

BRUK AV LOKASJONSDATA FOR Å SYNLIGGJØRE VIRKNINGER VED FRILUFTSLIVSTILTAK

PUBLIKASJON NR
M-2705 | 2024





Tittel Bruk av lokasjonsdata for å synliggjøre virkninger ved friluftslivstiltak	Dato 11.12.2023
Forfatter(e) Rune Aasgaard, Julia Olsson og Dan Ullén	PUBLIKASJON NR M-2705 2024
	Sider 57

Oppdragsgiver(e) Miljødirektoratet - Friluftslivsseksjonen	Kontaktpersoner hos oppdragsgiver Nils Yngve Berg
Sammendrag: Prosjektet utforsker bruk av mobile lokasjonsdata for å spore endringer i aktivitetsnivå i befolkningen som svar på friluftslivstiltak. Gjennom analyser av Smaalensbanen i Vestby og Øylibikk-hytta i Kristiansand, med data fra STRAVA og Telia samt tellerdata, kan vi se at lokasjonsdata kan gi prediksjoner om områdebruk. Resultatene indikerer metoder for trendidentifisering og fremtidsprognoser, men mer data og geografisk utvidelse trengs for bedre nøyaktighet. Prosjektet inkluderer også en POC for pedagogisk data-visualisering.	Utgitt av: Norkart



SAMMENDRAG

Prosjektets formål har vært å utforske i hvilken grad mobile lokasjonsdata og ulike former for mobilitetsdata kan benyttes som indikatorer for å måle aktivitet og endringer i aktivitetsnivået i befolkningen, som en følge av implementeringen av konkrete friluftslivstiltak. Studien har konsentrert seg om to områder: Smaalensbanen i Vestby kommune og Øyliblikk-dagsturhytta i Kristiansand kommune.

Resultatene fra analysene indikerer at det er gode muligheter for å utvikle indikatorer basert på data fra STRAVA og Telia, selv med visse iboende usikkerheter. Metoden for å beregne endringer i relativ trafikk har vist seg å være effektiv for å identifisere trender over tid. Disse datakildene kan potensielt være representative for faktisk trafikk, forutsatt at de behandles med egnede modelleringsmetoder. Det er imidlertid viktig å understreke at for å øke nøyaktigheten og påliteligheten i modellene, er det nødvendig å inkludere mer data og utvide analysen til flere geografiske områder.

For å utnytte potensialet i ulike lokasjonsdata er vi avhengige av kvalitetsrike tellerdata med bred geografisk spredning. Derfor er det essensielt at Miljødirektoratet, i samarbeid med andre relevante aktører som har tellere utplassert, samler eksisterende kunnskap og utarbeider en manual for metoder for utplassering, samt en veileder for bruk av tellere til forskjellige formål. Videre er det viktig å gjøre disse dataene mer tilgjengelige for forskning ved å etablere en nasjonal arena for innsamling, harmonisering og deling av tellerdata på nasjonalt nivå. Det bør også opprettes en nasjonal database for mer tradisjonelle stordata, som intervjuer, for å gjøre dem tilgjengelige for fremtidig arbeid.

Miljødirektoratet vil ha stor nytte av å initiere dialog med ledende mobiloperatører – Telenor, Telia og ICE – for videre utvikling av modeller. Det er nødvendig å etablere en avtale som sikrer tilgang til nasjonale data, som kan benyttes i ulike utviklingsprosjekter. Vi er også avhengige av at STRAVA lanserer et API og at det bygges opp en infrastruktur som gjør at disse dataene enkelt kan brukes i slike arbeider.

I prosjektet har vi også utviklet en Power BI-rapport med et interaktivt kart for pedagogisk visualisering av kildedataene. Denne løsningen kan relativt enkelt skaleres til å omfatte hele Norge, men dette forutsetter at STRAVA lanserer et API, og at det inngås en avtale som gir tilgang til landsdekkende mobilnettverksdata. De statistiske analysene er ikke inkludert i visualiseringsløsningen per nå, da de innebærer en del manuelle prosesser. Å integrere disse i et brukervennlig grensesnitt vil kreve ytterligere utvikling som ikke var mulig innenfor rammen av dette prosjektet.



1. Innhold

2. INNLEDNING	6
2.1 Bakgrunn og formål	6
2.2 Undersøkte områder	7
2.2.1 Øyliblikk-dagsturhytta	7
2.2.2 Smaalensbanen	8
3. DATA	9
3.1 Mobile nettverksdata	10
3.1.1 Telenor	10
3.1.2 Telia	10
3.1.3 ICE	11
3.2 Data fra smarttelefon	11
3.2.1 Polar	11
3.2.2 STRAVA	12
3.3 Passeringstellere	12
3.3.1 Smaalensbanen	13
3.3.2 Øyliblikk-dagsturhytta	13
3.4 Transaksjoner	14
3.5 Oppsummering av aktuelle datasett	14
4. ANALYSEMETODE OG BEARBEIDING AV DATA	16
4.1 Måling av friluftaktivitet	16
4.2 Passeringsteller	17
4.2.1 Utfordringer med data fra passeringstellere	18
4.2.2 Passeringsteller for Vestby	18
4.2.2.1 Langtidsvariasjoner	21
4.2.3 Oppsummering - Passeringsteller for Vestby	21
4.2.4 Passeringsteller for Kristiansand	22
4.2.5 Oppsummering - Passeringsteller for Kristiansand	22
4.3 STRAVA	22
4.3.1 Utfordringer med data fra STRAVA	22
4.3.2 Sammenligningsområde for STRAVA i Vestby	25
4.3.3 Oppsummering fra sammenligningsområde i Vestby	33
4.3.4 STRAVA i Kristiansand	33
4.3.5 Oppsummering STRAVA i Kristiansand	35
4.4 Sammenligning av STRAVA-data og tellerdata	35
4.4.1 Data fra Vestby alene	36



4.4.2	Data fra Vestby og Oslo	37
4.4.3	Tidsavhengig modell	42
4.4.4	Bruk av en periode med tellerdata for å ekstrapolere med STRAVA	43
4.4.5	Oppsummering fra sammenligning av STRAVA-data og tellerdata	45
4.5	Telia-data	45
4.5.1	Utfordringer med data fra Telia	45
4.5.2	Sammenlikning av Telia og tellerdata i Vestby	46
4.5.3	Oppsummering sammenlikning av Telia og tellerdata i Vestby	48
4.5.4	Bestilling av en spesialtilpasset analyse fra Telia – Trip analyse	48
5.	VISUALISERINGSLØSNING	48
5.1	Verktøy for visualisering	48
5.1.1	STRAVA analyse	49
5.1.2	STRAVA kilddata	49
5.1.3	Telia kilddata	51
5.1.4	Tiltaksområde	52
5.1.5	Skalerbarhet	52
6.	RESULTATER	53
6.1	Passeringsteller	53
6.2	STRAVA	53
6.2.1	Sammenligning av STRAVA og Tellerdata	53
6.2.2	Sammenligning av Telia og Tellerdata	53
6.3	Studiens evne til å besvare spørsmålene i forespørselen	53
7.	KONKLUSJON: ANALYSENE	55
8.	VEIEN VIDERE OG ANBEFALINGER	56



2. INNLEDNING

I juni 2020 lanserte regjeringen en nasjonal handlingsplan for fysisk aktivitet, '**Sammen om aktive liv – Handlingsplan for fysisk aktivitet 2020–2029**'¹, som strekker seg over en periode på 10 år. Et av de sentrale tiltakene i denne planen er å «gjøre lokasjonsdata og andre digitale datakilder mer tilgjengelig for kommunene for bruk i planlegging».

Hvor mye innsikt kan vi egentlig hente fra såkalte mobile lokasjonsdata/Stordata ²? Per i dag har Miljødirektoratet, kommunene og en rekke andre aktører utplassert passeringstellere, og det har også blitt utført feltstudier der man har intervjuet turgåere. Disse representerer de primære og konkrete dataene som brukes for å evaluere effekten av tiltak for å fremme fysisk aktivitet. En utfordring oppstår når det mangler tellere i områder der man vurderer å implementere tiltak for å øke fysisk aktivitet. Det blir da vanskelig å sette opp nyttekostnadsberegninger før tiltakene iverksettes. Det blir også utfordrende å sammenligne ulike alternativer, og vurderinger må derfor ofte baseres på skjønn. I denne studien vil vi undersøke om det finnes andre type mobile lokasjonsdata, som kan brukes som et supplement til de tradisjonelle metodene som tellere og intervjuer, som er mer arealdekkende og kostnadseffektive, for å bedre vurdere effekten av konkrete friluftslivstiltak og deres bidrag til økt fysisk aktivitet.

2.1 Bakgrunn og formål

Miljødirektoratet og fylkeskommunene er ansvarlige for tildelingen av statlige midler til prosjekter som fremmer friluftsliv og fysisk aktivitet. For å forbedre denne prosessen, er det viktig å øke forståelsen for hvordan slike prosjekter faktisk påvirker folks aktivitetsnivå. Miljødirektoratet utlyste sommeren 2023 et prosjekt knyttet til sin rammeavtale på GIS-tjenester. Prosjektet hadde som mål å undersøke i hvilken grad mobile lokasjonsdata og ulike former for mobilitetsdata kan benyttes som indikatorer for å måle aktivitet og endringer i aktivitetsnivået i befolkningen, som en følge av implementeringen av konkrete friluftslivsinitiativer. Norkart ble gjennom et avrop tildelt oppgaven med å gjennomføre prosjektet.

Formålet med prosjektet er å finne ut hvordan man på en mest mulig effektiv måte kan produsere et datagrunnlag som kan anslå effekten av tiltak på endringer i aktivitet som en del av en vurdering av friluftslivstiltak. Vurderingene skal kunne brukes som en del av kommunenes beslutningsgrunnlag og for å gjøre det enklere for Miljødirektoratet å prioritere søknader fra kommunene og for å måle eventuell helsegevinst i en kalkulator for nyttekostnadsberegninger.

Konkrete målsettinger for utlysningen var å svare ut følgende spørsmål.

1. *Hvor stor aktivitet (antall personer) det er i området som undersøkes (case) og hvordan fordelingen er i løpet av døgn/ukedag/uke/måned/år?*

¹ Sammen om aktive liv, Handlingsplan for fysisk aktivitet 2020–2029, Helse- og omsorgsdepartementet, publisert: 04.06.2020, <https://www.regjeringen.no/contentassets/43934b653c924ed7816fa16cd1e8e523/handlingsplan-for-fysisk-aktivitet-2020.pdf>

² Mobile stordata brukes ofte som et samlebegrep om store mengder digitale data, teknologi og analysemetodikk knyttet til store mengder data. Vi bruker begrepet lokasjonsdata eller mobilitetsdata om data som beskriver hvor folk befinner seg eller som beskriver hvordan folk beveger seg.



2. *I hvilken grad det er mer aktivitet i området det er gjort friluftstiltak i enn før tiltak ble gjort. Dette kan være antall brukere, antall turer, lengden på turen eller hvor lenge folk oppholder seg i området.*
3. *I hvilken grad friluftslivstiltaket samlet sett skaper mer aktivitet, eller flytter aktivitet fra et annet område. Hvor stort er influensområdet i casene?*
4. *Aldersintervall på personer som er i aktivitet i det aktuelle området, dersom informasjonen er tilgjengelig. Er det konkrete tiltak som tiltrekker seg nye brukergrupper (for eksempel i forhold til alder, kjønn, etc.)? Hvilken informasjon om brukerne er det mulig å fremskaffe innenfor rammene til personvernforordningen?*
5. *Influensområde: Er det hovedsakelig de lokale som tar i bruk friluftsområdet? Avstand fra bolig til friluftsområder? Har aktiviteten flyttet seg fra et nærområde til et annet? Vi ønsker å kunne avdekke om folk flytter sin fysiske aktivitet fra et annet område og til det aktuelle området som det har vært gjort tiltak på, eller hvorvidt tiltaket bidrar til en reell økning i aktivitetsnivået for befolkningen lokalt og/eller regionalt.*

2.2 Undersøkte områder

Kommunene Kristiansand og Vestby er inkludert i studien, begge ble valgt ut av Miljødirektoratet før prosjektstart. Vi har gjennomført flere møter med representanter fra kommunene for å identifisere to egnede turområder. De utvalgte områdene skal være gode eksempler for å kunne analysere om ulike tiltak som har blitt gjennomført faktisk fremmer friluftsliv og aktivitet. Kriteriene som ble lagt til grunn for valg av områder inkluderer:

- Det må ha blitt gjennomført et spesifikt friluftslivstiltak i området i løpet av de siste tre årene.
- Det skal finnes tilgjengelige data fra tellere i området før eller etter tiltaket.

Det var kun ett område i hver kommune som oppfylte disse kriteriene. De to områdene som ble valgt for analysen er **Smaalensbanen i Vestby kommune** og **Øyliblikk-dagsturhytta i Kristiansand kommune**.

Ved å undersøke og sammenligne disse områdene, var målet å forstå hvordan de har bidratt til å oppfylle målene om økt aktivitet eller ikke. Områdene skiller seg ift. sentralitet - Smaalensbanen går gjennom sentrum av en kommune, mens Øyliblikk-dagsturhytta strekker seg ut i periferien av en kommune. Dette gir oss muligheten til å undersøke hvordan datagrunnlaget egner seg ift. sentrale eller perifere områder.

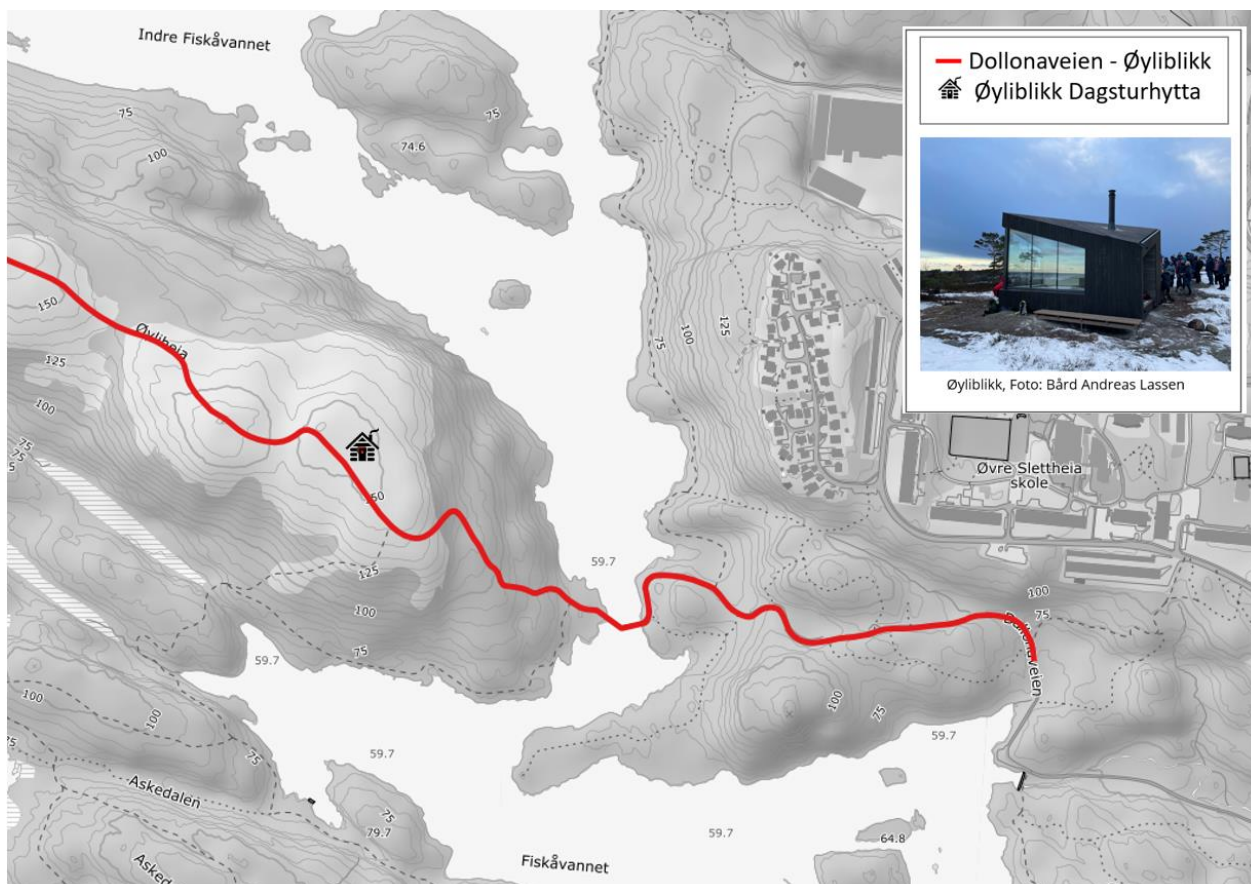
2.2.1 Øyliblikk-dagsturhytta

Øyliblikk-dagsturhytta i Kristiansand representerer et eksempel på tiltak som har som mål å fremme friluftsliv og aktivitet i regionen. Hytta er strategisk plassert på Øyliheia, og tilbyr en fantastisk utsikt (Figur 1). Turen til hytta er omtrent 2,5 kilometer hver vei, og tar deg 161 meter over havet. På returen kan en velge å følge samme rute som på vei opp, eller velge en noe lengre rute som går på vestsida av Fiskåvannet. **Hytta ble offisielt åpnet den 9. mars 2023**

Dagsturhyttene, som Øyliblikk, finnes utplassert mange plasser i Norge og har som mål å stimulere til friluftsliv og turglede i trygge omgivelser. De er gratis og tilgjengelige for alle, døgnet rundt, hele året, og



tilbyr et sted hvor besøkende kan ta en pause og nyte naturskjønne omgivelser. Prosjektet med dagsturhyttene representerer et samarbeid mellom de regionale fylkeskommunene, Sparebankstiftelsen og lokale kommuner i de respektive fylkene.



Figur 1: Øyliblikk Dagsturhytta og Dollonaveien opp til Øyliblikk.

2.2.2 Smaalensbanen

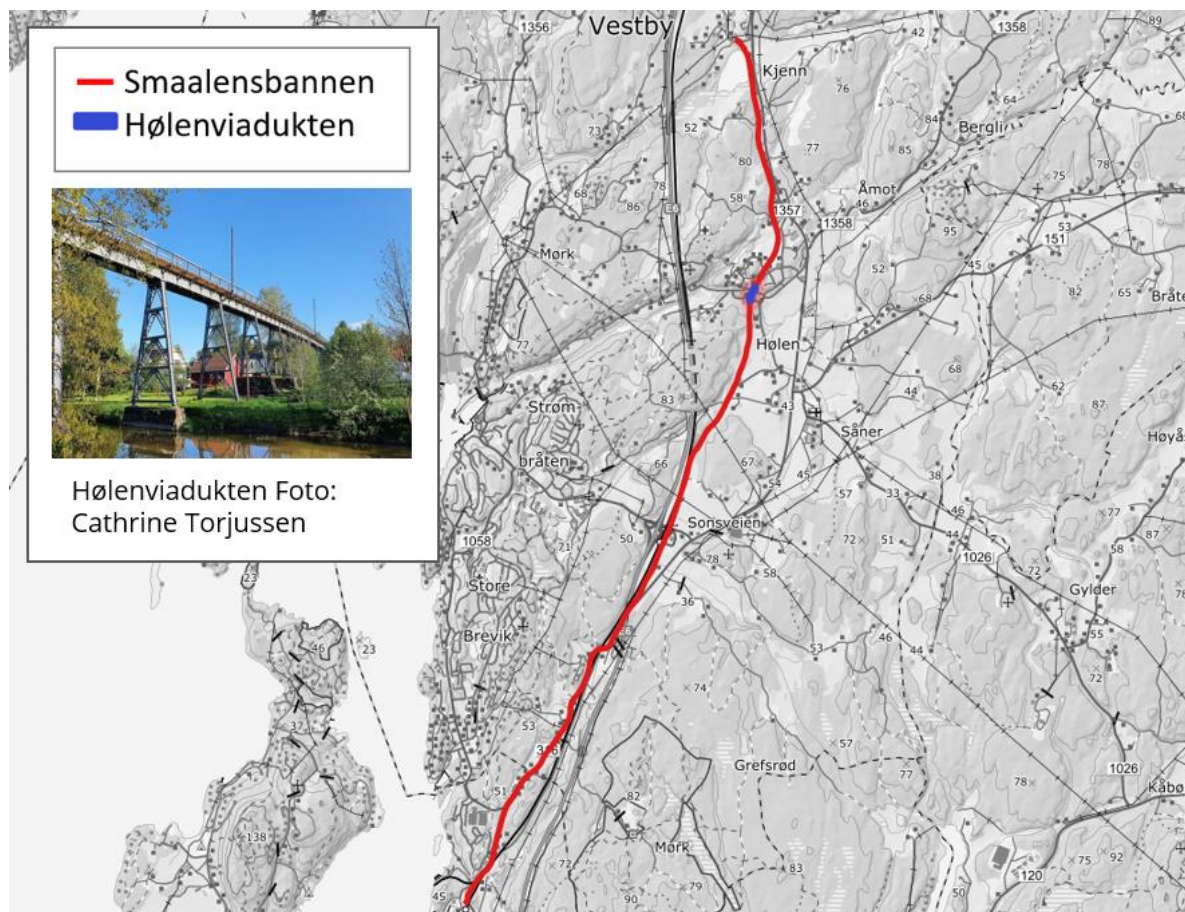
Smaalensbanen i Vestby er et eksempel på tiltak rettet mot å forbedre tilgjengelighet og oppmuntre til friluftsliv i kommunen. Denne stien er en tidligere jernbanetrase som har blitt oppgradert og transformert til turvei, som åpner for muligheter for både fot- og sykkelturner, og fungerer også som en transportkorridor gjennom Vestby kommune. Smaalensbanen som tursti sto klar allerede i starten av 2000-tallet. Ruten strekker seg fra Kjenn og Kjennstjernet i nord, gjennom Hølen tettsted og over Hølenviadukten, til Sonsveien stasjon i sør, med en total lengde på omtrent 5 km (Figur 2). Den flate traseen gjør Smaalensbanen tilgjengelig for alle og fremmer et inkluderende friluftsliv.

Hølenviadukten, en viktig del av Smaalensbanen, er en gammel jernbanebro som lenge har vært stengt for trafikk, og stått som en «missing link» for en sammenhengende turveitrasé. Vestby kommune fikk et tilbud fra Bane NOR om å overta eierskapet og forvaltningen av broen. Dette tilbudet ga kommunen en unik mulighet til å revitalisere Hølenviadukten, ikke bare som et kulturminne, men også som en viktig ferdselsåre for gående og syklende, samt et attraktivt friluftslivsområde



I 2021, 25 år etter nedleggelsen av jernbanen, ble Hølenviadukten gjenåpnet, denne gang for fotgjengere og syklist. Denne gjenåpningen knytter sammen ulike deler av ruten og forventes å øke bruken av turstien. Viktige mål for prosjektet inkluderte etablering av en tryggere skolevei, tilrettelegging for friluftsområder som tilbyr naturopplevelser, og å fremme økt fysisk aktivitet blant befolkningen.

Hølenviadukten ble offisielt åpnet den 26. juni 2021.



Figur 2: Smaalensbanen og Hølenviadukten.

3. DATA

Det finnes mange ulike stordata som er mer eller mindre egnede for å anslå effekten av et friluftslivstiltak, disse er vurdert med perspektiver som f.eks. tilgjengelighet, presisjonsnivå, kostnad og personvern. Barton et al. 2021³ har beskrevet potensielle stordata for å måle fysisk aktivitet i befolkningen. I Figur 3, hentet fra Barton et al. 2021, representeres de primære datatypene ved blå bokser, mens de brune boksene indikerer de datatypene som har størst potensial for utvikling av indikatorer til måling av fysisk aktivitet i befolkningen. De boksene som er plassert til høyre, viser de

³ Barton. David N., Gundersen. Vegard., Venter. Zander. S. Bruk av stordata i arbeidet med å tilrettelegge for fysisk aktivitet. NINA Rapport nr. 1937. Publisert januar 2021. <https://brage.nina.no/nina-xmlui/bitstream/handle/11250/2722733/ninarapport1937.pdf>



tradisjonelle datakilder som kan benyttes for kalibrering.



Figur 3: Barton et al. 2021, Bruk av stordata i arbeidet med å tilrettelegge for fysisk aktivitet.

For å måle effekter og virkninger av friluftslivstiltak ved hjelp av mobile lokasjonsdata og mobilitetsdata har vi hatt som mål å bruke flere ulike datakilder. Dette gir oss muligheten til å dra nytte av det beste fra flere datakilder som utfyller hverandres svakheter. Vi har valgt å fokusere på data fra kommunikasjon i form av mobile nettverksdata fra telekommunikasjonsselskaper, data innsamlet fra smarttelefoner via treningsapper, data fra bedrifter i form av transaksjoner, samt sensorer som tellere. Tellere er også brukt for kalibrering i prosjektet.

3.1 Mobile nettverksdata

Mobilnettverksdata kan brukes til å kartlegge folks bevegelser. Dette gjøres ved å bruke data fra mobiltelefoner som kommuniserer med nærliggende mobilsendere. Når en mobiltelefon er aktivert, det vil si at den enten foretar eller mottar samtaler, sender tekstmeldinger eller bruker datatjenester, registrerer den nærmeste mobilbasestasjonen telefonens tilstedeværelse.

Vi har vært i kontakt med tre telekommunikasjonsselskaper som er aktive i Norge; Telenor, ICE og Telia.

3.1.1 Telenor

Telenor har ikke ferdige data til bruk for å kartlegge mobilitet. I et møte med Telenor ble det opplyst at de har erfaring med lignende prosjekter, der de har bearbeidet anonymiserte data for forskningsformål for å analysere bevegelsesmønstre. De har ikke umiddelbart tilgjengelige data som kan tilbys eller selges for spesifikke prosjekter som dette.

3.1.2 Telia

Telia tilbyr mobilitetsdata gjennom et produkt kalt «Crowd Insights», som selges som en standardløsning. Dette produktet innebærer levering av anonymiserte data i et GRID-format. GRID-oppløsningen varierer fra 250 til 4000 meter, avhengig av befolkningstetthet og dekningskvalitet i de respektive områdene. Hver enhet kan generere mellom 200-400 datapunkter daglig, avhengig av



brukerens aktivitetsnivå. I dialog med Telia ble det informert om at de registrerer mellom 500-600 millioner hendelser hver dag. Telia har en markedsandel på 23,6 prosent i Norge, men ekstrapolerer dataene for å representere hele befolkningen. De kan levere data tilbake i tid fra 2019 og dekker nesten hele Norge. I dette prosjektet har vi kun hatt tilgang på uttrekk av data for Vestby og Kristiansand kommune.

Dataene oppdateres daglig, og det er mulig å generere statistikk for time, dag, måned eller år. Produktet tillater også analyse av varigheten av en aktivitet innenfor et bestemt GRID, noe som muliggjør differensiering mellom tre typer mobilitetsatferd: aktiviteter, reiser og personstrømmer. En aktivitet indikerer at abonnementet er stasjonært, mens reiser og personstrømmer identifiseres når abonnementet er i bevegelse. Nøyaktigheten av denne identifikasjonen påvirkes av mobilnettets celledørrelse, som varierer; i urbane områder kan bevegelser på noen få hundre meter identifiseres, mens i tynt befolkede områder kreves lengre bevegelser.

Innenfor hvert GRID kan en hente ut informasjon om opprinnelse, inndelt i tre kategorier: «home», «work» og «other». «home» defineres som det området hvor mobilen er i ro mellom kl. 23 og 07, «work» er der mobilen er mellom kl. 07 og 16, og «other» er andre tidsrom og bevegelse. Ved å bruke denne informasjonen er det mulig å se hvor aktivitetene finner sted og hvor de kommer fra, på et nivå som sikrer at enkeltpersoner ikke kan identifiseres. Datasettet er kommersielt tilgjengelig med en forretningsmodell fra Telia.

3.1.3 ICE

ICE tilbyr for øyeblikket ikke ferdigbearbeidede data for salg eller distribusjon for kartlegging av mobilitet. I et møte med ICE ble det nevnt at de har mulighet til å tilrettelegge data etter behov, men de har ingen datasett som er klare for direkte levering. Vi har hatt ytterligere dialog med ICE på deres initiativ, og de uttalte at de er positive til å gå i videre dialog for å se hva de eventuelt kan tilby.

3.2 Data fra smarttelefon

I dag benytter mange seg av applikasjoner som lagrer informasjon om våre aktiviteter, inkludert geografisk lokasjon. Et typisk eksempel på dette er ulike treningsapper som sporer og registrerer fysiske aktiviteter. Vi har vært i kontakt med 2 leverandører av treningsapper som mange bruker i Norge; Polar og STRAVA.

3.2.1 Polar

Polar er framfor alt kjent for sine treningsklokker og relatert teknologi. De tilbyr også tjenester og funksjoner som gir personlig treningsovervåkning som også kan kobles til lokasjon i kartet, hvor du kan se på dine treningsaktiviteter som løping, sykling og annen fysisk trening. Vi har prøvd å komme i kontakt med Polar Norge, for å se om de er mulig å få tilgang til mobilitetsdata som blir innsamlet gjennom deres apper, men vi har dessverre ikke klart å komme i kontakt. Det antas at deler av disse dataene er representert gjennom STRAVA.



3.2.2 STRAVA

STRAVA er en populær plattform for registrering av fysisk aktivitet, som løping, sykling, og annen fysisk trening. Opprinnelig populær blant syklister, men etter hvert også blant løpere og for andre utendørsaktiviteter. STRAVA har utviklet en tjeneste kalt STRAVA Metro. Denne plattformen benytter data innhentet fra STRAVA-brukere på en anonymisert måte for å analysere bevegelsesmønstre i ulike områder. For å få tilgang til disse dataene, må man inngå et partnerskap med Metro. Tilgang gis kun til organisasjoner som planlegger, eier eller vedlikeholder infrastruktur for aktiv transport, eller som driver med lignende planleggingsaktiviteter. Hvis en organisasjon blir godkjent av Metro, får de gratis tilgang til dataene. I denne studien har Norkart fått tilgang til dataene gjennom Miljødirektoratet, som har inngått et partnerskap med Metro og dermed kan gi Norkart tilgang på deres vegne.

De dataene som kan lastes ned er: Opprinnelses- og destinasjonsdata (Origins & Destinations), som viser polygoner for hvor en tur starter og avsluttes, og Gate-data (Streets Data) som linjedata. Gate-data er det som gir den mest detaljerte informasjonen om mobilitet ned til individuelle stier, og er også det vi har valgt å bruke i dette prosjektet. Stinettverket har en god topologi, der endepunkter som skal henge sammen samsvarer helt nøyaktig, og veistykker som ikke skal henge sammen ikke møtes i endepunktene. Antallsaggregeringen er avrundet til nærmeste 5-tall for at dataene skal holdes anonymisert. Data kan lastes ned per time, dag, måned eller år.

Som grunnkart brukes OpenStreetMap (OSM) fra februar 2023, dette oppdateres periodisk cirka hvert tredje år 3 år (forrige grunnkart som ble brukt var 2019-2022). Aktiviteter blir koblet til veier og stier som er representert i grunnkartet. Dette innebærer at bevegelser utenfor det eksisterende nettverket automatisk justeres ('snappes') til de nærmeste veiene og stiene. I et møte med STRAVA den 30. oktober 2023, ble det informert at det er satt en maksimal avstand på 300 meter. Om det ikke er en sti utenfor denne avstanden blir aktiviteten slettet. STRAVA informerte også om at de fra og med 2024 planlegger å øke frekvensen av oppdateringer fra OSM til 1-2 ganger per år. Videre ble det informert om at det fra og med 2024 vil bli mulig å laste ned data via API.

3.3 Passeringstellere

En passeringsteller, også kjent som en besøksteller, er et verktøy som plasseres i terrenget for å registrere antall personer eller kjøretøy som passerer et spesifikt punkt eller område over en bestemt tidsperiode. Ved valg av egnede turstier til denne studien, ble det stilt krav om at det allerede skulle være samlet inn passeringsdata fra tellere i terrenget før eller etter et spesifikt tiltak. Disse dataene er planlagt å brukes både for å vurdere effekten av friluftslivstiltak og for å se om de kan anvendes til å kalibrere andre stordata som vi har tilgang på for det samme området.

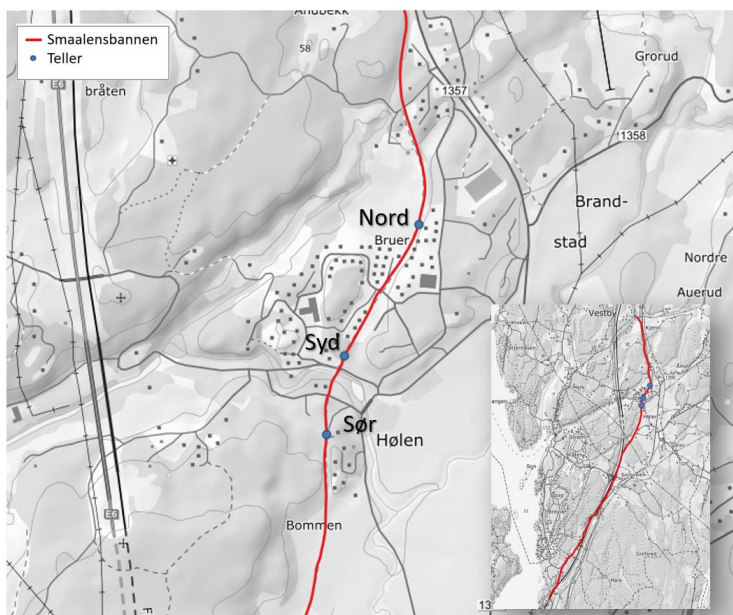
Selv om passeringstellere kan gi en indikasjon på besøkstall i et område, ligger det en utfordring i at de kun er installert på begrensede steder. Oslo Economics⁴ gjennomførte en kartlegging i bruken av passeringstellere og registreringer for å fange opp endret bruk. Konklusjonen var at passeringstellere

⁴ Oslo economics 2022: Hva gjør norske kommuner for å måle og registrere aktivitetsendringer som utløses av tiltak for å fremme friluftsliv? Kartlegging utarbeidet på oppdrag for Miljødirektoratet. Rapportnummer: M-2399 | 2022. OE-rapportnummer 2022_104: <https://www.miljodirektoratet.no/publikasjoner/2022/november/hva-gjor-norske-kommuner-for-a-male-og-registrere-aktivitetsendringer-som-utloses-av-tiltak-for-a-fremme-friluftsliv/>

brukes lite systematisk til dette formålet. I tillegg påpeker de at det ofte er vanskelig å måle effekten av friluftstiltak, på grunn av tekniske feil med tellerne eller at tellerne er utsatt for hærverk. Tellere settes ofte opp etter at et tiltak er gjennomført, og gir dermed ingen data tilbake i tid. Videre kan det være flere veier inn i området, og det er ofte ikke tellere ved alle veiene. Det er derfor ønskelig å finne ut om mobilitetsdata kan være et alternativ til passeringstellere for å finne antall turgående.

3.3.1 Smaalensbanen

På Smaalensbanen har det blitt utplassert en teller over en periode på tre år. I løpet av denne tidsperioden har telleren vært plassert på tre forskjellige steder, vist i Figur 4, under 4 tidsperioder, som er vist i Tabell 1.



Figur 4: Plasseringen av telleren på Smaalensbanen.

Tabell 1: Viser plasseringen og tidsperioden for telleren på Smaalensbanen.

Dato	Plassering	Registrering
28.06.2019 - 25.07.2019	Syd for Hølenviadukten	Gående og syklende
26.07.2019 - 28.05.2020	Nord for Hølenviadukten	Gående og syklende
29.05.2020 - 07.10.2021	Syd for Hølenviadukten	Gående og syklende
07.10.2021-til dags dato	Nord for Hølenviadukten	Gående og syklende

3.3.2 Øyliblikk-dagsturhytta

Det er en teller utplassert på veien opp mot Øyliblikk-dagsturhytta, som illustrert i Figur 5, under en tidsperiode av 1,5 måneder, Tabell 2.



Figur 5: Plasseringen av telleren på Dollonaveien, opp mot Øyliblikk-dagsturhytta.

Tabell 2: Viser plasseringen og tidsperioden for telleren for Øyliblikk-dagsturhytta.

Dato	Plassering	Registrering
02.08.2023 -18.09.2023	Dollonaveien	Gående

3.4 Transaksjoner

Ved å analysere transaksjonsdata fra lokale næringsdrivende, inkludert kafeer, butikker og andre virksomheter i området rundt turstiene, kan vi evaluere eventuelle endringer i aktivitetsnivået før og etter at et tiltak ble iverksatt. Som datakilde er det mulig å bruke Proff.no. Dette er en nettside, hvor en kan søke etter detaljert informasjon om selskaper og få opp informasjon om driftsinntekter, resultat før skatt, lønnsomhet, selskapsform, etableringsår, antall ansatte og mer. Vi har mulighet til å analysere data som strekker seg tre år tilbake i tid, til 2020.

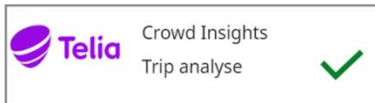
Etter et møte med representanter fra Vestby og Kristiansand viste det seg at det ikke finnes noen aktuelle bedrifter i nærheten av Øyliblikk-dagsturhytta. Det ble imidlertid nevnt at det er en kafé i nærheten av Hølenviadukten som kan være relevant å undersøke: Hølen Kafé AS. Men på grund av for store innvirkninger fra pandemien og nedstengninger i 2020 og 2021 kan vi ikke bruke dataene. Det er likevel en potensial kilde til informasjon.

3.5 Oppsummering av aktuelle datasett

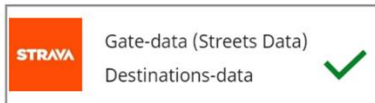
Etter en innledende vurdering av tilgjengelige datakilder, har vi identifisert fire relevante kilder som vil bli brukt i analyser for prosjektet. Dette er oppsummert i Figur 6. De utvalgte datakildene inkluderer mobile nettverksdata fra Telia, data fra smartenheter via STRAVA, transaksjonsdata, samt tellerdata fra de to spesifikke turområdene. Figur 7 og Figur 8 viser tilgjengelige data og for hvilken tidsperiode disse finnes tilgjengelig for respektive område.



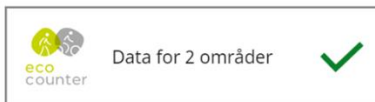
Mobile nettverksdata



Data fra smarttelefon



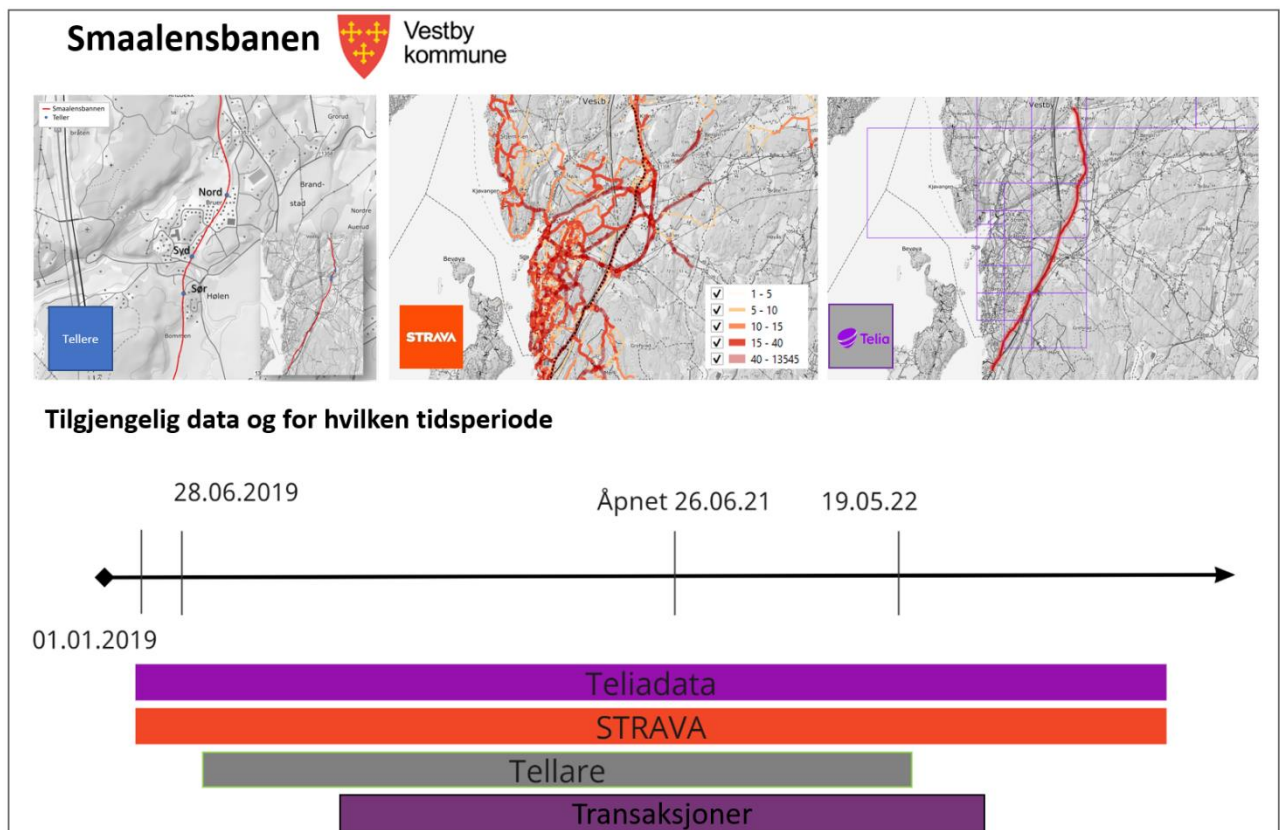
Automatiske tellere



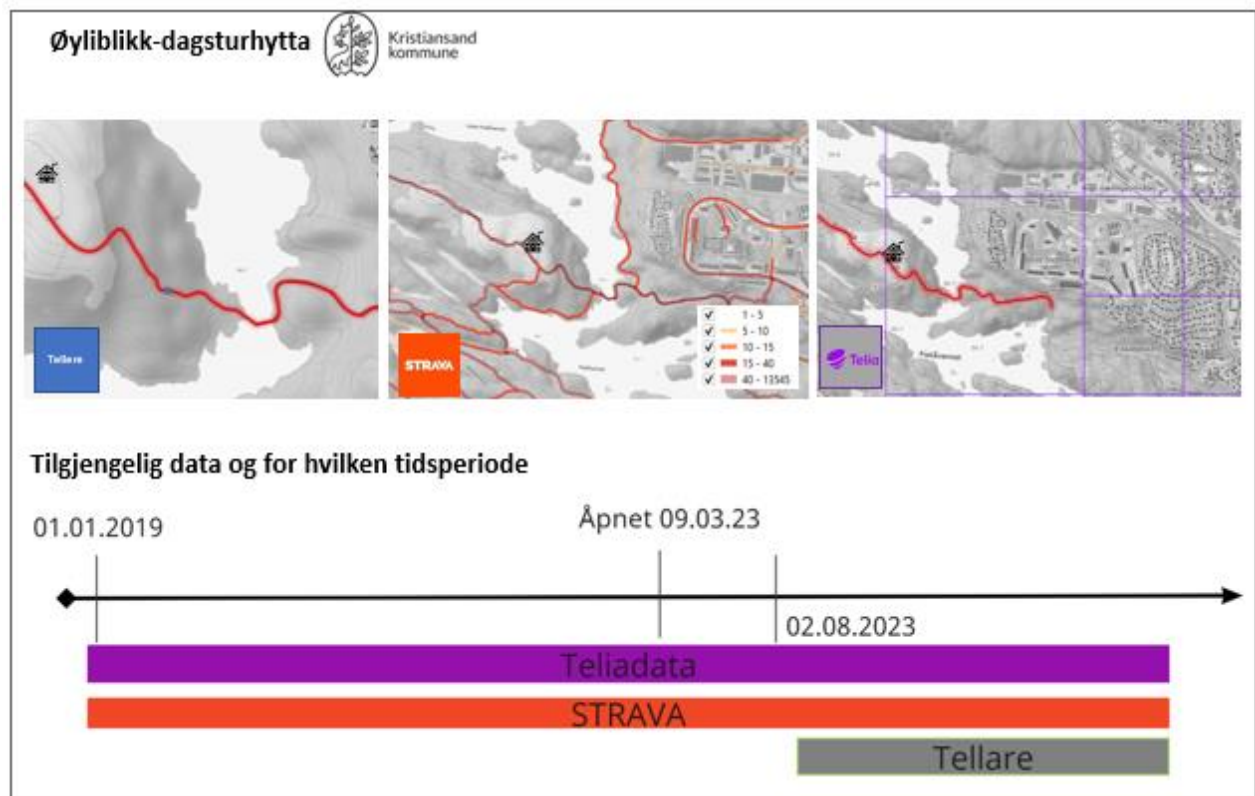
Transaksjoner



Figur 6: Viser dataleverandører vi har vært i kontakt med, og hvilke data vi har tilgang til for dette prosjektet.



Figur 7: Data vi har brukt i prosjektet og for hvilken tidsperiode disse finnes tilgjengelig for Smaalensbanen.



Figur 8: Data vi har brukt i prosjektet og for hvilken tidsperiode disse finnes tilgjengelig for Øyliblikk.

4. ANALYSEMETODE OG BEARBEIDING AV DATA

4.1 Måling av friluftaktivitet

I prosjektet ønskes det å måle friluftaktivitet i et område. Et mål på aktivitet kan være antall personer som befinner seg i et område egnet for friluftaktivitet. Antall personer kan måles og analyseres på forskjellige måter, vi vil her beskrive enkelt-metoder, samt hvilke feil og usikkerheter de har.

Vi må se om vi finner endringer i hyppigheten av passeringer, og om disse endringene kan være relatert til et gjennomført tiltak. Målinger av antall passeringer har en naturlig variasjon, en variasjon som også gir utslag i en usikkerhet i om en endring er en del av de naturlige variasjonene eller ikke. I dette arbeidet vil vi anvende ulike statistiske metoder som tar høyde for usikkerheten i målingene.

En variabel som dette, der enhver verdi kan assosieres med en sannsynlighet kalles en stokastisk variabel. I en del tilfelle kan vi gjøre antagelser om at den usikre variabelen følger en velkjent sannsynlighetsfordeling. Det vil si at vi har en formel for å kunne beregne sannsynligheten for en verdi, og kan svare på spørsmål av typen «hva er sannsynligheten for at det skal passere mindre enn 20 personer på stien i løpet av en dag». Slike sannsynlighetsfordelingsformler styres normalt av noen få



(relativt) enkle parametere som gjennomsnittsverdi⁵ og standardavvik⁶. Så snart vi kjenner disse kjenner vi også sannsynlighetsfordelingen.

Mye av analysearbeidet går ut på å vurdere om sannsynlighetsfordelingen for en variabel har endret seg, om vi kan si at de målinger vi får etter et tiltak er så forskjellige fra de vi fikk før at det er helt usannsynlig at de kan tilhøre samme sannsynlighetsfordeling. Et vanlig kriterium er å si at når sannsynligheten for at de kan tilhøre samme sannsynlighetsfordeling er under 5% (0.05) så er de for forskjellige. Da er nullhypotesen (en hypotese om at ingenting er endret) brutt, og vi sier at målingene er signifikant forskjellige.

Så lenge vi kan bruke en velkjent sannsynlighetsfordeling med veletablerte testmetoder som for eksempel normalfordelingen (også kjent som «klokkekurven» eller Gaussfordelingen) så kan dette løses med relativt enkle formler, men når vi har data som avviker fra dette så kan det bli mere utfordrende.

Antall personer som passerer et sted (som et punkt på en sti) kan modelleres matematisk som en poisson-prosess. Der regnes ankomstene av personer som uavhengige av hverandre, men de kommer i en fast rate, i gjennomsnitt over et langt tidsforløp vil det komme omtrent like mange personer pr tidsenhet. Sannsynligheten for at et bestemt antall mennesker (k) skal passere i løpet av en tidsenhet kan da uttrykkes med formelen $P(k|\lambda) = \lambda^k e^{-\lambda} / k!$, der poisson-parameteren λ er forventningsverdien for antall personer som kan passere pr tidsenhet, f.eks pr. døgn.

Målt mot den virkelige verden har en modell som dette sine svakheter. Mennesker beveger seg ofte i grupper (en familie, en trimgruppe, en kameratgjeng) og er derfor ikke uavhengige, antall personer ute på tur kan variere med tid på døgnet, dag i uken, værforhold; så poisson-parameteren λ blir i praksis ikke konstant, men en funksjon avhengig av mange forhold som vi ikke har kontroll over. Det vil allikevel være av verdi å ha dette som et matematisk fundament for det videre arbeidet. Samtidig så vil en poisson-fordeling nærme seg en normalfordeling ved store verdier av λ , noe som åpner opp for å kunne benytte tradisjonelle statistiske metoder for normalfordelinger.

Vi analyserte først data fra passeringstellere da disse antas å gi et resultat nærmest virkeligheten. Videre så vi på endringer i antall passeringer over tid, samt vurderte mer kostnadseffektive og arealdekkende data fra STRAVA og mobiltelefonoperatører for å se i hvilken grad disse samsvarte med tellerne. Denne metodikken ble etablert i samspill med oppdragsgiver.

4.2 Passeringsteller

Vi antar at tellere er fasit for bruken av turstiene, og disse dataene vil i hovedsak bli brukt som en referanse som vi kan kalibrere mobilitetsdata mot, derfor starter vi med å analysere disse.

⁵ Gjennomsnittet forteller hvor tyngdepunktet av sannsynlighetsfordelingen er plassert langs x-aksen og beregnes som: $\bar{x} = 1/n \sum x_i$

⁶ Standardavviket beskriver spredningen i dataene, hvor stort intervall rundt gjennomsnittet hvor verdiene stort sett befinner seg, og beregnes som: $\bar{\sigma} = \sqrt{1/(n-1) \sum (x_i - \bar{x})^2}$

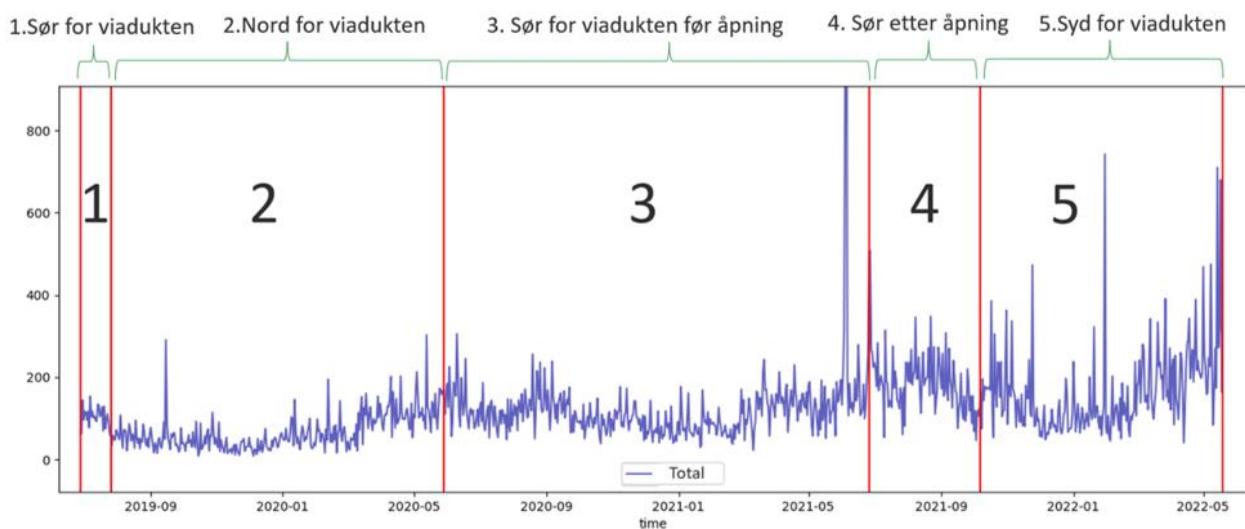
4.2.1 Utfordringer med data fra passeringstellere

Når vi analyserer data fra passeringstellere, er det flere faktorer som kan påvirke resultatene, blant annet plasseringen og monteringen av tellerne. Som tidligere nevnt var telleren i Vestby utplassert på tre forskjellige steder, fordelt på fire perioder, telleren står altså ikke i ro (Figur 4). Dette gir noen utfordringer når vi skal analysere dataene. En av de største utfordringene med å sammenligne data fra ulike plasseringer langs stien er sesongvariasjoner. Vi observerer en generell trend der flere mennesker er utendørs i sommerhalvåret sammenlignet med vinterhalvåret. Dette fører til at når vi ser på tellere som har stått på to forskjellige steder og i ulike årstider, kan det virke som om det er mer eller mindre aktivitet på enkelte steder. Denne forskjellen kan i realiteten skyldes årstiden. I datasettet observerer vi også noen ekstreme topper, som vi ikke kan fastslå om skyldes en spesifikk hendelse eller registreringsfeil. Disse toppene påvirker selvsagt analysene.

I Kristiansand har det vært tellere utplassert på Dollonaveien opp mot Øyliblikk (Figur 5). Den største utfordringen her er at vi kun har data for 1,5 måneder, og ingen data før åpning av hytta.

4.2.2 Passeringsteller for Vestby

Telleren i Vestby var utplassert på tre forskjellige steder, fordelt over fire perioder. I løpet av den fjerde perioden ble Hølenviadukten åpnet, noe som førte til at vi delte denne perioden i to for den videre analysen. Den første delen representerer tiden før åpningen av Hølenviadukten, fra 26. juli 2020 til 26. juni 2021, og den andre delen etter åpningen, fra 27. juni 2021 til 19. mai 2022. Figur 9 viser tellerens plasseringer og de tilsvarende tidsperiodene. Her ser vi at telleren var plassert på tre ulike steder før åpningen og på to ulike steder etter åpningen, totalt inndelt i fem perioder.



Figur 9: Passeringer pr dag for Smaalensbanen og alle utplasseringer.

For å finne ut om åpningen av Hølenviadukten har resultert i en økning i bruk av Smaalensbanen har vi beregnet enkelte viktige statistiske parametere for hver av periodene, som gjennomsnitt, standardavvik



og resultat fra en Shapiro-Wilk normalitetstest⁷. Resultater for dette er oppsummert i Tabell 3. Disse statistiske parameterne blir vist i de første tabellene for å illustrere variasjonen og utfordringene med dataene, men da de er mindre vesentlige for presentasjonen av analysen vil de senere bli utelatt.

Tabell 3: Statistiske parametere beregnet for de ulike periodene.

Periode	Dato	Gjennomsnitt	Standardavvik	Tilnærmet normalfordelt?
1	28.06.2019 – 25.07.2019 Sør for viadukten	106.3	23.4	Ja, $p=0.832$
2	26.07.2019 – 28.05.2020 Nord for viadukten	67.4	44.4	Nei, $p=10^{-15}$
3	29.05.2020 – 25.06.2021 Sør for viadukten, før åpning	120.9	104.4	Nei, $p=10^{-34}$
4	26.06.2021 – 06.10.2021 Sør for viadukten, etter åpning	187.9	76.5	Nei, $p=10^{-6}$
5	07.10.2021 – 19.05.2022 Nordenden av viadukten	166.0	105.6	Nei, $p=10^{-16}$

Vi ser i Tabell 3 at bare den første perioden kan antas å være normalfordelt. Dette er også den korteste perioden og den som må antas å være minst påvirket av variasjoner i underliggende faktorer som vær og sesong.

Vi sammenlikner endringen fra periode til periode. Siden de fleste periodene har en sannsynlighetsfordeling som ikke likner på normalfordelingen skulle man tro at en ikke-parametrisk test gir sikrest resultater, men en ikke-parametrisk test som Mann-Whitney U rank test⁸ sammenlikner forskjellen i hele sannsynlighetsfordelingene, og ikke bare forventningsverdien som vi er mest interessert i. Samtidig så har vi erfaring med at en Welch's t-test⁹ kan gi en fornuftig sammenlikning av gjennomsnittsverdier også når dataene avviker en del fra normalfordeling. Vi viser derfor resultatene fra både t-test og ikke-parametriske metoder i Tabell 4. Vi utfører testene på etterfølgende par av perioder og ser at ingen av parene kan forsvare en nullhypotese om likhet med en $p>0.05$.

Tabell 4: Signifikans av endringer.

Dato	Endring	Mann-Whitney $p=$	Welch's t-test $p=$
26.07.2019	Fra sør til nord for viadukten	$4.31 \cdot 10^{-8}$	$8.99 \cdot 10^{-10}$
29.05.2020	Tilbake til sør for viadukten	$1.02 \cdot 10^{-41}$	$9.56 \cdot 10^{-19}$
25.06.2021	Åpning av viadukten	$2.73 \cdot 10^{-22}$	$6.37 \cdot 10^{-12}$

⁷ SciPy v1.11.4 Manual, shapiro: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>

⁸ SciPy v1.11.4 Manual, mannwhitneyu:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html>

⁹ SciPy v1.11.4 Manual, ttest_ind: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

Welch's test avviker fra Student's ved at den også håndterer data med forskjellig varians.



07.10.2021	Til nordenden av viadukten	0.000170	0.0345
-------------------	----------------------------	----------	--------

Dette viser at de differanser vi ser i gjennomsnitt og standardavvik i Tabell 3 er signifikante. Men, vi kan ikke ut fra dette direkte slutte at disse differansene skyldes nettopp de endringer som definerer periodeinndelingen. Vi må også ta hensyn til at forholdene endrer seg i løpet av perioden, som sesongvariasjoner og generelle trender.

Vi forsøker derfor å se på forskjellene i gjennomsnittsverdi mellom måleperioder på en måned før og etter en endring. På denne måten forsøker vi å redusere avhengigheten av langtidstrender og sesongvariasjoner.

Vi ser i Tabell 5 et signifikant fall fra 106 til 58 passeringer pr døgn når måleren flyttes fra sør til nord 26.07.2019, og en signifikant økning fra 144 til 183 når den flyttes fra sør til innpå viadukten 07.10.2021. Imidlertid har vi ingen signifikant endring når viadukten åpnes, eller når telleren flyttes tilbake fra nord til sør.

Tabell 5: Endring fra måned før til måned etter tiltak.

Periode	Gjennomsnitt	Std. Avvik	t-test p=	Endring
28.06.2019 – 25.07.2019	106.3	23.4	$1.1 \cdot 10^{-11}$	Sør til nord
26.07.2019 – 26.08.2019	57.5	20.0		
28.04.2020 – 28.05.2020	127.9	49.2	0.38	Nord til sør
29.05.2020 – 29.06.2020	139.7	55.5		
25.05.2021 – 25.06.2021	236.8	316.5	0.64	Åpning av Hølenviadukten
26.06.2021 – 26.07.2021	209.2	96.5		
06.09.2021 – 06.10.2021	144.38	51.5	0.025	Sør til viadukt
07.10.2021 – 07.11.2021	182.5	77.6		

Fredag 4. og lørdag 5. juni 2021 har svært høye tellerverdier på hhv, 955 og 1764. Selv etter telefonsamtaler med Vestby kommune har vi ingen god forklaring på disse verdiene, og vet ikke om det kan ha vært et arrangement eller en feilmåling. Over en så kort måleperiode som en måned så gjør de uansett et stort bidrag til statistikken, og vi merker oss det ekstremt høye standardavviket for perioden like før åpning av viadukten. Vi forsøker derfor på nytt der de to dagene er utelatt, resultatet er presentert i Tabell 6.



Tabell 6: Endring fra måned før til måned etter tiltak, hvor to ekstremverdier er utelatt.

Periode	Gjennomsnitt	Std. Avvik	t-test p=	Endring
25.05.2021 – 25.06.2021	162.0	55.7	0.022	Åpning av Hølenviadukten
26.06.2021 – 26.07.2021	209.2	96.5		

Vi ser nå at standardavviket er nede på samme nivå som i de øvrige periodene, og at økningen i trafikk ved åpning av viadukten er signifikant.

4.2.2.1 Langtidsvariasjoner

Vi ønsker også å se på endringer i et lengre perspektiv. Vi vil da gjerne sammenlikne flere perioder som dekker mest mulig like årstider, for at vi skal få de samme sesongvariasjoner i alle. Problemet her er at måleren har blitt flyttet fra år til år, og vi kan ikke være sikker på at vi får sammenliknbare verdier på forskjellige steder. Tvert imot så viser forrige test (Tabell 5) at vi fikk en signifikant økning i antall passeringer da måleren ble flyttet.

Vi har et datointervall der måleren har stått i samme posisjon både i 2019, 2020 og 2021, nemlig 26.06-25.07. Da har måleren stått i posisjon «sør for Hølenviadukten» alle tre årene, resultatet fra denne testen kan ses i Tabell 7. Ut fra denne testen så har høysommermåneden 2019 og 2020 vært lik, mens vi har en sterk økning fra 2020 til 2021.

Tabell 7: Endring i målingene for måneden 26. juni til 25. juli i hvert år hvor måleren har stått i samme posisjon.

Periode	2019	2020	2021
Gjennomsnitt	105.7	104.0	210.9
Standardavvik	22.7	29.2	97.7
T-test p=	0.79		$1.8 \cdot 10^{-6}$

4.2.3 Oppsummering - Passeringsteller for Vestby

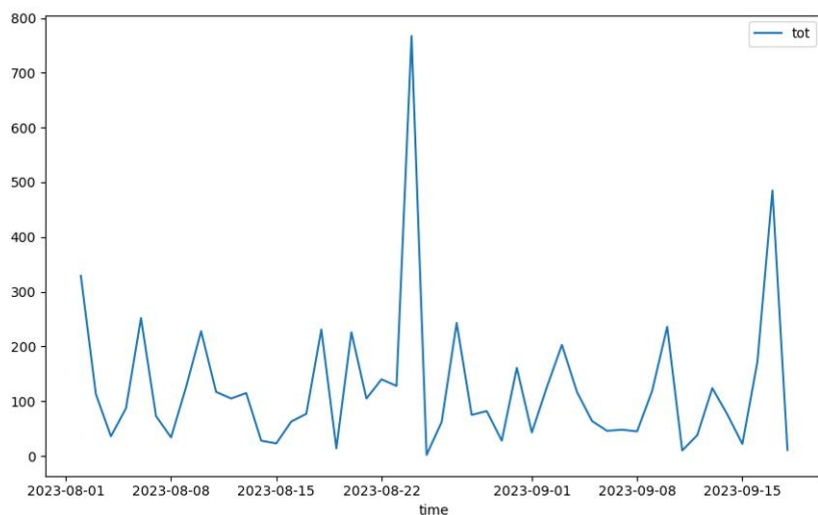
Ut fra de målinger som er gjort kan vi si at det har skjedd en økning i bruken av Smaalensbanen. Mye av denne økningen skjedde mellom sommeren 2020 og 2021, men noe av det skjedde allerede før Hølenviadukten ble åpnet. Vi kunne bare se en signifikant endring akkurat rundt åpningstidspunktet når to dager med ekstremt høye tellinger et par uker før åpning ble utelatt. Det er derfor vanskelig å konkludere helt sikkert med at dette er direkte forårsaket av åpningen av Hølenviadukten, men samtidig vet vi ikke om omtale av en framtidig åpning har ført til økt oppmerksomhet om Smaalensbanen, og derfor indirekte ført til økt trafikk allerede før åpning. Disse analysene viser at det kan være vanskelig å dra noen direkte konklusjoner, da det er mange parameter som kan påvirke resultatet. Faste plasseringer på tellere vil gi enklere analyser og antatt mer stabile resultater. Tellerdata er allikevel det datasettet



som best antas representere den virkelige bruken. Derfor har vi valgt å benytte tellerdata som en «baseline» (fasit) for å kalibrere mobilitetsdata.

4.2.4 Passeringsteller for Kristiansand

Vi har kun 1½ måned med passeringstellerdata i Kristiansand (fra 02.08.2023 til 18.09.2023), og siden august er den siste måneden vi har månedsaggregerte data fra STRAVA så har vi bare ett datapunkt for sammenlikning av disse (4070 tellerpasseringer og 70 STRAVA brukere), og dette gir ikke tilstrekkelig grunnlag for en sammenkobling av disse. Vi har totalt for hele den aktive perioden 6055 passeringer, eller 128.8 pr dag. Telleren er satt opp lenge etter at dagsturhytta Øyliblikk ble åpnet (august vs. mars 2023) så den gir ingen informasjon om endringer i ferdselsmønsteret. Figur 10 viser passeringer pr dag, ved Øyliblikk – dagsturhytta. Vi kan se at det er flest folk på tur i helgene og relativt få i løpet av ukedagene. Vi kan også se en større topp i løpet av sommeren.



Figur 10: Passeringer pr dag, teller ved Øyliblikk – dagsturhytta.

4.2.5 Oppsummering - Passeringsteller for Kristiansand

Måleperioden er for kort for å kunne gi noe bidrag til dette prosjektet: den dekker ikke tidspunktet for åpningen av dagsturhytta, kan ikke si noe om årstidsvariasjoner og kan ikke si noe om en trend fra år til år.

4.3 STRAVA

4.3.1 Utfordringer med data fra STRAVA

STRAVA-dataene tilbyr høyoppløselige posisjoner for bevegelser basert på GPS-sporing, noe som gjør det mulig å skille mellom aktiviteter på to nærliggende stier. Imidlertid representerer disse dataene ikke nødvendigvis den generelle befolkningen. Det er en overrepresentasjon av fysisk aktive personer, hovedsakelig middelaldrende menn. I tillegg brukes appen av en relativt liten del av Norges befolkning.



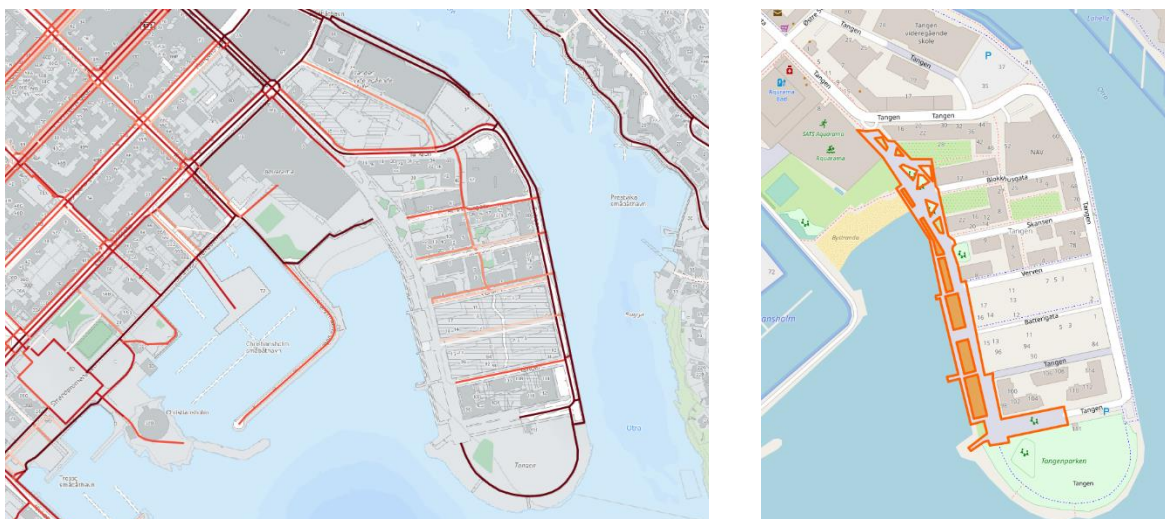
Ifølge tall fra STRAVA Metro bruker omtrent 6,5 % av befolkningen STRAVA til registrering av gåing/løping og 2,5 % til sykling i 2023. Det er viktig å merke seg at det kan forekomme dobbeltregistrering av brukere som registrerer både gåing/løping og sykling, noe som kan bety at den faktiske andelen av befolkningen som bruker appen kan være noe lavere. Selv om antallet personer som registrerer sykling er lavere, ser vi at i dataene aktivitetsnivået for disse er generelt høyere. Andelen som bruker STRAVA varierer også fra område til område. I tillegg har det under analyseperioden 2019-2023 skjedd en generell økning i STRAVA-bruken; i 2019 var det 2,8 % som brukte appen til å registrere gåing/løping og 1,5 % for sykling (Tabell 8).

Tabell 8: Antall personer som bruker STRAVA i Norge, hentet fra STRAVA Metro.

Aktivitet	2019	2020	2021	2022	2023
Gå/løping	149 938	222 760	279 368	315 309	359 312
Sykkkel	82 190	113 298	128 351	127 146	135 946

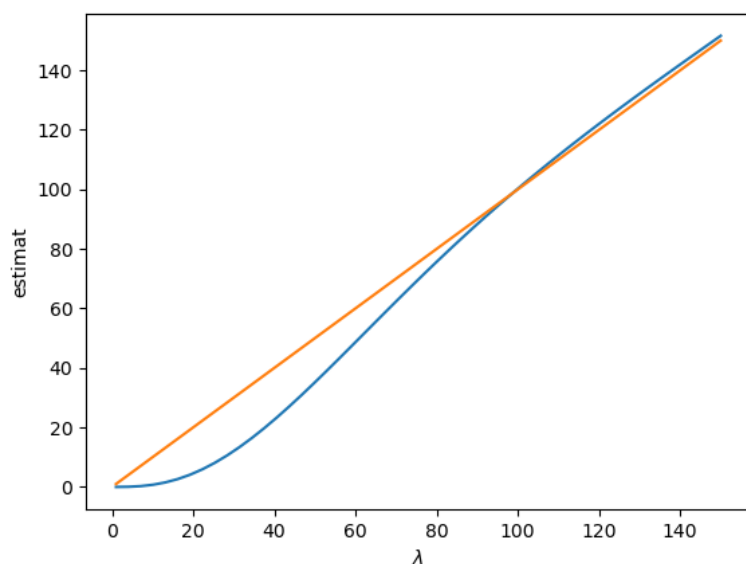
For at datasettet skal kunne brukes til å identifisere bruksmønstre for en ny sti, må denne stien først registreres i OSM. Deretter må STRAVA oppdatere grunnkartet og koble sine data til det nye kartet. Grunnkartet oppdateres periodisk, noe som betyr at det kan ta lang tid før en ny sti blir inkludert. Dette er ikke et problem for dette aktuelle prosjektet, da de to stiene som analyseres ligger inne i OSM, men kan bli en utfordring for andre områder. Med dagens oppdateringsfrekvens kan det i verste fall ta 3 år for en ny sti å komme inn. STRAVA planlegger å øke nedlastingsfrekvensen for OSM til halvår, noe som vil ha en positiv effekt. Innleggingen i OSM, som er forutsetningen for STRAVA uthenting, blir gjort som «crowd sourcing», men slik OSM fungerer i dag, med tusenvis av aktive bidragsytere i Norge, så vil en ny sti bli tegnet inn i løpet av noen uker. Det er også fullt mulig for ansvarlig for et stiprosjekt å oppdatere OSM så fort stien står klar.

En annen utfordring med STRAVA er at det kan forekomme brudd i topologi. I Kristiansand har vi identifisert et område hvor dette er tilfelle. I Figur 11 observerer vi flere høyt trafikkerte gangveier som ender blindt, vist i kartet til venstre. Når vi undersøker dette nærmere i OSM, vist til høyre, oppdager vi at det i dette området finnes et arealobjekt kategorisert som «pedestrian». Dette indikerer at STRAVA primært håndterer linjebaserte trafikkområder. Selv om dette ikke utgjør et problem for dette konkrete prosjektet, er det en viktig faktor å være oppmerksom på ved fremtidig bruk av STRAVA-data.



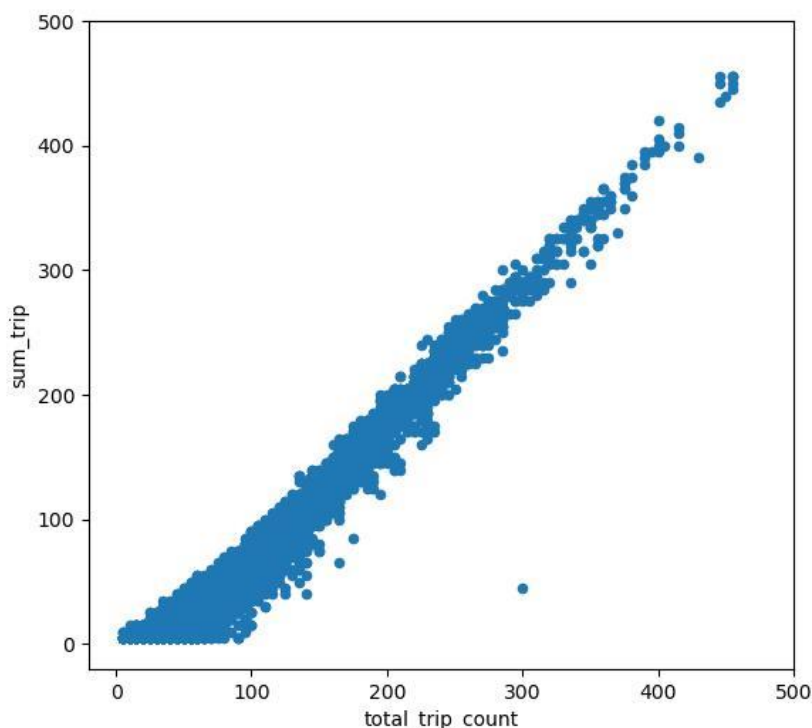
Figur 11: Til venstre, brudd i bevegelsesmønsteret, til høyre arealobjekt "pedestrian".

En annen utfordring er avrundingsfeil. Av personvern hensyn er den numeriske oppløsningen i antall STRAVA brukere redusert ved å avrunde til nærmeste 5; 12 personer blir avrundet til 10, 13 personer til 15 og verdier mindre eller lik 2 blir 0. Siden poisson-fordelingen (se avsnitt 4.1) er sterkt asymmetrisk ved små verdier av forventningsverdien λ så vil denne avrundingen kunne gi store utslag når vi forsøker å summere opp antall passeringer per dag til lengre tidsperioder. Dette er visualisert i Figur 12, hvor den blå kurven viser estimert verdi for gjennomsnittlig besøk pr måned ved å summere opp 30 avrundede dagsverdier plottet mot virkelig forventningsverdi λ .



Figur 12: Den blå kurven viser estimert verdi for gjennomsnittlig besøk pr måned ved å summere opp 30 avrundede dagsverdier plottet mot virkelig forventningsverdi λ .

I Figur 13 ser vi den praktiske virkningen av den teoretiske beregningen fra Figur 12. Når vi summerer opp dagstall for å finne antall passeringer pr. måned så vil veldig mange turer forsvinne fordi de har blitt avrundet til 0. Vi ser også at dette bedrer seg ved høyere verdier, men avrundingen bidrar fortsatt til støy i dataene. Dette har medført at vi har vært nødt til å bruke data på månedsbasis i stedet for daglig, som opprinnelig tenkt.



Figur 13: Månedssdata langs X-aksen og summerte dagsdata langs Y-aksen.

For å redusere støyen som blir innført med avrunding til nærmeste produkt av 5 så bruker vi her bare ferdigaggregerte månedssummer. Vi vil da få tilstrekkelig høy tallverdi til at denne ekstra usikkerheten blir mindre problematisk. På den annen side så vil denne månedsgrupperingen redusere tidsoppløsningen på dataene, og umuliggjør kortperiodiske analyser, for eksempel endringer i trafikk over døgn, ukedag og uke. Denne utjevningen av kortperiodisk variasjon gir også et kunstig «glattere» datasett, dataene vil virke mere forutsigelige enn de egentlig er.

4.3.2 Sammenligningsområde for STRAVA i Vestby

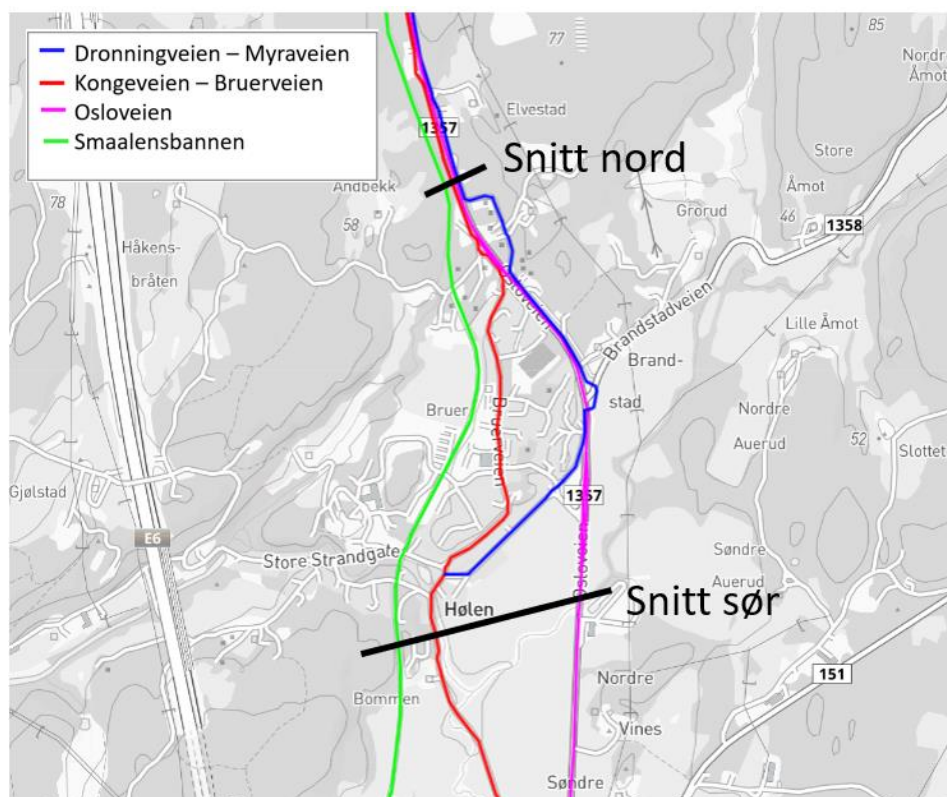
For å minimere påvirkningen fra naturlige variasjoner som værforhold, pandemirestriksjoner og fridager, samt endringer i antall STRAVA-brukere vil vi gjøre en analyse av den relative bruken av turveien. Ulempen ved denne analysen er at vi ikke kan se om tiltaket endrer den totale aktiviteten, men bare endringer i den relative fordelingen. Vi kan ikke si om det har vært en reell økning eller om det kun flytter aktivitet fra et område til et annet.

For å måle hvordan trafikken fordeler seg over veiene ser vi på bevegelsen gjennom gitte snitt. Disse er valgt ut fordi all trafikk i en hovedretning må gå gjennom en av veiene som krysser dette snittet, og vi kan se på trafikken gjennom Smaalensbanen vs. parallellveiene. Siden Hølenviadukten og Smaalensbanetraseen har hovedretning nord-sør så velger vi også våre snitt slik at de krysser veier med samme hovedretning. Vi velger ett snitt nord for tettbebyggelsen i Hølen, og ett like sør for Hølenviadukten, vist i Figur 14.

Snitt sør skjærer Osloveien («gamle E6», vei utenom tettbebyggelse, fartsgrense 80), Kongeveien (som går igjennom tettbebyggelsen, fartsgrense 40) og Smaalensbanen (jernbanetraseen, samme sted som



tellerposisjon «sør for viadukten»). Snitt nord skjærer gjennom Osloveien (fartsgrense 80), Myraveien (stikkvei til boliger med gjennomgående gang-sykkelvei), Bruerveien (gang-sykkelvei) og jernbanetraseen.



Figur 14: Viser snitt nord for tettbebyggelsen i Hølen, og ett sør for Hølenviadukten, disse er brukt for å analysere sammenligningsområde.

For disse dataene har vi store sesongvariasjoner, relativt få samplingspunkter (siden vi har aggregert på månedsnivå for å redusere støyen fra avrunding til nærmeste 5) og vi har en generell økning i bruken av STRAVA over perioden. For å forsøke å nøytralisere noe av dette har vi valgt å se på endringer i den relative fordelingen over veier i hvert snitt.

For hver måned har vi beregnet relativ trafikk for hver vei som «trafikk på denne veien» / «sum av trafikk gjennom snittet». Hver vei får da en verdi tilsvarende den relative andelen av trafikken som går gjennom denne, og summen for alle veier i et snitt blir 1. Videre har vi delt inn datasettet i to tidsperioder lagt før og etter bruåpning, en fra juni 2019 til mai 2021 og en fra juni 2021 til mai 2023. Der har vi beregnet differansen mellom relativverdien for hver måned etter bruåpning med samme måned to år tidligere. På denne måten har vi forsøkt å nøytralisere virkningen av årstidsvariasjoner.

Siden disse beregningene gir oss stokastiske data med en noe ubestemmelig sannsynlighetsfordeling, og vi har for lite data til å kunne anta at «det går mot normalfordeling når datamengden går mot uendelig» så har vi valgt å teste disse videre med «bootstrap»-algoritmen¹⁰. Denne algoritmen gjør repeterte tilfeldige utvalg fra samplingssettet vårt (altså listen av differanser mellom før og etter), og beregner den

¹⁰ SciPy v1.11.4 Manual, bootstrap:

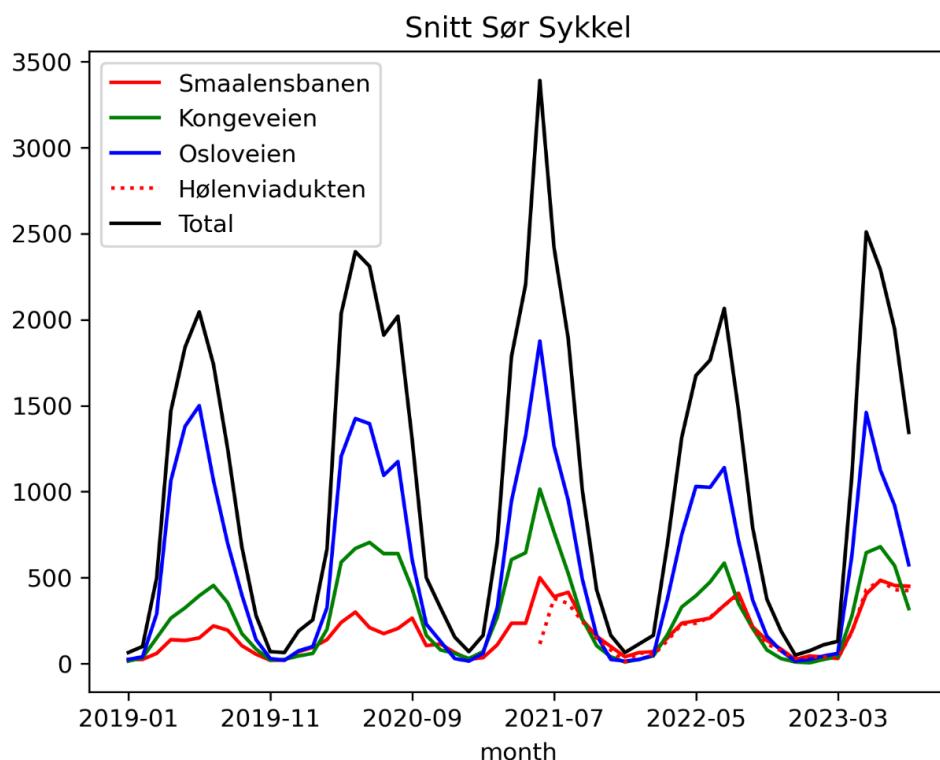
<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.bootstrap.html>



relevante statistiske parameteren (gjennomsnitt) for hvert utvalg. De variasjoner vi får mellom de forskjellige gjennomsnittsberegningene er forventet å tilsvare variasjonene vi ville ha fått ved å gjøre flere observasjoner, og kan brukes til å anslå usikkerhet i gjennomsnittsberegningen ved å beregne et standardavvik og konfidensintervall.

Vi ser fort at bruk av sykkel er framtreddende for STRAVA-brukere i analyseområdet, og i snitt sør for Hølenviadukten, foretrekker de å bruke Osloveien («gamle E6»), Figur 15. Veien er her lite tilrettelagt for sykling, og valget av denne veien tyder på at vi her har syklister med ønske om fart, som i mindre grad føler seg sjenert av biltrafikken, men som kanskje vil føle seg hindret av turgåere på Smaalensbanen.

I Figur 15 er Smaalensbanen vist med rød strek, og sum for all trafikk over snittet er vist med sort strek. For sammenlikning er sykkeltrafikk over Hølenviadukten (ikke tatt med i sum over snittet) vist med rød prikkete strek. Vi ser at kort tid etter at denne er åpnet blir kurven for viadukten nær sammenfallende med kurven for snittet sør for viadukten, og det ser ut til at de som sykler traseen sør for Hølen nå har gått over til å krysse Hølen ved å fortsette på viadukten.



Figur 15: STRAVA-sykkeldata, snitt sør for Hølenviadukten.

Vi ser at det er en økning i den relative bruken av Smaalensbanen fra toårsperioden før til toårsperioden etter bruåpning, vist i Tabell . Konfidensintervallet for estimatet av differansen inkluderer ikke null, så vi kan konkludere med at økningen er signifikant. Tilsvarende ser vi at den relative bruken av de to parallelle veiene har gått ned.

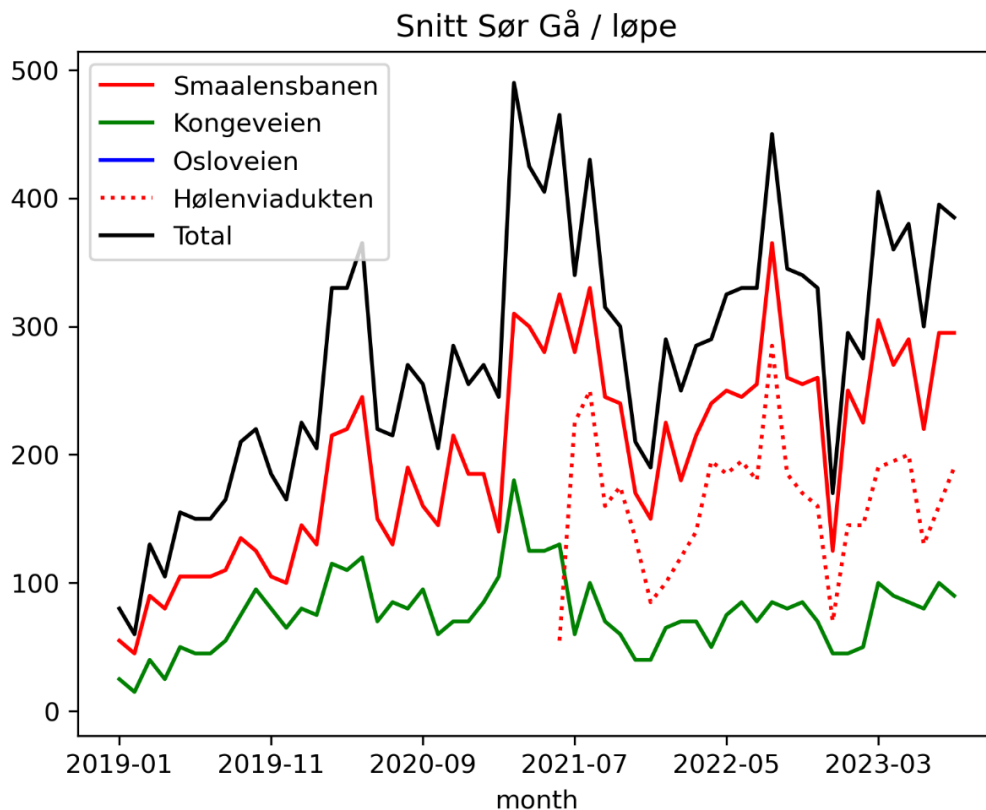
Tabell 9: Konfidensintervall for STRAVA-sykkeldata ved Snittpunkt Sør for Hølenviadukten.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.205	0.317	0.112	[0.084, 0.152]



Kongeveien	0.310	0.239	-0.070	[-0.115, -0.038]
Osloveien	0.485	0.444	-0.041	[-0.080, -0.007]

Gående og løpende bruker derimot ikke Osloveien, men heller Smaalensbanen, vist i Figur 16. Vi ser en økning i bruken av Smaalensbanen i 2021, men som i passeringstellerdataene (og i STRAVA-sykkeldata) starter denne økningen et par måneder før åpningen.



Figur 16: STRAVA-gå/løpe, snitt sør for Hølenviadukten.

Når vi igjen sammenlikner toårsperiodene før og etter bruåpning, vist i Tabell 9 så ser vi at den relative bruken av jernbanetraseen har økt signifikant.

Tabell 9: Konfidensintervall for STRAVA- gå/løpe ved Snittpunkt Sør for Hølenviadukten.

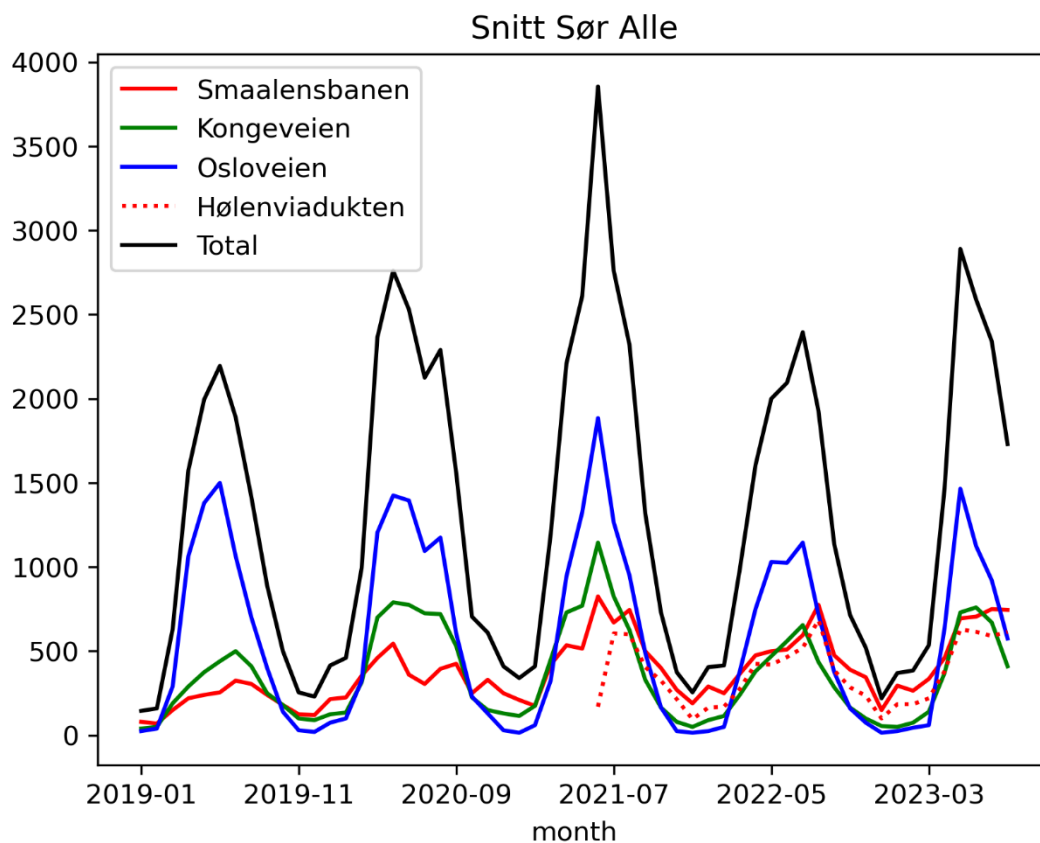
Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.659	0.775	0.116	[0.090, 0.144]
Kongeveien	0.342	0.222	-0.119	[-0.146, -0.093]
Osloveien	0	0.003	0.003	[0.001, 0.006]

Summen av alle STRAVA-aktiviteter (sykkel, gå, løpe) er vist i Figur 17, og konklusjonene er de samme som tidligere. Vi ser at syklister er overrepresenterte i dataene og Osloveien er mest brukt, men vi ser også i Tabell 10 en signifikant økning i bruken av Smaalensbanen i 2021, med en økning litt før åpningen av viadukten.



Tabell 10: Konfidensintervall for alle STRAVA-aktiviteter ved Snittpunkt Sør for Hølenviadukten.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.332	0.470	0.138	[0.110, 0.167]
Kongeveien	0.318	0.240	-0.078	[-0.108, -0.049]
Osloveien	0.350	0.291	-0.060	[-0.084, -0.037]



Figur 17: summen av alle STRAVA-aktiviteter, snitt sør for Hølenviadukten.

Vi kan bruke samme teknikk for å se på endringen fra år til år. Vi bruker fortsatt månedsskiftet juni – juli som avgrensing mellom periodene. I Tabell 11 over vises de relative fordelingene for hver «juli til juni neste år»-periode, med differanser og konfidensintervall mellom periodene.

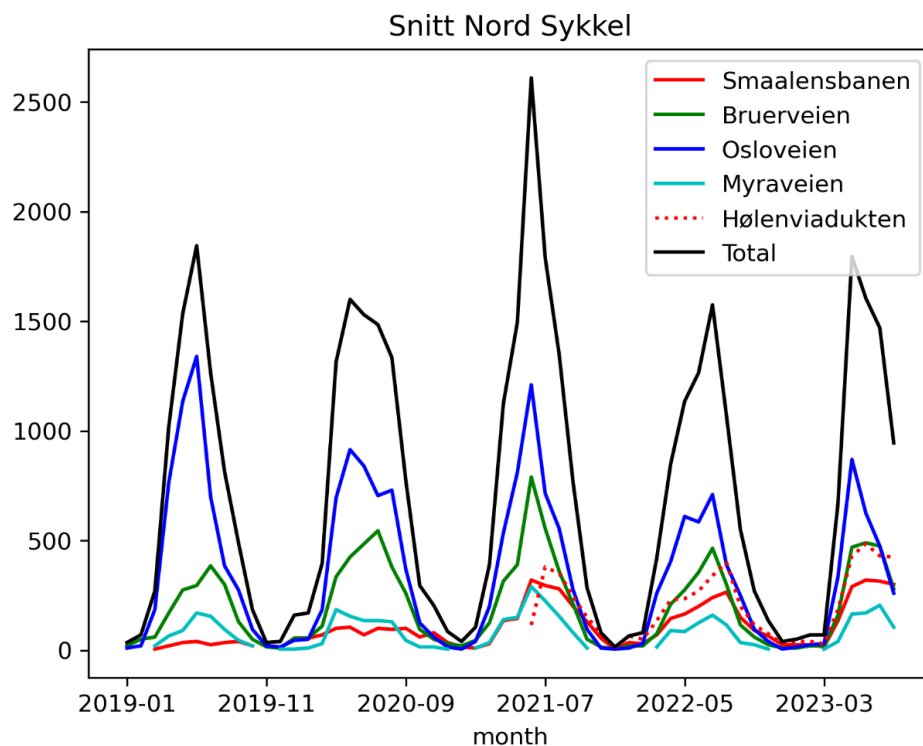
Vi ser i Tabell 11 at det største skiftet i preferanse av vei var en overgang til å bruke Smaalensbanen i 2021-22, og at dette er signifikant (konfidensintervallet dekker ikke 0). Vi ser også en økning i den relative bruken av Smaalensbanen 2020 og 2022, men her er det bare i 2022 at økningen er stor nok til å regnes som signifikant.



Tabell 11: Analyse av endringer fra år til år, alle STRAVA-aktiviteter ved Snittpunkt Sør for Hølenviadukten.

Navn	2019-20	Diff	Konfl	2020-21	Diff	Konf	2021-22	Diff	Konf	2022-23
Smaalensbanen	0.325	0.014	[-0.011, 0.043]	0.339	0.111	[0.080, 0.141]	0.450	0.039	[0.004, 0.102]	0.489
Kongeveien	0.306	0.023	[-0.028, 0.059]	0.329	-0.082	[-0.106, -0.058]	0.247	-0.015	[-0.039, 0.006]	0.232
Osloveien	0.368	-0.037	[-0.073, 0.002]	0.331	-0.029	[-0.064, 0.012]	0.302	-0.024	[-0.099, 0.006]	0.279

I snitt nord ser vi noe av det samme, STRAVA-syklistene, visualisert i Figur 18, foretrekker den «raske» veien, men her ser vi at også Bruerveien er et godt alternativ til «gamle E6». Tabell 12 viser at bruken av Smaalensbanen har økt fra 2021, og også her startet økningen litt før åpningen av viadukten. Antall som passerer snittet på Smaalensbanen er her noe lavere enn antall over selve viadukten, så noen bruker bare traseen fra Hølen tettsted og sørover.



Figur 18: STRAVA-sykkeldata, snitt nord.

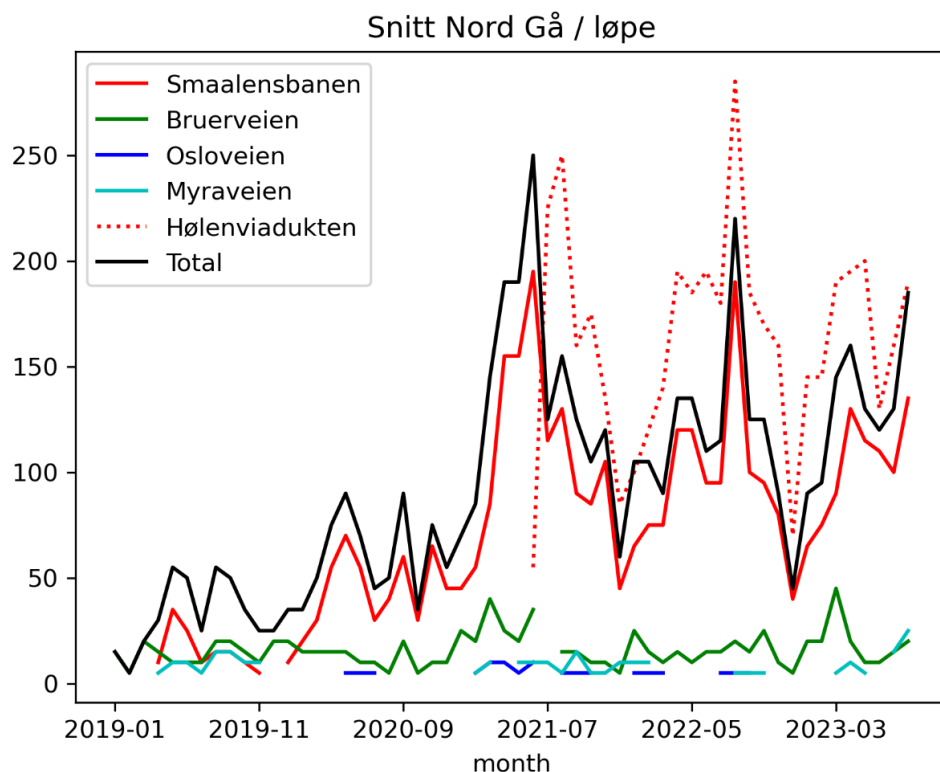
Tabell 12: Konfidensintervall for STRAVA- sykkel ved Snittpunkt nord.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.145	0.306	0.156	[0.119, 0.235]
Bruerveien	0.314	0.253	-0.061	[-0.099, -0.026]



Osloveien	0.454	0.382	-0.072	[-0.128, -0.026]
Myraveien	0.083	0.060	-0.023	[-0.041, -0.009]

Gående og løpende STRAVA-brukere har blitt langt mere aktive på Smaalensbanen fra 2021, dette vises i Figur 19, men igjen har økningen startet et par måneder før åpningen. De andre gangveiene er lite brukt.

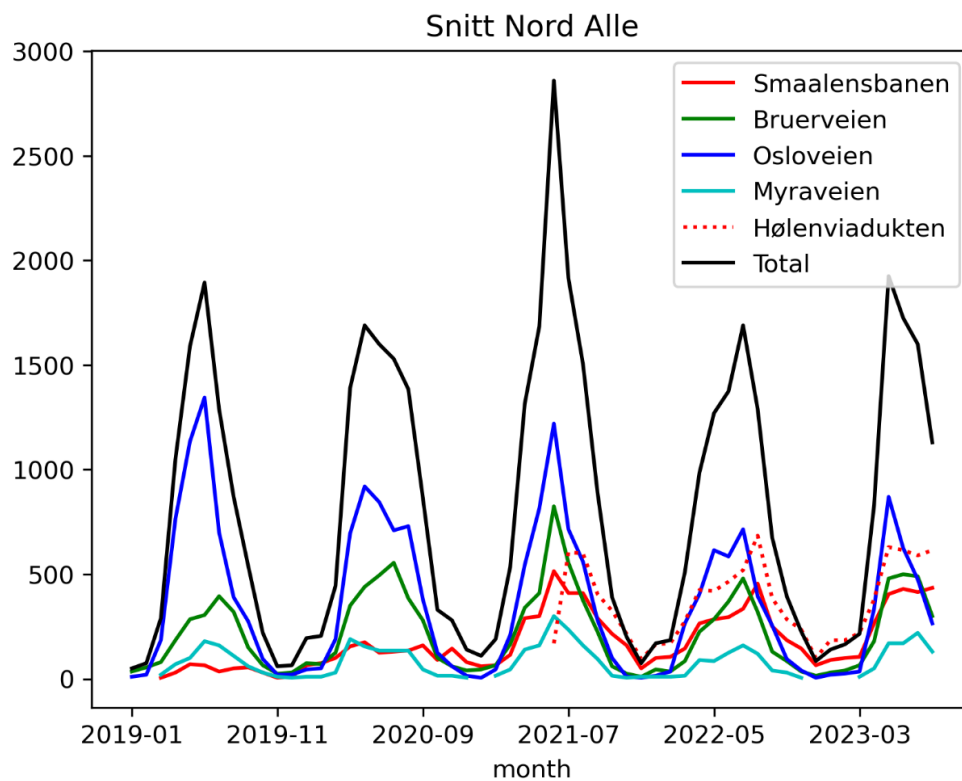


Figur 19: STRAVA-gå/løpe, snitt nord.

I Tabell 134 ser vi dette tallfestet, vi har en signifikant økning i andelen av brukere av jernbanetraseen, og tilsvarende tydelige fall for alle andre veier. Denne er og vist i Figur 20.

Tabell 13: Konfidensintervall for STRAVA- gå/løpe ved Snittpunkt nord.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.579	0.807	0.228	[0.149, 0.333]
Bruerveien	0.288	0.131	-0.157	[-0.243, -0.099]
Osloveien	0.044	0.023	-0.021	[-0.047, -0.002]
Myraveien	0.090	0.040	-0.050	[-0.101, -0.008]



Figur 20: STRAVA alle brukergrupper, snitt nord.

Når vi ser på alle trafikantgrupper summert, Tabell 14, så er igjen STRAVA-trafikken dominert av syklister. Vi ser det samme som tidligere, at den relative andelen som velger jernbanetraseen har økt signifikant, og at alle andre har falt.

Tabell 14: STRAVA alle brukergrupper, snitt nord.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Smaalensbanen	0.210	0.426	0.216	[0.170, 0.295]
Bruerveien	0.310	0.222	-0.088	[-0.130, -0.054]
Osloveien	0.394	0.286	-0.108	[-0.146, -0.067]
Myraveien	0.086	0.065	-0.020	[-0.039, -0.006]

Når vi ser på år-til-år analysen, Tabell 15, så ser vi det har vært en økning hele tiden, og ikke bare i 2021-22. Økningen i 2020 var vel så stor.



Tabell 15: Analyse av endringer fra år til år, alle brukergrupper, snitt nord.

Navn	2019-20	Diff	Konf	2020-21	Diff	Konf	2021-22	Diff	Konf	2022-23
Smaalensbanen	0.144	0.131	[0.068, 0.241]	0.276	0.136	[0.095, 0.189]	0.412	0.028	[-0.024, 0.074]	0.440
Bruerveien	0.317	-0.014	[-0.076, 0.033]	0.303	-0.089	[-0.120, -0.059]	0.214	0.017	[-0.009, 0.049]	0.231
Osloveien	0.436	-0.083	[-0.124, -0.038]	0.353	-0.057	[-0.098, -0.003]	0.296	-0.019	[-0.122, 0.026]	0.277
Myraveien	0.103	-0.034	[-0.058, -0.013]	0.068	0.01	[-0.012, 0.037]	0.078	-0.026	[-0.056, -0.005]	0.053

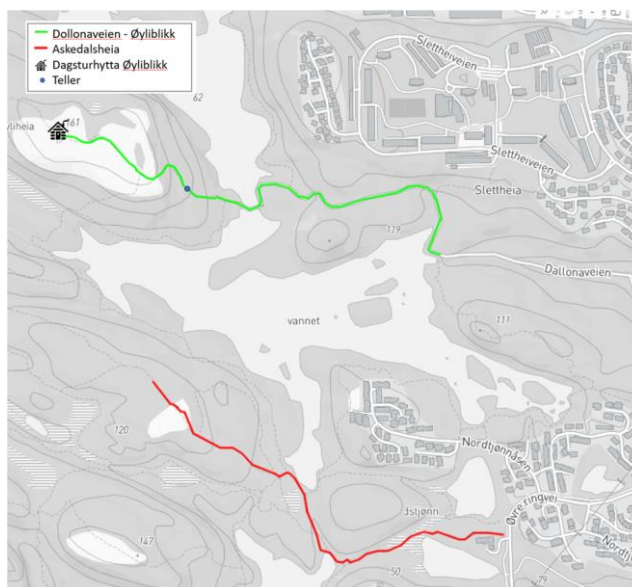
4.3.3 Oppsummering fra sammenligningsområde i Vestby

Samlet sett viser analysen av STRAVA-data en signifikant overføring av trafikk til Smaalensbanen etter bruåpningen. Denne omfordelingen er spesielt synlig blant gående og løpere. Trendene er konsistent over flere år, noe som tyder på en vedvarende endring i trafikkmønsteret. Vi ser også en økning blant syklister, men disse foretrekker i større grad andre parallelle veier hvor de kan sykle raskt. Vi drar konklusjonen at dette skyldes at syklister som bruker STRAVA ofte er ute og treningssyklere, hvor det er et ønske om fart, og da er en tursti mindre egnet. Dette er også den store utfordringen med STRAVA-dataene; de representerer ikke «folk flest». Dette gjelder også for gående og løpere, men i analysen blir dette ekstra tydelig for syklister. Selv om dataene gir verdifull innsikt, må de tolkes med forståelse for at de primært representerer en spesifikk gruppe entusiaster, ikke den brede befolkningen.

4.3.4 STRAVA i Kristiansand

Vi bruker samme hovedteknikk for å analysere STRAVA-data for Øyliblikk dagsturhytta som vi gjorde ved Smaalensbanen, og velger ut turveier der vi kan sammenlikne utviklingen av den relative trafikken på disse. Vi vil fortrinnsvis også sammenlikne perioder som dekker de samme årstider. På denne måten forsøker vi å nøytralisere virkningen av årstidsvariasjoner og trender i penetrasjonen av STRAVA i befolkningen.

Som sammenlikningspunkt velger vi stien over Askedalsheia, like sør for Øyliheia, vist i Figur 21. Vi har ikke den samme parallelle føringen som i Vestby, der man bare kan velge en av veiene i snittet om man ønsker å bevege seg i den transportkorridoren. Vi har her to stier i liknende turområder, tilstrekkelig nær til å være alternative valg, tilstrekkelig fjern til at det er mindre sannsynlig at noen vil passere begge på samme tur. De har også omtrent samme STRAVA-trafikk. Ingen av disse stiene er populær blant STRAVA syklister (0-15 besøk pr måned), så vi bruker her bare aggregerte data for alle kategorier (Figur 22).

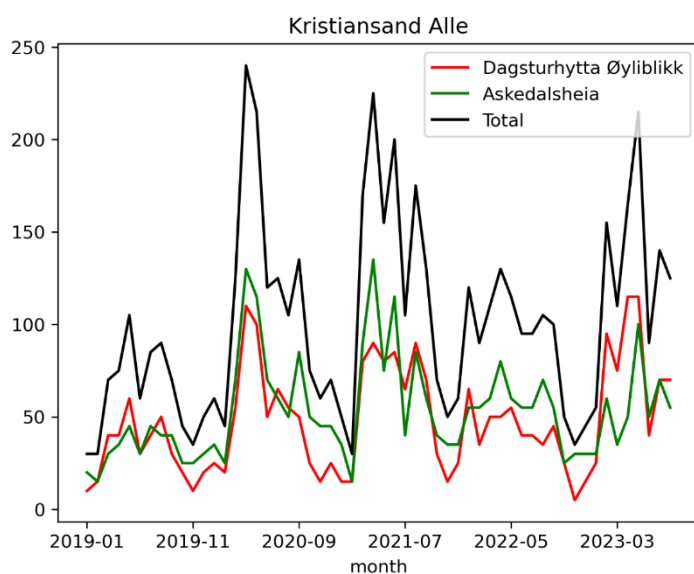


Figur 21: Kartet viser Dollonaveien opp til Øyliblakk og stien over Askedalsheia som er brukt som sammenlikningspunkt.

Siden vi bare har STRAVA data fram til august 2023 så velger vi i første omgang å sammenlikne perioden for et halvt år før åpning (2022-09 – 2023-02) med perioden et halvt år etter (2023-03 – 2023-08). Tabell 16 viser at endringen er signifikant (konfidensintervallet dekker ikke null).

Tabell 16: Endring fra et halvt år før til et halvt år etter.

Navn	Før	Etter	Differanse	Konfidensintervall (95%)
Dagsturhytta Øyliblakk	0.416	0.570	0.154	[0.049, 0.275]
Askedalsheia	0.584	0.430	-0.154	[-0.275, -0.049]



Figur 22: Månedssum for STRAVA, alle kategorier, Øyliheia og Askedalsheia.



Her kan man innvende at vi sammenlikner to perioder som ikke dekker samme årstid, så vi har ikke kompensert tilstrekkelig for årstidsvariasjoner, og vi har bare sammenliknet perioden før og etter, og kan derfor ikke si om dette er en del av en generell trend. Vi går derfor videre på samme måte som før, med å sammenlikne endringer fra år til år. Denne gang bruker vi bare perioden fra mars til august i alle år.

Vi ser av Tabell 17 at den største endringen var fra 2022 til 2023 da vi hadde en signifikant økning av andelen som går forbi Øyliblikk vs. Askedalsheia. Vi ser ingen signifikant økning andre år, heller fall både for 2019-20 og 2021-22.

Tabell 17: Endring år for år.

Navn	2019	Diff	Konf	2020	Diff	Konf	2021	Diff	Konf	2022	Diff	Konf	2023
Dagsturhytta	0.53	-0.06	[-0.10, -0.00]	0.47	0.02	[-0.02, 0.06]	0.49	-0.08	[-0.16, -0.02]	0.42	0.15	[0.07, 0.24]	0.57
Askedalsheia	0.47	0.06	[-0.00, 0.10]	0.53	-0.02	[-0.06, 0.02]	0.51	0.08	[0.02, 0.16]	0.58	-0.15	[-0.24, -0.08]	0.43

4.3.5 Oppsummering STRAVA i Kristiansand

Metoden med å beregne endringer i relativ trafikk ser ut til å fungere godt også for STRAVA-data for Kristiansand. Analysen viser at åpningen av Øyliblikk dagsturhytta har hatt en positiv effekt på trafikken i området relativt aktiviteten over Askedalsheia. Vi er i stand til å påvise endringer i trafikken forbi Øyliblikk dagsturhytta, både når vi sammenligner periodene et halvt år før og etter åpningen av Øyliblikk, og når vi analyserer de samme periodene over forskjellige år. Vi begynner derfor å føle oss trygge på at denne metoden kan brukes for å se utvikling over tid der vi kan nøytralisere utenforliggende faktorer som påvirker de analyserte stiene i samme grad.

4.4 Sammenligning av STRAVA-data og tellerdata

Vi vil forsøke å se om det er en sammenheng mellom tellerdata og STRAVA. Oppsetting og drift av tellere er ganske kostbart, og de gir kun målinger for ett punkt i veinettverket. STRAVA-data er derimot rimelige, de er innsamlet av brukerne selv, og de dekker dessuten et stort stinettverk. Det ville derfor kunne være nyttig å se om det er mulig å finne en funksjonell sammenheng mellom STRAVA og passeringstallere. Alle forbehold om snever brukeropulasjon og avrunding i aggregering av data fra kapittel 4.3.1 er imidlertid fortsatt gyldige.

For å se på sammenhengen mellom passeringstallere og STRAVA så summerer vi opp tellerdata for det veistykket der telleren var plassert for hver måned, og forsøker med en regresjonsanalyse¹¹. Vi setter da opp en regresjonsmodell, en hovedstruktur for den funksjonen vi vil tilpasse til dataene, og bruker denne

¹¹ SciPy v1.11.4 Manual, linregress:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html>



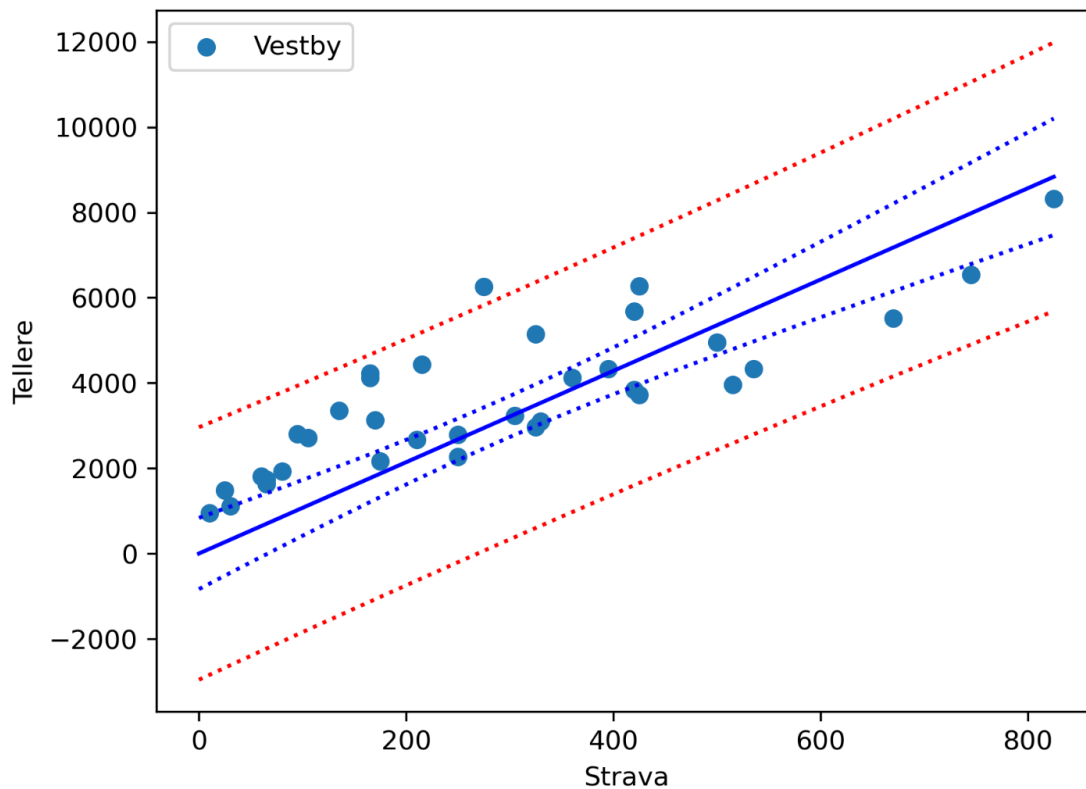
i en regresjonsanalyse. Der vil parametere i funksjonen bli justert slik at differansen mellom funksjonen og måledata blir minst mulig.

For våre formål vil det være naturlig å velge en funksjon som er garantert å gå gjennom origo, at null STRAVA-brukere også betyr null passeringer. Vi starter da med en regresjonsmodell av formen $y = a \cdot x$ der x er antall STRAVA-målinger, y er prediktert antall passeringer og a er parameteren som skal tilpasses til dataene.

Normalt vil man «ta til side» noe av dataene, for å kunne teste modellen med andre data enn hva man har for å tilpasse den. Her har vi dessverre ikke stort mere data enn hva vi behøver for å tilpasse modellen, og mangler luksusen med uavhengig kontroll. Vi har da sterke avhengigheter mellom modellen og tekstdataene, som kan få modellen til å se bedre ut enn hva den egentlig er. Vi gjør allikevel et forsøk, der vi bruker data fra første halvdel av den perioden vi har data til å kalibrere modellen, og deretter tester denne modellen mot data fra siste halvdel av perioden. Ulempen med dette er at vi får litt lite data til hver enkelt av disse del-analysene.

4.4.1 Data fra Vestby alene

Det første forsøket ble gjort med kun data for Vestby. Figur 23 viser målinger fra passeringstelleren (y-aksen) plottet mot STRAVA (x-aksen). Regresjonslinjen har formen $y = 10.7 \cdot x$. Et 95% konfidensintervall for trendlinjen er vist med de blå prikkete linjene og viser at vi har en ganske veldefinert lineær sammenheng. De røde prikkete linjene viser et 95% konfidensintervall for prediksjonsevnen, og vi ser at spredningen i grunnlagsdataene gjenspeiles i en ganske stor usikkerhet.

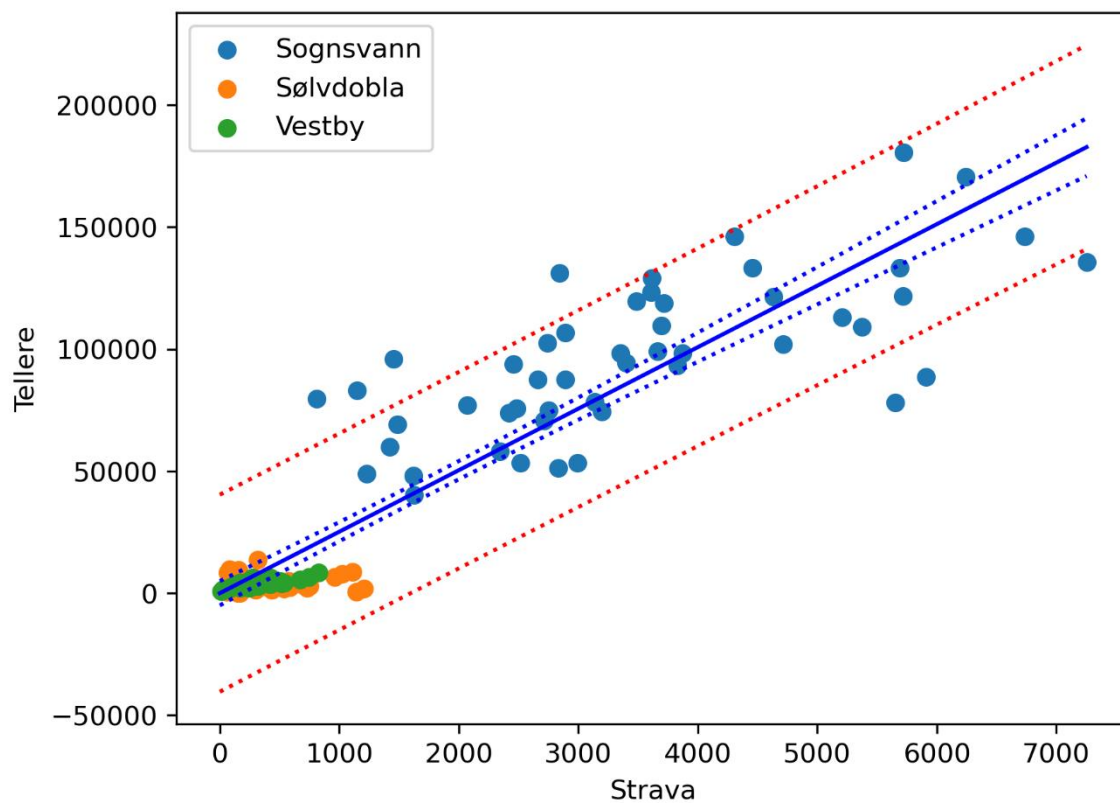


Figur 23: Regresjonsanalyse for Smaalensbanen. Antall STRAVA-passeringer på x-aksen og antall tellerpasseringer på y-aksen. Regresjonslinje med hel blå strek, og et 95% konfidensintervall (usikkerhet i plasseringen av regresjonslinjen) for denne i prikkete blå strek. De prikkete røde linjene viser usikkerheten i å bruke denne regresjonslinjen til å forutsi totalt antall passeringer.

Ut fra disse dataene kan det derfor ikke anbefales å bruke denne trendlinjen til å forutsi totalt antall passeringer for enkeltmåneder ut fra STRAVA-målinger. Men det er mulig de predikerte dataene kan brukes for aggregert statistikk (sum og gjennomsnitt over lengre perioder). Det er også mulig at vi kan oppnå bedre resultater med et større datasett.

4.4.2 Data fra Vestby og Oslo

For å få et bedre datagrunnlag har vi fått tilgang til ytterligere to tellere. En ved Sognsvann like i utkanten av bebyggelsen, mot Nordmarka i Oslo, og en ved Sølvdobla, i Østmarka nær Klemetsrud i Oslo. Disse sammenkobles med STRAVA-data for tilnærmet samme vei-lenke som der telleren var oppstilt. Vi starter med samme analyse som for Vestby, men nå med data fra alle tre områdene.

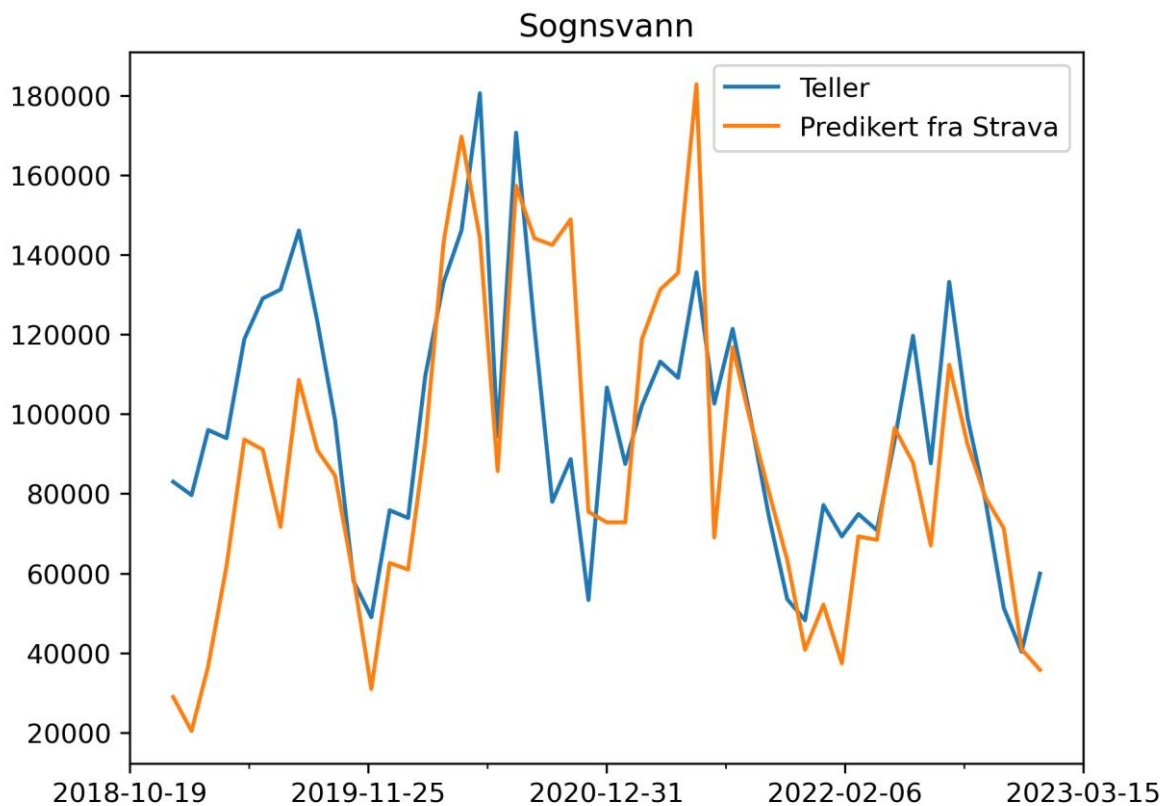


Figur 24: Regresjonsanalyse med konfidensintervaller: Tellerpasseringer som en funksjon av STRAVA-brukere basert på data fra Sognsvann, Sølvdobla og Vestby.

Som i forrige forsøk med bare data fra Vestby ser vi at vi har en klar lineær sammenheng, men at vi også har en ganske stor spredning rundt en trendlinje (Figur 24). Her har vi funnet en sammenheng med 25.2 antatte passerende pr. STRAVA-bruker. Vi får en korrelasjonskoeffisient på 0.916, og kan konkludere



med en god lineær sammenheng. Når vi plotter predikerte og målte passeringer sammen så ser vi at resultatet ikke blir så galt ved Sognsvann, se Figur 25.



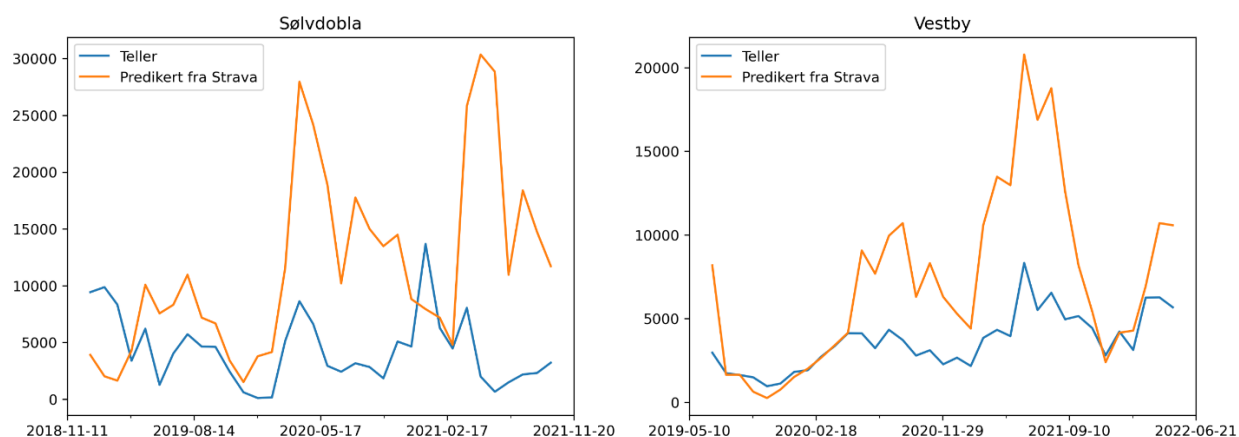
Figur 25: Prediksjon av antall passeringer pr måned basert på STRAVA-data, etter regresjonsanalyse av data fra Sognsvann, Sølvdøbla og Vestby.

Men når vi forsøker det samme for Sølvdøbla og Vestby så vil denne prediksjonen overestimere antall passeringer ganske kraftig. Oppsummert så får vi da rundt dobbelt så mange predikerte passeringer som målt, vist i Tabell 18 og Figur 26.

Tabell 18: Sum av tellerdata vs. predikerte data.

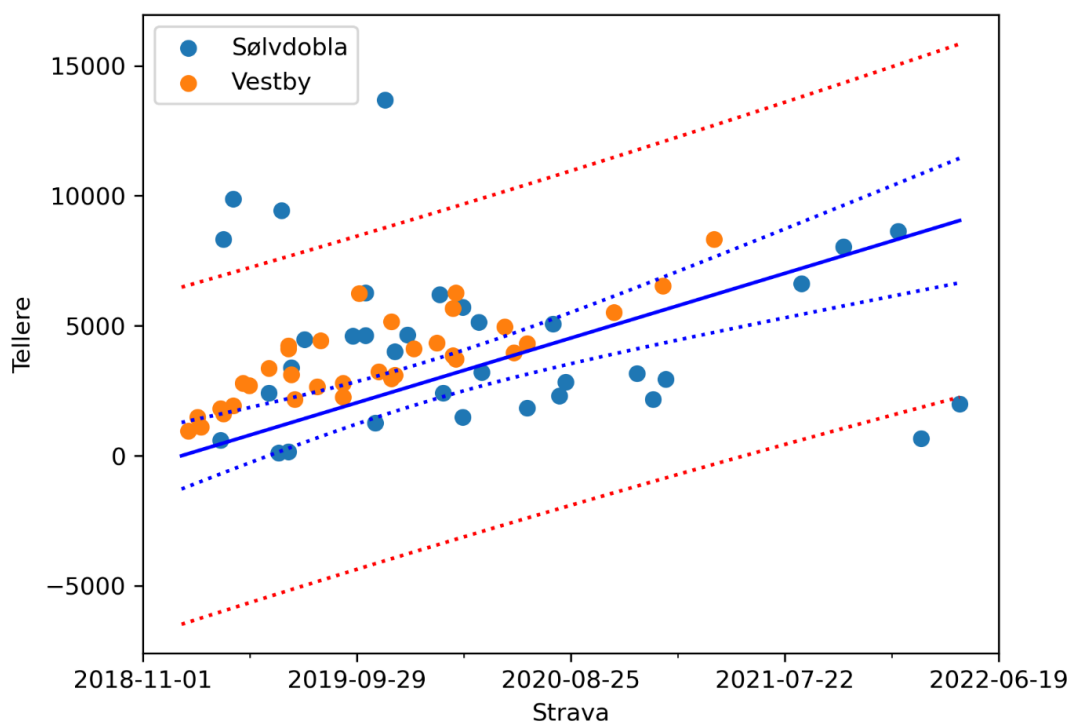
Navn	Sum tellerdata	Sum predikert fra STRAVA-data	Feil (%)
Sognsvann	4 739 875	4 298 475	-9.3 %
Sølvdøbla	148 374	398 557	168 %
Vestby	127 633	253 570	98 %

Dette forteller oss at andelen av STRAVA-brukere kan variere fra sted til sted, og at det kan være nødvendig å ha flere modeller, tilpasset forskjellige områder og populasjoner av brukere.



Figur 26: Prediksjon av antall passerende pr måned for Sølvdobla og Vestby ut fra en modell som også inneholder data fra Sognsvann gir sterk overestimering.

En ny analyse basert på bare data fra Sølvdobla og Vestby gir naturlig nok en bedre tilpassing til akkurat disse dataene (Figur 27).



Figur 27: Regresjonslinje med konfidensintervaller for Sølvdobla og Vestby.

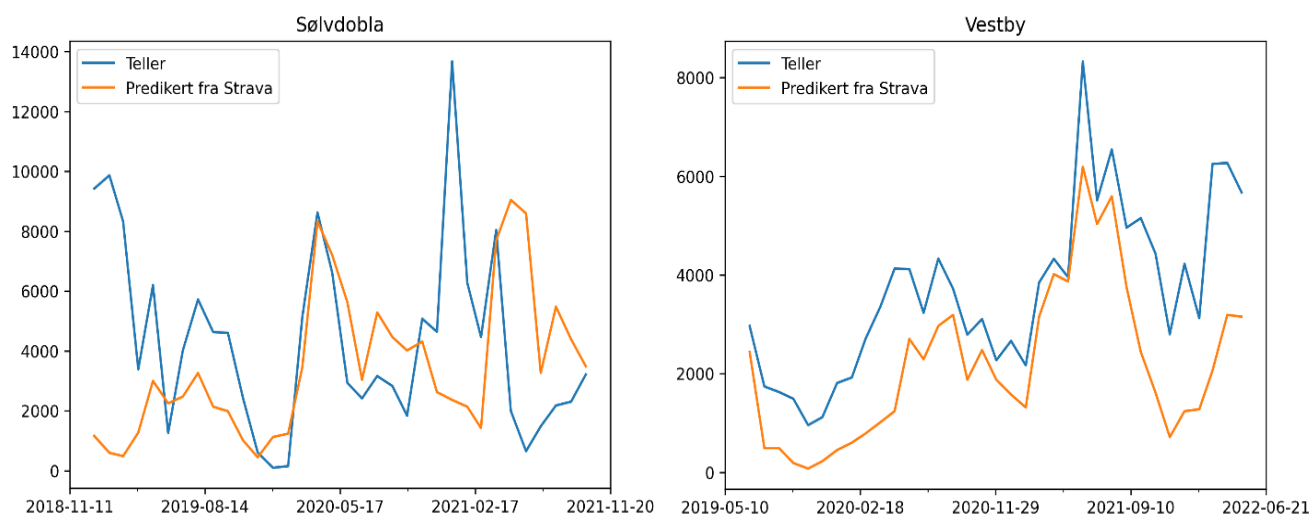
Vi estimerer her 7.5 antatte passerende pr. STRAVA-brukerpassering, altså rundt 1/3 av hva vi hadde med data fra Sognsvann, men heller ikke disse to gir helt samstemte resultater. En korrelasjonskoeffisient på 0.216 forteller at den lineære sammenhengen er mindre tydelig når vi utelater Sognsvann.



Tabell 19: Sum av tellerdata vs. predikerte, proporsjonal og logaritmisk regresjonsmodell, data fra Vestby og Sølvdobla.

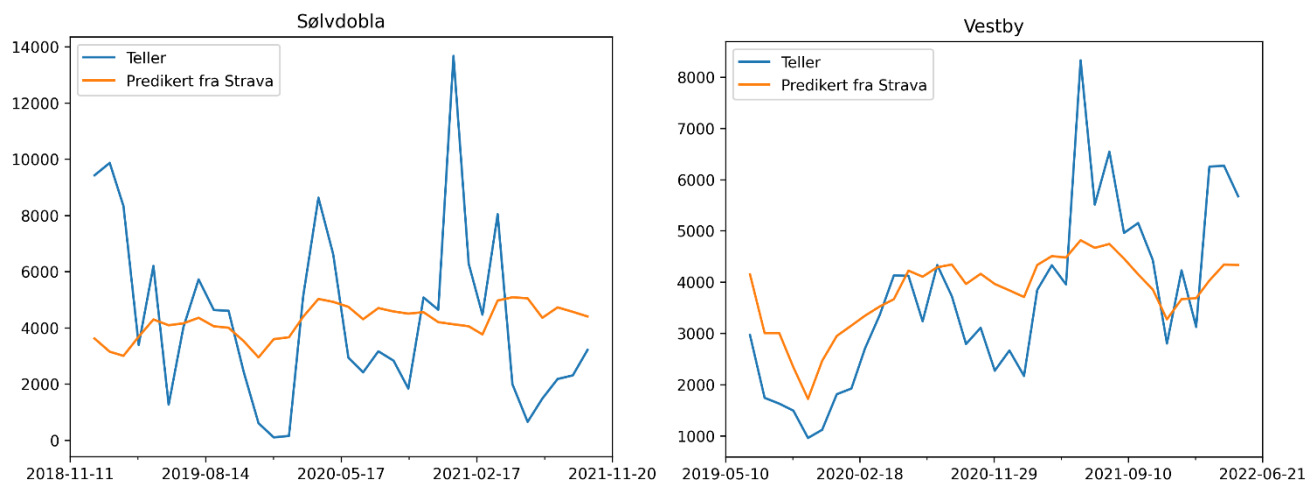
	Sum tellerdata	Sum predikert fra STRAVA, lineær regresjon	Sum predikert fra STRAVA, logaritmisk regresjon
Sølvdobla	148 374	118 795	143 252
Vestby	127 633	75 580	133 236

Vi ser av Tabell 19 at prediksjonen med en lineær modell gir ca. 40% for lave verdier for Vestby og 20% for lave for Sølvdobla (Figur 28).



Figur 28 Prediksjon av antall passeringer pr måned basert på bare data fra Vestby og Sølvdobla, lineær regresjon

Et forsøk med logaritmisk regresjon (ferdig tilpasset regresjonsmodell: $y = 717 \cdot \ln(1 + x)$) gir aggregerte resultater som er langt nærmere forventningsrett (Tabell 19), men har større problemer med å gjenskape noe som likner persontellergrafen (Figur 29).

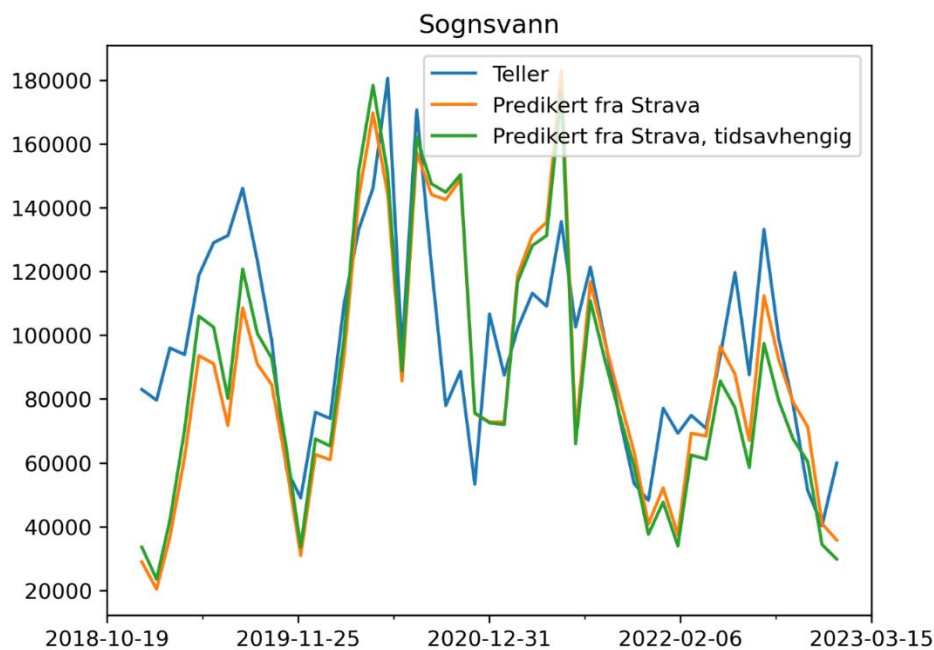


Figur 29: Prediksjon med logaritmisk regresjonsmodell, data fra Sølvdøbla og Vestby.

4.4.3 Tidsavhengig modell

Vi er allerede kjent med at andelen av befolkningen som bruker STRAVA har økt med tiden, og vil gjøre et forsøk på å beregne hvilken virkning dette har på regresjonsmodellene våre. Vi forsøker derfor med en bi-lineær tilpassing, der prediksjonen også er avhengig av tidspunktet. Den tilpassede modellen er her: $y = a_1 \cdot x + a_{12} \cdot x \cdot d$, der y er prediktert verdi, x er STRAVA-verdien, d er tiden i antall dager siden 01.01.2019, og a_1 og a_{12} er parametere som her er estimert til $a_1 = 29.2$ og $a_{12} = -0.00563$. Standardavviket for a_1 og a_{12} er henholdsvis 1.73 og 0.00220.

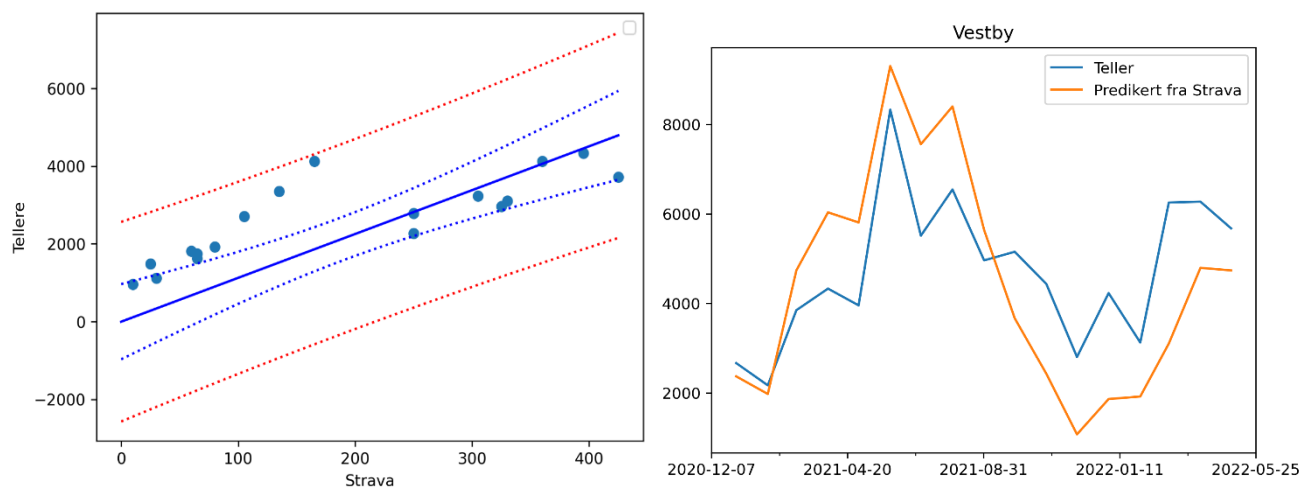
Parameteren a_1 kan tolkes som antall predikterte passeringer pr passerte STRAVA-bruker med den andelen STRAVA-brukere som var i 2019. Parameteren a_{12} forteller hvor mye antall predikterte passeringer faller pr. tidsenhet. Siden antall brukere av STRAVA har økt over perioden så forutser vi færre passeringer av ikke-STRAVA-brukere pr STRAVA-bruker. Omregnet til årsverdier så faller den med 2.05 pr år (Figur 30).



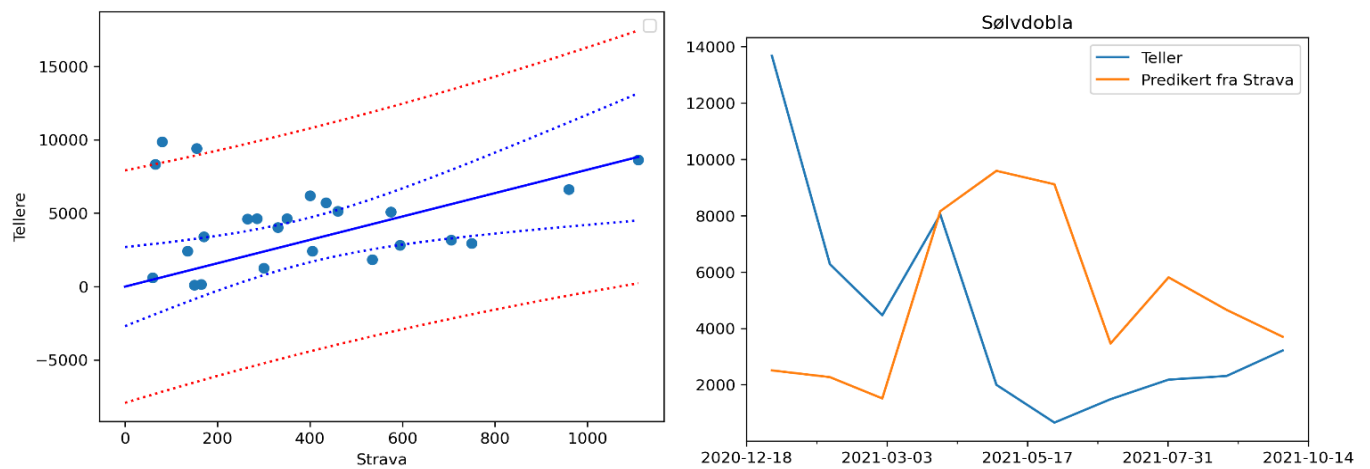
Figur 30: Enkel lineær modell vs. tidsavhengig bilinear modell.

4.4.4 Bruk av en periode med tellerdata for å ekstrapolere med STRAVA

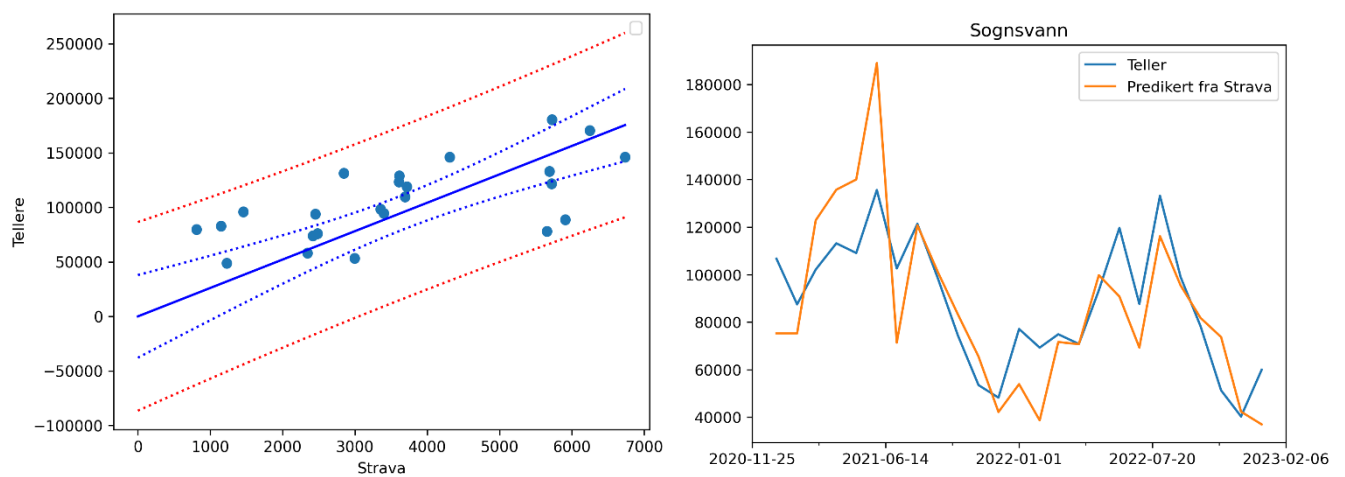
Vi har også forsøkt å bruke regresjonsanalyse over en kortere periode (01.01.2019 – 31.12.2020), og deretter brukt denne regresjonsmodellen sammen med STRAVA-data for å predikere tellerdata for resten av den tiden vi har data (Vestby Figur 31, Sølvdobla Figur 32 og Sognsvann Figur 33).



Figur 31: Regresjonsanalyse og prediksjon for Vestby.



Figur 32: Regresjonsanalyse og prediksjon for Sølvdobla.



Figur 33: Regresjonsanalyse og prediksjon for Sognsvann.

Tabell 201: Oppsummering regresjonsbasert STRAVA-prediksjon.

Navn	Proporsjonalitet (antall personer pr STRAVA-bruker)	Standardavvik	Summert tellerpasseringer	Summert STRAVA- prediksjon
Vestby	11.3	1.1	80 233	75 376
Sølvdobla	7.96	1.49	44 307	50 795
Sognsvann	26.1	1.86	2 207 476	2 163 472



Vi har kjørt denne analysen på Vestby, Sølvdobla og Sognsvann, og ser en etter forholdene god sammenheng (feil på ca. 10% av sum over lengre periode (Tabell 20) mellom predikerte og tellerbaserte verdier.

Vi gjorde også et forsøk på regresjon med en tidsavhengig bilineær funksjon her, men fant at treningsdatasettet er for lite og perioden for kort til å gi en god bestemmelse av utviklingen i STRAVA- bruk over tid.

4.4.5 Oppsummering fra sammenligning av STRAVA-data og tellerdata

Analysen av sammenhengen mellom passeringstellere og STRAVA-data for Smaalensbanen viser en tydelig lineær trend, men med betydelig usikkerhet. Dette indikerer at vi ikke kan forutsi det totale antallet passeringer med høy grad av sikkerhet i et område med lite data, basert på kun STRAVA-data. Det er mulig at vi kan oppnå bedre resultater med et større datasett, derfor har vi også analysert to områder i Oslo. Disse analysene viser at det er mulig å bestemme andelen av STRAVA-brukere av det totale antall brukere når vi har en større datamengde, og dermed finne en funksjonell sammenheng mellom antall STRAVA-brukere og totalt antall besøkende. Imidlertid kan denne andelen variere sterkt fra område til område, og man kan ikke umiddelbart bruke en modell kalibrert et sted på et annet. Vi har ikke tilstrekkelig data her til å kunne anslå influensområdet til en kalibrert modell, og må foreløpig råde til forsiktighet. På den annen side ser andelen av STRAVA-brukere på samme sted ut til å variere lite over tid, og vi fikk gode resultater med å bruke en modell kalibrert på et par års målerdata til å anslå summen av totalt antall besøk i påfølgende år ut fra STRAVA-data.

4.5 Telia-data

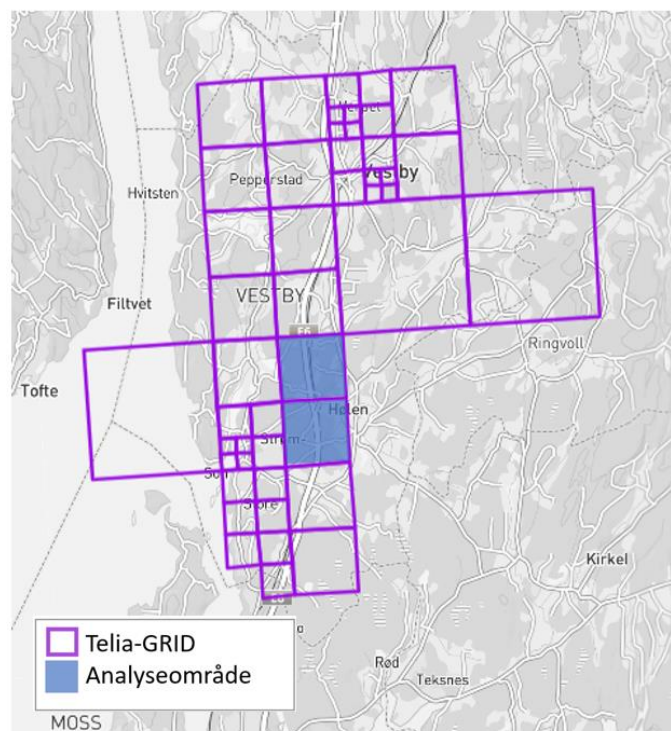
4.5.1 Utfordringer med data fra Telia

En av de primære utfordringene med data fra Telia er at de er for grove til å kunne brukes effektivt på enkelte turområder. Dataene er organisert i et GRID-system. Vår opprinnelige forventning var at størrelsen på disse rutene ville variere dynamisk basert på antall besøkende – større ruter i tider av dagen med færre mennesker ute, og mindre ruter under perioder med høy aktivitet, som for eksempel rushtiden. Dessverre er dette ikke tilfellet. Rutene i GRID-systemet er statiske og endrer ikke størrelse basert på antall personer i et område.

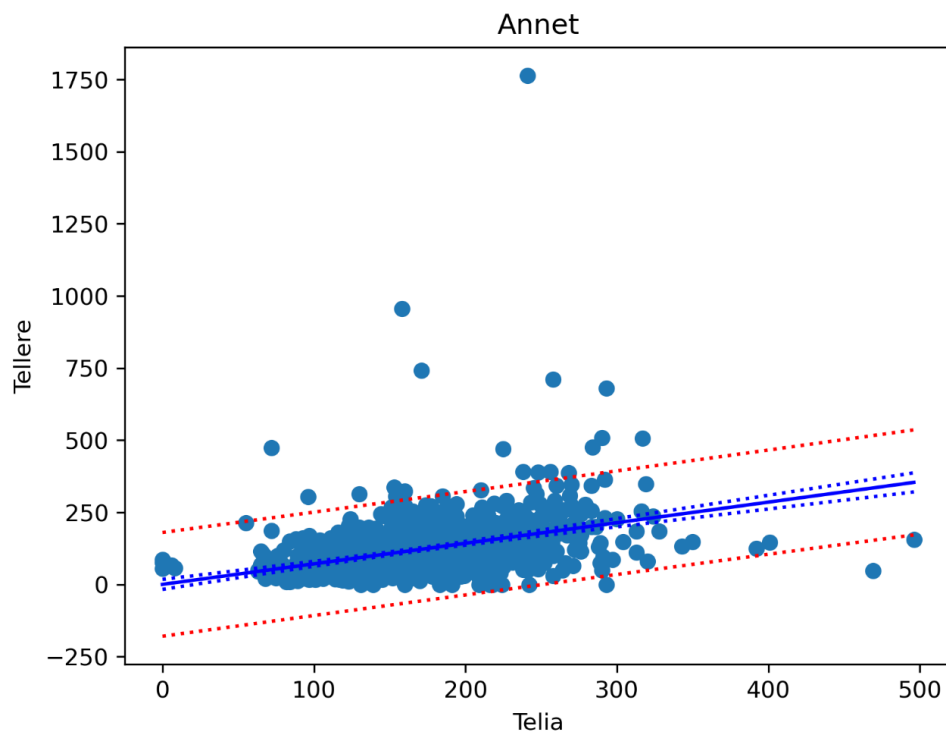
Rutenes størrelse, også de minste, gjør at det ikke er mulig å skille mellom aktiviteter på nærliggende stier. Dataene inneholder heller ikke informasjon som kan skille en type aktivitet fra en annen. Det er mulig å skille aktiviteter basert på hastighet, for eksempel å differensiere mellom om en person beveger seg i et kjøretøy eller til fots. Men igjen, det blir vanskelig å skille mellom ulike typer gangaktiviteter, som for eksempel å gå for å kjøpe melk, eller å være ute på en spasertur. Det er heller ikke mulig å skille mellom ulike typer aktiviteter som gang, løping og sykling. Nøyaktigheten i å kunne skille på om noen er i aktivitet eller sitter i ro påvirkes også av mobilnettets cellestørrelse, som varierer. I analyseområdet har vi ruter som er fra 500 til 4000 meter, noe som betyr at det kreves lengre bevegelser for at en aktivitet skal bli registrert i de større rutene, noe som kan generere ujevne tall.

4.5.2 Sammenlikning av Telia og tellerdata i Vestby

Vi forsøker her med en liknende metode som med STRAVA-dataene, der vi forsøker å finne en funksjonell sammenheng mellom Telia og tellerdata. Her forsøker vi å koble tellerdata til antall personer som har vært i den Telia-gridrute som dekker telleren. I Vestby betyr dette rutene «Hølen» og «Bylterudveien» (for den perioden da telleren var i posisjon «Nord»), vist i Figur 34. Begge disse rutene har et areal på 4 km². Telia kategoriserer besøk etter brukeroppførsel som «home», «work» eller «other». Etter å ha forsøkt korrelasjon mot disse fant vi at brukere av kategorien «annet» hadde best samsvar med telleren, visualisert i Figur 35. Vi har også her valgt å tvinge regresjonsformelen gjennom null og får en tilpasset regresjonsmodell: $y = 0.713 \cdot x$, der regresjonsparameteren har standardavvik 0.0161. Regresjonskoeffisienten er på 0.350, noe som forteller oss at den lineære sammenhengen er der, men den er ikke så sterk.

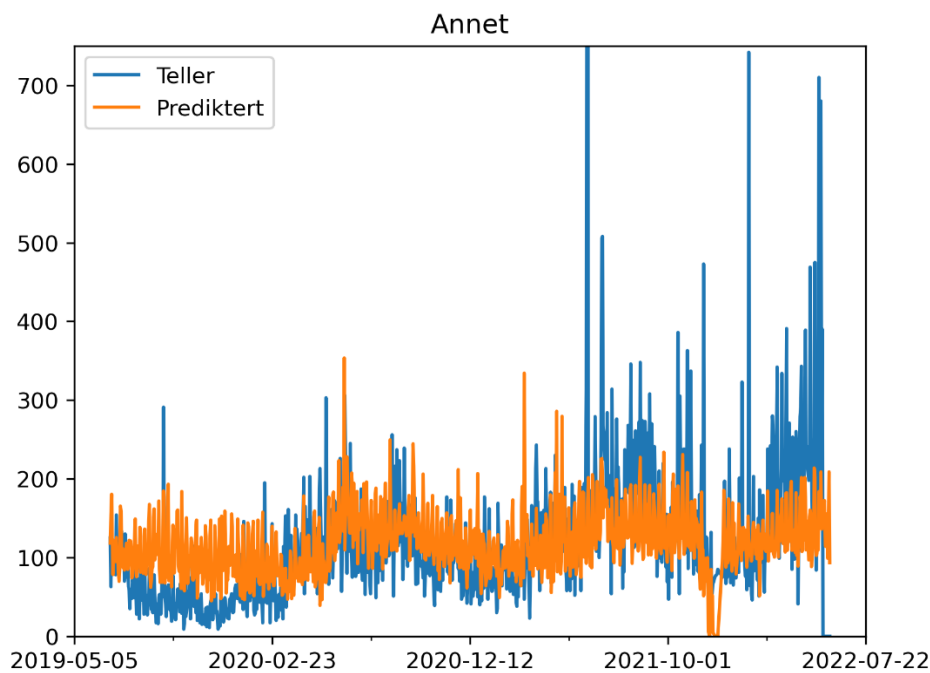


Figur 34: Kartet viser Telia-GRID rutenett og de to rutene som dekker plasseringen for tellere; «Hølen» og «Bylterudveien».



Figur 35: Regresjonsanalyse mellom Telja brukere av kategori "annet" og tellerdata i Vestby.

Vi sammenlikner de predikerte data vi får fra Telja og regresjonsformelen med tellerdata, og ser at prediksjonen faktisk kan følge noen trender i datamaterialet, Figur 36.



Figur 36: Tellerdata og prediksjon fra Telja, kategori "annet".



4.5.3 Oppsummering sammenlikning av Telia og tellerdata i Vestby

Selv om sammenhengen ikke er særlig sterk, viser våre resultater at prediksjonene fra Telia-dataene følger observerte trender i tellerdataene. Disse funnene understreker potensialet for å bruke telekommunikasjonsdata som et hjelpemiddel for å forstå og analysere bevegelsesmønstre. Dette gjelder selv om GRID-området som brukes i analysen dekker et areal på 4 km². For å kunne trekke mer konkrete konklusjoner, er det imidlertid nødvendig å utvide analysen til å omfatte flere områder.

4.5.4 Bestilling av en spesialtilpasset analyse fra Telia – Trip analyse

Vi ønsket å utforske muligheten for å innhente mer detaljerte data fra Telia og inviterte dem til en dialog. Telia var positive til forespørselen og foreslo å bruke en uke med en av deres konsulenter for å undersøke muligheten for å estimere antall personer som passerer spesifikke punkter på stiene og samtidig forholde seg til GDPR-reglene. Telia påpekte at de ikke kunne garantere konkrete resultater, men at formålet var å utforske mulighetsrommet. Detaljene og resultatene fra dette arbeidet er beskrevet i **Vedlegg 1 - Rapport på leveranse fra Telia.pdf**.

Oppsummeringen fra analysen indikerer at det er utfordrende å innhente data ned til enkeltpunkter. Vi observerer også at vår tilnærming med å sammenligne Telia-data og tellerdata, ved å koble tellerdata til antall personer i den Telia-dekkede gridruten som omfatter telleren, viser lovende resultater.

5. VISUALISERINGSLØSNING

Dataene er visualisert i en Proof of Concept (POC). Tidlig i prosessen med dette arbeidet innså vi at det fantes en unik mulighet for dyptgående analyser for å virkelig forstå dataenes potensial. Dette arbeidet var imidlertid ikke enkelt å vise i en kartløsning. Dette førte til at rapporten fikk prioritet, mens arbeidet med POC-en ble nedprioritert. POC-en viser i hovedsak grunnlagsdataene og ikke de statistiske analysene.

Den opprinnelige metoden vår for visualisering var å arrangere de mer detaljerte linjedataene fra STRAVA med de grove Telia-dataene i et felles GRID-system. Vi så imidlertid tidlig at dette ville bli en utfordring, ettersom gridene til Telia var grovere enn estimert. Det som til slutt ble løsningen, var å berike STRAVA-data med Telia-data.

5.1 Verktøy for visualisering

Power BI er valgt som verktøy for visualisering. Vi har utviklet et interaktivt dashbord med fire sider for å forenkle analyser og evaluering av kildedata. Kildedataene består av linjer med månedsdata fra STRAVA Metro, ruter med månedsdata fra Telia, samt demografiske data fra SSB for å gi en referanse til

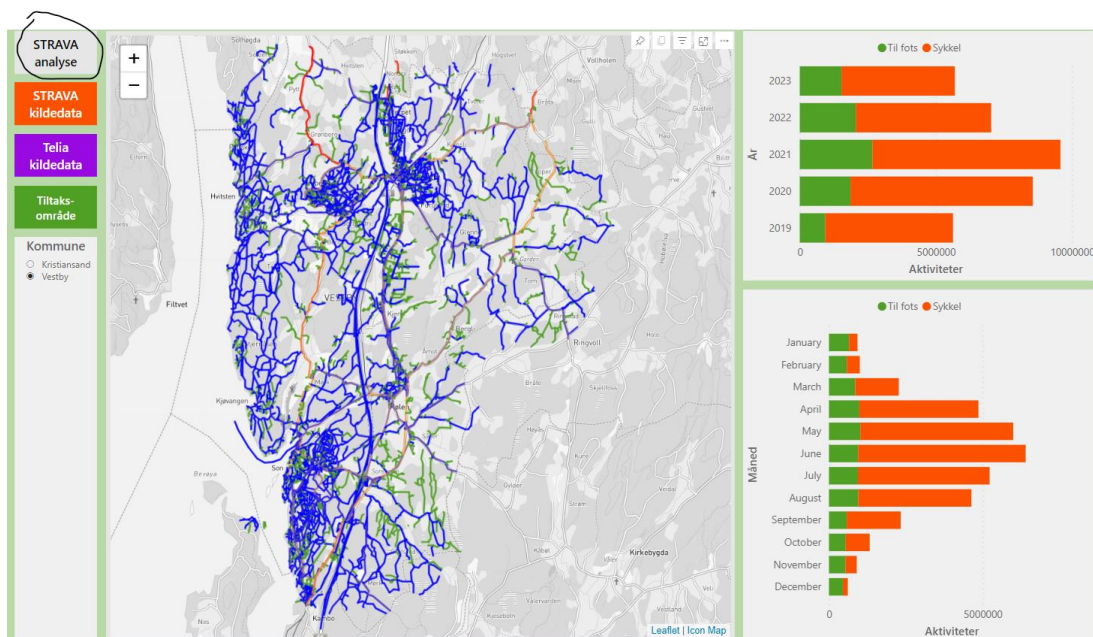
aldersfordeling i STRAVA-dataene. Disse dataene er forbehandlet i FME¹² for optimal nøyaktighet og effektivitet i Power BI, og det er gjennomført tilpasninger i PowerQuery under importen til Power BI (se flytskjema i **Vedlegg 2**).

Kildedataene for STRAVA Metro er i form av linjer, kalt "edges" av STRAVA, som representerer vei- eller stisegmenter. Til disse er det koblet månedsdata, slik at man kan få frem statistikk for én eller flere valgte edges. For Telia brukes flate-objekter som er koblet til månedsdata på samme måte.

Power BI-dashbordet er delt inn i fire sider, hvor hver side er beskrevet i teksten under.

5.1.1 STRAVA analyse

Denne siden (Figur 37) tillater brukere å velge en eller flere STRAVA-edges for å analysere summen av aktiviteter. Egne diagrammer for år og måned viser aktivitetstrender langs en edge, inkludert når på året den er mest brukt. Brukerne kan også se fordelingen mellom sykling (elsykkel og vanlig) og aktiviteter til fots (vandring og løping).

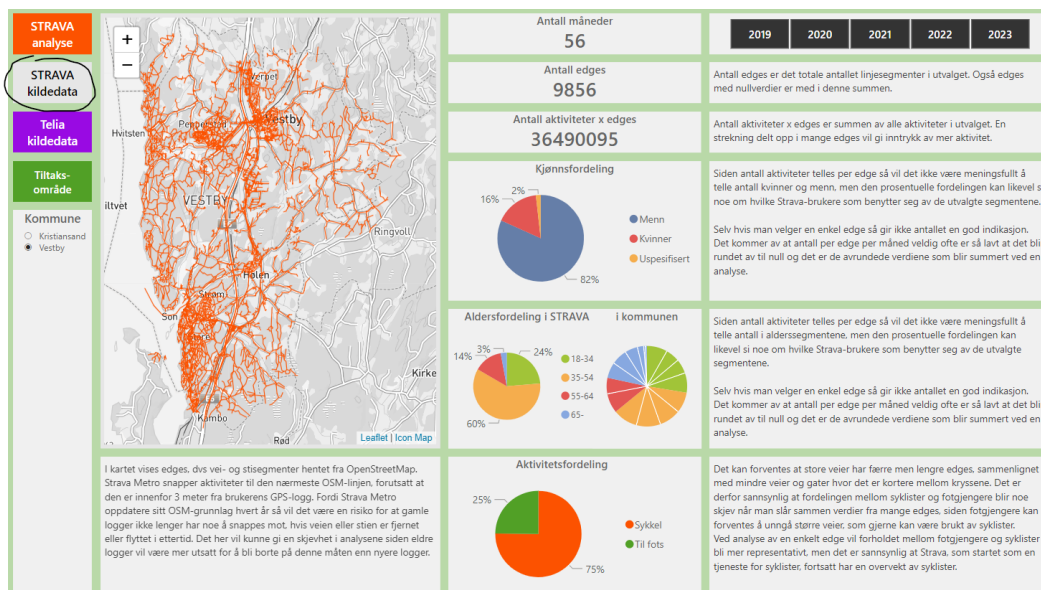


Figur 37: Viser siden STRAVA analyse i Power BI-dashbordet.

5.1.2 STRAVA kildedata

Her presenteres STRAVA-kildedataene (Figur 38) på ulike måter for å illustrere deres potensial og begrensninger i å gi klare analyseresultater om stibruk før og etter et friluftstiltak. Hvert diagram er ledsaget av en beskrivende tekst som forklarer aspektet av dataene det representerer.

¹² FME (Feature Manipulation Engine) er en programvareplattform utviklet av Safe Software, spesialisert på integrering og transformasjon av romlige data.



Figur 38: Viser siden STRAVA kildedata i Power BI-dashbordet.

Muligheter

Så lenge dataene blir brukt i relasjon til samme type data, dvs for å sammenligne én edge med en annen, så kan dataene si noe om forholdet i bruk mellom de to og også om forholdet mellom de to med tanke på hvem som bruker dem. Analysemetodene brukt i Kapittel 4.3 med sammenligning av relativ trafikk bør gis et brukergrensesnitt og inkorporeres i løsningen.

Begrensninger

Fordi antall passeringer telles per edge så kan man ikke bruke tallene f.eks. for å si noe om hvor mange som har vært i området eller hvor mange som har brukt en lengre strekning. Det blir også misvisende å sammenligne antall mellom to lengre strekninger, da den ene kan bestå av flere edges enn den andre og dermed telle den samme aktiviteten flere ganger.

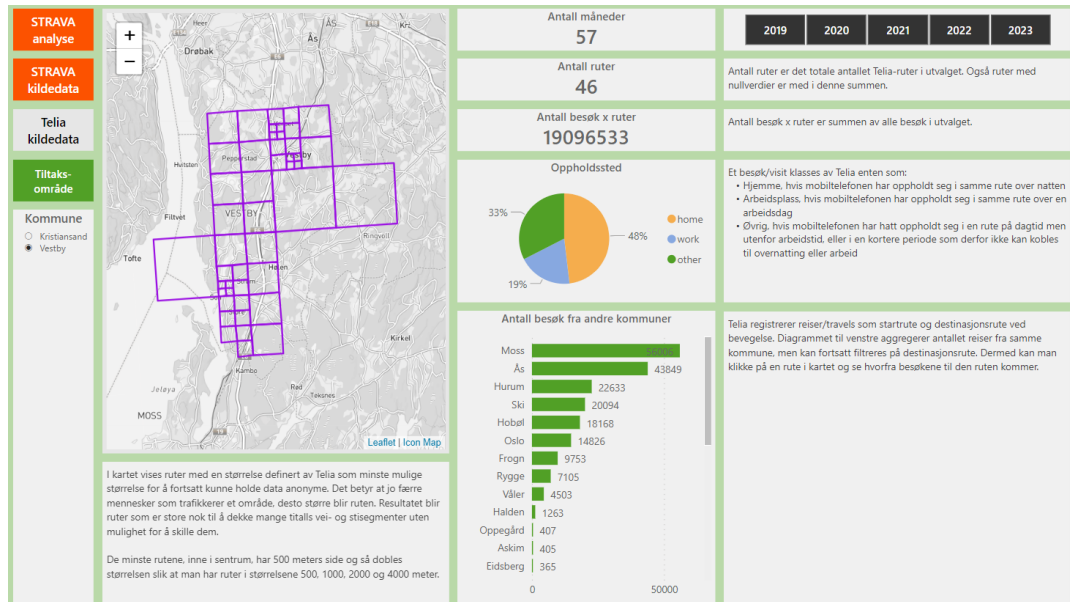
Kjønns- og aldersfordeling viser seg å ha et tydelig avvik fra SSB sine data for den aktuelle kommunen. STRAVA-brukerne har en markant overvekt av menn og aldersgruppen 35-54 år som ikke gjenspeiles i befolkningen. Også fordelingen mellom syklist og fotgjenger ser ut til å ha en falsk overvekt mot syklist, sannsynligvis grunnet at STRAVA startet som en tjeneste for syklist før også andre sporter og aktiviteter ble mulige å loggføre. En indikasjon på dette er at prosentandelen syklist synker over tid samtidig som brukermassen øker.

En annen grunn til at kjønns- og aldersforhold må brukes varsomt er at disse gruppene er avrundet hver for seg til nærmeste femtall. Det betyr at der det f.eks. har passert seks menn og fire kvinner en måned så vil resultatet si fem menn og null kvinner. Jo mer disse tallene summeres (flere måneder eller flere edges), desto høyere vil tallet for menn bli mens tallet for kvinner i verste fall blir stående på null, selv om det reelle forholdet skulle være 60-40%. Det samme gjelder for aldersinndelingen, hvor de mindre aldersgruppene risikerer å avrundes til null mens gruppen 35-54 år summeres til høyere tall. Det skal sies at aktivitetstallet som blir brukt på siden STRAVA analyse isteden bruker en verdi som er totalen utregnet før anonymisering og derfor gir et mer riktig tall ved summering.



5.1.3 Telia kildedata

Her presenteres Telia-kildedataene (Figur 39) for å illustrere deres potensial og begrensninger i å gi klare analyseresultater. Hvert tall og diagram er ledsaget av en beskrivende tekst som forklarer aspektet av dataene det representerer.



Figur 39: Viser siden Telia kildedata i Power BI-dashbordet.

Muligheter

Telia-dataene er inndelt i ruter som er for store for å si noe om enkelte stier og veier. Men de kan fortsatt gi nyttig innsikt i et område. For oppholdssted så viser dataene om mobiltelefonen er hjemme (i ro over natten), på jobb (i ro gjennom normal arbeidstid) eller om den kan antas å være utenom disse stedene.

Det finnes også informasjon for hver rute om hvor mange som den måneden har beveget seg fra en annen rute (i dashbordet summert på kommune) og til den valgte ruten eller rutene. På den måten kan man se om et område har mye trafikk fra nabokommuner eller om det er overveiende lokale aktiviteter.

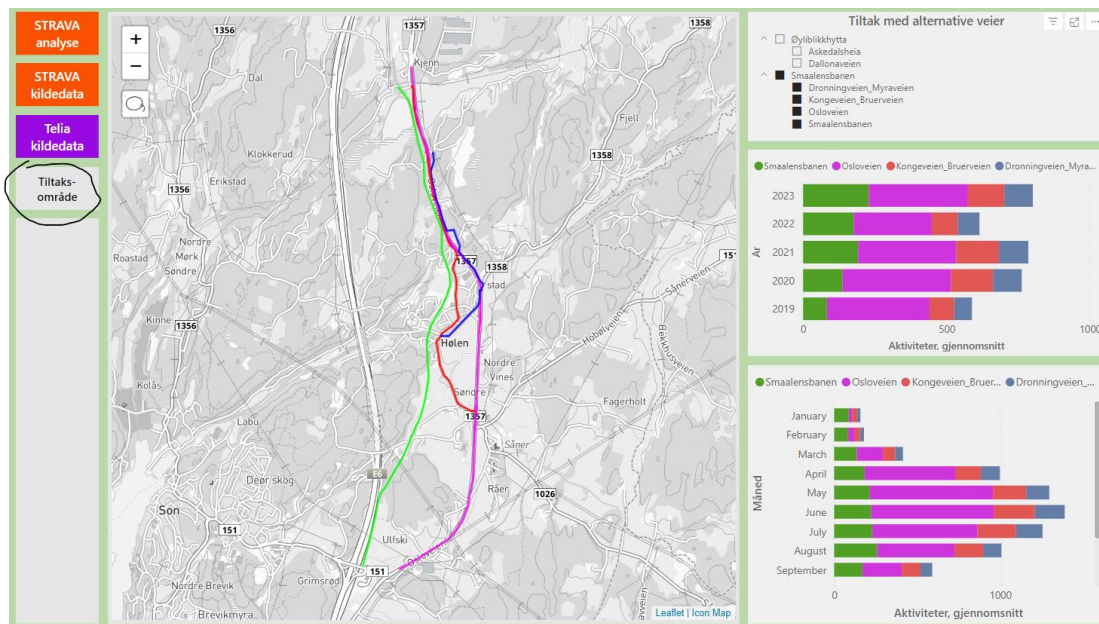
Begrensninger

Siden rutene er 500, 1000, 2000 eller 4000 meter i kvadrat, og noen ganger større hvis det trengs for å anonymisere brukerne grunnet lave besøkstall, så kan de ikke brukes for å si noe om forholdet mellom forskjellige stier eller veier. Det er også lite hensiktsmessig å bruke absolutt antall i en analyse, siden rutenes størrelsesforskjell vil gjøre en sammenligning vanskelig, hvis det i det hele tatt er besøk per flateenhet som er relevant. Det kan derimot fungere å sammenligne prosentuell fordeling mellom to eller flere ruter og dermed si noe om hvilken bruk et område har og hvor brukerne kommer fra.



5.1.4 Tiltaksområde

Denne siden i dashbordet (Figur 40) er ment mer som en oversikt over utvalgte tiltaksområder og mindre som analysegrunnlag. Der det finnes alternative veivalg mer eller mindre parallelt med tiltaket er disse skillett ut med egen farge og navn. Tallene som blir brukt i diagrammene er gjennomsnittet per edge for den aktuelle strekningen, siden det absolutte antallet vil være avhengig av hvor mange edges en slik strekning består av.



Figur 40: Viser siden Tiltaksområde i Power BI-dashbordet.

5.1.5 Skalerbarhet

Power BI-dashbordet inneholder i dag kun Kristiansand og Vestby kommuner, men i prinsippet kan man legge inn alle landets kommuner. For noen kommuner vil man måtte dele inn dataene i mindre enheter enn hele kommuner, siden Power BI har en begrensning på 30000 objekter til samtidig visning i kartet. Til sammenligning har STRAVA 9856 edges i Vestby og 34166 i Kristiansand. I denne første versjonen av dashbordet er derfor Kristiansand-data klippet ned til 30000 objekter med et snitt i nordenden av kommunen.

Workspacet i FME som blir brukt for innlesing av alle de forskjellige kildedataene vil enkelt skalere opp til å kjøre hele landet hvis det skulle bli aktuelt. Det er kun noen få tilpasninger som må gjøres for å tolke kommunenavn. I tillegg må tiltaksstrekninger med sine alternativer defineres i form av en liste med edge-id, for å gi disse edges informasjon om at de inngår i en slik strekning.



6. RESULTATER

6.1 Passeringsteller

Analyser av tellerdata fra Vestby viser en økning i bruken av Smaalensbanen, spesielt mellom sommeren 2020 og 2021. Selv om det er vanskelig å direkte knytte denne økningen til åpningen av Hølenviadukten, indikerer dataene en mulig indirekte effekt. Mangelen på data over lengre tid og gjennom ulike sesonger begrenser vår evne til å identifisere trender. Når det gjelder Kristiansand, gjør den begrensede måleperioden at det er utfordrende å trekke konkrete konklusjoner. Det er viktig å merke seg at analysen er begrenset av tilgjengelige data og at en mer kontinuerlig datainnsamling kunne ha gitt mer robuste resultater. Dette understreker en tidligere påvist utfordring: Bruk av data fra tellere kan være problematisk på grunn av stor variasjon i tilgjengeligheten av disse dataene.

6.2 STRAVA

STRAVA-data viser en tydelig økning i aktivitet på Smaalensbanen etter bruåpningen. I tillegg er det en omfordeling av aktiviteter fra parallelle veier til Smaalensbanen, spesielt blant gående og løpere. Analyser av STRAVA-data i Kristiansand indikerer en positiv effekt av åpningen av Dagsturhytta Øyliblakk på trafikken i området. Metoden for å beregne endringer i relativ trafikk synes å være effektiv for å identifisere trender over tid.

6.2.1 Sammenligning av STRAVA og Tellerdata

Sammenligningen viser en lineær trend mellom STRAVA-data og passeringsteller for de tre områdene vi har undersøkt: Sognsvann, Sølvdobla og Vestby, men med betydelig usikkerhet, og det observeres en markant variasjon mellom de ulike områdene. Dette indikerer at bruk av STRAVA-data alene ikke er tilstrekkelig for å gi en nøyaktig prediksjon av total trafikk, særlig i områder hvor tilgjengelige data er begrensede. Faktorer som varierende bruk av STRAVA og restvariasjon i dataene bidrar til usikkerheten i disse estimatene. Resultatene viser at det kan være nødvendig å ha flere modeller, tilpasset forskjellige områder og populasjoner av brukere. Der vi har hatt en teller stående en periode så kan en tilpasset modell brukes for noen tid etter til å anslå totalt antall passeringer ut fra STRAVA-data. Man bør gjøre en større undersøkelse for å anslå gyldighetsområde i tid og rom for en beregnet sammenheng mellom STRAVA og tellerdata.

6.2.2 Sammenligning av Telia og Tellerdata

Selv om korrelasjonen ikke er særlig sterk, indikerer resultatene at Telia-data kan speile observerte trender i tellerdata. Dette understreker potensialet for å bruke telekommunikasjonsdata som et verktøy for å forstå og analysere bevegelsesmønstre over større områder.

6.3 Studiens evne til å besvare spørsmålene i forespørselen

Hvor godt kan vi besvare de fem spørsmålene som ligger som grunnlag til studien?



1. Hvor stor aktivitet (antall personer) det er i området som undersøkes (case) og hvordan er fordelingen i løpet av døgn/ukedag/uke/måned/år?

STRAVA-data og Telia-data kan gi innsikt i aktivitetsnivået i de to analyseområdene. Til tross for at Telia-dataene er grove, tyder analysene på at det er mulig å benytte datasettet til å lage prediksjoner om observerte trender i tellerdata for aktiviteter. Dette gjelder selv om GRID-området som brukes i analysen dekker et areal på 4 km². For å kunne trekke mer konkrete konklusjoner, er det imidlertid nødvendig å utvide analysen til å omfatte flere områder. Et problem med bruk av STRAVA-dataene er at vi er nødt til å basere oss på månedlig oppløsning, ettersom det registreres for få daglige aktiviteter i de to områdene. Dette betyr at vi ikke har mulighet til å detaljere aktiviteten ned til dag-, ukedag- eller ukes nivå, og må derfor holde oss til måneds- og årsbasis. I områder med høyere aktivitetsnivå ville dette ikke utgjøre et problem.

Det er også verdt å merke seg at vi ikke kan fastslå det totale antallet unike personer, bare det totale antallet aktiviteter per linjesegment. Likevel er det mulig å sammenligne aktivitetsnivåene fra en periode til en annen, som gir oss innsikt i endringer i bruksmønstre over tid.

2. I hvilken grad det er mer aktivitet i området det er gjort friluftstiltak i enn før tiltak ble gjort. Dette kan være antall brukere, antall turer, lengden på turen eller hvor lenge folk oppholder seg i området.

Analysene av STRAVA-dataene har vist at det er en signifikant økning i bruken av de to områdene etter gjennomføringen av tiltaket. Det er imidlertid ikke mulig å fastslå det eksakte antallet brukere, men vi kan observere en generell økning i antall aktiviteter. Ingen av datakildene gir mulighet til å utlede lengden på turene eller hvor lenge folk har oppholdt seg i området. I STRAVA-dataene er hastigheten for aktiviteter oppgitt, men dette representerer et gjennomsnitt for STRAVA-brukere, noe som gir begrenset informasjon og bare er representativt for den del av befolkningen som allerede trener aktivt. Det er mulig i Telia-dataene å se hvor lenge en aktivitet har vært på en rute, men ettersom rutene er svært store (opp til 16 km²), gir også dette begrenset informasjon. Men metoden for å beregne endringer i relativ trafikk er effektiv for å identifisere trender over tid.

3. I hvilken grad friluftslivstiltaket samlet sett skaper mer aktivitet, eller flytter aktivitet fra et annet område. Hvor stort er influensområdet i casene?

Vi kan observere i analysene av STRAVA-dataene fra de to områdene at tiltaket ikke bare har skapt mer aktivitet, men også at aktiviteten har flyttet seg fra nærliggende stier til den aktuelle stien. I området rundt Øyliblakk kan vi også se at tiltaket ikke bare har økt aktiviteten på den spesifikke stien, men ser også ut til å ha hatt positiv effekt på området som helhet. Det er selvsagt vanskelig å konkludere definitivt om dette skyldes tiltaket eller andre faktorer. STRAVA-dataene for Smaalensbanen viser at aktiviteter har flyttet seg fra nærliggende veier, som er mindre egnet for gåing, løping og sykling, til Smaalensbanen. Vi kan imidlertid ikke fastslå hvor stort influensområdet er, da vi har begrenset oss til å bare se på noen utvalgte veier. Telia-dataene gir informasjon om influensområdet, men disse dataene blir lite representative for den aktuelle stien og sier mer om folks bevegelsesmønstre for større områder i kommunen.



- 4. Aldersintervall på personer som er i aktivitet i det aktuelle området, dersom informasjonen er tilgjengelig. Er det konkrete tiltak som tiltrekker seg nye brukergrupper (for eksempel i forhold til alder, kjønn, etc.)? Hvilken informasjon om brukerne er det mulig å fremskaffe innenfor rammene til personvernforordningen?**

Telia- og STRAVA-dataene inneholder informasjon om aldersintervall og kjønn. Utfordringen med STRAVA-dataene er at de ikke er representative for befolkningen. For Telia-dataene er denne informasjonen basert på statistikk fra SSB og ikke på faktiske abonnenter, i tillegg gjør rutestørrelsen det vanskelig å knytte informasjonen til en enkelt sti. Men man kan fortsatt observere fordelingen innenfor en rute. Analysene viser at antallet aktiviteter i kategorien "other" (i motsetning til de stillestående "home" og "work"), korresponderer med antallet personer som passerer en teller innenfor ruten. Aldersgruppene i Telia-dataene er basert på data fra SSB, det betyr det at vi kun får et overordnet estimat av befolkningens alderssammensetning, og ikke informasjon om de enkelte turgåernes alder.

- 5. Influensområde: Er det hovedsakelig de lokale som tar i bruk friluftsområdet? Avstand fra bolig til friluftsområder? Har aktiviteten flyttet seg fra et nærområde til et annet? Vi ønsker å kunne avdekke om folk flytter sin fysiske aktivitet fra et annet område og til det aktuelle området som det har vært gjort tiltak på, eller hvorvidt tiltaket bidrar til en reell økning i aktivitetsnivået for befolkningen lokalt og/eller regionalt.**

Telia-dataene avslører brukernes opprinnelsesruter, men er mindre representative for spesifikke stier og reflekterer heller bevegelsesmønstre i større områder. Selv om de ikke nøyaktig angir avstand fra bolig til friluftsområde, indikerer de om besøkende kommer fra forskjellige deler av kommunen eller utenfra. STRAVA-data er nyttige for å analysere aktivitetsflyt i mindre områder, men det er utfordrende å bestemme om endringer i fjerne områder påvirker besøkstall i det aktuelle området. Rundt Øyliblikk tyder tiltakene på økt aktivitet ikke bare på den aktuelle stien, men også en generell positiv effekt på området som helhet.

7. KONKLUSJON: ANALYSENE

Konklusjonen av analysene av STRAVA- og Telia-data med ulike statistiske modelleringsmetoder og korrelasjonstester, både individuelt og med tellerdata som en baseline og referansepunkt viser at vi kan få ut gode prediksjoner for å si noe om bruken av et område. Selv om tellerdataene har visse begrensninger på grunn av svakheter og støy, har de vist seg å være en verdifull fasit for modelleringen.

Resultatene fra analysene tyder på at det finnes gode muligheter for å utvikle indikatorer, fra STRAVA og Telia-data, til tross for iboende usikkerheter. Metoden for å beregne endringer i relativ trafikk har vist seg effektiv for å identifisere trender over tid. Data fra Telia i kategorien "other" innenfor en rute, og data fra STRAVA på en sti, korresponderer relativt godt med tellerdata. Dette indikerer at disse datakildene kan være representative for faktisk trafikk, gitt at de blir behandlet med passende modelleringsmetoder. Det er imidlertid viktig å påpeke at for å forbedre nøyaktigheten og påliteligheten i modellene, er det nødvendig å inkludere mer data og å utvide analysen til å omfatte flere geografiske områder. Vi ser også tendenser til å kunne estimere antall brukere fremover i tid, basert på dagens bruk



av et område. Med en teller som har stått i ro, kan vi etter en periode bygge en modell som, basert på data fra STRAVA, estimerer relativ endring i tidsserier fremover.

Influensområde kan man delvis trekke ut fra Telia-data. Men når det gjelder demografi, står vi overfor betydelige utfordringer. Det kan være nødvendig å søke etter alternative datakilder for å få innsikt i disse aspektene. Selv om spørreundersøkelser og lignende metoder kan tilby noe informasjon, antas det at de ikke vil være tilstrekkelige for bruk i større, mer komplekse modelleringsprosjekter.

Til slutt, POC-en viser at vi kan visualisere kildedataene på en pedagogisk måte. Det er utviklet et workspace i FME som gjør det relativt enkelt å skalere dataene til å dekke hele Norge, men det forutsetter en filtrering som begrenser visningen til maksimalt en kommune eller grunnkrets om gangen. En skalering forutsetter også at STRAVA lanserer et API, da det per nå er for tidkrevende å laste ned dataene manuelt, samt at det inngås en avtale som gir landsdekkende mobilnettverksdata.

De statistiske analysene inkluderer en del manuelle prosesser og de er derfor ikke inkludert i POC-en eller i FME workspace. Implementering av de manuelle prosessene i et brukervennlig grensesnitt vil kreve en utvikling som ikke er mulig å gjennomføre innenfor rammen for dette prosjekt.

8. VEIEN VIDERE OG ANBEFALINGER

De ulike statistiske modelleringsmetodene og korrelasjonstestene har vist seg å være lovende. Variasjon, omfang og svakheter i de ulike datasettene er primære kilder til ikke-konklusive resultater. For å ytterligere forsterke nøyaktigheten og påliteligheten av disse modellene, anbefaler vi en utvidelse av analysen til å inkludere et bredere spekter av geografiske områder og et større volum av data. Fremtidig arbeid bør konsentrere seg om utvikling av mer generelle modeller som kan integrere data fra både STRAVA og Telia for å oppnå en mer omfattende forståelse av trafikkmønstre.

Vi ser at tellerdataene er en verdifull fasit for modelleringen. For å kunne utnytte potensialene i ulike lokasjonsdata er vi avhengige av gode tellerdata med en bred geografisk utspreiding. Det er derfor essensielt at Miljødirektoratet, i samarbeid med fylkeskommuner, kommuner og andre relevante aktører som har tellere utplassert, samler eksisterende kunnskap og lager en manual for metode for utplassering, og en veileder for bruk av, tellere til ulike formål. Videre er det viktig å gjøre disse dataene mer tilgjengelige for forskning ved å etablere en nasjonal arena for å samle, harmonisere og dele tellerdata på nasjonal basis.

I vårt arbeid har vi observert at det er fordelaktig å ha sammenhengende data over minimum to år. Dette gjør at vi kan nøytralisere støy eller ujevnheter i dataene, slik som sesongvariasjoner. Med slike data kan vi utvikle robuste modeller som ikke bare kan anvendes for å estimere dagens bruk av et område, men også tidsserier fram og tilbake i tid og potensielt på tvers av lokasjon.

Det bør også etableres en nasjonal base for mer tradisjonelle stordata som intervjuer, slik at de blir tilgjengelige i et eventuelt videre arbeid. Vi er også avhenge av at STRAVA lanserer et API og at det bygges opp en infrastruktur som gjør at disse dataene enkelt kan anvendes i denne type arbeid. Det vil også være av verdi å lagre historiske data fra STRAVA for å kunne bruke i kobling for ikke å miste data ved oppdatering til ny versjon av OpenStreetMap.



Miljødirektoratet har mye å vinne ved å gå i dialog med de ledende mobiloperatørene – Telenor, Telia og ICE. For videre utvikling av modeller må det etableres en avtale som sikrer leveranse av nasjonale data, som kan tilgjengeliggjøres for bruk i ulike utviklingsprosjekter.

POC-en som er utviklet i dette prosjektet har potensial til å skaleres opp. For å realisere dette, er det imidlertid nødvendig at STRAVA lanserer et API, og at det inngås en avtale for tilgang til landsdekkende mobilnettverksdata. I en eventuell utvidelse av prosjektet, kan vi også integrere noen av modellene og de statistiske analysene. Det kan også være fordelaktig å vurdere alternative programvarer, som for eksempel GeoNode, for å se om de er bedre egnet for nasjonal skalering.

Vedlegg 1 - Leveranserapport fra Telia



Standard documentation

Telia Crowd Insights data statement

ACCOUNTABLE

We put our customers' integrity at the centre and all our services are created according to privacy by design principle. Our data is collected, stored, anonymized and aggregated according to GDPR and ePrivacy requirements.

TRANSPARENT

We are using data from our own mobile network end to end. We know the source of the data and we are experts at using and analysing it. This means that we have governance of the data from the moment it's generated until the final insights are delivered.

EXPLAINABLE

Our data is a statistically trustworthy ground for decision making. The data is continuously and consistently produced from the same data source and with the same processes. Telia has a large subscriber base in each country which helps make the insights highly representative for the population at large.

Methodology

Mobile network generates continuously signalling data which get processed in the Telia data lake for several types of normal business process, such as invoicing and network optimization. By using this same data Telia can produce anonymized and aggregated mobility insights through a GDPR compliant method. Crowd Insights does not source any new or additional data from the subscribers besides the data already being used for normal business purposes by Telia. However, the data is stripped down so that it only contains anonymized subscription identifiers, time stamps and mobile network cell identifiers. The method for producing mobility insights ensures no users can be identified by never including data which represent less than 5 individuals and re-anonymizing all users every 24 hour. Telia Crowd Insights also complies with local regulation on data protection.

From the time series of signalling data three types of mobility behaviour are identified: activities, trips and people flows. Activity means that the subscription is dwelling stationary, while trips and people flows are identified as the subscription moves. The subscription is always either stationary or in motion and the cell size of the cellular network affects how accurately this can be identified. In general, in urban areas already movement of a few

hundred meters can be identified, whereas in sparsely populated areas longer movement is required.

As a result, Telia Crowd Insights provides insights where the crowds spend time, what kind of trips they make and how they move on different routes. Figure 1 illustrates, how movement behaviour is identified from the mobile network data. As stated earlier, observations are always grouped into groups of at least five similar observations, so that the information provided by the service cannot identify an individual subscription or even small groups.

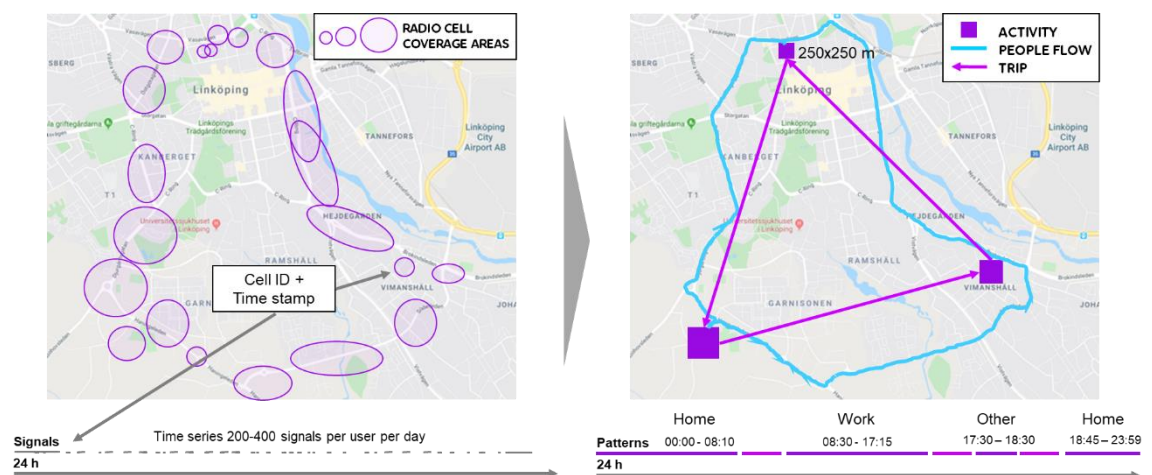


Figure 1. Identifying mobility patterns from mobile network data. Illustrative example – no individual insights can be exposed.

Comment:

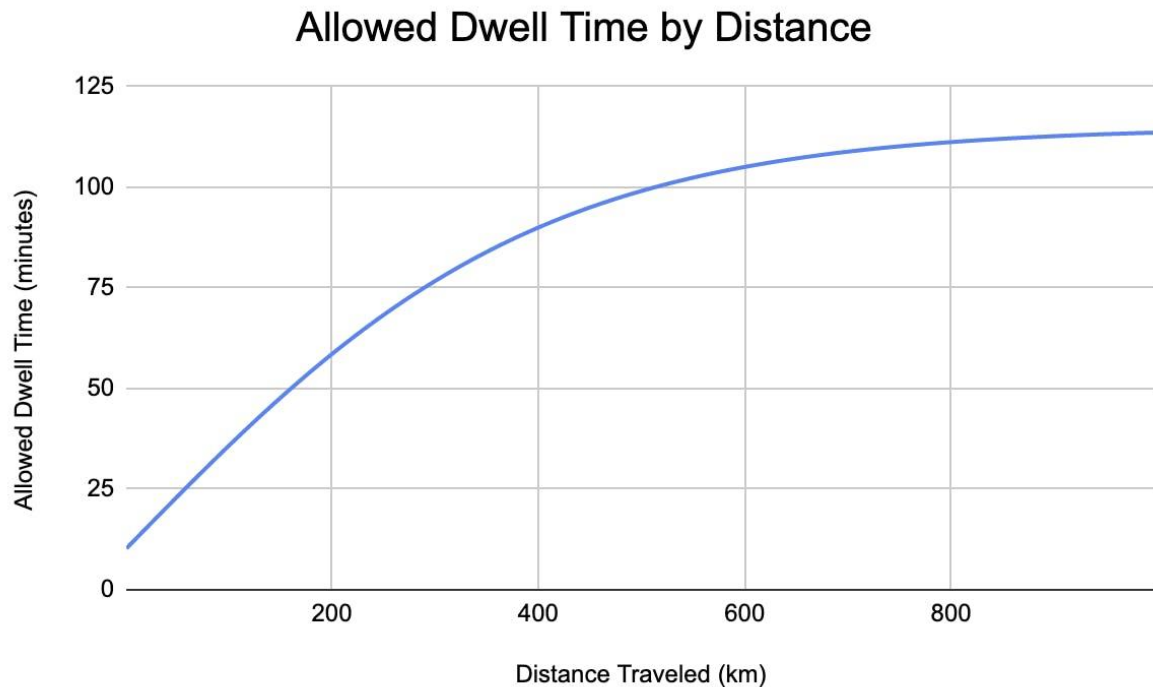
For Norkart data delivery the trips passing the area of interest (trails) has been used, as trail in Øylibikk for the most part is located in the area without infrastructure that could be connected to demographics data (like home or work areas), therefore making it impossible to count activity within given area attributable to Øylibikk only. To have datasets with similar logic applicable to both trails the trips approach is therefore in use.

Trip report parameters

Trip definition

Trips are defined with a break parameter and a direction change parameter. The break parameter means the minimum duration of stationary dwelling that breaks an ongoing trip and starts a new trip when the subscription starts to move again. The shortest break parameter is just 10 min of stationary signalling and that only applies to short trips of around

2 km. For long trips above 200 km stationary signalling can be tolerated up to 70 min. To complement the temporal break parameter there is also a directional change parameter which is good at detecting rounds trips. If the trip has a bearing change of more than 155 degrees between the waypoints furthest away from each other than the trip is broken into 2 trips.



Data quality

Even though Telia Crowd Insights has several strengths it also has certain weaknesses. For transparency, we have included the following information on the four most important factors for data consistency and resolution.

Consistency over time – People counts

The stability of the data from the sourcing systems has impact on the consistency of insights over time. The major aspects to consider are explained below.

Inconsistency in data volumes

Maintenance, updates, and bugs could cause certain parts of the country to see a drop-in number of signals and devices which could impact the ability to compare some periods with each other. These drops are however known and can be avoided. The signal quality is

reported on municipality level for each day based on the percentage drop from the long-term baseline calculated as the 6-week moving median.

Table 6 introduces the ratings (A-D, S).

Table 6. Signal quality classification

Class	Definition	Interpretation
A	less than -5 % drop	OK
B	-5 to -10 % drop	OK, minor outlier
C	-10 to -15 % drop	Use with caution
D	more than -15 % drop	Avoid
S	no data	Data is not available

Comment:

Bad rated days have been dropped from the data and calculations. Counts of days removed from data available in table below:

year	bad rated days
2019	8
2020	11
2021	11
2022	1
2023	1

Faulty metadata

There is a small percentage of signals and antennas that we discard due to faulty metadata. These are mostly assumed to be evenly distributed across the country but can in some cases impact certain areas more than other which can result in higher or lower counts of people over time.

Extrapolation methodology

The current extrapolation methodology is negatively impacted by people overnighing in other locations than where they are registered, such as summer houses or hotels. This has higher impact on areas where people go or visit during certain periods of the year. In practise this means that the seasonality effects are not fully reflected in absolute counts since the

extrapolation assumes that the night population should be corresponding to the census counts. The trend, however, captures the increase and decrease in visitors.

Comment:

As mentioned on the introduction meeting, the extrapolation is relying on the home area for device. Given current setup of 1 day hashing the home area is sensitive to overnight visitors travelling to areas other than their home. For Norkart case it means that any sudden increases or drops in visits from inner and outer areas could be result of travel within Norway, resulting in changes to home area devices. To account for it, total numbers from other areas (i.e. area E – remaining) should also be considered.

Consistency over time – Placement

The antenna infrastructure of mobile network is constantly being worked on, either by installing new antennas, changing old or faulty ones or upgrading to higher frequency bands such as 4G or 5G. These changes are likely to affect the placement of devices over time due to following reasons:

- New or removed antennas may change the coverage areas.
- In case of faulty antennas, devices use neighboring antennas until faulty antenna is fixed or replaced.

Telia Crowd Insights updates the coverage areas continuously on a weekly basis. Areas which have few radio cells such as rural or confined areas can experience larger errors in placement due to faulty cells or antennas.

Comment:

Important aspect affecting scalability of solutions with custom areas especially as small as trails, since such areas are potentially more affected by infrastructure changes. An improvement in antenna could add or remove counts seen within custom area due to different placement. This means that each custom area needs to be analyzed separately.

Resolution – Temporal

Accuracy

The temporal resolution is affected by the underlying signal density that devices have over a given period. This affects the accuracy of determining the start and end time and duration for activities and trips. Shorter time buckets will, therefore, have a negative impact on:

- The proportion of short dwells being captured for the given time bucket

- The proportion of start and end time that falls within the given bucket

Sample size

Smaller temporal resolution, such as, hourly insights will lead to a smaller sample of devices. This leads to higher uncertainty in the extrapolated numbers due to a smaller sample size. The same effect can impact late evenings and early morning when mobile phone usage is very low in the population and signals will be registered more sparsely. This will impact the size of the confidence interval. Small sample sizes also increase the privacy loss caused by the privacy preserving K5 filter.

Comment:

Given that scope of this research is to understand general trends, the data on daily level has been used and then aggregated to Monthly & Yearly level to retain as much observations as possible. However, even if sample is above K5 filter while remaining small, uncertainty should be considered and those observations taken with precaution.

Resolution – Spatial

Accuracy

The spatial resolution is affected by the underlying network of antennas and the sizes of the coverage areas that each antenna provides. Coverage tends to be larger in areas close to water and in rural areas, thus decreasing spatial accuracy in those environments.

Moreover, the placement of observations from the mobile network cells to real-world locations uses a logic where the observations are placed closer to mobile network base stations and in the areas with registered population and workplaces. For most use cases this logic generates more accurate insights since the assumptions fit the general behavioural patterns of the population. However, in some use cases, such as in large events organized in recreational areas, the placement logic leads to decreased spatial accuracy.

In general, the limited spatial accuracy creates problems when the results are reported with high spatial resolution and in lower resolutions in the borders of area division, such as in the borders of municipalities.

Comment:

Both trails are located near the other infrastructure like roads, railway networks and households. This inevitably results in noise in the data, as both the placement and movement seen within

network is only approximate and while on average is still quite accurate on grid level or larger, won't be as accurate on trails level.

Sample size

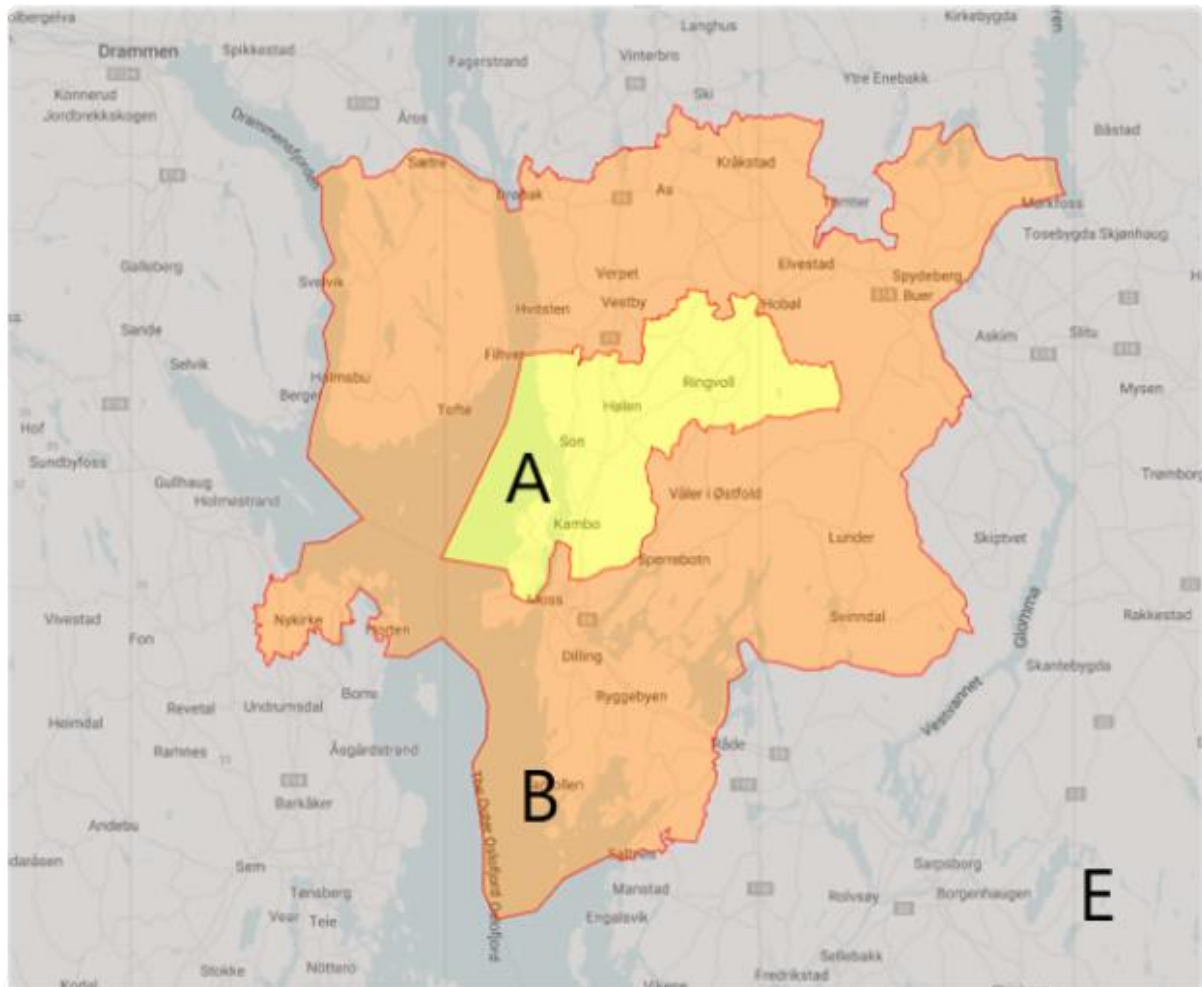
Smaller geographical areas, for example grids, lead to a smaller sample of devices. This leads to higher uncertainty in the extrapolated numbers resulting in larger confidence intervals.

The opposite is also true when looking at larger geographical areas, for example municipality level, or when looking over a longer time period, for example several days instead of just one day. Due to this effect we recommend avoiding placing too high trust in absolute counts for specific grids and specific time periods as they typically have large confidence intervals.

Specific documentation

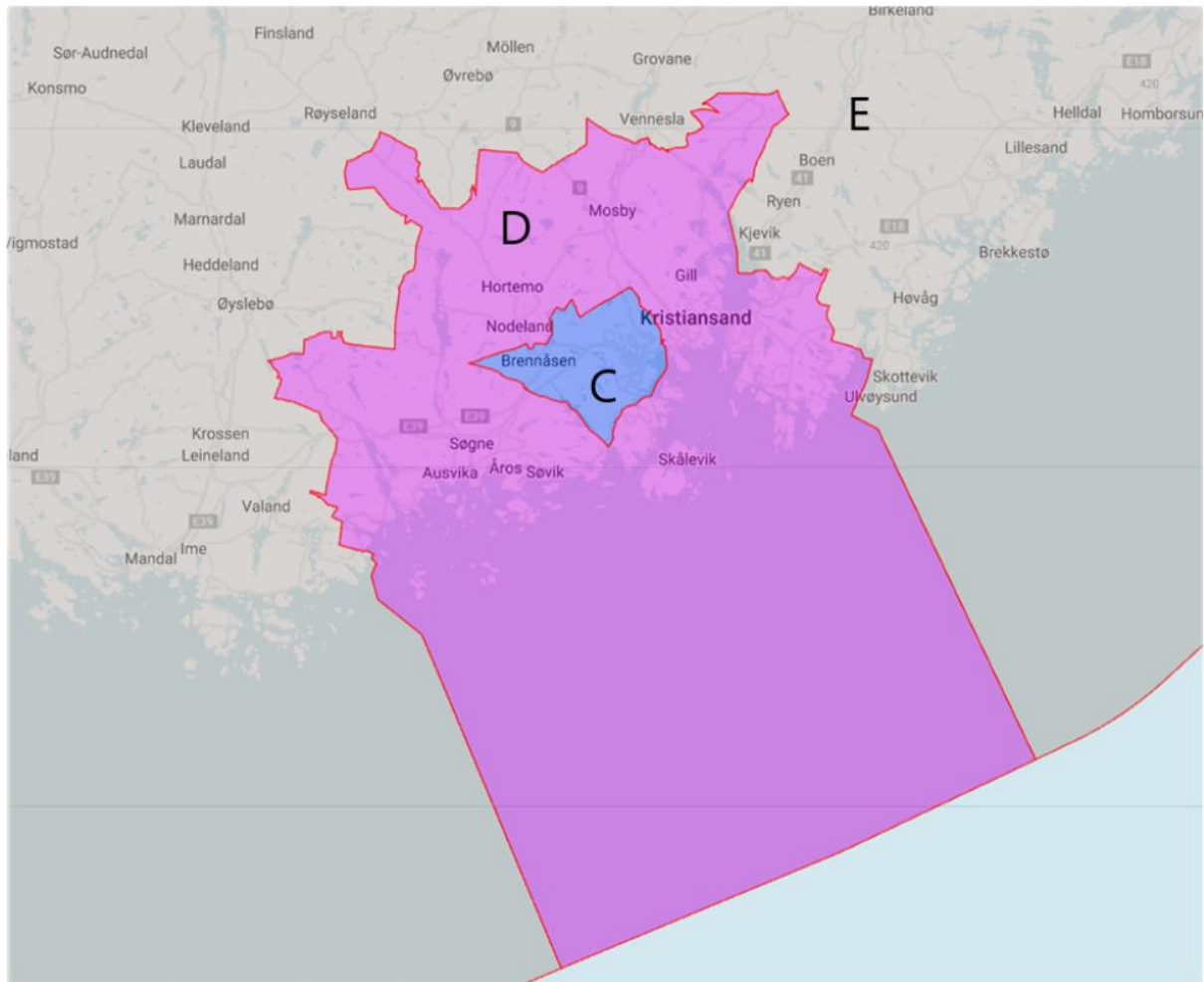
Custom areas

On the picture 1 the area around Smaalensbanen can be seen. The aggregation is done to inner area (A) and outer area (B), where multiple delområde areas are combined to negate the privacy filtration effect. Everything outside inner and outer area is defined as remaining parts of Norway.



Picture 1. Area around Track Smaalensbanen

Similar approach is used for trail Øyliblikk, where inner and outer grouped delområde areas are used. Both aggregations were used simultaneously, meaning that area definitions are shared between the trails of Smaalensbanen and Øyliblikk.



Picture 2. Area around Trail Øyliblikk

Breakdown of the area naming is as follows:

- A. Smaalensbanen_inner
- B. Smaalensbanen_outer
- C. Øyliblikk_inner
- D. Øyliblikk_outer
- E. remaining

Delomrade list shaping the areas mentioned above can be found with breakdown of unaggregated areas and provided separately in shapefile format. Overall, this approach provides future flexibility if aggregate areas are to be shrunk or expanded. The aggregation is done to overcome minimum requirement of 5 people to be seen within given geographical area/timeframe.

Catchment areas

To capture trips passing through trail but minimize the noise from other trips placed in same area, relatively small gates have been chosen. Contrary to small gates and potential loss of trips placed slightly away from them, the trail itself has been used as alternative catchment area. Finally, all gates aggregated together have been used as proxy for total unique trips captured by either of the gates.

The capture logic on example of Øyliblikk is working as follows:

Scenario A – Trip passing Gate 1 & 3 and trail itself.

Scenario B – Trip passing Gate 2 only.

Scenario C – Trip passing Trail only.

Gate	Scenaro A	Scenario B	Scenario C
Øyliblikk_gate_1	1	0	0
Øyliblikk_gate_2	0	1	0
Øyliblikk_gate_3	1	0	0
Øyliblikk_gate_total	1	1	0
Øyliblikk	1	0	1

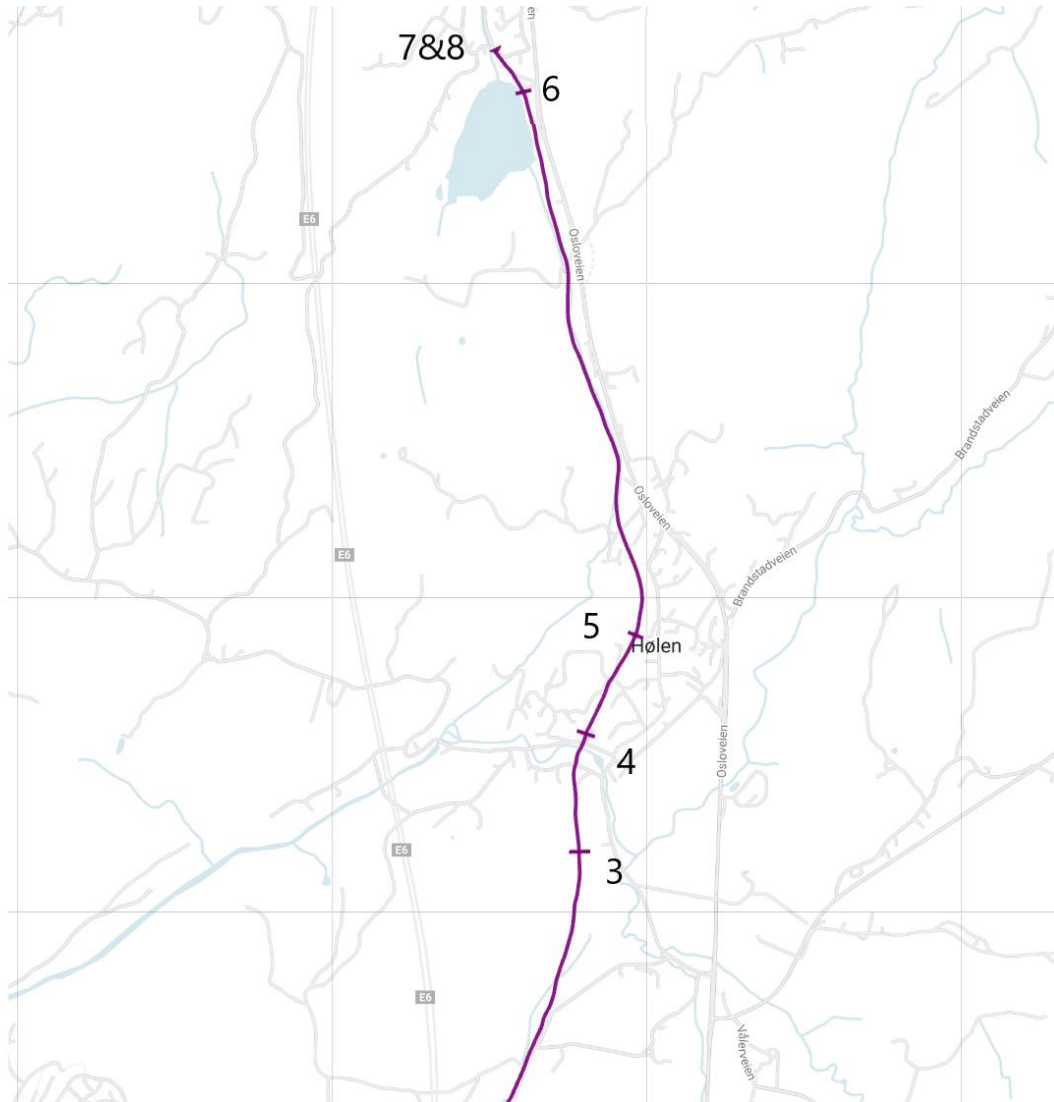
This way Gate total is proxy for Total Unique trips passing any of the gate, while individual Gate would work as proxy for any particular area on the trail. The Trail overall catchment area is alternative proxy for trips passing trail, however by its nature it will capture all the trips that go not only along but also across (and potentially in completely different direction), resulting in substantially larger numbers. The idea is that trips that are believed to be of “walking” type would remain somewhat relevant to people visiting or at least crossing the trail.

Catchment area overview

For Trail in Smaalensbanen 6 gates have been used, together with trail and aggregate for all gates, breakdown as follows:



Picture 3. Catchment area around Trail Smaalensbanen



Picture 4. Catchment area around Trail Smaalensbanen

For Trail in Smaalensbanen 6 gates have been used, together with trail and aggregate for all gates, breakdown as follows:

- 1 Smaalensbanen_gate_1
- 2 Smaalensbanen_gate_2
- 3 Smaalensbanen_gate_3
- 4 Smaalensbanen_gate_4
- 5 Smaalensbanen_gate_5
- 6 Smaalensbanen_gate_6
- 7 Smaalensbanen_gate_total
- 8 Smaalensbanen



Picture 5. Catchment area around Trail Øyliblikk

For Trail in Øyliblikk 3 gates have been used, together with trail and aggregate for all gates, breakdown as follows:

- 1 Øyliblikk_gate_1
- 2 Øyliblikk_gate_2
- 3 Øyliblikk_gate_3
- 4 Øyliblikk_gate_total
- 5 Øyliblikk

Custom & Catchment area mapping

In datasets provided, custom areas are used to showcase the Trips Origin, Destination and Home Area (home location where we believe people are from). While the catchment area is used to showcase the via area or which area was passed in the given Origin – Destination matrix.

Mode of transport

The mode of transport shows transportation method based on mode with longest distance within the trip. If the trips consists of multiple means of transportation, only the longest distance one is displayed. Mode of transport itself is assigned based on the distance travelled and speed during the trip.

Primary mode for this delivery to look at is Walking means of transportation or people who have walked via the given areas. Included are also Rail and Road transportation means, but for general

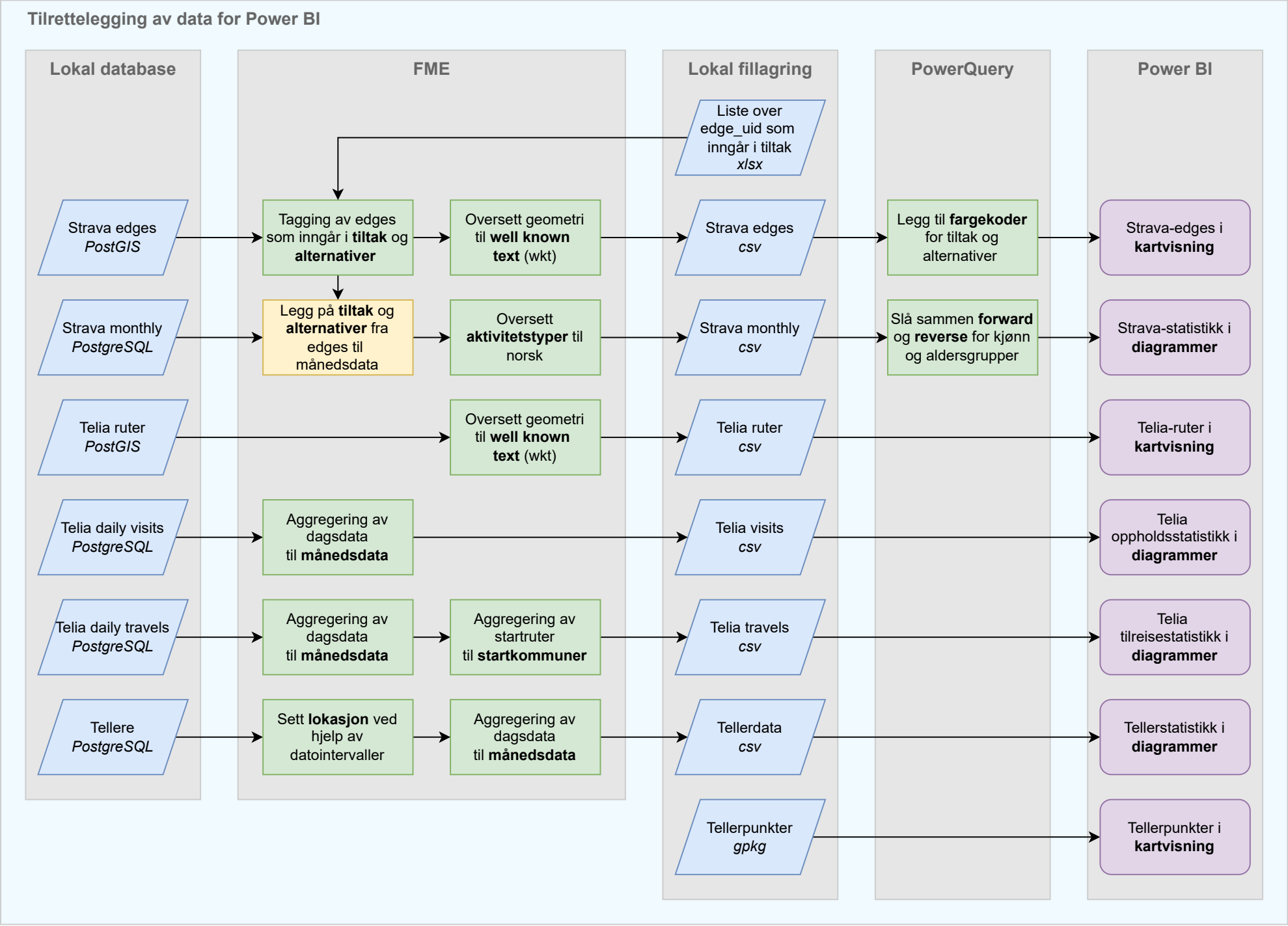
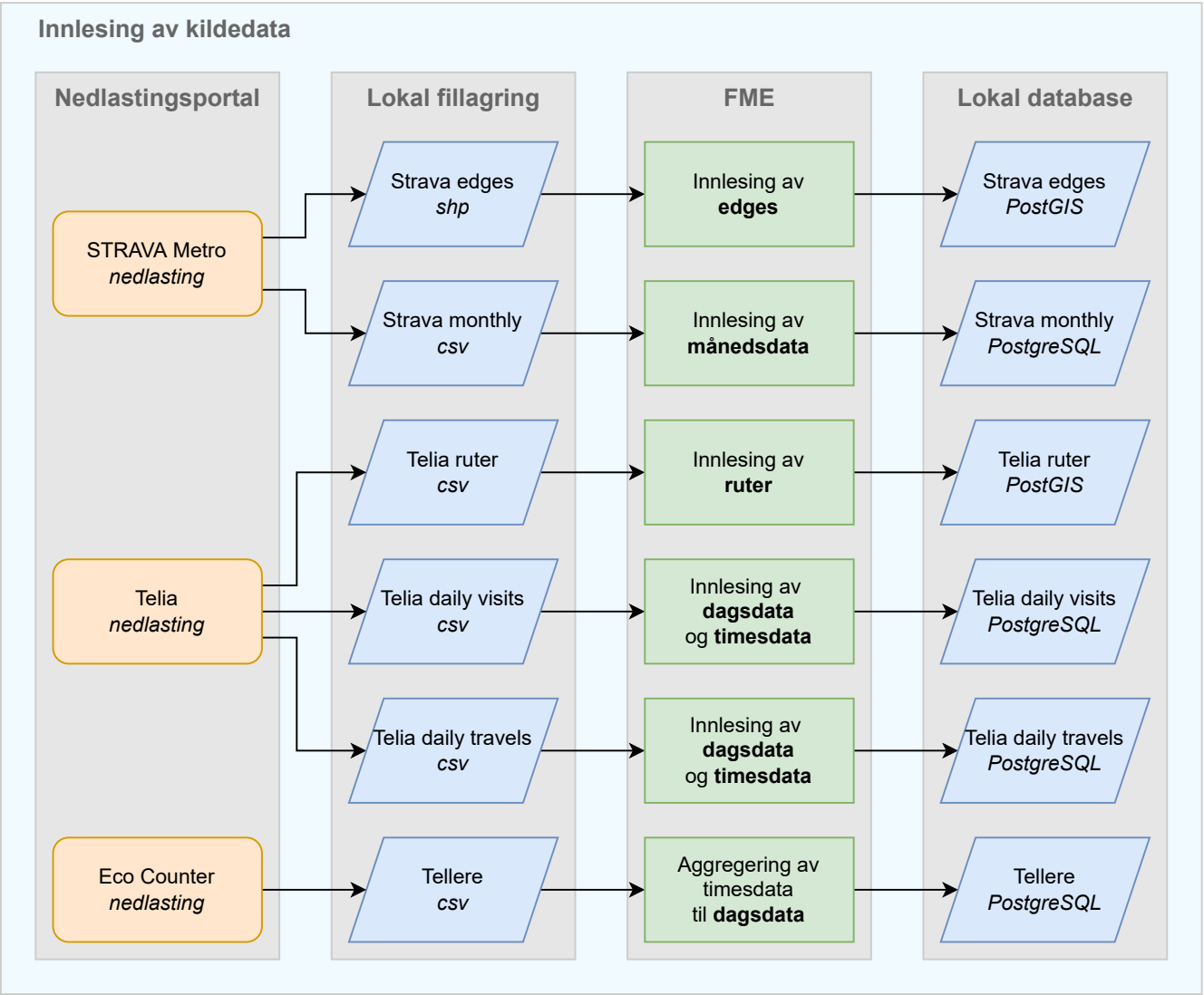
informational purposes. As the trail areas are located next or near to rail & road related infrastructure, and it is hard to separate those who came by i.e. car and walked via trail, from those who simply passed it via nearest highway.

Given the numbers for RAIL & ROAD it is clear that most of trips are indeed passers by and not people visiting the trails. UNKNOWN mode of transport is used when none of predefined modes of transport can be assigned to trip in particular – reasons lie in abnormal speeds or distances, high uncertainty or other factors making these unapplicable for given case. Again, included for informational purpose as they could include Walking type trips also.

Vedlegg 2 – flytskjema

NORKART

The logo for NORKART features the word "NORKART" in a bold, black, sans-serif font. A small green triangle is positioned below the letter "A".





NORKART



Sammen skaper vi smartere samfunn

Norkart er et norsk teknologiselskap, eid av [Ferd Capital AS](#) og de ansatte. Vi tilbyr markedsledende løsninger innen kommunalteknikk, plan og geodata, og øvrig kart- og eiendomsinformasjon til offentlig og privat sektor.

Telefon: +47 67 55 14 00
E-post: info@norkart.no
norkart.no