

De Logistieke Voetafdruk 2024: Semi-Geautomatiseerde Monitoring van Distributievastgoed in Nederland

Scriptie Uitbreiding Steyn Bosch 2822302

Datum 12-5-2026

Begeleider: Eric Koomen

Inhoud

Doel van de Modelleer- en Programmeerwerkzaamheden	2
Beschrijving Ruwe Data	2
Methode en Verwerking	3
Voorlopige Resultaten.....	5
Reflectie op betrouwbaarheid en aannames	5
Conclusie en Gebruik in Scriptie.....	6
Bronnenlijst	7
Bijlagen	8
A. Visuele Verificatie Logistiek Vastgoed	8
B. Audit Tabel Resultaten	10
C. GIS/Google Interface voor Potentials Checken	11

Doel van de Modelleer- en Programmeerwerkzaamheden

Het hoofddoel van deze scriptie-uitbreiding is het actualiseren en waar het kan automatiseren van de Dutch Distribution Centers geodataset (Nefs, 2023). De oorspronkelijke set bood een overzicht van het logistieke complex tot en met november 2021, waarbij de economische typering van de panden was gebaseerd op LISA-data uit 2020. Gezien de sterke post-pandemische dynamiek (e-commerce boom) en de groeiende maatschappelijke discussie over het ruimtebeslag en de verdozing door distributiecentra, is een update van de set noodzakelijk.

De kern van deze uitbreiding ligt in het maken van een robuuste geautomatiseerde workflow in R. Hierin worden alle stappen waar mogelijk gescript, wat de wetenschappelijke reproduceerbaarheid van het onderzoek verhoogt.

Beschrijving Ruwe Data

Voor de actualisering van de monitor zijn drie primaire bronnen geïntegreerd:

- **Basisregistratie Adressen en Gebouwen (BAG, 2026):** Gebruikt voor de extractie van gebouwgeometrieën (footprint/voetafdruk), de bouwstatus, het oorspronkelijke bouwjaar en de verblijfsobjecten van alle panden groter dan 500m² (kleinere panden zijn er eerst uit gefilterd).
- **LISA (Vestigingsregister, 2024):** Deze bron levert de noodzakelijke economische data op bedrijfsniveau, waaronder het aantal banen en de typeringen via SBI-codes. Nu is alleen nog de 2024 data beschikbaar, als de 2025 data beschikbaar is wordt deze gebruikt. (Aangeleverd door EUR)
- **IBIS (Bedrijventerreinen 2026):** deze database vormt de basis voor de ruimtelijke afbakening van officiële bedrijventerreinen in Nederland. (Aangeleverd door EUR)

Daarnaast wordt de dataset van Nefs (Nefs, 2023) gebruikt als referentiekader om de handmatige werklust te beperken. Panden die nu als potential worden geïdentificeerd, maar in de dataset van 2021 zijn gevalideerd, worden automatisch overgenomen in de nieuwe kernvoorraad.

Methode en Verwerking

Het programmeerproces in R is opgedeeld in een gelaagde trechter die de ruwe brondata zuivert tot een betrouwbare voorraad. Dit gebeurt in 5 stappen:

Vorbewerking BAG:

vanwege de enorme omvang van de landelijke BAG (miljoenen objecten verdeeld over duizenden XML-files), zijn twee een voorafgaande stappen uitgevoerd buiten R zodat mijn laptop niet kan vastlopen:

- **Vectorisatie (GDAL/ogr2ogr):** In plaats van duizenden losse XML-bestandjes te openen, heb ik met een geautomatiseerd batch-script (.bat bestand) de rauwe XML-data in OSGeo4W Shell (de gespecialiseerde commandcentre van QGIS) samengevoegd tot een GeoPackage (.gpkg). Hierdoor kan R makkelijker direct naar de juiste plek springen wat uren aan rekenkracht bespaart in de vervolgstappen.
- **VBO-Extractie (PowerShell Regex):** Een gebouw (Pand) kan verschillende functies hebben, die in de BAG verblijfsobjecten (VBO) genoemd worden. Om in R later slimmer te kunnen filteren is het handig om ook dit per pand te weten. Voor de koppeling tussen VBO en Panden is een PowerShell-script (.bat) ontwikkeld, die deze data in een CSV-bestand heeft gezet.

Verwerking in R:

Nadat de BAG data door de vorbewerking is gegaan zijn er vijf stappen uitgevoerd in R om tot de uiteindelijke dataset te komen.

Stap 1: SQL-Filter en De-duplicatie (Script 1):

De eerste stap in R was het opschonen van de BAG-data. De gebouwenregistratie is een enorme dataset waarin de ruwe bestanden vaak de volledige administratieve geschiedenis van een pand staat, waardoor een gebouw meerdere versies kan hebben. Om dubbeltellingen te voorkomen, is er eerst gefilterd om de actuele status (“in gebruik” of “in bouw”). Door te kijken naar het ID van een pand is de ruis van de oudere versies verwijderd. Na deze schoonmaak zijn de VBO’s na de vorbewerking aan de panden toegevoegd om al een eerste beeld te krijgen van de functie. Tot slot is op basis van de exacte geometrie de oppervlakte berekend, waardoor elk pand direct al een officiële grootteklasse (S tot XXL) is ingedeeld.

Stap 2: LISA Koppeling (Script 2):

In de tweede fase is de economische laag over de fysieke gebouwen heen gelegd door de LISA-dataset (2024) te koppelen aan de overgebleven panden. Het doel hiervan is om

te bepalen of een gebouw daadwerkelijk een logistieke functie vervult. Hierbij zijn bedrijven geclassificeerd in drie hoofdgroepen:

- Groep A Handel, Import en Export
- Groep B Transport en Logistieke Diensten
- Groep C: XXL Retail Centers

gebaseerd op de Standaard Bedrijfsindeling (SBI) en de classificatie van Nefs (2023). Om onzuiverheden van de dataset te bewaken is een dominantieregel toegepast: een pand wordt pas als logistieke kernvoorraad beschouwd als meer dan 50% van de totale werkgelegenheid in het gebouw uit deze sector afkomstig is. Dit zorgt ervoor dat we bijvoorbeeld een fabriek met een kleine opslagruimte niet onterecht als distributiecentrum aanmerken.

Stap 3: Spookpanden Killer (Script 2):

Deze derde stap pakt een bekend probleem in ruimtelijke data aan: overlappende panden. Hoewel in theorie overlap in stap 1 eruit gefilterd wordt op actuele status blijkt dit in de praktijk nog niet genoeg te zijn om alle overlappende panden eruit te halen. Door een ruimtelijke overlap-check uit te voeren, identificeert het script deze spookpanden. In het geval van een overlap wordt automatisch het pand met het meest recente bouwjaar behouden. Deze stap is essentieel om een eerlijk beeld te krijgen van het werkelijke ruimtebeslag in het landschap zonder de groei te overschatten.

Stap 4: Potentials en Checken (Script 2 en 3):

De LISA data valideert een groot deel van de dataset, maar bestaat uit punten waardoor niet alle gebouwen meegenomen worden. Gebouwen die geen LISA punt bevatten, maar wel voldoen aan de ruimtelijke criteria (gelegen op IBIS-terrein en groter dan 2500m²), worden als potentials aangemerkt. Voordat deze panden gecheckt worden wordt er eerst gekeken of er overlap is met de gevalideerde set van 2021, als een potentieel gebouw toen als logistiek is geclassificeerd wordt dit overgenomen om zo de werklust te beperken. De gebouwen die verder overbleven zijn opgedeeld in 4 werksets en handmatig gecheckt op de volgende wijze (Bijlage A)

Stap 5: Samenvoegen (Script 4)

Na het checken worden de werksets weer samengevoegd en worden de panden die logistiek waren toegevoegd aan de al gevalideerde set om zo een dataset te creëren van de totale logistieke voetafdruk in Nederland.

Voorlopige Resultaten

Omdat de handmatige controle van de potentials nog in volle gang is, zijn de resultaten nog alleen van de gevalideerde set uit de geautomatiseerde workflow. Deze cijfers geven de eerste indicatie van de logistieke groei in Nederland. En zullen later nog worden aangevuld met de potentials en de 2025 LISA data.

Audit van de dataset (12 mei 2026, volledige audit Bijlage B.):

- Gevalideerde Kernvoorraad: 29.637 panden
- Lisa 2024: 20.001 panden
- Erfenis 2021: 9.636 panden
- Potentials op IBIS: 5.558 panden

In de bijlage is de volledige audit-tabel opgenomen, waarin de huidige kernvoorraad wordt vergeleken met een snapshot van hoe deze getallen in 2021 zouden zijn en de voorraad van 2021 uit de oude monitor.

Reflectie op betrouwbaarheid en aannames

De methodiek is ontworpen om een balans te vinden tussen automatiseren en precisie. Bij de interpretatie van de resultaten moet echter rekening worden gehouden met de volgende methodologische beperkingen en aannames:

Administratieve versus ruimtelijke zuiverheid:

Een fundamenteel punt uit de workflow is de eliminatie van administratieve dubbeltellingen via de 16-digit pand-ID's. Door deze te dwingen tot uniekheid, is zijn de historische versies uit de BAG verwijderd. Een belangrijke aanname hierbij is dat de geometrie met het grootste oppervlak binnen een ID-groep de meest actuele en representatieve weergave is van het fysieke gebouw. Hoewel dit administratief sluitend is, blijft de "Spookpanden-killer" uit stap 3 essentieel om ook ruimtelijke dubbeltellingen (verschillende ID's op dezelfde locatie) op te lossen.

Tijdsconsistentie en het verschil:

Er bestaat een onvermijdelijk verschil tussen de vastgoeddata van de BAG (2026) en de economische LISA data (2024). In een dynamische sector als de logistiek, waar distributiecentra in minder dan een jaar gerealiseerd kunnen worden, betekent dit dat panden gebouwd in 2025 of 2026 nog geen afspiegeling hebben in de huidige LISA-set. De aanname dat de functie van een pand tussen 2024 en 2026 constant blijft is noodzakelijk voor de stabiliteit van het model, maar kan leiden tot een lichte onderschatting van de meest recente verschuivingen in werkgelegenheid.

Kwalitatieve validatie van Potentials:

Een bewuste methodologische keuze is het opnemen van Potentials waarvoor geen LISA-data beschikbaar is. Dit heeft als beperking dat voor deze groep de validatie niet kwantitatief is op basis van banen als bij de kernvoorraad, maar kwalitatief en visueel is op basis van een korte handleiding (Bijlage A). Met als gevolg dat de dataset een hybride karakter heeft en niet voor elk pand alle variabelen beschikbaar zullen zijn. Deze keuze zorgt er wel voor dat de totale fysieke voetafdruk van de sector accurater zal worden.

Conclusie en Gebruik in Scriptie

De ontwikkelde workflow in R heeft aangetoond dat de monitoring van het Nederlandse logistieke complex te automatiseren is, mits er scherpe filters worden toegepast op de administratieve ruis. De uitbreiding leidt tot een dataset die niet alleen actueler is dan de versie uit 2021, maar bevat methodologische upgrades. Hierdoor kan deze een zuiver beeld geven van de schaa sprong naar XXL-panden.

Gebruik van de data

Deze eindrapportage vormt het officieel het sluitstuk van de data-fase. De potentials worden nog verder gevalideerd met een geüpdatete interface op basis van een Google API (Bijlage C). De uitkomsten hiervan zullen worden ingezet als input voor het Multinomiaal Logit Model (MNL). De zuivere onderverdeling in Grootteklassen (S tot XXL) stelt mij in staat om statistisch te toetsen of de ruimtelijke drijfveren, zoals bijvoorbeeld nabijheid van multimodale knooppunten, significant verschillen tussen de verschillende schaalniveaus van logistiek vastgoed.

Bronnenlijst

BAG. (2026, april 9). <https://service.pdok.nl/lv/bag/atom/bag.xml>

Nefs, M. (2023). *Dutch Distribution Centres 2021 Geodata* (Versie 2) [Geopackage, with copies in Shapefile, CSV and Geojson. Method in Pdf.]. 4TU.ResearchData.
<https://doi.org/10.4121/19361018.V2>

Bijlagen

A. Visuele Verificatie Logistiek Vastgoed

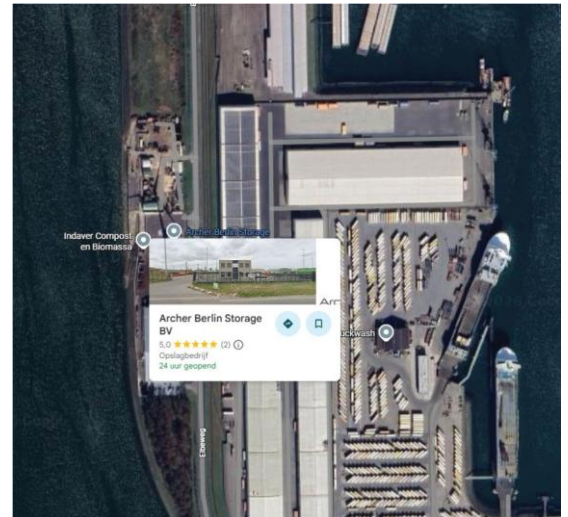
Categorieën

1. (Groot)handel, import, export
2. Transport, opslag en andere logistieke dienstverlening
3. XXL Retail
4. Tankopslag
5. Bulkopslag

Stap 1. Identificatie van bedrijf via Google Maps

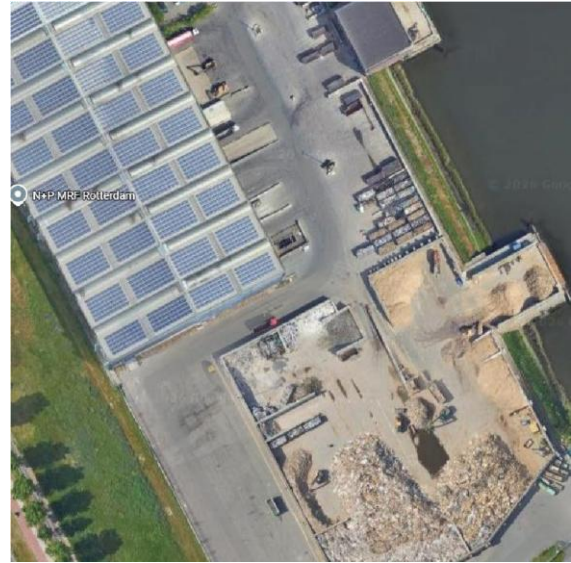
(Naam en link) of streetview (naam of logo op gevel)

- Naam van bedrijf in groothandel, e-commerce, import of export -> categorie 1
- Naam van bedrijf in transport, opslag, distributie of logistieke dienstverlening -> categorie 2
- Naam van winkel met zelfbedieningsmagazijn (Ikea, Gamma etc.) -> categorie 3
- Ronde tank met evt. logo van oliemaatschappij of tankopslag (Vopak) -> categorie 4
- Naam van bedrijf niet actief in bovenstaande activiteiten -> categorie 0 (te verwijderen) • Geen bedrijfsnaam -> ga naar Stap 2



Stap 2. Identificatie van activiteit via fysieke context

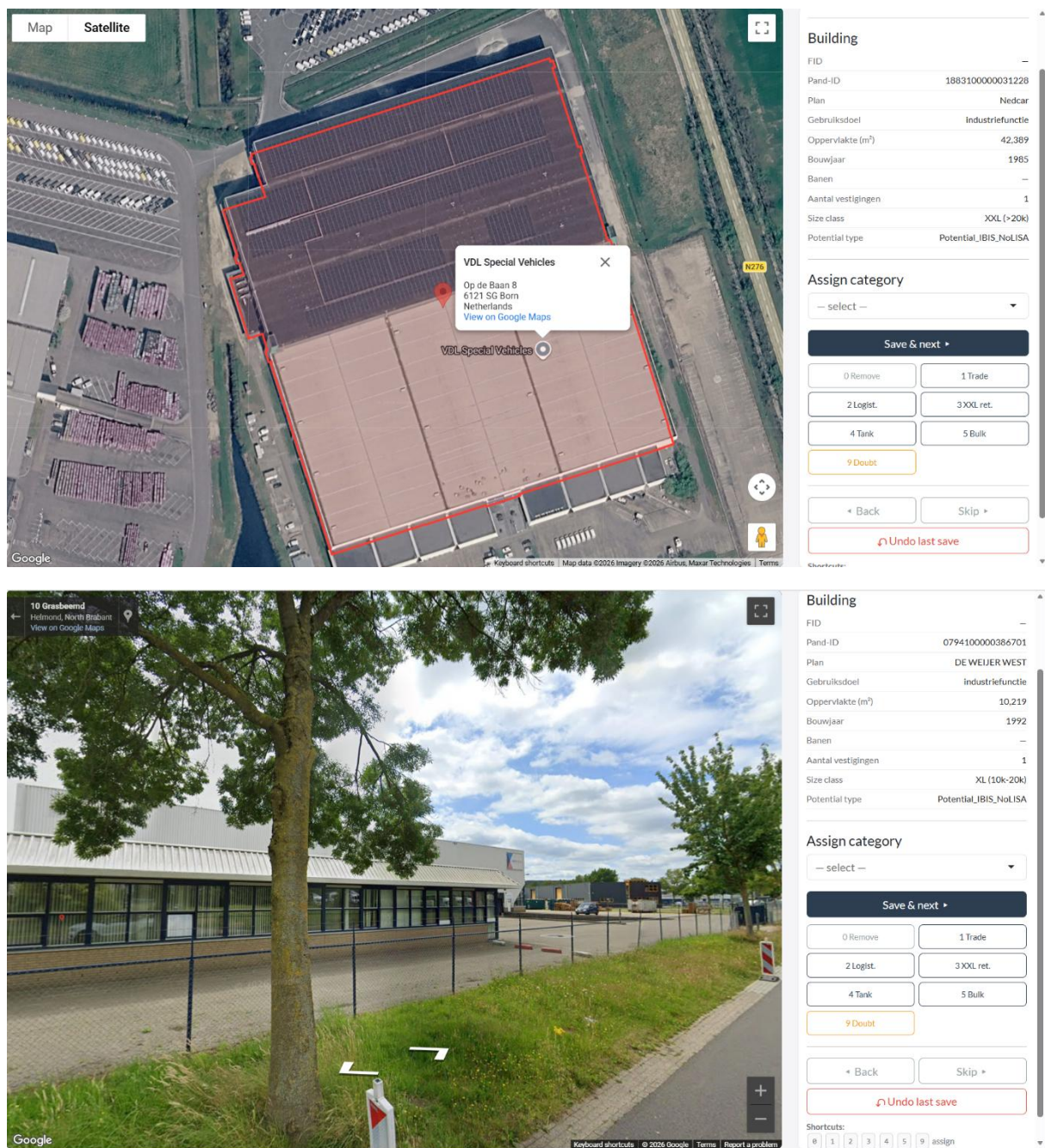
- Pand met opstelruimte voor trucks en weinig loading docks -> categorie 1
- Pand met opstelruimte voor trucks en veel loading docks -> categorie 2
- Pand met opslag van bulkmaterialen op terrein (Bv. voor recycling, geen bouw) -> categorie 5
- Pand zonder loading docks of bulkopslag -> categorie 0 (te verwijderen)
- Twijfel -> categorie 9 (2nd opinion)



B. Audit Tabel Resultaten

Bron & Functie	S (500<2,5K)	M (2,5k- 10k)	L (10k<20k)	XL (20k- 40k)	XXL (>40k)	Totaal
LISA 2024						
Groep A: Trade	8442	2363	1010	404	189	12408
Groep B: Transport	1557	518	474	347	314	3210
Groep C: Retail	2256	1336	464	159	50	4265
Erfenis 2021						
Groep A: Trade	4198	1055	399	133	32	5817
Groep B: Transport	894	759	480	372	151	2656
Groep C: Retail	887	278	86	27	3	1281
Potentials						
Nog te checken	0	3381	1330	598	249	5558
TOTAAL	18234	9690	4243	2.040	988	35195

C. GIS/Google Interface voor Potentials Checken



In de Interface kan je in- en uitzoomen, in QGIS en Street View. Daarnaast wordt alles automatisch opgeslagen in GeoPackage-formaat elke 1/2 minute terwijl jij clickt. Wel moet nog opgelost worden dat de FID hetzelfde blijft als die van het origineel, al is dat misschien niet eens nodig omdat alles al correct als GeoPackage wordt opgeslagen. Als een gebouw een bedrijf of iets is, kun je erop klikken. Dan word je naar Google Maps doorgestuurd wordt, zodat je extra informatie kunt bekijken.