

Innhold

Introduksjon	1
Dataprosessering	1
Nøkkeltall	2
Bingewatching	4

Introduksjon

Hensikten med dette dokumentet er å besvare de to delspørsmålene i case casebeskrivelsen. Disse er:

- 1) Hvilke nøkkeltall vil du trekke frem for å oppsummere Unge Lovende i NRK TV? Visualiser der det er nyttig.
- 2) Alle episodene av Unge Lovende ble sluppet samtidig. Hva kan du si om bingewatching knyttet til slippet av Unge Lovende? Visualiser.

Dokumentet inneholder enkelte utklipp generert fra R-scriptet «Unge_Lovende» som også ligger vedlagt på Git.

Dataprosessering

Vi starter med å ta en titt på dataeksporten fra Google Analytics. Datafilen inneholder opprinnelig 4 kolonner. Disse kolonnene gir informasjon om når ulike seere (userId) har startet ulike sesjoner, hvilke episode som er sett i sesjonen og tidspunkt brukeren så ferdig de ulike episodene. En håndfull rader av rådataen vises i utklippet under.

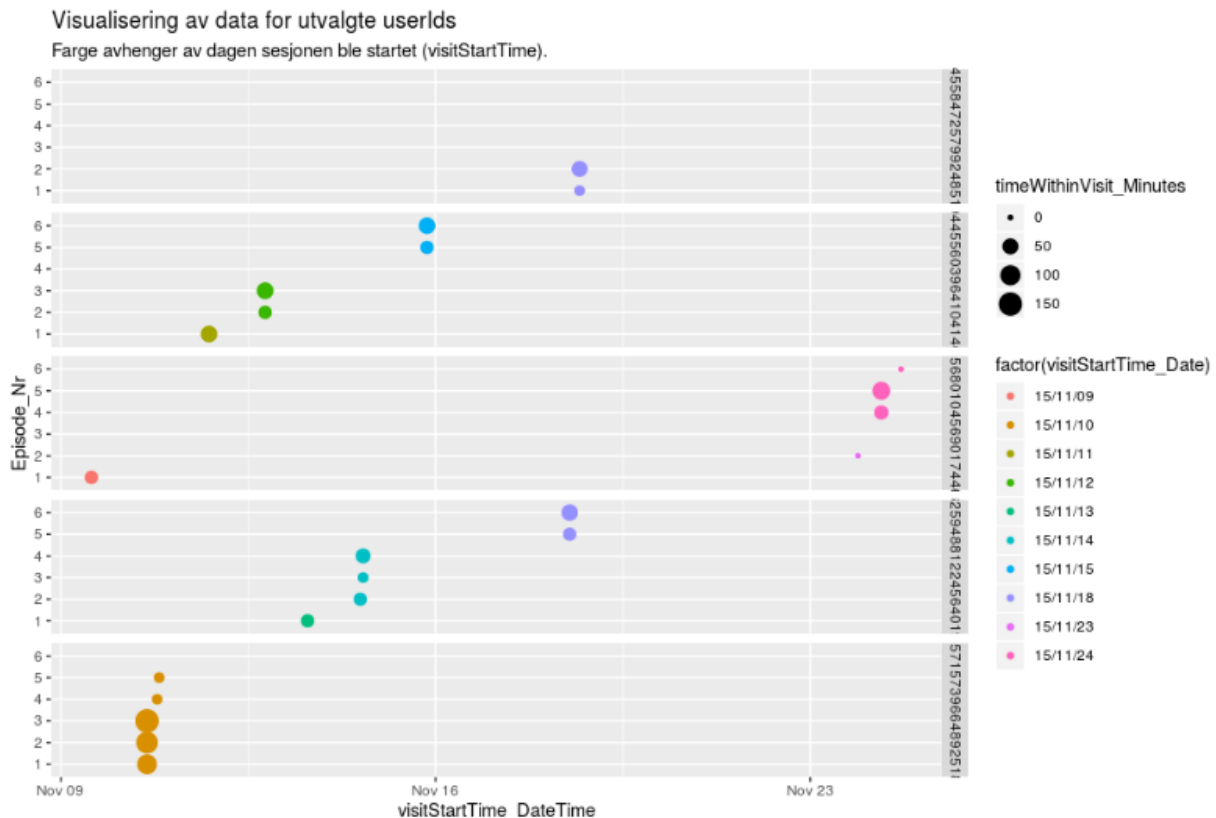
userId	programId	visitStartTime	timeWithinVisit
2.172883e+18	KMTE20000514	1447617021	1614261
2.214427e+18	KMTE20000314	1447578419	1782834
3.261021e+18	KMTE20000114	1447180923	206935
3.300131e+18	KMTE20000114	1447139006	1630400

Det er hensiktsmessig å gjøre litt prosessering av rådataen, mer spesifikt:

- Trekke ut episodenummer
- Transformere timeWithinVisit fra millisekund til minutt
- Konverterer fra unix timestamp til et lettere forståelig tidsformat

For å få en bedre forståelse av hvordan dataen henger sammen, kan det være nyttig å studere dataen for enkelte seere nærmere. Dette kan både fungere for å kvalitetssikre at man har forstått dataen rett, samtidig som det kan generere hypoteser om mønstre i dataen som kan være interessant å studere nærmere. Figuren under illustrerer dataen for 5 tilfeldige seere. Av figuren

fremgår det at seere konsumerer serien ulikt: noen ser mange episoder rett etter hverandre, mens andre fordeler sesjonene over flere dager.



Dataen til de utvalgte userId'ene illustrerer at enkelte brukere typisk kan se mer enn 1 episode pr dag. Rent intuitivt virker dette rimelig med tanke på hvordan folk ser serier, og det gir noen ideer til hvilke nøkkeltall som kan være hensiktsmessig for å oppsummere Unge lovende og andre serier.

Nøkkeltall

Det er flere nøkkeltall som kan være interessant for å oppsummere Unge Lovende. For å utforske hvilke som er mest nyttig, starter vi med de mest nærliggende nøkkeltallene, og legger deretter til nye nøkkeltall for å dekke informasjonsbehovet man ønsker i oppsummeringen.

Først og fremst er det interessant å få et overordnet bilde av volumet av visninger, både totalt antall sesjoner og unike seere. I koden under hentes følgende nøkkeltall

- Antall episoder (dynamisk utifra dataen som er tilgjengelig),
- Antall unike visninger,
- Antall unike seere (userId),
- Snitt antall visninger pr userId,
- Ukedag med flest visninger.

Tabellen under viser de nevnte nøkkeltall for «Unge lovende». Det vil typisk være interessant å presentere tilsvarende nøkkeltall for andre serier i samme tabell for å muliggjøre sammenligning av ulike serier.

Serie	episoder_n	visninger_unike	userId_unike	visning_pr_user_id	Dag_1
Unge lovende	6	589774	148375	3.974	Sunday

Fra tabellen ser vi at det er rett under 600.000 unike visninger fordelt på ca 150.000 unike seere, med andre omtrent 4 visninger i snitt pr userId. Disse nøkkeltallene gir et overordnet bilde av volumet av antall seere og visninger. Umiddelbart fremstår det som nyttig å legge til ytterligere nøkkeltall for å belyse *hvordan* serien blir sett. Dette kan være f.eks.

- Andel userId som har sett alle episodene,
- Andel userId som har sett episode nr. 1, men ikke flere («churn»)
- Andel "binge watchere"
- Median-antall episoder hver userId ser

Nøkkeltallene over beregnes ved å først finne relevante indikatorer på userId-nivå, og videre aggregere dette opp på serie-nivå. Tabellen under viser slike indikatorer på userId- nivå.

userId	episode_forste	episode_siste	episode_unike	sesjoner_unike	sesjoner_unikeDager	BW_flg
3.549506e+13	1	6	6	4	2	TRUE
3.852729e+13	4	6	3	2	2	FALSE
6.781039e+13	1	1	1	1	1	FALSE
9.538408e+13	1	6	5	3	1	TRUE

Her er «Bingewatching» (BW_flg) for én gitt seere definert som

True hvis seeren/userIden i snitt ser minst tre episoder pr dag $h \cdot n$ ser serie
False ellers

Denne definisjonen av Binge-watching er litt vilkårlig valgt, men gir dog en indikasjon på om hver enkelt userId typisk ser flere episoder i slengen. Ved å aggregere indikatorene fra userId-nivå til serie-nivå, finner vi nøkkeltall a)-d) som legges til opprinnelig tabell med nøkkeltall:

Serie	episoder_n	visninger_unike	userId_unike	visning_pr_user_id	Dag_1	sett_alle_episoder_andel	BW_andel	episode_unike_median	churn_etter_1_episode
Unge lovende	6	589774	148375	3.974888	Sunday	0.39	0.42	4	0.23

Utifra tabellen ser man at andelen userIder som har sett alle episodene er 39%. Dette fremstår som en relativt høy andel, og hadde vært interessant å sammenligne med andre serier. At median-antallet episoder hver userId ser er 4 fremstår også som høyt.

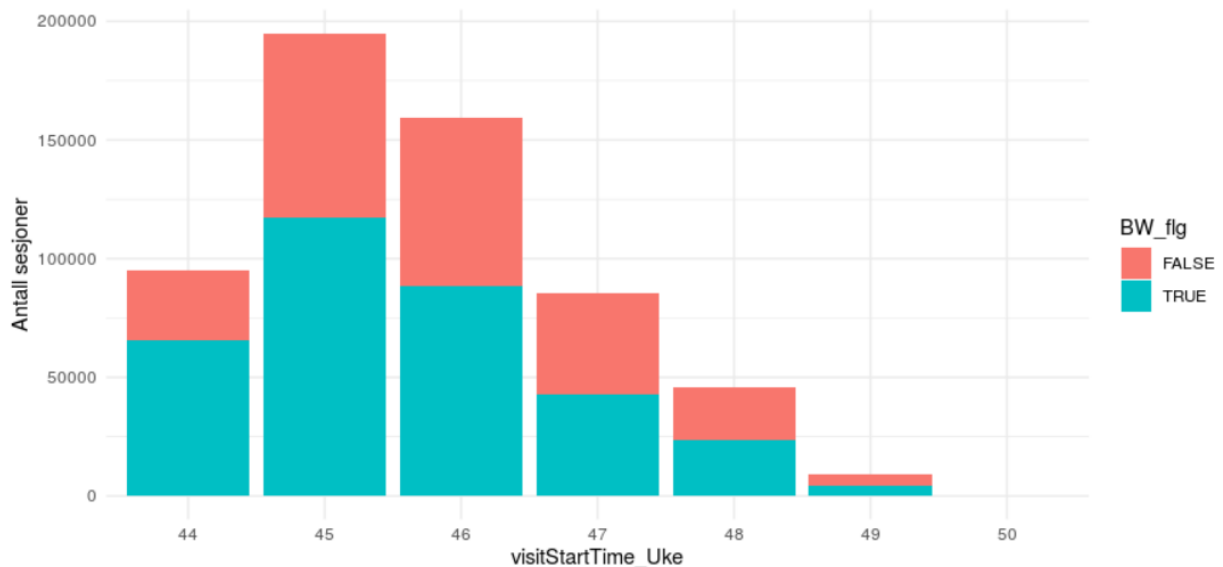
Tilsvarende hadde det vært interessant å sammenligne nøkkeltall b) «churn_etter_1_episode» med andre serier for å avgjøre om 23% er usedvanlig lavt eller ikke. Dersom andelen er høy kan man argumentere for innholdet i første episode ikke treffer seerne godt nok.

Basert på den valgte definisjonen av "Bingewatcher" ser vi at en betydelig andel (42%) i snitt ser minst 3 episoder de dagene vedkommende ser på Unge lovende.

Nøkkeltallene vi har sett på hittil i kapitelet gir overordnet informasjon om serien Unge Lovende. I det påfølgende kapittelet tar vi en nærmere titt på aspektet bingewatching for å få et klarere bilde av i hvor stor grad seere ser mange episoder etter hverandre.

Bingewatching

Utifra vår definisjon av bingewatching ble 42% av userId'ene definert som «Bingewatcher». Det kan være interessant å få et inntrykk av hvor mange av Google Analytics-sesjoner som ble gjennomført av hhv. «Bingewatcher» og «Ikke-bingewatcher». Figuren i utklippet under viser antallet sesjoner pr uke for de to ulike typene seere:



Av figuren fremgår det at «Bingewatcher» jevnt over har flere sesjoner enn «Ikke-bingewatcher». Dette er interessant: 58% av sesjonene ble gjennomført av «Bingewatcher», selv om denne typen seere er i mindretall (42%). Bingewatcherne står med andre ord for over halvparten av alle sesjonene. Av figuren ser man at andelen bingewatcherne tilsynelatende er størst de to første ukene serien er aktiv, altså uke 44 og 45.

Utifra vår definisjon av bingewatcherne fremstår det som om en relativt stor andel av seerne konsumerte hele eller deler av serien Unge lovende ved å se flere episoder kort tid etter hverandre. Vi så også at bingewatcherne står for over halvparten av alle GA-sesjonene, noe som tyder på at bingewatching utgjør en stor del av konsumet av serien Unge lovende.