

DAT120 øving 10: Prosjekt del 2

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 fortere 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 for prosjektet

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

Øving 4 og øving 10 (denne øvingen) skal gjøres i grupper på inntil 4 studenter. Øvingene 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.

I utgangspunktet skal øving 10 (denne øvingen) gjøres i de samme gruppene som øving 4.

Filformatet

For øving 10 er to filer delt ut. Begge filene er hentet fra nettsida seklima.met.no og inneholder data fra Meteorologisk Institutt. Data er lisensiert under lisensen CC BY 4.0. Skriv inn «license CC BY 4.0» i Google søkefeltet for å finne den om dere er interessert.

Den ene av filene, «snoedybder_vaer_en_stasjon_dogn.csv», er obligatorisk. Den inneholder data fra en enkelt værstasjon. Den andre, «snoedybder_vaer_fem_stasjoner_dogn.csv» er frivillig og inneholder data fra fem ulike værstasjoner.

Begge filene viser snødybde og vær hver dag over flere år på følgende format:

Navn; Stasjon; Dato på formen DD.MM.ÅÅÅÅ; Snødybde i centimeter; Nedbør i millimeter; Middelttemperatur i grader celcius; Gjennomsnittlig skydekke; Høyeste middelvind i meter pr. sekund

Et lite utdrag av fila:

```
Venabu;SN13420;01.08.1980;0;0;13,8;5,7;2,6
Venabu;SN13420;02.08.1980;0;0,5;13,1;7,3;4,6
Venabu;SN13420;03.08.1980;0;26,3;12,2;6;2,6
Venabu;SN13420;04.08.1980;0;1,9;13,3;7;4,6
Venabu;SN13420;05.08.1980;0;0,4;11,8;7,3;4,6
Venabu;SN13420;06.08.1980;0;12,7;11,1;6,3;9,8
Venabu;SN13420;07.08.1980;0;0,7;6,9;3,3;9,8
Venabu;SN13420;08.08.1980;0;0;5,1;7;9,8
```

Deloppgaver for del 2 av prosjektet

Dere skal bygge på koden dere skreiv i del 1 av prosjektet.

- a) Skriv en funksjon som leser inn ei fil på følgende format, som inneholder data fra en værstasjon for hvert døgn over mange år:

Navn; Stasjonsid; Dato på formen DD.MM.ÅÅÅÅ; snødybde i centimeter; nedbør i millimeter; middelttemperatur; gjennomsnittlig skydekke i en skala fra 0 (skyfritt) til 8 (helt overskyet);høyeste middelvind i meter pr. sekund.

Merk at ikke alle dataene eksisterer for alle dagene, det vil være et «-» tegn for slike manglende data.

Funksjonen skal lagre disse dataene i en eller flere lister eller dictionaries. Funksjonen skal returnere disse listene eller dictionaryene. Det er en del av oppgaven at dere finner ut hvilke datastrukturer som er mest hensiktsmessige til dette.

- b) Dere kan regne med at det er skiføre om snødybden på værstasjonen er 20 centimeter eller mer. Regn ut antall dager med skiføre for hver vintersesong i datasettet. En vintersesong strekker seg fra oktober ett år til juni året etterpå. Bruk funksjonen fra del 1 oppgave d) til å beregne antall dager med skiføre i hver skisesong. Dette vil kreve at dere lager egne lister for hver skisesong som dere kan bruke som input til funksjonen fra del 1 oppgave d)
- c) Beregn trenden for antall dager med skiføre med å bruke funksjonen fra del 1 oppgave g). Året skisesongen starter skal være x-akse og antall dager med skiføre den skisesongen skal være y-akse.
- d) Plott antall dager med skiføre for hver skisesong (fra oppgave b) og trend (fra oppgave c) i samme plott, med året skisesongen starter på x-aksen og antall dager med skiføre på y-aksen. For å plote trenden, bruk formelen $y = ax + b$ til å regne ut to punkter, ett for året datasettet starter og ett for året datasettet slutter. Inkluder bare år hvor det er data om snødybde for mesteparten av skisesongen, det må være data for minst 200 dager i hver skisesong.

- e) Beregn veksten for den tenkte planten for hvert år i datasettet med bruk av funksjonen fra del 1 oppgave h). Plott dette for hvert år i datasettet. Inkluder bare år hvor det er temperaturdata for mesteparten av året, det må være data for minst 300 dager for at et år skal være gyldig. Dette vil kreve at dere lager separate lister for hvert år som kan brukes som parameter til funksjonen fra del 1 oppgave h)
- f) Bruk funksjonen fra del 1 oppgave f) til å finne den lengste perioden med tørke (ingen nedbør) for hvert år i datasettet. Plott resultatet. Inkluder bare år hvor det er nedbørsdata for mesteparten av året, det må være data for minst 300 dager for at et år skal være gyldig.
- g) Finn antall penværsdager for hvert år og plott dette. Man kan finne antall penværsdager ved å sjekke gjennomsnittlig skydekke. Hver dag med verdi 3 eller lavere er en penværsdag. Inkluder bare år hvor det er data om skydekke for mesteparten av året, det må være data for minst 300 dager for at et år skal være gyldig.
 - a. **Frivillig:** Bruk andre kategorier slik som «skyfritt» (verdi 0), dager med overskyet vær (verdi 6 eller høyere) og flere.
- h) For hvert år finn høyeste middelvind samt medianen for vindstyrke. For å finne medianen, lag ei liste over alle verdiene for det året, sorter lista, og plukk ut det midterste elementet i den sorterte lista. Plott dette for hvert år. Inkluder bare år hvor det er data om vind for mesteparten av året, det må være data for minst 300 dager for at et år skal være gyldig.
- i) En måte å sjekke trender i temperatur gjennom året er å ta gjennomsnittet av temperaturen for en tidsperiode (for eksempel ei uke eller en måned) og så sjekke om snittet endrer seg. Ved å ta gjennomsnittet så fjerner man at temperaturen går opp og ned fra en dag til den neste. Regn ut gjennomsnittstemperaturen for hver måned (for eksempel april 2007) og legg disse gjennomsnittene i ei ny liste. Bruk funksjonen fra del 1 deloppgave e) for å regne ut ei liste med differanser. Plott både lista over gjennomsnittstemperaturer og lista over differanser med måned og år på x-aksen.
- j) **Frivillig, avansert:** Ei fil med samme format for flere værstasjoner er inkludert. Dere skal gjennomføre alle tidligere deloppgaver og plote resultatet separat for hver værstasjon. Data fra de ulike værstasjonene skal være del-plott i samme hoved-plott slik at dere kan enkelt sammenlikne resultatene fra de ulike værstasjonene.