

Intervjuoppgave - Data Scientist

Victoria Troland

January 19, 2016

I dette dokumentet har jeg analysert et GA-datasett som registrerte aktiviteten av serien Unge lovende (UL) på NRK.no¹.

Serien ble i november lagt ut på NRK.no og består av seks episoder som alle ble lagt ut samtidig.

1 Oppsummering av Unge lovende - nøkkeltall

Tallene for å oppsummere UL kan deles inn i to deler: seere og avspillingstidspunkt. Seer-delen fokuserer på antall ganger seerne har sett på UL. Avspillingstidspunktene oppsummerer når og hvor ofte UL ble spilt av.

1.1 Seerne av serien

I løpet av perioden 6. november 2015 og frem til 8. januar 2015 ble et registrert 589 775 unike visninger av serien Unge lovende.

Av 589 775 unike visninger ble det registrert 148 376 unike seere. I gjennomsnitt så hver unike seer på UL 4 ganger ($=3.98$). Seerne så på UL mellom 1 til 240 ganger. Størst andel seere så UL 6 ganger (34.2 %).

I tabellen 1 vises antall ganger UL er sett per seer. Tabellen er delt inn i antall ganger seerne så på UL fra 1 til 10 ganger, hvor mange unike seere som hadde sett UL n-antall ganger, hvor mye dette utgjorde i prosent og til sist den kumulative prosenten fra 1 til 10 ganger.

¹I forbindelse med denne oppgaven ble det skrevet 4 skript: 2 shell-skript, 1 Perl-skript og 1 R-skript. Skriptene genererte tallene og figurene som er blitt brukt i oppgaven og er lagt i samme *repository* som dette dokumentet.

Tabell 1: Distribusjon av antall ganger sett UL per seer

Ganger sett	Antall brukere	Prosent (%)	Kummulativ (%) %
1	37252	25.1	25.1
2	17468	11.8	36.9
3	11104	7.5	44.4
4	10141	6.8	51.2
5	12571	8.5	59.7
6	50685	34.2	93.8
7	5273	3.6	97.4
8	1411	1.0	98.3
9	687	0.5	98.8
10	462	0.3	99.1
11-240	1322	0.9	100

Resultatene viser at 93 % av alle de unike seerne så UL mellom 1 og 6 ganger og under 7 % ser UL mer enn 6 ganger. Omtrent 1/4 av seerne er registrert med 1 registrert avspilling av serien. Den største andelen seere registrert i forhold til antall avspillinger (34.2%) ble registrert med 6 avspillinger av serien. Dette tyder på at disse seerne så alle episodene i serien når de lå op NRK.no.

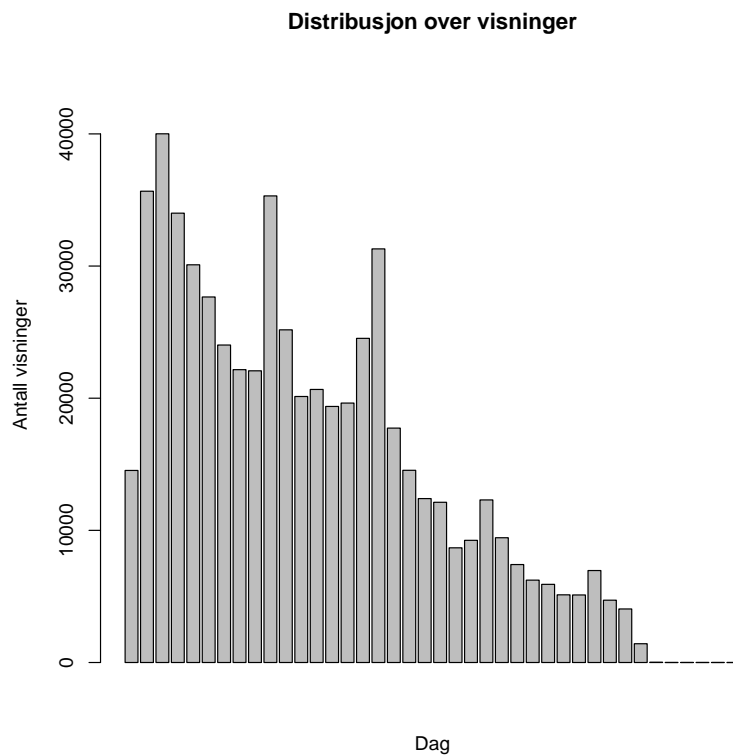
1.2 Avspilingstidspunkt

De registrerte datoene seerne så på UL var i intervallet 6.11.2015 - 20.12.2015.

Figuren 1 antall visninger per dag som ble registrert. Gjennomsnittlig ble UL spilt av 14 744 ganger per dag. På det høyeste var det registrert 40 007 visninger (søndag 8. november) og på det laveste var det 1 registrert visning (søndag 20. desember). 92 % av alle visninger av UL forekom i november og 8 % av visningene forekom i desember.

Figuren 1 viser en nedengående trend fra utgivelsesdato og frem til slutten av registreringen. De siste 6 dagene av registrerte dager ble UL kun spilt av 29 ganger.

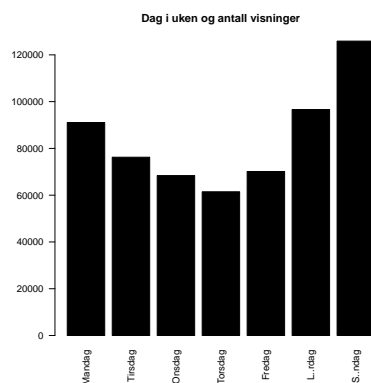
Figur 1: Antall visninger av UL i november og desember - kronologisk



Figur 2 viser at UL ble sett oftest på søndager og færrest på torsdager. Serien ble spilt av flest ganger i løpet av helgen og hadde deretter en nedgang i ukedagene, før det økte igjen til helgen.

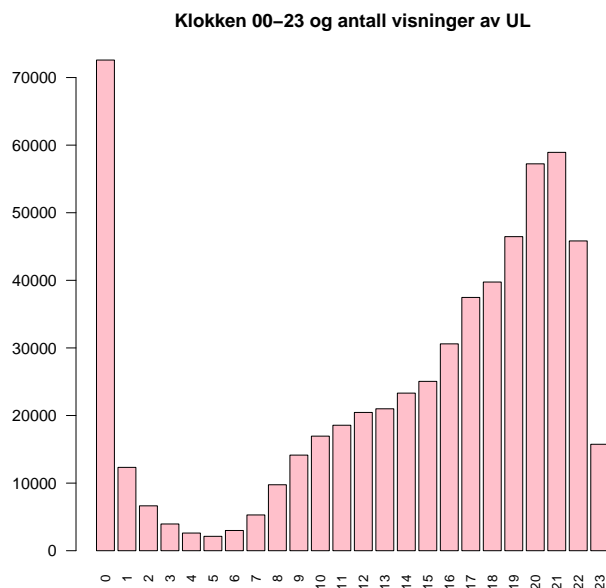
Dette reflekteres også i figur 1 som inneholder flere "hopp" som viser hvordan visningsantallet øker i helgene.

Figur 2: Antall visninger av UL per uke dag



Figuren 3 illustrerer når UL ble spilt av sortert etter time på døgnet. Figuren starter med kl. 24 og avslutter med kl. 23. Serien ble flest ganger spilt av kl. 24. Grafen er u-formet og synker raskt etter kl. 24, og stiger gradvis frem til kl. 21, før den synker frem til kl. 24.

Figur 3: Avspillingstidspunkt per time i døgnet



Figur 3 har den samme U-formen som figur 2. Figurene illustrerer samlet at mulige ytre begrensninger påvirker avspillingstidspunkt av UL. Dette vises ved at det er færre avspillinger på ukedager, dager når folk ofte er på jobb eller skole og på tidspunkt som er typisk forbundet med jobb, skole eller søvn.

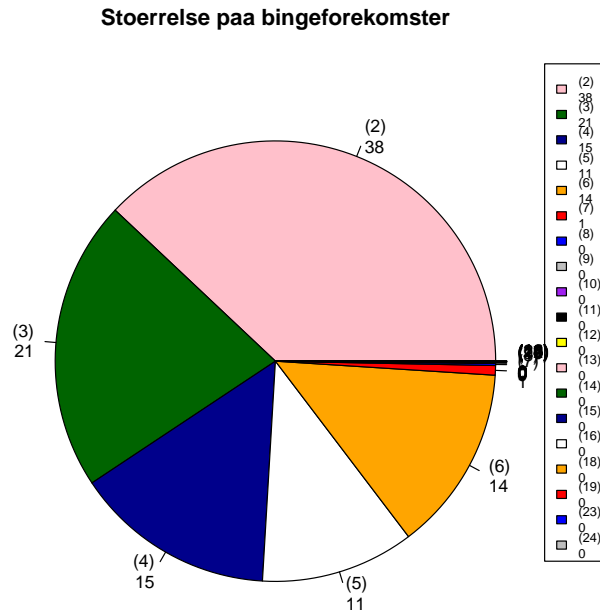
2 Bingewatching og Unge lovende

Bingewatching kan defineres som det å se flere episoder av samme TV-serie etter hverandre. I denne oppgaven velger jeg å tolke bingewatching som å se 2 eller flere episoder i løpet av samme dag.

Datasettet viste at unike seerne hadde sett UL mellom 1-24 ganger i løpet av av et døgn. Når forekomstene av 1 episode UL per dag ble fjernet, gjensto 63.6 % (214 434) av det som kan kalles "binge"-forekomstene. Dette viser at i 63.6 % av tilfellene en seer vil se på UL, vil de "binge". I kun 36.4 % av tilfellene vil en seer nøye seg med å kun se 1 episode i løpet av en dag. Gjennomsnittlig vil en seer se 3.45 episoder i løpet av en "binge".

I figur 4 vises antall forekomster av "bingewatching" og hvor mange episoder som ble sett i løpet av en "binge". Tallene i parentes viser til antall episoder sett når seerne "bingewatched".

Figur 4: Antall episoder sett per døgn av unike seere



Tallene under er prosentandel av alle forekomstene av "binging". I kakediagrammet er det vises det at 2 episoder per "binge" forekommer oftest. Ellers er det tydelig at 2-6 episoder er mest typisk for en "binge" med UL. Etter 6 episoder faller forekomstene med "binging" i stor nok grad til at de ikke gir utslag i en prosentandel.

Det er en liten økning i kakestykket som representerer 6 episoder. Dette tyder på at denne andelen seere (14 %) så hele serien i løpet av en "binge". Noe som også reflekteres i tabell 1 der en større andel av unike seere (34.2 %) ble registrert med totalt 6 avspillinger av serien.

Det tyder ikke på at det finnes en korrelasjon mellom "binge" datoer og størrelsen på "bingewatching". Det er heller ingen korrelasjon mellom størrelsen på "bingewatching" og hvilken dag seerne "binger".

3 Konklusjon

Nøkkeltallene og undersøkelsene rundt bingewatching viser at de fleste seerne har sett UL mer enn en gang. I begge tilfelle har seerne i høy grad sett seriene innenfor et intervall mellom 1 og 6 ganger eller forekomster.

Tidsmessig finns det tydelige mønstre i avspillingene av UL. Avspillingene avtok gradvis jo lengre serien hadde ligget på nett. UL var mer populær rundt helgen, og på tidspunkt som ikke er typisk forbundet med arbeidstid eller søvn.

Ut ifra den valgte definisjon på "bingewatching" her viste datasettet at nesten $2/3$ av seerne "bingewatched" da se så på UL.