

# DAT120 øving 4: Gruppeprosjekt del 1

## Om prosjektet

Dette er et prosjekt i to deler hvor dere på slutten av del 2 har et Python-script som kan lese og gjøre data-analyser på filer med data fra værstasjoner fra meteorologisk institutt. I del 1 utvikler og tester dere en del av funksjonene for å gjøre data-analysen. For å lage disse funksjonene trenger dere litt basiskunnskap om lister. Derfor vil to videoer om lister være lagt ut sammen med temaet versjonskontroll. Mer om lister kommer i temaet samlingsobjekter. I del 2 leser dere dataene fra fila, bruker funksjonene fra del 1 til å gjøre data-analysene, og plotter dataene.

Når hele prosjektet (del 1 og 2) er ferdig vil programmet deres kunne gjøre følgende analyser for filer fra alle værstasjoner som støtter dataene hver analyse trenger:

- For værstasjoner som registrerer snødybde kan dere regne ut antall dager med skiføre hver skisesong.
- For værstasjoner som registrerer snødybde kan dere regne ut om det er en trend over mange år. Blir det flere eller færre dager med skiføre?
- For at en plante skal vokse krever den som regel at det er en viss minimumstemperatur, og så vokser den forttere om det er varmere. For å forenkle så kan veksten telles i antall grader over minimum for hvert døgn. For værstasjoner som registrerer temperatur, beregn antall dager hvert år hvor temperaturen er over minimum for planten og beregn total vekst slik som beskrevet.
- For værstasjoner som registrerer nedbør, finn lengste periode med tørke (ingen nedbør) hvert år
- For værstasjoner som registrerer skydekke finn antall dager med fint vær
- For værstasjoner som registrerer vind finn høyeste vind og median for vinden det året. Medianen er den midterste verdien i ei liste som er sortert på verdi.

## Læringsmål

I prosjektet som helhet skal dere lære hvordan samarbeide for å løse en større programmeringsoppgave. Dere skal lære enkel bruk av versjonskontrollsystemet Git og Github. I del 2 skal dere erfare hvordan det er å bygge på kode som dere selv skrev for en stund siden.

## Godkjenning

Denne øvingen og øving 10 skal gjøres i grupper på inntil 4 studenter. Øvingen skal godkjennes ved å demonstrere det gruppa har lagd på samme vis som for de tidligere øvingene. I tillegg skal dere vise studentassistenten Github repo-et deres. Dere trenger bare å demonstrere en gang for at alle i gruppa skal få godkjent.

## Deloppgaver for del 1 av prosjektet

- a) En av dere lager et Github repository for denne øvingen. De andre inviteres med i dette repoet.
- b) Hver student i gruppa skal klonе dette Github repositoryet til sin lokale datamaskin og jobbe i denne mappa med øvingen. Koden dere skriver i denne øvingen skal lastes opp til Github repoet.

- c) Lag en fordeling av oppgaver til de ulike medlemmene i gruppa. Fordel resten av deloppgavene i denne øvingen slik at alle i gruppa har noen deloppgaver de skal gjøre.
- d) Skriv en funksjon som tar inn ei liste med flyttall og en enkeltverdi. Funksjonen skal telle antall elementer i lista som er større enn eller lik den oppgitte verdien og returnere dette.
- e) Skriv en funksjon som tar inn ei liste med tall. For hvert tall i lista unntatt det siste skal funksjonen regne ut differansen mellom neste tall i lista og dette tallet. Differansene skal legges inn i ei ny liste.
  - a. **Frivillig 1:** Hvis du har to lister, en med x-koordinater og en med y-koordinater som representerer verdier av en funksjon  $y = f(x)$ , så kan du regne ut den numeriske deriverte gjennom å regne ut  $f(x+h) - f(x) / h$ , hvor h er avstanden mellom to etterfølgende innslag i lista over x-koordinater
  - b. **Frivillig 2:** For å regne ut om temperaturen er stigende eller synkende som trend over tid er man gjerne interessert i å se på om gjennomsnittsverdier stiger heller enn hver verdi siden temperaturen gjerne kan gå både opp og ned fra en dag til den neste. Lag en funksjon som tar inn ei liste og et antall elementer e. For hvert element i lista unntatt de siste e elementene skal den regne ut gjennomsnittet av elementet og de e neste elementene og lagre disse gjennomsnittene i ei ny liste. Deretter bruker du funksjonen for å regne ut numerisk derivert på denne lista over gjennomsnitt heller enn på den opprinnelige lista.
- f) Lag en funksjon som tar inn ei liste med tall. Funksjonen skal finne og returnere lengden til den lengste sammenhengende sekvensen av 0-ere i lista.
  - a. **Frivillig, avansert:** Ta inn ei liste med heltall. Returner lengden og verdien til den lengste sekvensen av samme verdi.
  - b. **Frivillig, avansert:** Det samme for flyttall, men med en oppgitt toleranse.
- g) Skriv en funksjon som regner ut hva trenden i et datasett er. Datasettet skal bestå av to lister, ei liste med x-koordinater og ei liste med y-koordinater. Elementene på samme indeks i de to listene hører sammen. Trenden skal beregnes som to tall a og b, som skal fungere som parametere i en lineær formel:  $verdi = ax + b$ . Implementer følgende formler, hvor n er til men ikke med lengden til lista,  $x_i$  er i-ende element i lista x, og  $\bar{x}$  er gjennomsnittet av alle verdiene i lista x. Merk at disse formlene og hvorfor de ser slik ut er pensum i emnet STA100, temaet lineær regresjon, minste kvadraters metode.

$$a = \frac{\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a * \bar{x}$$

- h) Anta du har en plante som krever at temperaturen er +5 grader celsius for å vokse i det hele tatt, og så vokser forttere desto varmere det er, lineært med temperatur over 5 grader. Skriv en funksjon som regner ut summen av alle tall over 5 i lista. Så i lista [4, 7, 15] blir summen 0 (for 4) + 2 (for 7) + 10 (for 15).
  - a. **Frivillig:** Planten tar skade av minusgrader, så minusgrader gir «negativ vekst» og planten dør hvis den når 0.
  - b. **Frivillig avansert DAT200 pensum:** Skriv en funksjon som finner perioden i lista som gir maksimal vekst samt summen for denne perioden. Funksjonen skal returnere alle tre tallene. Dette er et eksempel på «maksimal delsekvens sum» problemet som brukes for å illustrere effektive algoritmer i emnet DAT200.
- i) Fila «lister\_for\_del\_1.py» er delt ut sammen med denne øvingen. Bruk funksjonen fra oppgave d) på lista «temperaturer» fra denne fila. Anta at hvert innslag er

maksimaltemperaturen for en dag. Finn antall sommerdager (over 20), høysommerdager (over 25) og tropedager (over 30).

- j) Bruk funksjonen fra oppgave e) til å finne ut om temperaturen er stigende eller synkende for hvert tidspunkt. Gå gjennom lista som dere får når dere kaller funksjonen fra oppgave e) på temperaturlista. For hvert element skriv ut indeksen og skriv ut «stigende» om differansen er over 0, «synkende» om den er negativ eller «uforandret» om den er 0.
- k) Bruk funksjonen fra oppgave g) for å finne trenden i temperaturlista. For å finne trenden bruker dere lista «temperaturer\_tidspunkter» som x-koordinater og lista «temperaturer» som y-koordinater. Skriv ut om trenden er stigende eller synkende. Trenden er stigende hvis a er positiv, synkende hvis a er negativ og det er ingen trend hvis a er 0.
- l) Bruk funksjonen fra oppgave h) for å finne total vekst for planten gitt lista med temperaturer
- m) Bruk funksjonen fra oppgave f) på den oppgitte lista «døgn-nedbør» for å finne den lengste perioden uten nedbør