

Power consumption Amsterdam

Data analyse en data visualisatie

KENAN AMATRADJAK, SOUFIANE BEN HADDOU, SATIGA GODRIE,
MIKKI RICHERT
Universiteit van Amsterdam
28 Juni, 2018

Inleiding

Wie aan Amsterdam denkt, denkt aan de hoofdstad van Nederland. Amsterdam daarentegen lijkt een land op zichzelf, als dichtsbevolkste stad in Nederland. Één van de meest multiculturele steden op de wereld, brengt het ook veel tourisme met zich mee. Hoewel de grote hoeveelheid mensen veel pluspunten met zich meebrengt, zoals de diversiteit en de grote werkgelegenheid, brengt het ook een ander heel groot punt met zich mee, namelijk veel stroom verbruik.

In dit project wordt er gekeken naar het stroom- en gasverbruik van Amsterdam. Waar wordt het meeste stroom en gas verbruikt en wat is de oorzaak hiervan. Hierbij wordt er gebruik gemaakt van de datasets van Liander, welke het jaarlijks verbruik van stroom en gas in Amsterdam bevatten. Deze informatie is relevant om te kijken waar en waardoor er veel energie verbruikt wordt. Om verklaringen te vinden voor de hoeveelheid energieverbruik en de veranderingen hierin, wordt er gekeken naar andere datasets die met Amsterdam te maken hebben. Hier is gekozen voor drie datasets. De eerste dataset heeft alle horecagelegenheden in Amsterdam geregistreerd op postcode en coördinaten. De tweede dataset is een dataset die het klimaat van Amsterdam bevat. De laatste set bevat het klimaat en de temperatuur per dag in Amsterdam. Hiervan zal het gemiddelde genomen moeten worden zodat deze vergeleken kan worden met het verbruik per jaar.

De extra gekozen datasets dienen het doel opmerkelijke data uit de Liander datasets aan te tonen door de correlatie tussen de onderlinge datasets te zoeken. Aan de hand van deze correlaties worden afwijkende resultaten opgemerkt en worden zodoende in kaart gebracht. De dataset over het klimaat toont inzicht in strengere en/of langere winters. Een gevolg van deze winters is dat het algehele verbruik van die jaren dus ook stijgt. Dit vanwege veel kachels die aan staan en het feit dat de dagen korter zijn, waardoor men eerder lichten aan zet.

Ook wordt er gekeken naar een correlatie tussen verschillende gebieden in Amsterdam. Dit houdt in dat er gekeken wordt naar de verschillende gebieden in Amsterdam, en hoe deze verschillen of juist hetzelfde zijn. Hier wordt verwacht dat in het centrum het stroomverbruik significant hoger is. De verklaring hiervoor is dat er veel horeca en tourisme in het centrum is waar van verwacht wordt dat er veel stroomverbruik mee gepaard gaat.

Verder wordt er gekeken naar de correlatie tussen stroom- en gasverbruik en of deze elementen dicht bij elkaar liggen. Gebruikt een postcode even veel gas als stroom?

1 Methode

Data pre-processing

De geanalyseerde dataset is afkomstig van Liander en omvat het stroom- en gasverbruik van heel het land van de afgelopen tien jaar. Om de dataset op te schonen zijn korte functies geschreven in de programmeertaal Python. Door middel van deze functies zijn de delen van de dataset die bruikbaar waren voor het beantwoorden van de vragen en het analyseren van de data behouden gebleven en is er een nieuw csv bestand gemaakt met alle bruikbare en opgeschoonde data. Door middel van het handmatig opschonen van de data worden fouten voorkomen. Zodoende kon er doorgewerkt worden met de opgeschoonde bestanden, in plaats van het elke keer opnieuw te specificeren welke data nuttig is. Hierbij zijn de straatnamen, postcodes, het gas- en elektriciteitsverbruik, het standaardjaarverbruik (SJV), het aantal aansluitingen per postcode en de teruggave geselecteerd. Ook zijn er aparte bestanden gemaakt van het gas- en elektriciteitsverbruik om deze los te kunnen analyseren. Er is voor gekozen om met de Liander datasets van 2009 tot en met 2016 te werken.

Als eerste stap is de data schoongemaakt, wat inhoudt dat de waarden die niet klopten zijn aangepast zodat dit bij het analyseren geen veranderingen zou opleveren. In het 'cleaning' proces hebben is er rekening gehouden met verschillende aspecten. De datasets bevatten veel data die voor dit project niet relevant is en dus verwijderd moet worden. Hieronder valt de woonplaats, deze is echter eerst gefilterd zodat de overgebleven rijen de gegevens van Amsterdam bevatten. Vervolgens kan deze kolom verwijderd worden. Ook de kolom met landcodes is verwijderd, gezien dit overal 'NL' was en irrelevant voor de analyses. Andere kolommen die uit de database gehaald zijn, zijn de meetverantwoordelijke, het netbeheer, het netgebied, het verbruikssegment, de definitieve aanslag, het soort aanslag, de soort aanslag naam, slimme meter en het gemiddeld aantal telwielen, omdat deze niet van toepassing zijn voor dit project.

Naast de datasets van Liander is er ook gebruik gemaakt van drie andere datasets, te weten een dataset met horecagelegenheden in Amsterdam, een dataset met postcodes en bijbehorende coördinaten van deze postcodes en een dataset die het klimaat in Amsterdam bevat. Hierna is de data geanalyseerd door middel van eigen geschreven functies in Python en is er gebruik gemaakt van de visualisatie bibliotheek Bokeh.

EDA basis

Bij het visualiseren van het verbruik is er gebruik gemaakt van een multivariate EDA. Dit is gedaan met een scatterplot. Meerendeel van de visualisaties zijn scatterplots.

1. Er zijn kaarten met het totale verbruik per postcode geplote, deze zijn opgedeeld in een kaart met het verbruik van gas en elektriciteit.
2. Ditzelfde is ook gedaan met het verbruik per aansluiting op de kaart geplote.
3. Daarnaast is er ook een kaart met de horecagelegenheden op de kaart geplote, deze is ook samen gevoegd met de eerder genoemde plots van het totale verbruik.
4. Voor het plotten van de klimaat grafieken is er gebruik gemaakt van een scatterplot die de temperatuur per dag weergeeft met een grafiek en de gemiddelde temperatuur per jaar met een punt aangeeft. Hier is ook de totale SJV per jaar als punt in geplote.
5. Ook is er een plot gemaakt die de temperatuur en het verbruik per dag voor het jaar 2009 als grafieken plot.
6. Het totale gas en het totale elektriciteit verbruik per postcode groep zijn bij elkaar opgeteld en zijn vervolgens als twee continue variabele tegen elkaar in een scatterplot geplote.
7. Tot slot is er een plot gemaakt die het totale verbruik per jaar plot.

De achtergrond van de scatterplot gebaseerd op coördinaten is een plattegrond van Amsterdam. De scatterplot plot de punten ten opzichte van de achtergrond met behulp van de coördinaten op de x- en y-as. De coördinaten van de postcodes zijn verkregen uit een postcode-coördinaten-bestand. De grootte en kleur van de geplote punten geeft de grootte van de SJV waarden aan. Van laag naar hoog worden de punten met licht groen tot rood aangegeven.

In-depth analysis

Voor de in-depth analysis wordt er gekeken naar de gevisualiseerde data zodat daar de resultaten uit afgeleid kunnen worden. Onder de gevisualiseerde data vallen de kaarten van het verbruik in Amsterdam en de horecagelegenheden. Als eerst hebben we het verbruik voor zowel elektriciteit als gas per postcode in één kaart verwerkt. Het werd al gauw duidelijk dat dit niet duidelijk was en de hoeveelheid postcodes voor onoverzichtelijkheid zorgde. Als oplossing is er overgegaan op clustering. Clustering houdt in dat er cluster centres worden bepaald waar de

dichtbijzijnde punten naartoe worden verschoven. De cluster centres werden gecreëerd door de letters van de postcodes te strippen, waardoor er groepen van vier-cijferige postcodes ontstaan. Hierdoor zijn de kaarten een stuk overzichtelijker geworden, waardoor de gebieden met meer stroom- en gasverbruik een stuk duidelijker zichtbaar waren. hebben we een scatterplot toegevoegd voor het klimaat grafiek, vervolgens hebben we het gemiddelde genomen om hier een duidelijke lijn te verkrijgen die makkelijker afleesbaar is.

2 Resultaten

De kaarten met de scatterplots tonen aan dat het hoogste SJV, wat inhoudt het gemiddeld energieverbruik per jaar van de aansluitingen per postcode, om het centrum heen zit. Het gaat hier om gebieden zoals Westpoort en Duivendrecht. Met behulp van het in kaart brengen van het energieverbruik in combinatie met de horecagelegenheden in één plot, is af te leiden dat bij de plaatsen waar veel horecagelegenheden zitten, namelijk in het centrum, er per aansluiting minder verbruikt wordt dan bij de plaatsen waar weinig horecagelegenheden zitten. Bij het maken van een kaart waar het totale verbruik gevisualiseerd is, is juist zichtbaar dat in het centrum totaal meer verbruikt wordt dan om het centrum heen. In het centrum wordt per aansluiting minder verbruikt dan verwacht, gezien het totale verbruik daar juist hoog is.

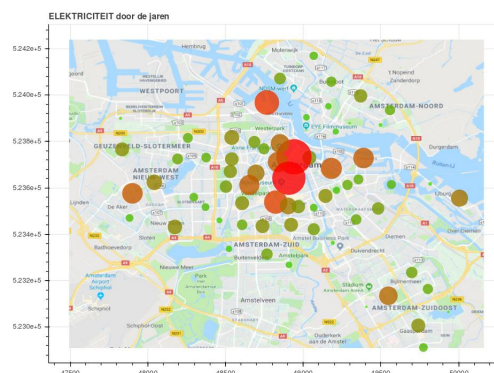


Figure 1: Totaal elektriciteitsverbruik per postcode 2016

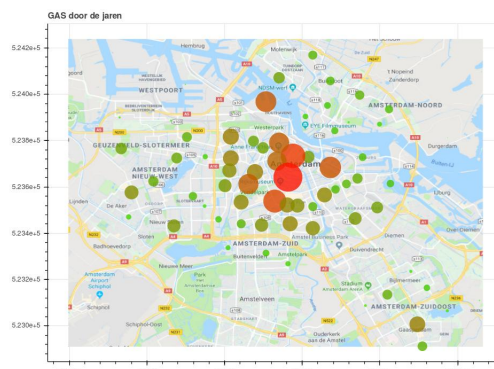


Figure 2: Totaal gasverbruik per postcode gas 2016

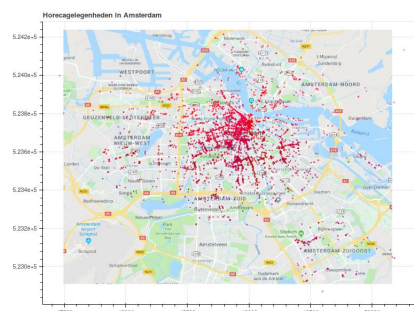


Figure 3: Horeca kaart

Het gas- en elektriciteitsverbruik is in de regio van het Amstel Business park per aansluiting is erg hoog in vergelijking met andere stadsdelen, de Westhaven buiten beschouwing gelaten, waar het elektriciteitsverbruik nog hoger is. Hier was het energieverbruik per aansluiting voor vijf postcodes significant hoger in vergelijking met de rest van de stad. Het energieverbruik in de westhaven is van 2009 tot en met 2016 gemiddeld hetzelfde, er is geen duidelijke lijn van verhoging of verlaging. Bij het Amstel Business park blijft het elektriciteits- en gasverbruik elk jaar grof omlaag gaan, van 2014 naar 2015 is er zelfs een afname van 51,14%. Als het energieverbruik per aansluiting op een kleinere schaal wordt bekeken, dit houdt in de uitschieters zoals Westhaven en Amstel Business Park buiten beschouwing latend, is er ook te zien dat er rond Centraal Station meer wordt verbruikt dan gemiddeld.

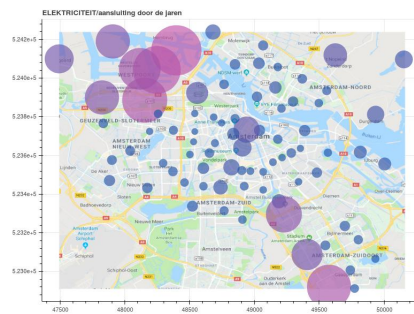


Figure 4: Elektriciteitverbruik per aansluiting 2016

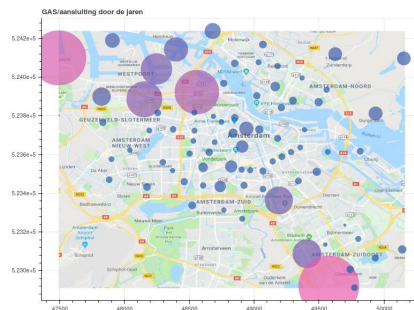


Figure 5: Gasverbruik per aansluiting 2016

In figuur 6 zijn de temperatuur per dag en het stroomverbruik per dag uit 2009 tegen elkaar afgezet. Hier is te zien dat de seizoenen een duidelijke invloed hebben op het stroomverbruik in Amsterdam. In de winter, wanneer de temperatuur lager is, is het verbruik significant hoger dan in de zomer. Zelfs is te zien dat bij meer extreme temperatuurstijging het verbruik nog verder daalt.

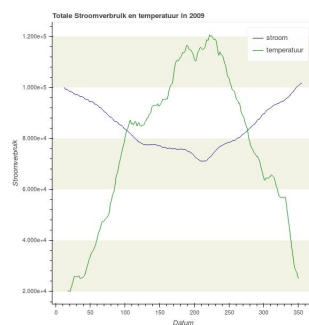


Figure 6: Temperatuur en SJV voor het jaar 2009

Bij het plotten van het stroomverbruik van 2009 en de dagelijkse temper-

atuur van 2009 is er zichtbaar dat er minder stroomverbruik is in de warme periode van het jaar. In de koude periodes is er significant meer stroomverbruik. In de eerste 75 dagen wordt er meer verbruikt hoe kouder het is. Vervolgens wordt het warmer rond dag 175 waar ook het stroomverbruik vermindert. Vervolgens wordt het rond dag 275 weer kouder waar ook het stroomverbruik sterk toeneemt.

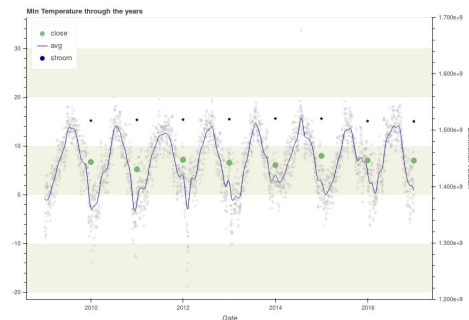


Figure 7: Minimale temperatuur over de jaren

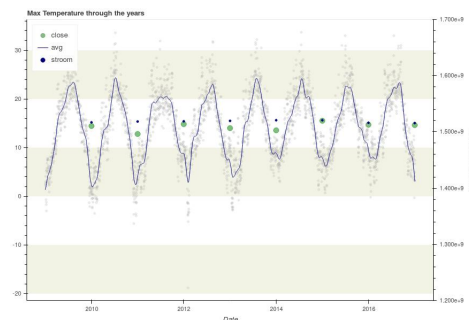


Figure 8: Maximale temperatuur over de jaren

Bij figuur 9 van elektriciteit afgezet tegen gas is er zichtbaar dat er per postcode er veel postcodes zijn die ongeveer even veel gas als elektriciteit gebruiken. Er zijn wel wat uitschieters die meer gas dan elektriciteit verbruiken maar niet veel uitschieters die veel electriciteit verbruiken en winig gas.

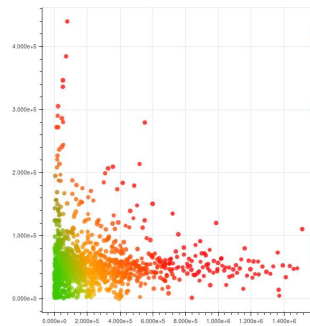


Figure 9: Gasverbruik tegenover elektriciteitsverbruik geplot per postcode met letter

Uit figuur 10 blijkt dat na het clusteren van de postcodes het gebruik van gas en elektriciteit lineair is. Postcode groepen die veel gas verbruiken verbruiken veel stroom en andersom.

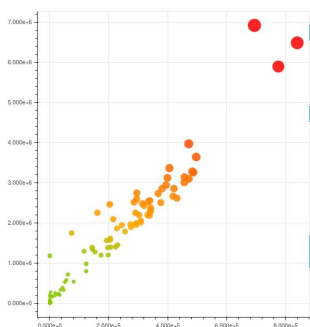


Figure 10: Gasverbruik tegenover elektriciteitsverbruik geplot per postcode zonder letters

Uit figuur 11 blijkt dat het totale gasverbruik per jaar ieder jaar lineair afneemt. Opvallend genoeg blijkt uit figuur 12 dat het totale stroomverbruik per jaar van 2009 tot en met 2013 juist toeneemt, waarna het een jaar stagneert en de volgende jaren erg sterk afneemt.

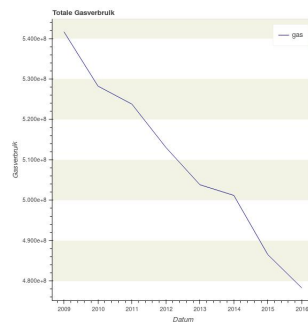


Figure 11: Totale gasverbruik per jaar

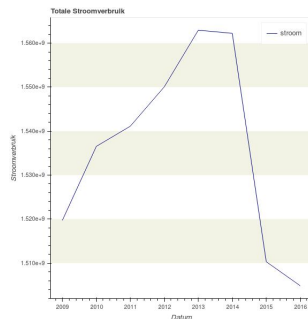


Figure 12: Totale elektriciteitsverbruik per jaar

3 Discussie

De plaatsen waar het SJV hoog is, zijn plaatsen waar veel bedrijven zitten of veel gezinnen wonen. De oorzaak hiervan is dat er juist in die gebieden veel kleine bedrijven zitten die nog net onder kleinverbruik vallen, maar relatief veel stroom verbruiken. De overige gebieden rondom het centrum waar veel verbruikt wordt zijn de gebieden waar veel gezinnen wonen. Vermoedelijk is dit het gevolg van het feit dat de panden in het centrum vaak meer zijn opgedeeld. De panden zijn opgedeeld in verschillende delen met als gevolg dat er per pand meerdere aansluitingen zijn. Hierdoor zou het verbruik per aansluiting lager zijn. Dit verklaart dat het algehele verbruik in het centrum hoger is, ondanks dat het energieverbruik per aansluiting een stuk minder hoog is. Hiermee is duidelijk aantoonbaar dat de horecagelegenheden aan veel verbruik doen.

Uit de resultaten van het dagelijks stroomverbruik en dagelijkse temperatuur is duidelijk geworden dat er warme periodes minder stroom verbruikt wordt. Dit heeft te maken met de langere dagen waardoor er minder licht van pas komt.

Wat voornamelijk als een verrassing kwam is het feit dat er in het centrum minder verbruik per aansluiting is dan aan de rand van de stad. In eerste instantie zijn we ervan uit gegaan dat door de hoeveelheid horecagelegenheden in het centrum van Amsterdam het energieverbruik erg hoog of misschien wel het hoogst zou zijn. Hieruit komt voort dat juist plaatsen waar veel kleine bedrijven zich gevestigd hebben het energieverbruik per aansluiting het hoogst is. Dit houdt in dat de aansluitingen nog voldoen aan klein verbruik maar net onder de grens liggen van groot verbruik. Ook is te zien dat er rondom het centrum, in de wijken waar grote gezinnen wonen, meer elektriciteit en gas per aansluiting verbruikt wordt in vergelijking tot het centrum, alhoewel deze hoeveelheden niet in de buurt komen van de hoeveelheid stroomverbruik in Westpoort het havengebied en Duivendrecht met het Amstel Business park.

Dit komt doordat de horecagelegenheden in dit gebied meestal langer open zijn dan in de rest van Amsterdam. Omdat Centrum Amsterdam langer open is betekent het ook dat het totale energieverbruik daar ook een stuk hoger is, in tegenstelling tot het energieverbruik per aansluiting. Dus is er een duidelijke correlatie tussen de hoeveelheid horeca en de hoeveelheid energie die verbruikt wordt. Hoe meer horecagelegenheden zich in een gebied bevinden, des te meer energie er wordt verbruikt.

Tijdens het analyseren van de data kwam naar voren dat het met de gegeven datasets lastig te bepalen is of er grote veranderingen per aansluiting plaats gevonden hebben of er groot verschil zit tussen het verbruik van aansluitingen op een postcode, gezien er alleen een gemiddelde beschikbaar was per jaar van alle aansluitingen op een postcode. Zo is het alleen mogelijk om zeer grote onregelmatigheden te detecteren. Hierdoor was het niet mogelijk om bijvoorbeeld bitcoin mining of marihuanateelt te ontdekken en hebben we deze vragen achter ons moeten laten. Met inzicht in het verbruik per aansluiting zou dit probleem kunnen worden opgelost. De datasets gaven ons alleen inzicht in aansluitingen met kleinverbruik. Dit zorgt ervoor dat er geen compleet beeld kan worden geschetst van het totale gebruik van gas en elektra in Amsterdam. Zo zou het ook mogelijk kunnen zijn dat een aansluiting overgaat op grootverbruik en deze dus verdwijnt uit de gebruikte dataset, maar er in werkelijkheid meer energie wordt verbruikt. Voor vervolgonderzoek zou het daarom goed zijn om ook een dataset van het grootverbruik in Amsterdam mee te nemen.

Bij de dataset van het gebruik per dag werd een voorspelling gemaakt van het gemiddelde gebruik en daarbij zou zo te zien zijn wat voor invloed het klimaat heeft op het stroom- en gasverbruik. Gezien dit voorspellingen zijn die gebaseerd zijn op de afgelopen tien jaar, kan je hier niet aan zien wat bijvoorbeeld een extreem koude of warme dag werkelijk teweeg brengt.

Tevens is dit gemeten bij 10000 willekeurige aansluitingen en is daarom niet erg specifiek. Dit probleem zou kunnen worden opgelost met een dataset die het verbruik per dag weergeeft en deze af te zetten tegen een dataset die van hetzelfde jaar het klimaat per dag bevat.

De plot in figuur 9 toonde geen duidelijk verband tussen het gasverbruik en het elektriciteitsverbruik, vandaar dat de postcodes gehergroepeerd zijn tot postcodes zonder letters. Vervolgens bleek er wel een sterk verband tussen beide waardes te zijn, het verbruik van gas en elektriciteit blijkt evenredig te zijn.