|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Torsdag | Fredag | Lørdag | Søndag | Mandag | Tirsdag | Onsdag | Torsdag | Fredag |
| Formiddag | Forelæsning | Forelæsning | * Scrape data / Hent data fra API. * Sorter og opsæt data til analyse. * Skriv intro-afsnit * Start på metode-afnist (e.g. Data, Ethics). | * Færdiggør metodeafsnit og beskrive metodevalg (ML? Scrape? Sentiment?) * Påbegynd analyse |  |  |  |  |  |
| Eftermiddag | * Færdiggør aflevering 2. * Første udkast til Problemformulering. * Lokalisering af data. | * Scrape data / Hent data fra API. * Sorter og opsæt data til analyse. * Skriv intro-afsnit * Start på metode-afnist (e.g. Data, Ethics). * (Peegrade). | HUSK FORELÆSNING SUPERVISION | HUSK SUPERVISION |
| SNACK/ KAGE-ANSVARLIG: |  |  |  | Naeim | (Mette-Marie) | Nikolaj | Mette-Marie | August | Naeim |
| Fraværende: |  | August (17-) | August (all day) | Nikolaj (16-) |  | Naeim (14-16)  Nikolaj (19-) | Naeim (-13) | (August (17-)) |  |

**Forventningsafstemning:**

* Vores ambition er som minimum at få 10 i karakter, men vi stiler efter 12.
* Vi mødes 8 timer hver dag (9-17), med 30-60 minutters frokostpause.
  + De sidste par dage (Onsdag, Torsdag, Fredag) skal der dog tages forbehold for ekstra arbejde om aftenen, sådan at projektet kan færdiggøres.
* Snak med gruppen hvis du er gået i stå eller er i tvivl om hvad du selv, eller gruppen, skal gøre som det næste skridt i processen.
* Pauser er vigtige! Og det samme er vand, mad og kaffe (og snacks)
  + Overspringshandlinger er derimod overhovedet ikke vigtige!

**Noter til scraping:**

* Forsøg at skaffe titlerne (eller anden type ID) på jobopslag når der scrapes, for at eliminerer dupletter på en nemmere måde.
  + Forsøg samtidigt at skaffe by eller postnummer på jobopslagene.
* Da nogle under-kategorier fra oversigten på hjemmesiden overlapper hinanden, men som deler samme url, er disse direkte dupletter fjernet (f.eks. fremgår underkategorien detailledelse både under ”Ledelse og personale” samt ”Handel og service”, men de har den samme url).
  + Som følge af overlappene vil der også være kategorier som indeholder de samme jobopslag, hvilket betyder at det samlede antal jobopslag ikke er en direkte sum af

Lørdag d. 25.08.2018:

* Vi har fået lavet en scraping procedure, sådan at vi kan scrape antallet af jobopslag fra JobIndex. Processen er bygget op sådan at vi selv kan vælge hvilke områder, sektorer, jobtyper og år vi scraper for.
* Vi har fået skabt den grundlæggende DataFrame der indeholder alle kombinationer af år, område, sektor og jobtype.

Næste skridt: Scrape antallet af jobopslag for hver karakteristika og merge det med vores DataFrame, sådan at der er én søjle per niveau af antal jobopslag (i.e. N\_jobtype, N\_sektor, N\_område, N\_år).