

# Obligatorisk læringsaktivitet i Flow 3, Dataanalyse



Afleveringsfrist:

Torsdag den 10. december 2024, kl. 20

Udarbejdet af:

Thorbjørn Wulf, adjunkt Cphbusiness

Thorbjørn Baum, adjunkt Cphbusiness

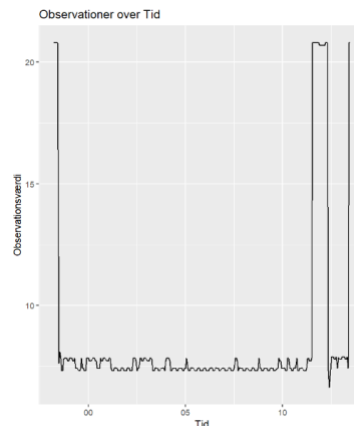
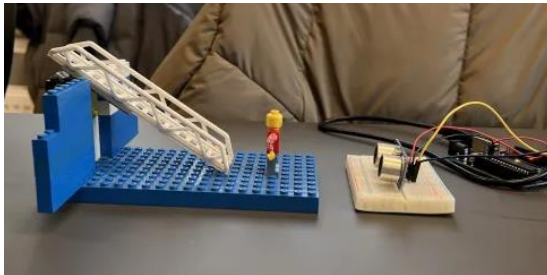
## Indhold

Opgave 1 – IoT og Fog .....	3
Opgave 1.1 – Opstille en model for måling af aktivitet.....	3
Opgave 1.2 – Giv en vurdering af mulighederne at realisere projektet i Fogs trælast.....	3
Opgave 2 – data flow på AWS .....	3
Opgave 2.1 – Opstille en cloud-arkitektur til systematisk indsamling af data.....	3
Opgave 3 – Spørgeskema og Cumulative Link Models .....	4
Opgave 3.1 – Illustration af spørgsmålet .....	4
Opgave 3.2 – Cumulative Link Models.....	4
Opgave 3.3 – Illustration af forklarende variable .....	4

## Opgave 1 – IoT og Fog

### Opgave 1.1 – Opstil en model for måling af aktivitet

Ud fra Fogs KPI om certificeret træ - <https://ipaper.johannesfog.dk/csr-2022/csr-2022/?page=34> - skal I forsøge at opstille en model i Lego for hvordan man kan indsamle data for kundernes adfærd i forhold til Fogs ønske om at fastholde deres opmærksomhed på informationsmateriale om certificering af træ (FSC og PEFC). Censoren er en afstandssensor - HC SR04 - som skal gemme data på et sd-kort. I skal desuden lave en test i målestok 1:1 som viser en opstilling hvor sensoren kan bruges til at registrere personaktivitet. Herunder et par eksempler fra sidste år.



### Opgave 1.2 – Giv en vurdering af mulighederne at realisere projektet i Fogs trælast.

Ud fra de erfaringer I gør jer med sensoren, skal I vurdere om det teknisk og data-analytisk set giver mening at fortsætte med projektet. Det vil sige leverer sensoren data som kan bruges til at levere troværdige analyser af kunders adfærd i trælasten

## Opgave 2 – data flow på AWS

### Opgave 2.1 – Opstille en cloud-arkitektur til systematisk indsamling af data.

I skal implementere nedenstående arkitektur på en EC2-instans på AWS. Instansen skal enten være Windows eller Ubuntu efter eget valg. Den skal bestykes med en MySQL-database og R med Rscript. I skal derpå hente data fra de fire lokationer som er angivet i OLA 4, hvor I skal hente miljødata fra H.C.Andersens Boulevard, Anholt, Banegårdsgade i Århus og Risø. Det er vigtigt at I inkluderer logging og at jeres script kan køre helt automatisk – det vil sige at det skal kunne eksekveres via Windows task-scheduler eller linux cron og resultatet gemmes i en egnet datastruktur i MySQL. I skal have registreret data fra mindst 2 kørsler med et døgns mellemrum.

I skal dokumentere jeres setup med skærmdumps fra AWS-konsollen (bl.a. jeres inbound rules), task-scheduleren, ER-diagram samt select's fra databasen, hvor man kan se på timestamps, at der er tale om data fra mindst to forskellige kørsler. I skal desuden vedlægge jeres R-kode samt logfil på github.



## Opgave 3 – Spørgeskema og Cumulative Link Models

### Opgave 3.1 – Illustration af spørgsmålet

Hent data fra filen "regnskaber\_industri\_transport\_byg\_5\_25000\_ansatte\_anonym" ind i R og lav en grafisk illustration af fordelingen af svarene på spørgsmålet: "Hvordan ser du mulighederne for at låne penge til din virksomhed?"

### Opgave 3.2 – Cumulative Link Models

Lav en Cumulative Link Model for spørgsmålet "Hvordan ser du mulighederne for at låne penge til din virksomhed?". I må selv vurdere, hvilke forklarende variable I ønsker i jeres model. Vær opmærksom på, at I kan finde inspiration til forklarende variable i artiklerne "Historisk nemt at låne penge" og "Gode muligheder for at låne penge i hele landet".

### Opgave 3.3 – Illustration af forklarende variable

Lav en grafisk illustration af den ene af jeres forklarende variable og spørgsmålet "Hvordan ser du mulighederne for at låne penge til din virksomhed?". (hint: se figur på side 6 i artiklen: "Gode muligheder for at låne penge i hele landet")