# Generell

Du skal lage et program som leser inn informasjon fra datasettet og presenterer ulike data i tre oversikter, en tekstlig og to grafisk. Du skal bruke datasettet fra filen Medier.csv som du finner i zip-filen vedlagt i oppgaven i TEAMS. *Du står fritt til å velge hvordan programmet skal presentere informasjon, så lenge presentasjonen er godt egnet til å vise det oppgaven spør etter*

***NB!. Oppgave 1 skal løses uten å bruke pandas bibliotek.*** *Oppgave 2 vil vise deg hvordan du kan løse samme oppgave med pandas. Krever ikke forkunnskap, bare å svare på et spørsmål etter at du har skrevet kode etter en gitt oppskrift. (Hvis du ikke rekker Oppgave 2 kan du gjøre det på tirsdag eller i studietid)*

## Praktisk info Last ned zip filen og pakk ut på din egen PC. Du vil finne to mapper: datasett og out. Du finner csv-fil under datasett-mappen. Du skal lagre bildefiler du lager i oppgaven under out-mappen.

Når du er ferdig, skal du pakke denne mappen og levere i TEAMS. Du skal levere en python fil med koden du lager for alle oppgavene, to bildefiler hver av de to grafiske framstillingene og en word dokument med output fra konsoll.

Hjelpemidler: Du kan bruke Aunivers, w3schools og andre oppslagsverk (sjekk med Miriam), egne notater, OneNote, tidligere kode.

# Oppgave 1 (LES HELE OPPGAVE FØR DU/DERE BEGYNNER Å KODE. LAG EN PLAN)

## Info om innhold (dette hadde du vanligvis måtte finne selv i forberedelses tid) Medier.csv inneholder informasjon er automatisk skapt fra SSB. Hvis du åpner filen Medier.csv i VS Code vil du se noe slik:

Et bilde som inneholder tekst, skjermbilde, programvare, Multimedieprogramvare

Automatisk generert beskrivelse

Det eret er tre problemer med denne datafila:

1. De to første radene inneholder ikke data vi er interessert i, og man må «hoppe over»/ignorere dem ved innlesing.
2. Det mangler tall flere steder. På slike steder står det enten ett eller to punktum, avhengig av hvorfor det mangler tall. I begge tilfelle må det settes np.Nan(not a number fra numpy bibliotek) som verdi ved innlesing. I tillegg leses resten av tall som streng, og må konverteres til int.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| .. = Tallgrunnlag mangler. |  |  |  |  |  |  |  |  |  |  |
| . = Ikke mulig å oppgi tall. |  |  |  |  |  |  |  |  |  |  |

1. Linje 3 har overskrifter. Kolonneoverskriftene er veldig lange. På bildet over ser du på kolonne B hele overskriften. Også de andre tallkolonnene har den samme lange overskriften. Vi er bare interessert i å ha selve årstallene fra overskriftene (f.eks. 2010 fra overskrift som viser på kolonne B). Du står fritt for å finne en løsning på utfordring av å få en liste med de årene vi er interessert i.

Fra linje 3 og til slutten av filen får vi en linje per medie vi skal analysere. Første elementen er navnet på media, mens resterende elementer i listen av antallminuter for den medie.

## Deloppgaver

* 1. **Lag et program som lager en grafisk framstilling der man ser hvordan utvikling har gått for de ulike medietypene fra 2010 til 2019.**
  2. **Utvid programmet slik at den lager en ny grafisk framstilling der man ser utvikling i same graf bare av kategoriene Hjemme-PC, Bøker og Internett. Program skal lagre resultatet som et bilde.**
  3. **Utvid programmet slik at den viser en tekstlig oversikt over de årene der man brukte mest og minst minutter på Internett, og hvor mye minutter ble brukt.** *Tips: For å kunne beregne dette må du tilpasse innholdet i det aktuelle datafeltet for at den regner størst og minst riktig.*

Husk å skrive passende kommentarer i koden og passe på brukervennlighet og gjenbruk i kode. Dekorere grafene med passende informasjon og farger.

Oppgave 2  
1. Lag et program som løser oppgaven med pandas ved hjelp av denne nettside: [Instruksjoner for å gjennomføre oppgave med pandas bibliotek](https://ndla.no/subject:1:f18b0daa-6507-4025-8998-b8a11c8ccc70/topic:1:edeb4b57-277d-45de-abe2-b96ee0d67ad8/topic:1:55397d8e-0259-4772-8c4a-26c25dc3fa36/resource:9d56d6be-26f8-4e4c-9438-f81bec176160) (husk å installere pandas hvis du ikke har gjort det allerede)

2. Skriv kort hvilke fordeler og eventuelt ulemper du ser ved å bruke pandas bibliotek for å lese store datasett.