

# Intervjuoppgave Data Scientist

Ingeborg

24. november 2019

- 1 Nøkkeltall for Unge lovende
- 2 Bingewatching av Unge lovende
- 3 Kommentarer til løsning av oppgaven

# Nøkkeltall for Unge lovende

Totalt antall eventer: 559 620

Totalt antall sesjoner: 254 040

Gjennomsnittlig antall eventer per sesjon: 2.2

Antall eventer med unike brukere: 146 567

Antall brukere som har sett alle 6 episodene: 43 487

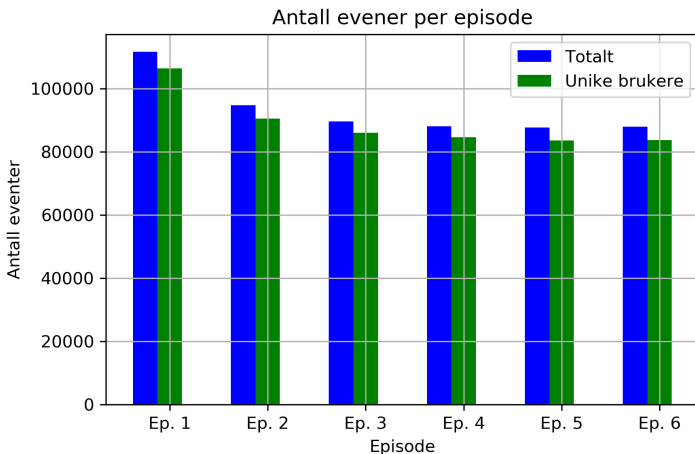
Antall eventer for hver episode:

	Totalt antall brukere	Unike brukere	Ikkeunike brukere
Episode 1	111 609	106 418	5 191
Episode 2	94 743	90 504	4 239
Episode 3	89 623	86 031	3 592
Episode 4	87 992	84 599	3 393
Episode 5	87 684	84 599	4 071
Episode 6	87 969	83 750	4 219

NB: Jeg har fjernet alle eventer der `timeWithinVisit = 0`.

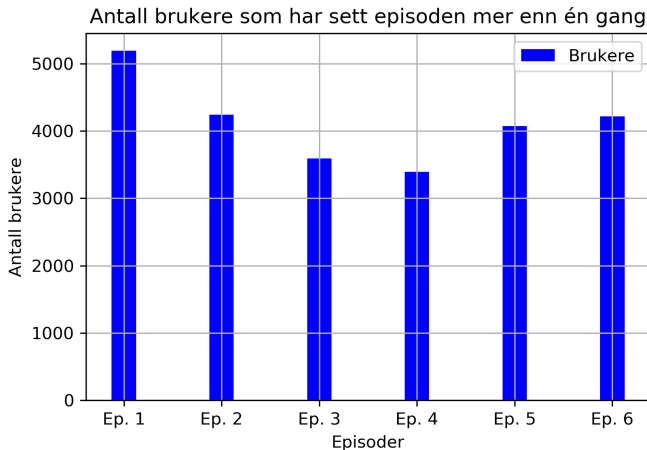
# Nøkkeltall for Unge lovende

Episode 1 er episoden som har blitt vist flest ganger.  
De fleste brukerne har sett hver episode én gang.



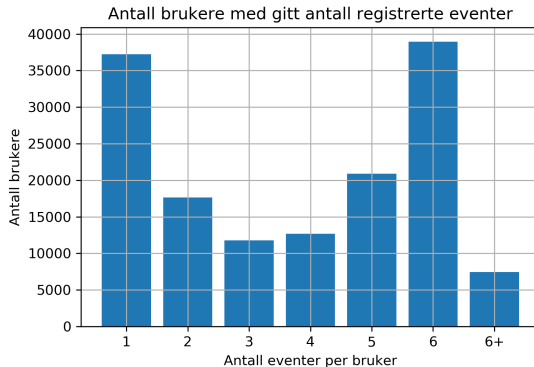
# Nøkkeltall for Unge lovende

De to første og de to siste episodene er de som har blitt sett omigjen av flest brukere.



# Nøkkeltall for Unge lovende

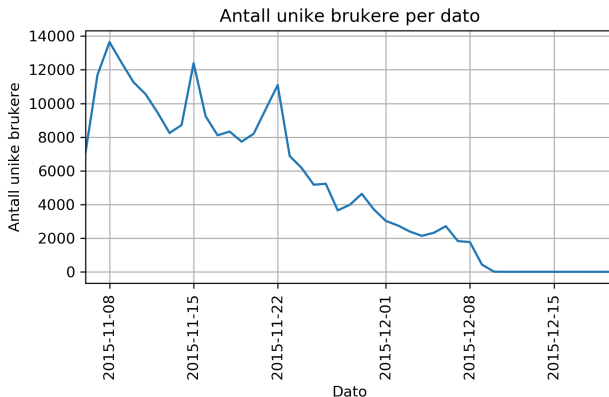
I gjennomsnitt er det registrert 3.8 eventer per bruker, mens medianen er 4 eventer per bruker. Grafen under viser antall brukere med et gitt antall registrerte eventer per bruker. Typetallet er 6 episoder, det vil si at det antallet episoder som går forekommer flest ganger.



# Nøkkeltall for Unge lovende

Det var flest brukere som starta sesjoner på datoene 8., 15. og 22. november 2015. Disse datoene er søndager.

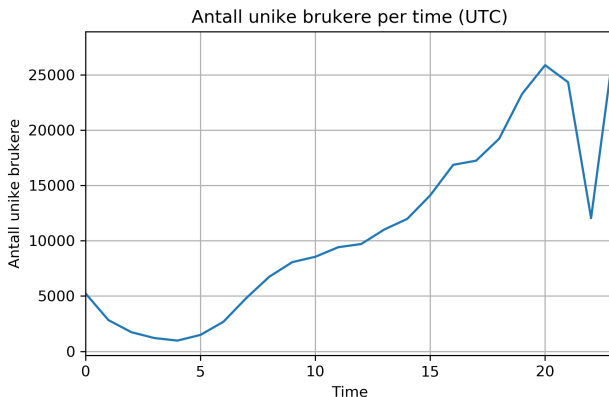
Serien hadde flest loggførte eventer rett etter at den ble tilgjengeliggjort.



# Nøkkeltall for Unge lovende

Det mest populære tidspunktet å starte en sesjon på er i tidsrommene kl 21-22 og 00-01 (norsk tid).

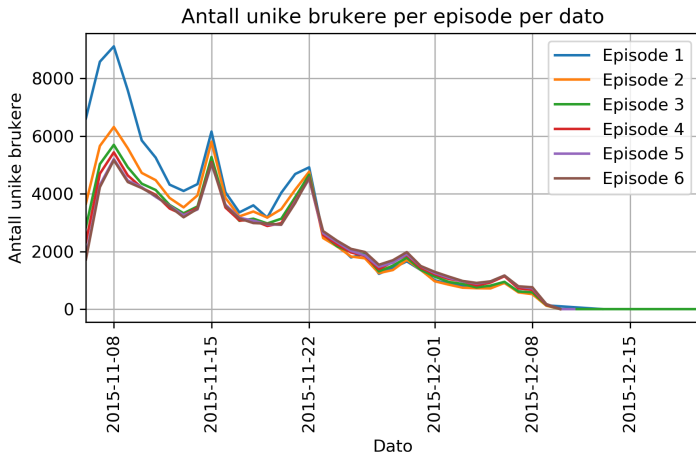
(Det siste tidsrommet ville jeg gjerne undersøkt nærmere om jeg hadde hatt bedre tid, ettersom dette bryter mønsteret i grafen og ser feil ut. Kunne sett på tidssemlene til eventene.)





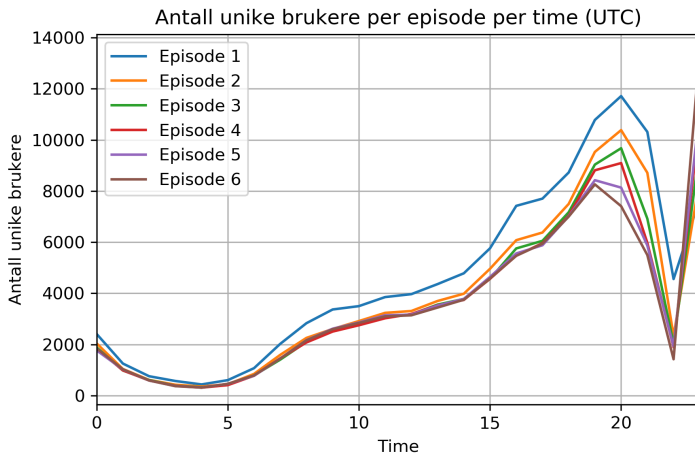
# Bingewatching av Unge lovende

Deler vi opp eventer per dato på episoder ser vi at disse linjene følger hverandre tett. Derfor er det rimelig å anta at mange brukere ser flere episoder etter hverandre.



# Bingewatching av Unge lovende

Det samme funnet får vi om vi ser på når på døgnet sesjonene har startet.



# Bingewatching av Unge lovende

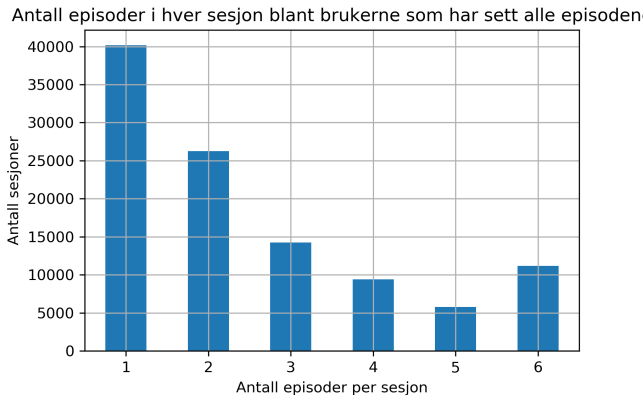
Totalt har 43 487 brukere sett alle seks episodene fordelt på 106 986 sesjoner.

Tabellen under viser hvor mange episoder hver bruker (som har sett alle episodene) har sett per sesjon.

<b>Antall episoder</b>	
Min	1
Median	2.0
Gjennomsnitt	2.5
Typetall	1
Maks	6

# Bingewatching av Unge lovende

Av brukerne som har sett alle seks episodene fordeler de seg slik over hvor mange episoder de har sett per sesjon:



**I overkant av 11 000 brukere har sett alle seks episodene i løpet av samme sesjon.**

Antagelser jeg har gjort:

- Har fjerna alle eventer der  $\text{timeWithinVisit} = 0$ , fordi jeg antar at episoden da ikke har blitt vist

Mulige forbedringer:

- Har ikke vasket datasettet
- Har ikke tatt hensyn til varighet på eventene annet enn å fjerne eventer med  $\text{timeWithinVisit} = 0$
- Kanskje skulle det vært satt ei minimumsgrense (nær varigheten av hver episode) for  $\text{timeWithinVisit}$  for å si at episoden har blitt vist

Videre arbeid:

- Undersøke nærmere de eventene der varigheten er kortere enn en episode. Har brukeren sett episoden ferdig seinere, f.eks.?
- Er det feil som gjør at en bruker er registrert med 239 eventer? Er det feil i datasettet som går igjen flere steder?
- Finne ut hvorfor så mange sesjoner starter mellom kl 00 og 01 (norsk tid).
- Finne ut mer om oppdelingen av episodene blant brukerne som har sett hele serien. Eks: har de sett først 3 episoder, så 1, deretter 2 episoder. Hvordan er mønsteret?

Andre nøkkeltall som kunne vært interessante med flere data tilgjengelig er for eksempel:

- Hvilken plattform er mest populær?
- Hvor mange ser lineært vs web/mobil?
- Demografien til seerne