonden, dan blijft de cel in Excel leeg.

4. Handmatig aanpassen van deze gegevens (via controlebestand).

Na stap 3 wordt een Excel-tabel gegenereerd en ter download aangeboden. Deze tabel bevat een selectie van de KvK-gegevens en vormt de basis voor de verdere analyses. Het is de verantwoordelijkheid van de

gebruiker om deze gegevens zorgvuldig te controleren en waar nodig aan te passen. Zie ook DPIA en IAMA.

Controle en correctie van websites

Vanwege de werkwijze van de tool kunnen onjuiste websites aan bedrijven worden gekoppeld. Daarom is het essentieel dat de gebruiker verifieert of de toegewezen website daadwerkelijk bij de onderneming hoort. Indien een website niet correct is, dient deze te worden verwijderd of handmatig vervangen door de juiste website.

Daarnaast kunnen ontbrekende websites worden toegevoegd en kunnen bedrijven indien nodig uit de tabel worden verwijderd. De aangepaste dataset wordt vervolgens gebruikt als input voor de vervolganalyses.

Kolomkoppen in de tabel

<i>Vestigingsnummer, kvkNummer, rsin, indNonMailing, statutaireNaam, eersteHandelsnaam, indHoofdvestiging, indCommercieleVestiging, voltijdWerkzamePersonen, totaalWerkzamePersonen, deeltijdWerkzamePersonen, materieleRegistratie, datumAanvang, formeleRegistratiedatum, namen, huisnummer, postcode, toevoeging, plaats, lat, long, website, sbi_all</i>
--

5. Het LLM stelt bedrijfsactiviteit vast en vergelijkt deze

Het Large Language Model (LLM) analyseert per onderneming de bedrijfsactiviteiten op basis van informatie die online beschikbaar is. Dit gebeurt in meerdere stappen, waarbij de activiteiten worden geïdentificeerd, samengevat en vergeleken met de geregistreerde gegevens in de Kamer van Koophandel (KvK). Op basis van deze analyse worden SBI-codes toegekend en vergeleken met kwetsbare bedrijfsactiviteiten.

5.1 Website onderneming openen en vaststellen bedrijfsactiviteit

Het LLM opent de vastgestelde website van de onderneming en zoekt naar de meest relevante informatie over de bedrijfsactiviteiten. Dit gebeurt door gerichte analyse van pagina's zoals *Over ons*, *Onze diensten* of *About us*. Indien deze informatie niet direct beschikbaar is, maakt het model een sitemap van de website om alsnog de meest geschikte pagina te identificeren.

Als de website geen of onvoldoende bruikbare informatie bevat, voert het LLM een aanvullende websearch uit. Dit gebeurt anoniem via verschillende zoekmachines, zoals Google en Bing, en via bedrijfsprofielen op sociale media zoals Facebook en Instagram. Dit zorgt voor een zo volledig mogelijk beeld van de bedrijfsactiviteiten.

In sommige gevallen kan het model beperkte of inconsistente informatie aantreffen, bijvoorbeeld door slecht functionerende websites of verouderde content. Hierdoor kan het voorkomen dat de bedrijfsactiviteit slechts summier wordt vastgesteld, wat de nauwkeurigheid van de verdere analyse kan beïnvloeden.

5.2 Extractie en samenvatting van bedrijfsactiviteit

Zodra de website is geanalyseerd, haalt het LLM de relevante informatie over de bedrijfsactiviteiten op en genereert het een samenvatting. Deze samenvatting wordt tijdelijk opgeslagen en na de analyse verwijderd. De broninformatie is na de analyse niet meer beschikbaar.

5.3 Vergelijking met SBI-codes

De gegenereerde samenvatting wordt vervolgens vergeleken met de officiële omschrijvingen van SBI-codes. De SBI-beschrijvingen zijn lokaal opgeslagen in een vectordatabase, waardoor semantische overeenkomsten efficiënt kunnen worden geïdentificeerd.

5.4 Keuze van semantische matches

Het LLM analyseert de tekstuele beschrijving van de bedrijfsactiviteit en bepaalt welke SBI-codes het best overeenkomen. Dit gebeurt op basis van semantische vergelijkingen, waarbij het model niet alleen exacte overeenkomsten zoekt, maar ook inhoudelijke gelijkenissen herkent.

5.5 Toewijzing van SBI-codes

De geselecteerde SBI-codes worden gekoppeld aan de onderneming en gebruikt voor verdere analyse. Hierbij wordt beoordeeld of de vastgestelde bedrijfsactiviteit overeenkomt met de bij de KvK geregistreerde SBI-code en of er sprake is van mogelijke risico-indicatoren.

Het gebruikte taxonomische model met SBI-codes en omschrijvingen is openbaar beschikbaar via: <https://sbi.cbs.nl/CBS.TypeerModule.TypeerServiceWebAPI/content/angular/app/#/>.

De prompt die gebruikt wordt is als volgt:

Het is jouw taak vast te stellen wat de bedrijfsactiviteit is van {organisation}. Deze dien jij vast te stellen op basis van de website van {organisation}, te vinden op {link}. Open deze website en stel vast of je op basis van deze website de bedrijfsactiviteit van {organisation} kan vaststellen. Als dat niet kan op basis van deze website, maak dan een sitemap van de website en gebruik deze sitemap om de pagina te vinden die waarschijnlijk de meeste informatie bevat over de bedrijfsactiviteit van {organisation}. Dit is waarschijnlijk een pagina met als titel bijvoorbeeld 'Over ons', 'Over {organisation}', 'Onze missie' of 'About us'. Als je deze pagina hebt gevonden, stel vast of je op basis van deze tekst de bedrijfsactiviteit van {organisation} kunt vaststellen. Als dit niet lukt op basis van de gekozen webpagina, kies dan een andere website

uit de sitemap en probeer het nog een keer. Als het na 3 pogingen nog niet lukt, probeer dan de volgende site: {link}. Als het na deze poging nog niet lukt, geef dan een lege string terug.
Als dit wel lukt, zoek dan mogelijk relevante SBI-codes op en bepaal vervolgens van deze SBI-codes welke er ook daadwerkelijk te relateren zijn aan de eerder geextraheerde bedrijfsactiviteit.
Als je klaar bent geef alleen een JSON terug te geven met twee keys:
\n -\'activity\': De geextraheerde bedrijfsactiviteit.
\n -\'sbicodes\': Een lijst met de meest relevante SBI-codes die te relateren zijn aan de geextraheerde bedrijfsactiviteit.

5.6 over het gebruik van het Large Language Model

Voor deze analyse wordt het GPT-4-o model gebruikt van het bedrijf OpenAI. Indien er wijzigingen komen in het model die invloed heeft op de functionaliteit, worden gebruikers hiervan op de hoogte gebracht.

Het model bezoekt websites en voert websearches uit via anonieme zoekopdrachten. Hierbij worden resultaten opgehaald uit diverse zoekmachines en webindexen, zonder afhankelijk te zijn van één enkele bron. Ook openbare bedrijfsprofielen op sociale media zoals Facebook en Instagram kunnen in de analyse worden meegenomen.

Door deze methode ontstaat een zo compleet mogelijk beeld van de bedrijfsactiviteiten, maar kunnen er beperkingen optreden, bijvoorbeeld wanneer websites weinig inhoud bevatten of niet toegankelijk zijn voor het model. Dit kan leiden tot afwijkingen tussen de vastgestelde en geregistreerde bedrijfsactiviteiten.

6. Weergave van adressen in kaart

De analyse wordt visueel gepresenteerd via een interactieve kaart met zeven kaartlagen, waarop panden rood of groen worden weergegeven. Deze kaartlagen zijn ook beschikbaar in het bijbehorende Excel-bestand. De eerste vier kaartlagen richten zich op kenmerken met betrekking tot de onderneming, terwijl de laatste drie kaartlagen zich focussen op adreskenmerken.

Kaartlaag 1: Ontbreken website onderneming

Query: Indien er geen website gevonden kan worden voor een onderneming, wordt deze indicator op 'ja' ingesteld.

Uitleg: Een pand bevat geen, één of meerdere adressen. Panden worden als vlakken weergegeven, terwijl adressen als punten binnen het pandvlak staan. Op een adres kunnen geen, één of meerdere ondernemingen bij de KvK als vestiging zijn ingeschreven. Als een onderneming op een adres geen website heeft, kleurt het adres (punt) rood. Daarnaast wordt vermeld hoeveel ondernemingen op dat adres geen website hebben. Het pand (vlak) kleurt hierdoor ook rood. Groene panden bevatten enkel adressen van ondernemingen met een vindbare website.

Kanttekening: Panden kunnen onterecht groen kleuren, bijvoorbeeld wanneer het vestigingsadres niet voorkomt in de BAG of wanneer er geen KvK-inschrijving is, terwijl er wel bedrijfsactiviteiten plaatsvinden. Dit komt voor bij bedrijven met een hoofdkantoor op een andere locatie, meerdere vestigingen of een foutieve KvK-registratie. Een veelvoorkomende fout is de verwisseling van het vestigingsadres en correspondentieadres.

Kaartlaag 2: Geregistreeerde bedrijfsactiviteit in KvK kwetsbaar voor ondermijning

Query: Als een SBI-code van een onderneming voorkomt op de gebruikte lijst met kwetsbare bedrijfsactiviteiten, wordt deze indicator op 'ja' ingesteld.

Uitleg: Het adres (punt) waar dit op van toepassing is kleurt rood, evenals het pand (vlak). Kwetsbare bedrijfsactiviteiten voor ondermijning zijn vastgesteld op basis van interviews met domeinexperts, criminologische literatuur en bestaande risicolijsten, zoals die van de Financial Action Taskforce, het Landelijk Dreigingsbeeld 2017, Graydon en Zicht op Ondermijning. Deze lijst is in 2022 samengesteld door Avans Hogeschool en Centerdata, in opdracht van de provincies Noord-Brabant en Gelderland.

Voor ondermijning kwetsbare bedrijfsactiviteiten zijn opgesteld op basis van interviews met domeinexperts, criminologische literatuur en bestaande risicolijsten (o.a. van de Financial Action Taskforce, het Landelijk Dreigingsbeeld 2017, Graydon en Zicht op Ondermijning). De lijst is samengesteld door Avans Hogeschool en Centerdata in opdracht van de provincies Noord-Brabant en Gelderland in 2022.

Kanttekening: Kwetsbare bedrijfsactiviteiten zijn bepaald op basis van een statische analyse van bedrijventerreinen, terwijl deze tool werkt op adresniveau. Daarnaast wordt onderscheid gemaakt tussen drie typen ondermijnende criminaliteit, die elk op een andere manier kwetsbaarheid aanduiden. Dit onderscheid wordt in de tool niet meegenomen, wat kan leiden tot een vertekend beeld.

Kaartlaag 3: Bedrijfsactiviteit volgens de website kwetsbaar voor ondermijning

Query: Als de door het LLM vastgestelde bedrijfsactiviteit van een onderneming voorkomt op de gebruikte lijst met kwetsbare bedrijfsactiviteiten, wordt deze indicator op 'ja' ingesteld.

Uitleg: Het adres kleurt rood indien van toepassing, evenals het pand. De bedrijfsactiviteiten zijn geïnterpreteerd door het LLM, waarbij deze de webpagina achterhaalt met de meest waarschijnlijke informatie over de bedrijfsactiviteit. De herkomst van de kwetsbare bedrijfsactiviteiten zijn beschreven bij kaartlaag 2.

Kanttekening: De vaststelling van bedrijfsactiviteiten door het LLM is afhankelijk van de beschikbare website-informatie en de werking van een LLM. Indien een website het LLM blokkeert of slechts beperkte inhoud heeft, kan de analyse onvolledig zijn en kunnen uitkomsten variëren (zie Hoofdstuk 5).

Kaartlaag 4: Bedrijfsactiviteit gepresenteerd op website kwetsbaar voor ondermijning en komt niet overeen met KvK

Query: Als de door het LLM vastgestelde bedrijfsactiviteiten voorkomen op de gebruikte lijst met kwetsbare bedrijfsactiviteiten, maar de bij de KvK geregistreerde bedrijfsactiviteiten niet, wordt deze indicator op 'ja' ingesteld.

Uitleg: De bedrijfsactiviteiten zijn geïnterpreteerd door het LLM, zie ook voorgaande kaartlagen. Wanneer de door het LLM vastgestelde bedrijfsactiviteiten overeenkomen met de lijst van kwetsbare bedrijfsactiviteiten, maar de KvK-inschrijving niet, wordt dit als afwijkend geregistreerd. De onderneming(en) waar dit op van toepassing is, wordt weergegeven als stip binnen het pandvlak. De criteria voor kwetsbare bedrijfsactiviteiten zijn dezelfde als beschreven bij kaartlaag 2.

Kanttekening: Dezelfde kanttekening als bij Kaartlaag 3 is van toepassing; er zitten beperkingen aan het vaststellen van de bedrijfsactiviteiten via de website door het LLM.

Kaartlaag 5: Adres in pand met kenmerken domicilie

Query: Indien de oppervlakte per onderneming op een adres minder dan 25 m² bedraagt, wordt deze indicator op 'ja' gezet.

Uitleg: Het vloeroppervlak per onderneming wordt berekend door het aantal ondernemingen op een adres te delen door het totale vloeroppervlak van het verblijfsobject (volgens de BAG). Het gehanteerde uitgangspunt is dat bij een vloeroppervlakte van minder dan 25 m² per onderneming sprake kan zijn van domicilieverlening. Dit houdt in dat een zakelijk adres (zoals een kantoor- of postadres) wordt gebruikt door partijen die er niet fysiek gevestigd zijn. De ondernemingen waar dit op van toepassing is, worden weergegeven als stip binnen het pandvlak. Indien sprake is van domicilieverlening, wordt ook het aantal betrokken ondernemingen vermeld.

Kaartlaag 6: Verhoogde kans op leegstand

Query: Als een adres volgens de BAG een bedrijfsfunctie heeft, maar er geen onderneming op dat adres is ingeschreven, wordt deze indicator op 'ja' ingesteld.

Uitleg: Er wordt gecontroleerd of het gebruiksdoel van het adres volgens de BAG een bedrijfsfunctie aangeeft. Indien er geen ondernemingen op het adres staan ingeschreven, wordt aangenomen dat het mogelijk leegstaat. Adressen met een verhoogde kans op leegstand worden weergegeven als stip binnen het pandvlak. De analyse wordt uitgevoerd op negen van de elf BAG-gebruiksdoelen; woon- en celfuncties worden uitgezonderd.

Kaartlaag 7: Afwijkend gebruiksdoel adres

Query: Als een bedrijf is ingeschreven op een adres waarvan het gebruiksdoel volgens de BAG geen bedrijfsmatige functie heeft, wordt deze indicator op 'ja' ingesteld.

Uitleg: De veronderstelling is dat wanneer een adres een gebruiksdoel heeft dat niet geschikt is voor bedrijfsmatig gebruik, ondernemingen hier mogelijk niet passend gevestigd zijn. Er wordt gecontroleerd of het gebruiksdoel van het pand volgens de BAG geen bedrijfsfunctie heeft, terwijl er wel een onderneming op dat adres is ingeschreven. Dit kan duiden op afwijkend gebruik. De BAG onderscheidt elf gebruiksdoelen. In deze analyse wordt specifiek gekeken of een bedrijf is ingeschreven op een adres met een woon- of celfunctie. Adressen met een afwijkend gebruiksdoel worden weergegeven als stip binnen het pandvlak.

7. Weergave in Excel bestand

De resultaten van de analyse worden gepresenteerd in een interactieve kaart en een bijbehorende Excel-tabel. De kaart is bedoeld voor weergave in de webbrowser, terwijl de Excel-tabel een gedetailleerde dataset bevat. Het Excel-bestand bestaat uit meerdere tabbladen:

Informatie Over Analyse

Bevat metadata over de analyse, inclusief de onderzochte postcodegebieden en aantallen.

Onderneming Output

Bevat de analyseresultaten met betrekking tot ondernemingen en correspondeert met kaartlagen 1, 2, 3 en 4. De beschrijvingen van deze kaartlagen zijn opgenomen in voorgaande secties. Hiernaast is in de aanvullende een ruwe data-export vanuit de KvK.

De eerste reeks kolommen is afkomstig uit het handelsregister (KvK):

EersteHandelsnaam – De geregistreerde handelsnaam van de onderneming.
Namen – Alle bekende namen bij inschrijving in de KvK.
Website – De vastgestelde website, afkomstig uit de KvK-registratie, vastgesteld via Google en het LLM, of handmatig ingevoerd.
Straatnaam, huisnummer, huisletter, toevoeging, postcode, plaats – Adresgegevens.
sbi_all – De opgegeven bedrijfsactiviteiten volgens de KvK.

Kolommen gekoppeld aan de kaartlagen:

Ontbreken website onderneming – Bron voor kaartlaag 1.
Geregistreerde bedrijfsactiviteit in KvK kwetsbaar voor ondermijning – Bron voor kaartlaag 2.
Bedrijfsactiviteit volgens de website kwetsbaar voor ondermijning – Bron voor kaartlaag 3.
Bedrijfsactiviteit gepresenteerd op website kwetsbaar voor ondermijning en komt niet overeen met KvK – Bron voor kaartlaag 4.

Overige kolommen:

KVK risico SBI Code – De SBI-code opgegeven bij inschrijving, indien deze voorkomt op de lijst zoals beschreven bij kaartlaag 2.
Website bron – De herkomst van de vastgestelde website (KvK, Google in combinatie met het LLM, of handmatige invoer).
Website risico SBI Code – De bedrijfsactiviteiten, uitgedrukt in SBI-codes, zoals vastgesteld door het LLM op basis van de website.
SBI beschrijving – De officiële omschrijving van de SBI-code(s) zoals bekend bij de KvK.

Adres Output

Geeft het resultaat van de analyses van de adressen, behorende bij kaartlaag 5, 6 en 7. De beschrijvingen van deze kaartlagen zijn opgenomen in voorgaande secties. Hiernaast is in de aanvullende een ruwe data-export vanuit de KvK.

De eerste reeks kolommen is afkomstig uit de BAG:

openbare_ruimte_id – in plaats van id is hier straatnaam ingevuld. huisnummer, huisletter, toevoeging, postcode, woonplaats_id – adresgegevens. gebruiksdoel oppervlakte bouwjaar

Kolommen gekoppeld aan de kaartlagen:

Adres kenmerken domicillie – Bron voor kaartlaag 5. Verhoogde kans op Leegstand– Bron voor kaartlaag 6. Afwijkend gebruiksdoel – Bron voor kaartlaag 7.

8. Beperkingen en bekende onnauwkeurigheden in de analyse

Door inconsistenties in de basisregistraties, zoals de KvK en de BAG, kunnen bedrijven soms aan het verkeerde adres worden gekoppeld. Dit kan leiden tot onjuiste verbanden tussen gegevens en ervoor zorgen dat locaties onterecht als 'rood' worden gemarkeerd.

Verschillen tussen KvK en BAG-adressen: De KvK maakt niet altijd gebruik van authentieke BAG-adressen. Hierdoor kunnen mismatches ontstaan door verschillen in spaties, hoofdletters en het gebruik van huisletters of toevoegingen.

Afwijkende inschrijvingen bij de KvK: Bedrijven staan regelmatig bij de KvK ingeschreven op een ander adres dan hun daadwerkelijke bezoekadres. Dit kan ervoor zorgen dat de postcode niet overeenkomt met het KvK-adres, waardoor een locatie ten onrechte als 'rood' of 'groen' wordt gemarkeerd.

Panden zonder verblijfsobject: Panden die in de BAG geen verblijfsobject en dus geen officieel adres hebben, worden niet weergegeven in de analyse en kunnen daardoor niet als 'rood' of 'groen' worden gemarkeerd.

Samen slimmer tegen ondermijnende criminaliteit

© 2024, Regionaal Informatie- en Expertisecentrum RIEC (uitgegeven in eigen beheer). Alle rechten voorbehouden. Niets uit deze uitgave mag zonder bronvermelding worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door print-outs, kopieën, of op welke andere manier zonder voorafgaande toestemming van de uitgever.

RIEC