Gas- en elektraverbuik Amsterdam door de jaren heen

J.H. Roelink (11903260) M. Pannekoek (11621958) R. van Woerkom (11297654) E. Jusic (11733586) J. Taal (11755075)

28-06-2018

Contents

1	Inleiding	2
2	Methode	4
3	Resultaten	6
4	Discussie	12

1 Inleiding

De database die onderzocht wordt in dit onderzoek is een verzameling van data over het verbruik van gas en elektriciteit per postcode in Amsterdam, van 2014 tot en met 2018, verzameld door het bedrijf Liander. Om een beter zicht te krijgen op het energie en gas verbruik in Amsterdam door de jaren heen, zijn verschillende onderzoeksvragen opgesteld. Deze vragen hebben betrekkin op de opkomst van Bitcoin mining, het verschil tussen strenge en milde winters, de verschillen tussen buurten van Amsterdam, het effect van slimme meters en het effect van teruglevering van energie. Eventuele patronen die in de loop van het onderzoek ontdekt worden, worden ook geanalyseerd. Er zijn vijf onderzoeksvragen opgesteld met betrekking op deze concepten.

- 1. Is de opkomst van Bitcoin mining detecteerbaar in de dataset?
- 2. Zijn strenge of milde winters detecteerbaar in de dataset?
- 3. Bestaan er verschillen tussen het energieverbruik van verschillende buurten in Amsterdam en zijn deze consistent door de jaren heen?
- 4. Welk effect heeft het percentage van de huishoudens, dat een slimme meter heeft, op het verbruik van een straat?
- 5. Zijn er opmerkelijke bevindingen gedaan in de dataset, zo ja welke?

Wij verwachten dat een evenement als Bitcoin minen niet detecteerbaar zal zijn in de dataset van Liander. Ten eerste is het energieverbruik van Bitcoin mining veel te klein in verhouding van het totale energieverbruik. Zo werd in een onderzoek door Karl J. O'Dwyer en David Malone[1] geschat dat de totale wereldwijde energieverbruik door Bitcoin mining van 2010 tot 2014 maximaal 10 GigaWatt is, oftewel 3.1×10^{17} Joule per jaar. In een jaarlijks verslag van het International Energy Agency[2] werd het totale energieverbruik van de hele wereld geschat op 5269 Megatoe per jaar, wat ongeveer $2*10^20$ Joule per jaar is. Er moet rekening gehouden worden met het feit dat dit getal ook energieverbruik buiten huishoudens meeneemt. Hierin is te zien dat Bitcoin mining ongeveer een duizendste van het energieverbruik van de wereld opmaakt. Ervan uitgaande dat deze verhouding ook in Amsterdam geld, verwachten wij niet een daling of stijging van het energieverbruik te zien in onze data ten gevolge van de Bitcoin mining opkomst. Daarnaast is het nooit met zekerheid te zeggen of een eventuele verandering in het energieverbruik als oorzaak de opkomst van Bitcoin mining heeft. Zo zouden eventuele patronen of afwijkingen in het verbruik van elektriciteit of gas veroorzaakt kunnen worden door meerdere evenementen.

Het opmerken van warme of koude winters kan in theorie gedaan worden door naar het gasverbruik te kijken, aangezien mensen in de wintermaanden vaker de verwarming aan hebben en dus meer gas verbruiken. Maar waarschijnlijk zullen strenge of zachte winters niet detecteerbaar zijn. Ten eerste is er het probleem dat de data per jaar is. Er wordt dus geen onderscheidt gemaakt tussen verschillende seizoenen.

Indien we het effect van koude winters willen meten, moet aangenomen worden dat het gasverbruik in de winter het merendeel van het jaarlijkse gasverbruik moet opmaken, waardoor verandering in het gasverbruik in de winter daadwerkelijk detecteerbaar is in een dataset waarin het verbruik per jaar gegeven is. Ten tweede is de dataset per kalenderjaar, de dataset splitst dus elke winter in tweeën. Hierdoor is het nog moeilijker om verschillen te zien per winter. Tot slot, de kolom "SJV" in de database is niet de daadwerkelijke waardes van het verbruik van dat jaar. De waardes uit die kolom zijn gebaseerd op het verbruik van het voorafgaande jaar[10]. Plotselinge evenementen die een grote stijging in het energieverbruik veroorzaken zullen dus dat jaar niet in de dataset vermeld staan. Zulke stijgingen zullen pas te zien zijn in de dataset van het volgende jaar.

Een patroon die naar verwachting wel zichtbaar zal zijn in de dataset is het verschil in verbruik tussen verschillende postcodes door de jaren heen. Vanwege de vele verschillen in gebieden in Amsterdam wordt verwacht dat er ook verschillen zichtbaar zijn in energieverbruik. Waarschijnlijk zijn de verschillen redelijk consistent door de jaren heen. Zo zijn er in het centrum en op de zuidas veel oude gebouwen met slechte isolatie, deze gebouwen zullen meer gas verbruiken voor de verwarming dan de nieuwbouwhuizen. Ook zal het terugleveren van energie eerder bij nieuwbouwhuizen voorkomen en het elektriciteitsverbruik lager liggen. Een ander duidelijk verschil van energieverbruik in de buurten kan liggen aan het soort gebouw. Een bedrijf verbruikt meer energie dan een huishouden. Het vermoeden is dat bedrijfsterreinen en winkelcentra gemiddeld meer energie verbruiken dan woonwijken.

De slimme meter is een digitale energiemeter die het energieverbruik van consumenten preciezer afmeet dan zijn voorganger de mechanische meter. Ondanks dit gegeven wordt er door verschillende consumenten en 'experts' beweert dat slimme meters foutieve meterstanden doorgeeft[11]. Echter wordt er ondanks deze beweringen verwacht dat de slimme meter geen verandering brengt in het verbruik per jaar van een postcode. Liander stelt dat de slimme meter zelf geen hoger of lager energieverbruik te weeg brengt, maar dat elke verandering in het energieverbruik na het installeren van een slimme meter komt doordat de vorige metingen foutief waren[4]. Wij gaan ervan uit dat Liander voor het installeren van de slimme meter gemiddeld het juiste energieverbruik afmeet, dus zou de slimme meter geen verschil in het gemiddelde verbruik van Amsterdam moeten veroorzaken.

2 Methode

Voor dat de dataset geschikt is om te analyseren moet deze nog worden bewerkt. Het proces dat hierbij wordt ingezet staat beter bekend als het 'preprocessen' van de dataset. Zo stond de dataset nog niet in de juiste bestandsvorm, was de dataset inconsistent genoteerd en moesten er variabelen aangepast worden om het analyseren en visualiseren van de data makkelijker te maken.

De data van Liander uit het jaar 2014 tot en met 2018 zijn gedownload. De data zijn verwerkt naar een bruikbaar formaat voordat ze geanalyseerd werd. Allereerst zijn ze omgezet werden naar csy bestanden. Met behulp van Python code zijn alle rijen met informatie over adressen buiten Amsterdam verwijderd. Hierna zijn de data van alle jaren in eenzelfde bestand onder elkaar gezet, met een extra kolom die het jaartal aangeeft. Enkele kolommen met overbodige of irrelevante informatie zijn verwijderd uit de database. Tot slot zijn bins toegevoegd voor waardes in een aantal kolommen, een legenda voor deze bins staat in bijlage 1. Deze bins zijn met de Python methode Pandas gemaakt. In de code worden de waardes gesorteerd op grootte en aan een bin toegewezen. Het resultaat is een kolom met een bereik van nul tot en met tien, in plaats van de oorspronkelijke waardes, die een bereik hadden van nul tot in de duizenden. Er is gekozen voor het gebruik van bins omdat dit voor verschillende analysemethodes zoals het berekenen van de covariantie erg van pas komt. Soms zal er gebruik worden gemaakt van zowel de bins als de daadwerkelijke waardes, afhankelijk van de analysemethode.

De kolommen die uit de database zijn verwijderd, omdat ze overbodig of niet relevant waren, zijn: Meetverantwoordelijke, Netbeheerder, Netgebied, Woonplaats, Landcode, Verbruikssegment, %Defintieve aansl (NRM), Gemiddeld aantal telwielen. De toegevoegde kolom is genaamd Jaartal, deze maakt een onderscheid tussen de metingen die in verschillende jaren zijn gedaan. De uiteindelijke database bestond uit de volgende kolommen: Jaartal, Straatnaam, Postcode_van, Postcode_tot, Productsoort, Aantal Aansluitingen, %Leveringsrichting, %Fysieke status, %Soort aansluiting, Soort aansluiting Naam, SJV, %SJV laag tarief, %Slimme Meter.

Om de onderzoeksvragen te beantwoorden zijn verschillende methoden gebruikt om de data te analyseren. Aangezien niet alle data zich leent voor elke analyse methode. Voor het plotten van histogrammen was het bijvoorbeeld nodig dat de data in bins zaten, terwijl het bij spreidingsdiagrammen juist nodig was dat de data niet in bins zaten.

Bitcoin mining en koude winters detecteren zijn vergelijkbaar. Voor Bitcoin mining moet gekeken worden naar het gemiddelde elektriciteitsverbruik per jaar en bij de hevigheid van de winters naar het gemiddelde gasverbruik per jaar. Deze getallen zijn behaald door respectievelijk het gemiddelden elektriciteitsverbruik en gasverbruik te nemen van alle postcodes in onze dataset. Deze waardes zijn uitgezet in een grafiek, zie figuur 5.

Voordat de buurten geanalyseerd konden worden, moest er bepaald worden wat precies een buurt was. Er is voor gekozen om elke postcode als buurt te zien en de letters van de postcodes weg te halen. Omdat niet alle buurten

over evenveel aansluiting beschikt, is het gemiddelde van alle aansluitingen van een postcode genomen. Hiervoor kan een spreidingsdiagram worden gebruikt. Om het overzichtelijk aan te geven, worden de productsoorten in gescheiden diagrammen weergegeven. Zo is duidelijk welke buurt veel gebruikt en waarvan.

Om een antwoord te vinden op de onderzoeksvraag: "Welk effect heeft de slimme meter op het verbruik per jaar van een postcode?" wordt lineaire regressie gebruikt om zo het verband tussen verbruik en het gebruik van slimme meters aan te geven. Eerst wordt door middel van een spreidingsdiagram de correlatie tussen het verbruik en het percentage slimme meter weergegeven. Verbruik van elektriciteit en gas zijn uiteraard gescheiden. Vervolgens wordt een formule voor een lijn opgesteld in de vorm y = ax + b. De variabelen kunnen berekend worden door middel van lineaire algebra. De formule van de lijn ziet er dan als volgt uit: y = Mp. Waarin M een tweedimensionale matrix is en p een tweedimensionale vector.

Nadat de lijn opgesteld is moet de foutmarge oftewel error berekent worden. Dit wordt gedaan door middel van de mean squared error. Deze functie berekent de gemiddelde van de som van de afstand tussen een datapunt en de lijn in het kwadraat. Vervolgens kan ook nog de root mean square error berekend worden, dit is simpelweg de wortel van de mean squared error. Een voordeel van de root mean squared error is dat het erg veel lijkt op de standaard deviatie. Het doel van de lineaire regressie is om de error zo klein mogelijk te krijgen. Dit betekend namelijk dat de afstand tussen de meetpunten en de lijn erg klein is en de lijn dus gebruikt kan worden voor voorspellen van datapunten.

Bij het berekenen van de covariantie tussen verschillende variabelen in onze database is iets opmerkelijks gevonden. Het gaat hier over de covariantie $\sigma(x,y)$ tussen de kolom x= "SJV" en de kolom y= "%Leveringsrichting". Eerst is de database gesplitst in een database met gegevens over gasverbruik en over elektriciteitsverbruik. Dit is gedaan door de data met behulp van de methode 'Pandas' te groeperen op basis van de waarde in de kolom 'Productsoort'. Vervolgens waren de waardes in de rijen met het productsoort elektriciteit in bins gesorteerd, dit is nodig voor het berekenen van een representatieve covariantie. Het berekenen van de covariantie is gedaan met de functie 'cov' uit de Python methode 'numpy'. Deze functie neemt twee argumenten, in dit geval de twee lijsten met de bins van "SJV" en "%Leveringsrichting". Bij het omzetten van de waardes in de kolom "SJV" in bins zijn de grenzen van de bins (zie bijlage 1) zo gekozen dat er van elke bin ongeveer dezelfde hoeveelheid bestaan. Dit zorgt voor een nauwkeurigere covariantie.

Er is ook gekeken naar het verband tussen het gemiddelde verbruik en de hoeveelheid aansluitingen die energie terug leveren. Aangezien alleen elektriciteit werd teruggeleverd, zijn alleen aansluitingen die elektriciteit verbruiken afgebeeld. De hoeveelheid aansluitingen, die energie terugleveren, werden genoteerd als een percentage van het totale aantal aansluitingen in de straat. De percentages zijn in bins opgedeeld. De verdeling is te zien in bijlage 1. Er is voor gekozen om 100% een aparte groep te maken aangezien we niet straten die energie terug leveren in dezelfde groep als straten die dat niet doen wilden hebben. Deze groepen zijn uitgezet tegen het gemiddelde verbruik van die groep.

3 Resultaten

Het gemiddelde elektriciteitsverbruik van Amsterdam neemt elk jaar af, zoals te zien is in figuur 1. De afname is elk jaar minder er is dus een omgekeerd evenredig verband tussen het jaartal en het elektriciteitsverbruik. Ook het gemiddelde gasverbruik nam door de jaren heen af. Het gasverbruik en elektraverbruik zijn in gescheiden grafieken gezet, zoals hieronder te zien is in figuur 1.

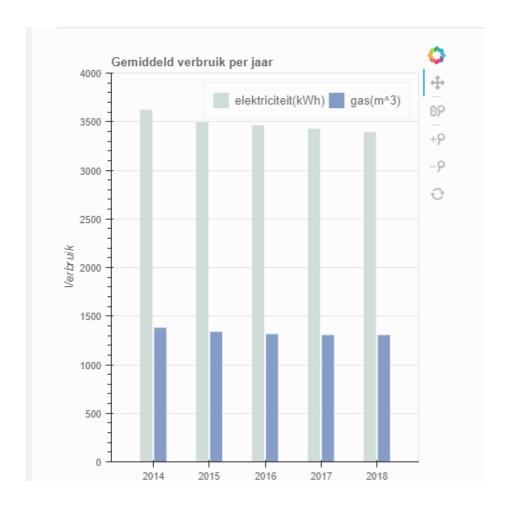


Figure 1:

Bij het elektriciteitsverbruik per postcode is er een duidelijke spreiding zichtbaar. Postcodes met nummers dicht bij elkaar liggen ook dicht bij elkaar. In de onderstaande grafiek is duidelijk te zien dat er twee grotere gebieden zijn die meer verbruiken dan de rest en een die lichtelijk meer verbruikt. Het merendeel van de postcodes verbruikt gemiddeld niet meer dan 5000 KwH per jaar zoals te zien is in figuur 2.

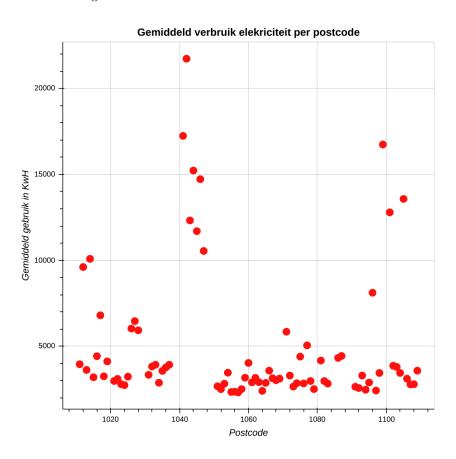


Figure 2:

Ook bij het gasverbruik zijn er duidelijk drie gebieden te zien die meer verbruiken dan de rest. Opvallend is dat deze drie gebieden overeenkomen met de drie gebieden van het elektraverbruik. De meeste postcodes verbruiken niet meer dan $2000\ m^3$ gas per jaar zoals te zien is in figuur 3.

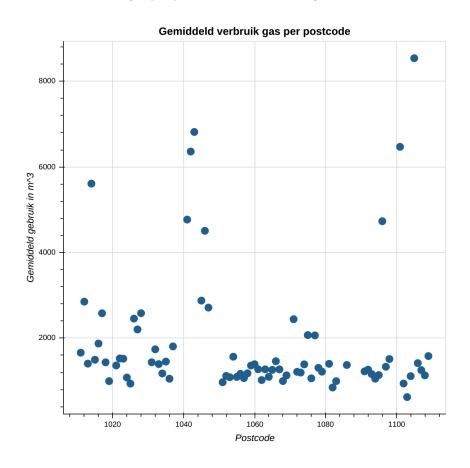


Figure 3:

Bij het opstellen van de lijn voor de correlatie tussen elektriciteitsverbruik en het gebruik van slimme meters zijn de de volgende variabelen uitgekomen: a = -56.49 en b = 3594.87. De formule voor de lijn ziet er dus als volgt uit: y = -56.49x + 3594.87. Dit resulteert in de diagram te zien in figuur 4.

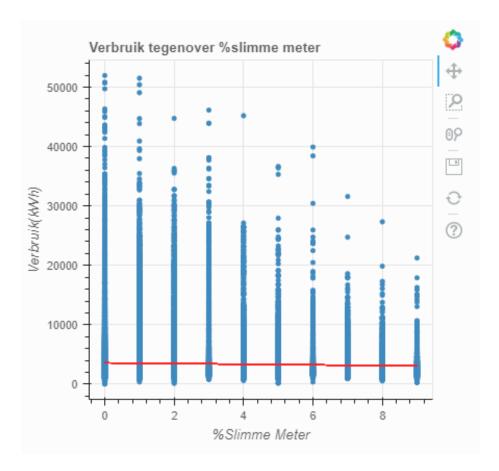


Figure 4:

De mean squared error van deze lijn is 10504698.5894, de root mean squared error is dus 3241.09527621. Bij het opstellen van de lijn voor de correlatie tussen gasverbruik en het gebruik van een slimme meter zijn de volgende variabelen uitgekomen: a = -16.11 en b = 1358.15 waardoor de formule van de lijn er als volgt uitziet: y = -16.11x + 1358.15. Dit resulteert in de diagram in figuur 5.

De mean squared error van deze lijn is 5689791.13501, de root mean squared error is dus 2385.3283076 In zowel figuur 5 als 6 lijkt het alsof de lijn een rechte lijn is. Uit de formules van de lijnen is echter af te leiden dat de lijn in beide gevallen licht daalt.

De covariantie $\sigma(x,y)$ die berekend is tussen de kolommen x = "SJV" en y = "%Leveringsrichting" heeft een waarde van -0.20. Op zichzelf zegt deze waarde niet veel, omdat dit nog niet significant hoeft te zien. Het bereik van de covariantie wordt gegeven door $-\sigma(x)\sigma(y) \leq \sigma(x,y) \leq \sigma(x)\sigma(y) = -1.66 \leq \sigma(x,y) \leq 1.66$.

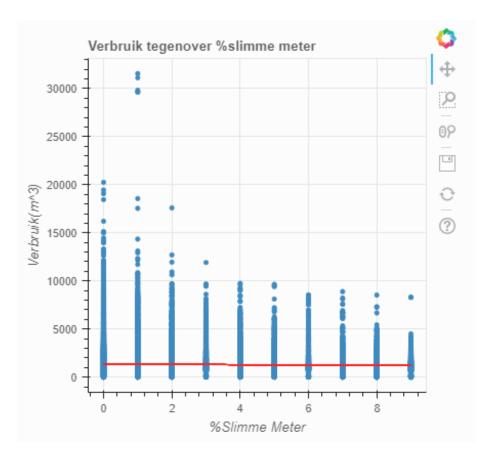


Figure 5:

De waarde is genormaliseerd door het te delen door het product van de standaarddeviaties $\sigma(\mathbf{x})\sigma(\mathbf{y})$ van zowel de kolom $\mathbf{x}=$ "SJV" als $\mathbf{y}=$ "%Leveringsrichting". Dit geeft de covariantie weer als deel van de maximaal bereikbare covariantie $\frac{\sigma(x,y)}{\sigma(x)\sigma(y)}=\frac{-0.20}{1.66}$. De waarde hiervoor is ongeveer -0.12. Het lijkt erop, op basis van de gegevens in deze situatie, dat deze laatste waarde significant is.

De bins van de %leveringsrichting kolom zijn uitgezet tegen het gemiddelde verbruik van die bin. Dit is te zien in figuur 6. Zoals te zien is er geen eenduidig lineair verband te vinden. Het gemiddelde verbruik stijgt naarmate er minder huishoudens energie terugleveren, maar er zijn veel uitzonderingen te vinden.

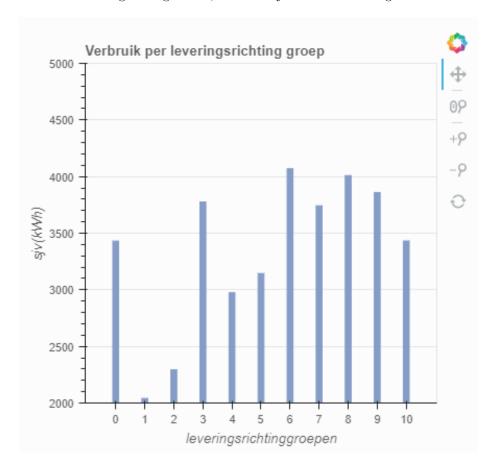


Figure 6:

4 Discussie

De eerste twee doelen van het onderzoek waren het vaststellen of Bitcoin mining en de heftigheid van winters te detecteren zijn in onze dataset. Hoewel er niet verwacht werd dat deze evenementen detecteerbaar zijn in de dataset, zorgt de data ervoor dat er met zekerheid geconcludeerd kan worden dat deze evenementen niet detecteerbaar zijn. Om deze evenementen te detecteren hadden veranderingen in het gemiddelde verbruik per jaar waargenomen moeten worden. De veranderingen in de figuur zijn echter niet groot genoeg om er een bepaald evenement als oorzaak aan toe te kennen . Bitcoin mining zou namelijk een stijging in het gemiddelde energieverbruik teweeg moeten brengen. De data laat echter zien dat het gemiddelde energieverbruik daalt. Dit resultaat spreekt niet het effect van Bitcoin mining tegen maar laat zien dat het effect niet detecteerbaar is.

Hetzelfde geld voor het gasverbruik in de wintermaanden. Om een strenge winter op te merken zou het gasverbruik moeten stijgen dat jaar. Het tegenovergestelde is waar. De winter van 2014-2015 en 2015-2016 waren historisch
erg zachte winters[3], maar het gas verbruik lag in die jaren juist hoger dan de
jaren daarop. Het is dus met zekerheid te zeggen dat de opkomst van het minen
van Bitcoin en de verschillen in hevigheid van de afgelopen vijf winters niet te
detecteren zijn in onze dataset. Een mogelijke manier om deze zaken wel op te
merken is het meten van de data op maandelijkse basis.

Er is een duidelijk verschil te zien tussen de buurten. Vooral drie gebieden vallen op vanwege het hogere gebruik. Na verder onderzoek, blijken deze gebieden haventerrein of winkelcentra te zijn. Deze gebieden gebruiken vaak meer elektriciteit omdat ze dag en nacht actief zijn en elektra en gas gebruiken. Ook zijn er door de jaren heen de verschillen niet erg vergroot of verminderd. Wel zijn er buurten die een hoge stijging of daling in verbruik in elektriciteit of/en gas zijn ondergaan. Dit kan aan nieuwbouw liggen of dat er een nieuw bedrijf is gevestigd in dat gebied.

De error van beide lijnen is dusdanig groot dat er geen sprake is van een correlatie tussen het verbruik van zowel gas als elektriciteit en het percentage slimme meters van een postcode. Bovendien kan de lijn ook niet gebruikt worden voor het voorspellen van nieuwe waarden aan de hand van één van de variabelen. Om terug te komen op de onderzoeksvraag: "Welk effect heeft de slimme meter op het verbruik per jaar van een postcode?" Aan de hand van de geanalyseerde data is er geen correlatie te vinden tussen het verbruik per jaar en het gebruik van een slimme meter. Het gebruik van de slimme meter heeft dus geen effect op het verbruik per jaar van een postcode.

De waarde -0.12 is significant, en geeft de conclusie dat als het percentage leverings richting hoog is, het vaker voorkomt dat de desbetreffende postcode relatief minder elektriciteit verbruikt. Ofwel, in het algemeen, hoe hoger het percentage leverings richting hoe minder elektriciteit er wordt gebruikt. Dit ontkracht onze hypothese. De covariantie blijft echter moeilijk te interpreteren. Ondanks dat de gevonden covariantie in dit geval redelijk significant is, is nog niet uitgesloten dat er geen verband bestaat tussen de twee variabelen. Voor een meer betrouwbare covariantie zou meer data getest moeten worden.

References

- [1] O'Dwyer, Karl J., and David Malone. "Bitcoin mining and its energy footprint." (2014): 280-285.
- [2] International Energy Agency. (2016). Key World Energy Trends Excerpt from: World energy balances. Retrieved from ftp://ftp.energia.bme.hu/pub/energetikaialapismeretek/KeyWorldEnergyTrends.pdf
- [3] Wintergek. Gemiddelde temperatuur in Nederland per maand en jaar. Geraadpleegd van https://www.wintergek.nl/data/lijst-gemiddelde-temperatuur-nederland
- [4] Liander. (2018, 8 juni). Feit of fabel: de slimme meter zorgt voor hogere eindafrekening [Blogpost]. Geraadpleegd op 28 mei 2018, van https://www.liander.nl/blog/feit-of-fabel-de-slimme-meter-zorgt-voor-hogere-eindafrekening/
- [5] Liander. (2018, 1 januari). Kleinverbruiksdata per jaar [dataset]. Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [6] Liander. (2017, 1 januari). Kleinverbruiksdata per jaar [dataset]. Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [7] Liander. (2016, 1 januari). Kleinverbruiksdata per jaar [dataset]. Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [8] Liander. (2015, 1 januari). Kleinverbruiksdata per jaar [dataset]. Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [9] Liander. (2014, 1 januari). Kleinverbruiksdata per jaar [dataset]. Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [10] Liander. Beschrijving Verbruiksbestand Kleinverbruik (versie 1.2). Geraadpleegd van https://www.liander.nl/partners/datadiensten/open-data/data
- [11] Radar Avro Tros. 'Slimme meter is onbetrouwbaar'. Geraadpleegd van https://radar.avrotros.nl/nieuws/detail/slimme-meter-is-onbetrouwbaar/

NAAM KOLOM	NIEUW BEREIK	BETEKENIS	
MEETVERANTWOORDELIJKE	[1,0]	1 = Liander N.V.	0 = N.v.t.
PRODUCT	[1,0]	1 = GAS	0 = ELK
LEVERINGSRICHTING	[0,1,2,3,4,5,6,7,8,9,10]	10 = 100%	4 = 40% - 49%
		9 = 90% - 99%	3 = 30% - 39%
		8 = 80% - 89%	2 = 20% - 29%
		7 = 70% - 79%	1 = 10% - 19%
		6 = 60% - 69%	0 = 0% - 9%
		5 = 50% - 59%	
FYSIEKE STATUS	[0,1,2,3,4,5,6,7,8,9]	9 = 90% - 100%	4 = 40% - 50%
		8 = 80% - 90%	3 = 30% - 40%
		7 = 70% - 80%	2 = 20% - 30%
		6 = 60% - 70%	1 = 10% - 20%
		5 = 50% - 60%	0 = 0% - 10%
%SJV LAAG TARIEF	[0,1,2,3,4,5,6,7,8,9,10]	0 = 0%	5 = 40% - 50%
		10 = 90% - 100%	4 = 30% - 40%
		9 = 80% - 90%	3 = 20% - 30%
		8 = 70% - 80%	2 = 10% - 20%
		7 = 60% - 70%	1 = 1% - 10%
		6 = 50% - 60%	
%SLIMME METER	[0,1,2,3,4,5,6,7,8,9]	9 = 90% - 100%	4 = 40% - 50%
		8 = 80% - 90%	3 = 30% - 40%
		7 = 70% - 80%	2 = 20% - 30%
		6 = 60% - 70%	1 = 10% - 20%
		5 = 50% - 60%	0 = 0% - 10%
SJV (Voor covariantie)	[0,1,2,3,4,5,6,7,8,9, 10]	10 = 6500 - 100000	4 = 2100 - 2300
		9 = 4300 - 6500	3 = 1900 - 2100
		8 = 3500 - 4300	2 = 1700 - 1900
		7 = 3000 - 3500	1 = 1500 - 1700
		6 = 2600 - 3000	0 = 0 - 1500
		5 = 2300 - 2600	