

# Assignment 4

Assignment 4 i kurset Data Science 2021

Karoline Midtbø

Morten Knutsen

## Oppgave 4

```
suppressPackageStartupMessages({  
  library(PxWebApiData)  
  library(tidyverse)  
  library(lubridate)  
})  
knitr::opts_chunk$set(echo = FALSE, include = FALSE)
```

##Loader inn data

I denne delen skal vi se på prisen per kvm.

**I denne delen har vi valgt ut hva vi skal ha som variabler, der vi har valt vekk *boligtyper* og *contentscode*.**

**Sjekke hvor mange NA verdier det er i pm2**

**Complete.cases fra 2006 til 2017**

Det er 197 complete cases i fra 2006 til 2017, det vil si antall kommuner som har data for alle årene f.o.m. 2006 t.o.m. 2017

Complete.cases fra 2008 til 2017

Det er 214 complete cases i fra 2008 til 2017, som vil si 214 kommuner om har data for alle årene f.o.m. 2008 t.o.m. 201

Vi ser at ved å velge perioden 2008-2017 istedenfor 2006-2017 får vi 17 ekstra complete.cases. Velger derfor å studere perioden 2008-2017.

Fjerner data

Datasett henting fra SSB

sjekker navn

Henter data fra SSB om 0-17 og 18+

merge data til pop\_08\_17\_ya\_p

Legger sammen navnene og finner dem i prosent

Sjekker

Merger data til pm2

Rydder opp

**Desiler**

Vi henter inn data fordelt opp i desiler

Rydder opp

**Utdanning**

**Handelsomsetning per innbygger**