



CAPSTONE ASSIGNMENT 2020

**Bootcamp Data Analysis Beginner
Apr-May in 2020**

Team 3: Mengwen Yang

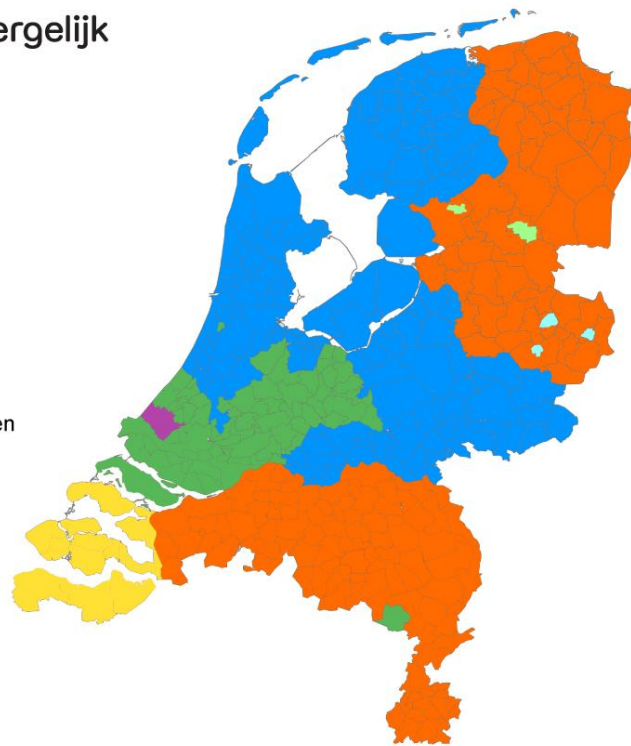
My idea for choosing dataset

The electrical use in the Netherlands are supplied from the three major network administrators : Enexis, Liander, and Stedin. Together, they provide energy to nearly the entire country.

This market is not competitive, meaning that the zones are assigned. This means that every year they roughly provide energy to the same zip codes.



- Liander
- Enexis
- Stedin
- Enduris
- Westland Infra
- RENDO Netwerken
- Coteq Netbeheer



3 Datasets and 14 Variables

- **3 Datasets** from <https://www.kaggle.com/lucabasa/dutch-energy>
 - **Enexis_electricity_01012020.csv**
 - **liander_electricity_01012020.csv**
 - **stedin_electricity_2020.csv**
- **14 Variables/Columns** in each file, which records the electricity consumption through the various connection for particular zip codes under the network administrator, ie. the energy provider, in year 2020.

My questions

- **Electricity consumption data of the Netherlands in 2020**
 1. How much electricity has Amsterdam consumed in 2020?
 2. Which city has consumed the most electricity in 2020?
 3. People in which city are more likely to consume electricity during low tariff hours (from 10 p.m. to 7 a.m. and during weekends)?
 4. Which company has offered the most net energy consumption?
(net energy consumption = Consumption_per_year x delivery_percentage_per_year)
 5. How many percentage of electricity have been consumed by the top 5 cities, respectively?



Thank you for watching!

CAPSTONE ASSIGNMENT 2020

Team 3: Mengwen Yang