

# Mandatory assignment DB

Michael S.  
Sebastian P.

## Air quality

23. april 2019

### Introduktion

I den moderne, digitale, verden er data vigtigere end nogensinde. Vi høster data i større mængder end nogensinde og til mange forskellige formål som f.eks., målrettet reklame eller klimaforandringer. Datamængderne der bliver indsamlet overskrider alt hvad man have forestillet sig for bare nogle år tilbage. Så store mængder data introducere nye spændende udfordringer som vi godt kunne tænke os, at forsøge at takle. Vi vil fokusere på, hvordan man kan gøre store mængder rå data mere spiselig for den almindelige person. Vi vil også se på de forskellige muligheder for opbevaring og behandling af data.

### Problemformulering

Opgaven er, at importere store mængder, luftkvalitets, måledata i XML format i en traditionel relationel SQL server. Det skal laves tabeller ud fra hvert XML dokument og dertil skal der også laves relationer mellem tabellerne ligesom der var i den originale database.

Efter dataene er blevet indlæst i databasen skal der laves løsninger til visning af dataene og dataanalyse. Visningen af dataene skal til, at starte med laves vha. en Microsoft ASP.NET MVC webapplication. Dataene skal vises i ren tekst og senere skal webapplikationen og indeholde et simpelt diagram over noget af dataen. Den store dataanalyse del skal laves med Tableau, hvor der bør udarbejdes komplekse diagrammer over dataene.

Når arbejdet med den relationelle database er færdigt skal dataen migreres til en NoSQL database. I denne opgave skal der tages brug af MongoDB. Man bør her gøres sig overvejelser

om, hvorvidt det er det rigtige valg at bruge en NoSQL database. Der skal også gøres tanker om, hvordan man skal strukturere sin data i et ikke relationel database miljø.

## Metode

For at opnå viden, som er relevant for denne opgave, har vi valgt at gøre brug af undervisningsmateriale. Som et supplement til undervisningsmateriale undersøge vi på nettet om der er bedre løsninger, til en given udfordring. Derudover vil vi gøre brug af vores underviser, hvis vi skulle være i tvivl.

Vi har også en intention om spare en masse med hinanden, da vi ofte har forskellige indfaldsvinkler til forskellige udfordringer.

Til sidst, vil vi prøve udarbejde forskellige løsningsforslag, til specifikke udfordringer. Da der kan være flere mulige løsninger, nogle bedre end andre.

## Planlægning

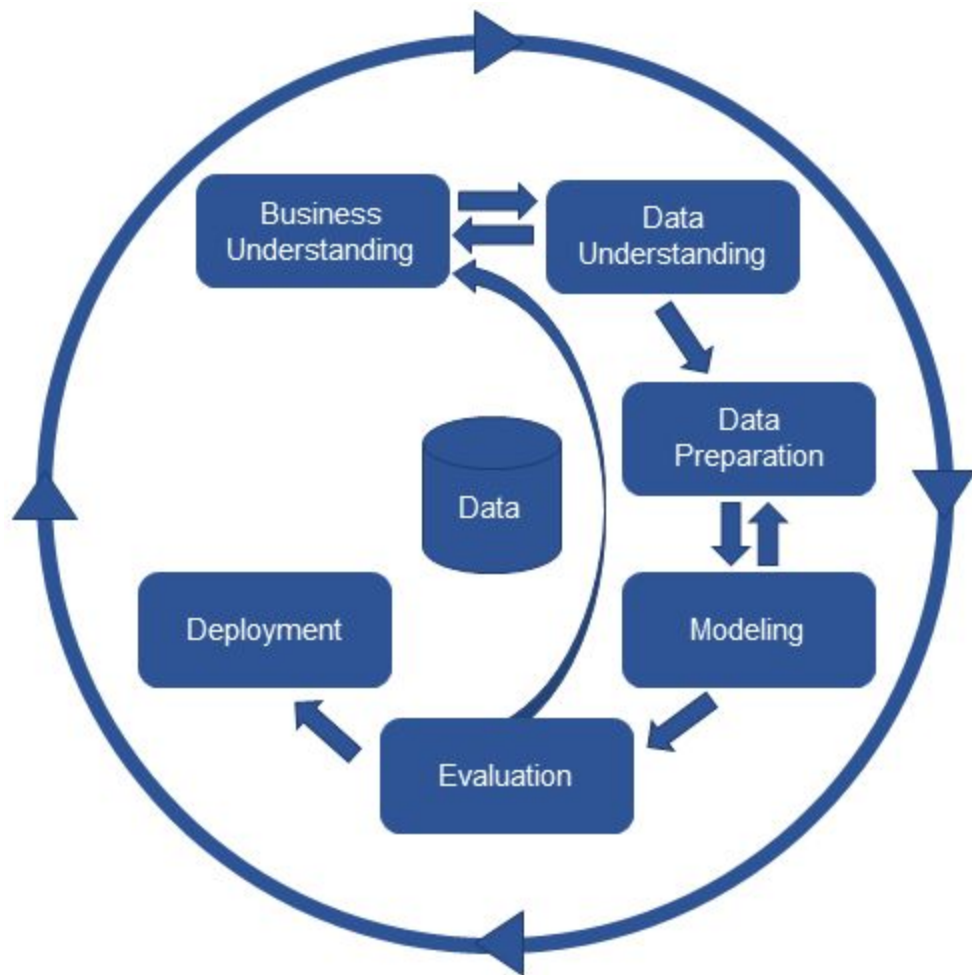
Da vi skal aflevere den 23. April og startede den 18 marts har vi 4 uger til projektet.

uge 1 vil vi, starte med at gøre brug af CRISP-DM modellen, da den er meget god til at skabe forståelse for projektet. den beskriver at vi, til en start skal skabe business og data forståelse. Da vi har haft et foredrag om luftkvalitet, er vi rimelig godt rustet til at håndtere den data vi har modtaget.

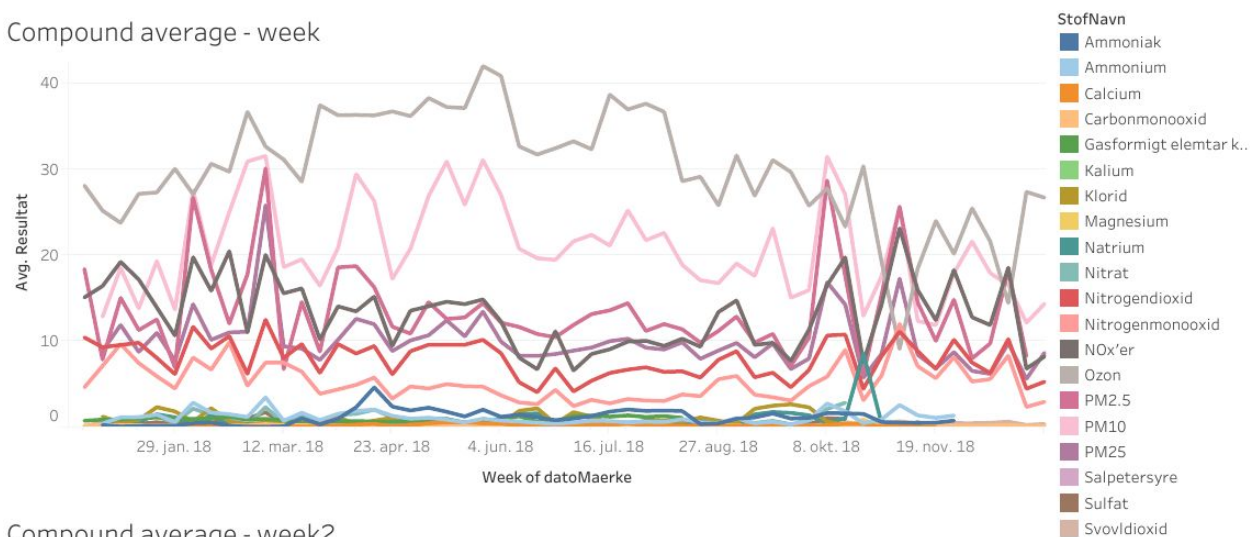
Uge 2-3 vil vi forberede dataen, som vi skal konvertere fra XML til SQL format med referencer og keys. Når alt det er klargjort og vi har konverteret dataen med referencer og keys, så skal vi vise det i vores ASP.NET MVC løsning. Grunden til at, vi vælger at, bruge 2 uger på databasen og modellering, er for at sikre at der er god tid til at håndtere data'en korrekt.

Uge 4 vil vi arbejde med tableau og vise data direkte fra vores database, i nogle flotte grafer, samt skrive vores synopsis.

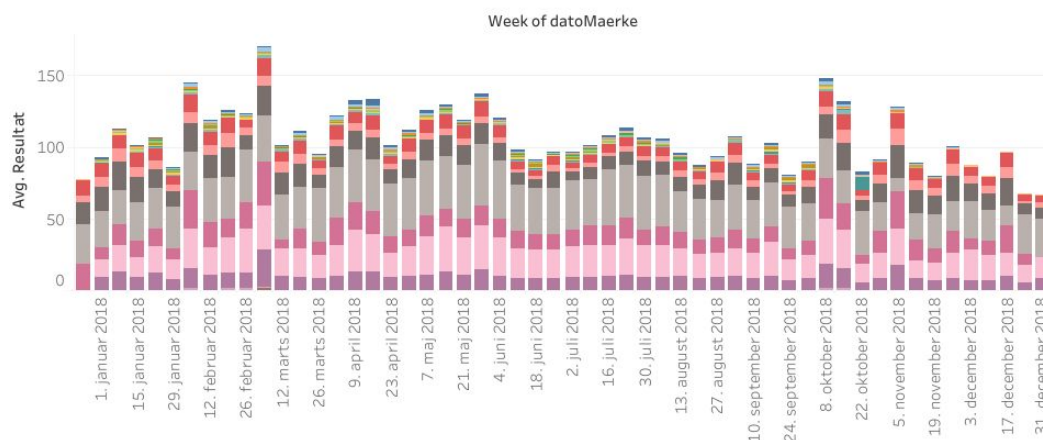
## Bilag



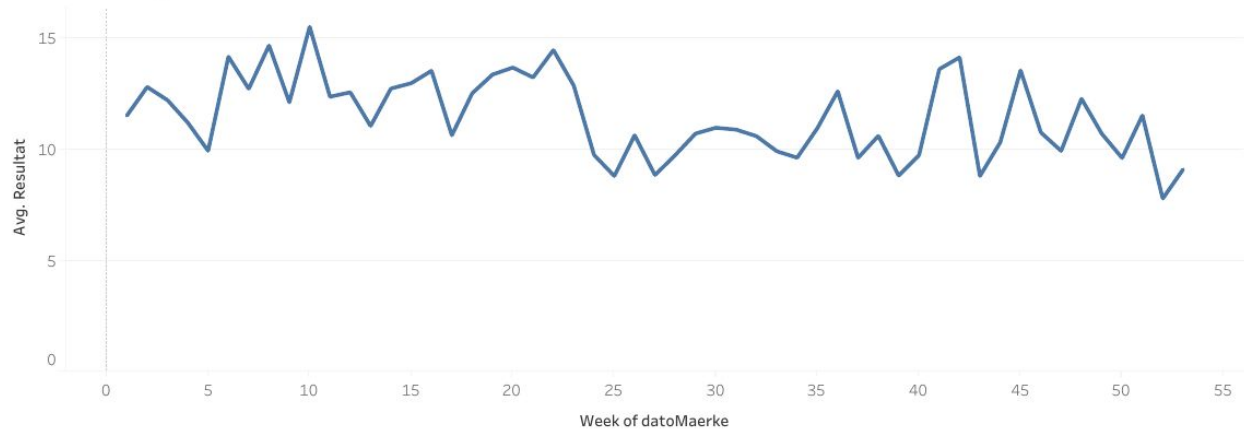
Compound average - week



Compound average - week2



Result average - week



Result sum - week

