

Energie-analyse en forecasting

Xander Vandooren

February 24, 2025

Contents

1 Week 1:	1
1.1 Inleiding:	1
1.1.1 Internet of Energy:	1
1.1.1.1 Data science (analyse) in Energie.	1
1.1.1.2 Weer voorspelling:	1
1.1.1.2.1 Voorspelling van energieverbruik:	2
1.1.2 Tools:	2
1.1.3 Data verkenning en data wrangling:	3
2 Week 2:	3
2.1 Statistiek:	3
2.1.1 verschillende maten voor meten/berekenen	3
2.1.1.1 Centrummaten	3
2.1.1.1.1 Gemiddelde	4
2.1.1.1.2 Mode (Nodus)	4
2.1.1.1.3 Mediaan	4
2.1.1.1.4 Skewness:	5
2.1.1.2 Spreidingsmaten	5
2.1.1.2.1 Spreidingsbreedte	5
2.1.1.2.2 Kwartielafstand (IQR)	5
2.1.1.2.3 Boxplot	6
2.1.1.2.4 Variantie	6
2.1.1.2.5 Standaardafwijking	6
2.1.1.2.6 Variatiecoefficient:	6
2.1.1.2.7 Relatieve spreidingsmaten	7
2.2 Data Visualisatie	7
2.2.1 Sankey Diagrams	9
2.3 Tools voor data Visualisatie	9
2.4 Verkennende data analyse	9
3 Week 3:	9

1 Week 1:

1.1 Inleiding:

1.1.1 Internet of Energy:

1.1.1.1 Data science (analyse) in Energie.

- Optimalisatie van energie productie met HEB
- Nationaal energiemarkt-werking
- Voorspelling van toekomstige E-vraag om productie en distributie aan te passen.
- Herkenning van energieverbruik patronen om aanbods op maat te maken en fouten te detecteren.
- Analyse van werking/status/fouten bij apparaten
- Data gedreven DSM

1.1.1.2 Weer voorspelling: Waarom:

- Optimalisatie van energie-stromen
- Verzekeren de decarbonisatie van de energiesector
- Kosten te verminderen
- Business decisions based on forecasting
- Beperken downtime en verbeteren van veiligheid en productiviteit
- Voorspel veranderingen in de vraag demand
- Trade-off: middelen - betrouwbaarheid - kosten
- Internet of Energy

Use-case:

- Voorspelling van energieverbruik
- operatie van een gebouw
- fout detectie

1.1.1.2.1 Voorspelling van energieverbruik: Voorbeeld:

- Input:
 - W-PV
 - Weer informatie
 - seizoen
 - Uur van de dag
 - WKK-status
 - Time of day (day-ahead)
- Model (machine learning, deep learning,...)
- output:
 - Afname elektriciteit

1.1.2 Tools:

Type problemen:

- Description:
 - Patronen of trends identificatie
 - Wat beïnvloedt energieverbruik
 - Wat beïnvloed energieopwekking
- Voorspelling:
 - Opwekking forecast
 - Optimale scheduling
- Detectie:
 - Zijn er fouten in het systeem
- Dynamische evaluatie:
 - Welke subsidies of tarieven stimuleren de efficiëntie en betrouwbaarheid

Bedrijven in energy analytics:

- 3E
- N-Side
- Energis
- Oktow
- Actemium
- Capgemini
- Teradata
- ...

1.1.3 Data verkenning en data wrangling:

2 Week 2:

2.1 Statistiek:

- Beschrijvende statistiek: (wij werken in de les met dit soort statistiek)
 - Methoden om samen te vatten en te beschrijven de belangrijkste kenmerken van een dataset zoals centrale tendens, variabiliteit en distributie.
- Inferentiele statistiek:
 - omvat van trekken van conclusies van de populatie.

2.1.1 verschillende maten voor meten/berekenen

2.1.1.1 Centrummaten Is een maat dat probeer een hele dataset te beschrijven. De waarde vertegenwoordigt het midden of de center van de dataset distributie.

2.1.1.1.1 Gemiddelde Het is de opsomming van twee of meerdere waarden gedeeld op de aantal waarden die opgesomd worden.

- Meest gebruikte maat van centrale tendens
- Sterk beïnvloedt door uitschieters
- Wordt gezien zoals een balanspunt

Gemiddelde → balanspunt (kwadratische fout)

2.1.1.1.2 Mode (Nodus) De meest voorkomende waarde in een dataset.

- Maat van centrale tendens
- Niet beïnvloedt door uitschieters
- In een dataset kunnen er geen of meerdere mode waarden aanwezig zijn
- Het kan gebruikt worden zowel voor kwalitatieve als kwantitatieve data
- Indien elke waarde maar 1 keer voorkomt in een dataset is er geen modus
- Indien er waarden zijn die meerdere keren voorkomen is de meest voorkomende de dataset (er kunnen meerdere modus aanwezig zijn)

2.1.1.1.3 Mediaan Het is de middelste waarde van een dataset. De mediaan komt overeen met de 50e percentile van een dataset, m.a.w. is het midden van een dataset met een half van de waarden kleiner en de andere half groter dan de mediaan

- Als n oneven is, is de mediaan de middelste waarde van de gesorteerde rij
- Als n even is, is de mediaan het gemiddelde van de twee middelste waarden van de gesorteerde rij
- positie van de median is $\frac{n+1}{2}$
- Wordt niet beïnvloedt door uitschieters

2.1.1.1.4 Skewness:

- negatively skewed (gemiddelde ligt meer naar links)
- Normal (no skew) (alle 3 de waarden liggen op elkaar)
- Positively skewed (gemiddelde ligt meer naar rechts)

Centrummaten alleen zijn niet genoeg data te beschrijven.

2.1.1.2 Spreidingsmaten

2.1.1.2.1 Spreidingsbreedte Het is het verschil tussen de grootste en de kleinste waarden in de dataset

- $x_{max} - x_{min}$
- Sterk gevoelig aan extreme waarden/uitschieters
- Geeft geen informatie over de distributie van de data
- Makkelijk te bepalen
- Neemt maar 2 waarden (min,max)
- Sterk gevoelig aan uitschieters

2.1.1.2.2 Kwartielafstand (IQR) Het wordt ook benoemd zoals middenspreiding, midden 50%, 4de spreading of H-spreading. Het wordt gedefinieerd zoals het verschil tussen de 75e en 25e percentielen van de data.

- splitst een gesorteerde dataset in 4 gelijke delen
- Eerste kwartiel is de 25e percentiel
- Tweede kwartiel is de 50e percentiel=mediaan
- Derde kwartiel is de 75e percentiel
- Eenvoudig te bepalen
- Vertegenwoordigt de spreiding van de middenste 50%
- niet gevoelig aan uitschieters

2.1.1.2.3 Boxplot Boxplot= box en whisker plot Geeft een 5 getallen samenvatting van een dataset weer.

- Minimum
- 1ste kwartiel
- mediaan
- 3de kwartiel
- maximum

Skewness kan je ook zien op een boxplot.

2.1.1.2.4 Variantie Telt de spreiding van de data t.o.v het gemiddelde. Het kan bepaald worden voor een sample (s^2) of een populatie (σ^2)

2.1.1.2.5 Standaardafwijking Net zoals de variantie kan bepaald worden voor zowel een sample (s) of een populatie (σ). In beide gevallen is de vierkantswortel van de variantie

- moeilijker te bepalen
- Gebasseerd op alle waarden
- Gevoelig aan uitschieters voor kleine datasets

2.1.1.2.6 Variatiecoefficient: Het wordt gedefinieerd zoals de verhouding van de Standaardafwijking tot het gemiddelde. V wordt ugedrukt in percentages.

- $V < 5\% \rightarrow$ heel kleine relatieve variatie
- $5\% < V < 10\% \rightarrow$ kleine relatieve variatie
- $V > 50\% \rightarrow$ heel grote variatie

2.1.1.2.7 Relatieve spreidingsmaten Ze worden gebruikt om de vergelijken te doen van de spreiding van twee of meer datasets. Relatieve betekent dat zonder grootheden is

- Beschrijft de positie van een datapunt t.o.v. de rest van de data
- Twee veel voorkomende maten zijn:
 - Percentiel
 - Z-score
- Een optie om uitschieters te detecteren
- Percentiel:
 - Waarde van 0 tot 100 dat de percentage geeft dat gelijk of onder is
 - * Mediaan is 50e percentiel
 - * K1 is de 25e percentiel
 - * K3 is de 75e percentiel
- Z-score:
 - Het beschrijft het verband van een datapunt met het gemiddelde van de groep. Het wordt gemeten in termen van de standaardafwijkingen van het gemiddelde.
 - Dataset met een absolute waarde van z-score groter dan 3 Voor heel skewed data distributies, een absolute waarde van z-score groter dan 2

2.2 Data Visualisatie

- Data wrangling
- Missing data and outliers
- Features and KPI
- Histogrammen:

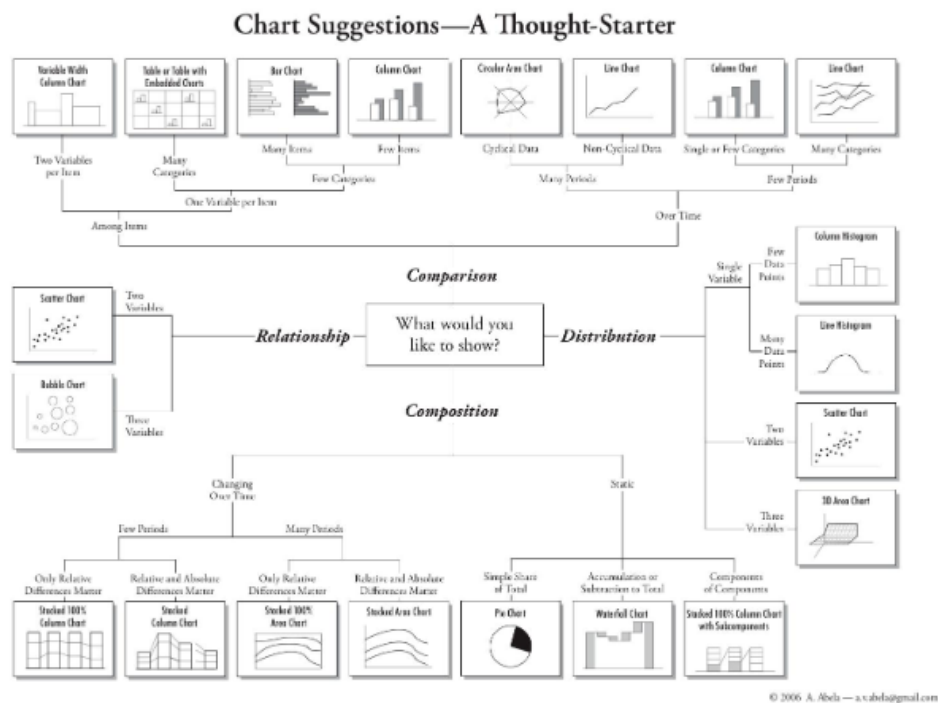


Figure 1: Verschillende charts voor specifieke situaties

- 1 variabele
- Box plot:
 - 1 variabele
- Scatter plots
 - 2 variabelen
- Lijn plots:
 - 2 variabelen
- Scatter plot matrix:
 - Indien meer dan 2 variabelen
- Heatmaps:

- Indien meer dan 2 variabelen

2.2.1 Sankey Diagrams

→ energie flow

2.3 Tools voor data Visualisatie

- Commerciele tools:
 - Power-BI, Looker, Tableau, Qlik sense, Google data studio
- Eigen tools:
 - Jupyter-voila, Streamlit, Plotly Dash, Flask, Django

2.4 Verkennende data analyse

- Profielaggregatie (jaarlijks, maandelijks, per week, per dag)
- Energieverbruik distributie - duration curve
- Correlatie tussen variabelen
- Heatmaps

3 Week 3:

Hoe meer