

# Stroomverbruik in de gemeente Amsterdam

Demi van Dijk  
Marlon de Jong  
Mathias Kirkeng  
Nina Verheijen

June 2018

# 1 Inleiding

Energiebedrijven maken tegenwoordig steeds meer gebruik van data die verschillende dingen vertellen over het stroomgebruik van een bepaald gebied of huishouden, omdat ze daarmee hun verdienmodel kunnen maximaliseren. Uit deze data kan men aflezen van welk product er gebruik wordt gemaakt, hoeveel gebruik er van een bepaald product wordt gemaakt en hoeveel aansluitingen er zijn op een stroom- of gasnet. Dit is belangrijke informatie, omdat hierdoor de producten beter op de wensen van de klant kunnen worden afgestemd.

Dit onderzoek richt zich op het stroomverbruik in de gemeente Amsterdam. Hoofdzakelijk richt het onderzoek zich op de vraag of er interessante informatie uit de data kan worden gehaald van de netbeheerder Liander. Wanneer blijkt dat deze data bruikbaar is wordt gekeken naar het stroomverbruik in bepaalde periodes. Gepoogd wordt om deze informatie te koppelen aan een bepaalde gebeurtenis, bijvoorbeeld *bitcoin mining*. Daarnaast wordt er ook gekeken naar ongewone patronen die uit de datasets zijn te halen. Zijn uit die patronen ook bepaalde gebeurtenissen af te leiden? Ten slotte wordt het stroomverbruik per gebied in Amsterdam nog geanalyseerd. Ook uit deze gegevens kunnen mogelijk een aantal conclusies worden getrokken. Heeft de bouw van nieuwe huizen in bepaalde gebieden invloed op het stroomverbruik daar? De verschillen per gebied moeten ook verklaard kunnen worden op basis van deze gegevens.

Wanneer er een antwoord kan worden gevonden op al deze vragen, zal de technologie waarvan gebruik wordt gemaakt bij het leveren van stroom zich ook verder kunnen ontwikkelen. Men kan hierdoor ook zorgen voor steeds meer duurzaam stroomverbruik, wat het milieu weer ten goede komt.

## 2 Methode

De datasets van 2011 tot en met 2017 worden gebruikt voor dit project. 2018 is niet compleet en zal dus niet gebruikt worden. 2011 tot en met 2017 staat in excel. Om bruikbare datasets te hebben worden de excel files omgezet in csf files. Ten tweede worden alle irrelevante kolommen verwijderd. Deze kolommen zijn: NETGEBIED, LANDCODE, Definitieve aansl (NRM). Verder moeten alle rijen die niet AMSTERDAM als woonplaats hebben verwijderd worden, hierna kan ook de kolom WOONPLAATS worden verwijderd. De originele datasets zijn nu "schoon". Om een completere dataset te maken wordt de dataset opgedeelt in stadsdelen. Amsterdam heeft acht stadsdelen. Ook worden de inwonersaantallen per stadsdeel toegevoegd aan de dataset. Hiermee kan de data worden gerelativeerd.

Om de dataset verder "schoon" te maken moeten de outliers worden verwijderd. Outliers worden gevonden middels box en whisker plot. De set wordt eerst geordend, vervolgens wordt de mediaan, het middelpunt, van de set gevonden (Q2). De mediaan van het deel van de set onder Q2 is Q1. De mediaan van het deel van de set boven Q2 is Q3. De minimale waarde is de laagste waarde van de set. De maximale waarde is de hoogste waarde van de set. De 50 procent in het midden, de box ofwel de IQR, is  $Q3 - Q1$ . Om te weten hoe groot je fences zijn bereken je  $1.5 * IQR$ . Alle datapunten die kleiner zijn dan  $Q1 - 1.5 * IQR$  zijn outliers. Alle datapunten groter dan  $Q3 + 1.5 * IQR$  zijn outliers. De 50 procent in het midden/de doos ofwel de IQR is  $Q3 - Q1$ . Om te weten hoe groot je fences zijn bereken je  $1.5 * IQR$ . Alle datapunten die kleiner zijn dan  $Q1 - 1.5 * IQR$  zijn outliers. Alle datapunten groter dan  $Q3 + 1.5 * IQR$  zijn outliers.

Bij het plotten van de histogram viel het op dat er een kleine piek is tot 450. Alle datapunten onder 450 zijn niet reële datapunten, het is immers zeer onwaarschijnlijk dat een verzameling van huizen minder dan 450 kilowattuur gemiddelt per jaar gebruikt. Bij het plotten van de histograms van de jaren 2011 tot en met 2017 is overal deze kleine piek te zien, dus de ondergrens wordt op 450 gezet.

Vervolgens is de dataset klaar voor exploratory data analysis. Er zijn vier categorieën, voor elke categorie moet, als het mogelijk is, minimaal twee soorten plots worden gemaakt. De categorieën zijn: univariate graphical, multivariate graphical, univariate non-graphical, en multivariate non graphical. Aan de hand van de gemaakte plots kunnen een deel van de vragen worden beantwoord. Niet elke categorie is geschikt voor het beantwoorden van elke vraag. Bijvoorbeeld als men de stroomrichting in kaart wil brengen is univariate graphical in eerste opzicht geschikt, maar als men de data bekijkt ziet men dat minder dan 3 procent van de data onder de 100 procent ligt. Deze data is dus niet geschikt voor een histogram, maar wel voor technieken die passen bij univariate non graphical.

Om de data analyse af te sluiten hoort er in-depth analysis plaats te vinden. Met andere woorden wordt toekomstige data voorspelt aan de hand van de dataset. Dit kan gedaan worden middels algoritmes, bijvoorbeeld clustering, echter zijn de beschikbare algoritmes beperkt en is de dataset niet geschikt voor de meeste aanpakken. Lineaire regressie kan wel worden toegepast. Met lineaire regressie wordt het optimale model gevonden. Het optimale model is een lijn waar de optelsom van de verticale afstand van de gegeven datapunten tot de lijn het kleinst zijn. Het optimale model kan gebruikt worden om toekomstige datapunten te voorspellen.

### 3 Resultaten

-Is er een verschil in gebruik tussen wijken waarvan bekend is dat er meer mensen wonen met een technische achtergrond en wijken waar mensen wonen met een geesteswetenschappen achtergrond?

-Gebruiken mensen die veel stroom gebruiken ook meer gas?

#### **Kunnen evenementen zoals Bitcoin mining of koude/warme winters worden gevonden binnen de data?**

De data geeft alleen het gemiddelde energie gebruik over een jaar. de winter is hierdoor verspreidt over twee jaren. hierdoor is het niet mogelijk om een uitspraak te doen over warme of koude winters. Het is immers zo dat als een koude winter een milde winter opvolgt het gemiddelt een normale winter is qua gebruikt.

Mining van crypto currency is te recent gebleken om hier data over te vinden. Als hier wel een causaal verband was gevonden was het echter nog steeds niet mogelijk om met zekerheid te zeggen dat het door bitcoin mining wordt veroorzaakt.

-Zijn er onverwachte patronen?

-Zijn er duidelijke verschillen tussen wijken in Amsterdam?

-Hebben welvarender wijken een hoger stroomconsumptie per huishouden.

-Zijn er “groenere” wijken, wijken die een andere leverings richting hebben zijn aannemelijk “groener”, want om een leverings richting te hebben die niet 100 procent is zijn bijvoorbeeld zonnepanelen nodig.

## 4 Discussie

## 5 Literatuurlijst

[dataset stadsdelen]. (2017, 1 januari). Geraadpleegd op 20 juni 2018, van <https://www.ois.amsterdam.nl/download/01b-kerncijfers-22-gebieden-en-stadsdelen-1-januari-2017>