

## **2. semesters eksamensopagve 2025**

*Vejleder: Simon Bjerrum Eilersen*

### **Forfattere:**

*Line A. Adolph, Maria B. A. Hitz, Maria Cristiana Maxim, Martin E. Bindner og Abdikadir A. M. H.*

*Omar*

**Afleveringsdato: 9. maj 2025**

Antal tegn: 88.710 + Shiny App (80.320)

## Indholdsfortegnelse

# 1 Resumé

Business Viborg arbejder for at skabe optimale rammer for erhvervslivet i Viborg Kommune og har en ambition om at nå 700 medlemmer i 2025. Medlemsafgang truer imidlertid både organisationens økonomiske fundament og dens rolle som erhvervspolitisk talerør. For at imødekomme denne udfordring er der i dette projekt udviklet en datadrevet prototype, der forudsiger churn og identificerer centrale risikofaktorer på medlemsniveau.

Løsningen kombinerer brugervenlig formidling med avanceret maskinlæring i et R-baseret workflow. Seks modeller blev afprøvet, hvor Random Forest blev valgt som slutmodel på baggrund af gennemsigthed og forklaringskraft. Feature engineering inddrager bl.a. kontaktfrekvens, eventdeltagelse og modtaget erhvervshjælp.

Modellen er operationaliseret i et interaktivt dashboard, som understøtter medlemskonsulenternes op søgende arbejde. Projektet er udviklet med respekt for GDPR og dataetik og illustrerer, hvordan en lokal medlemsorganisation med begrænsede ressourcer kan anvende data strategisk og ansvarligt til at styrke fastholdelsen og engagementet blandt sine medlemmer.

# 2 Indledning

Business Viborg er en medlemsorganisation, der arbejder målrettet for at skabe optimale vilkår for erhvervslivet i Viborg Kommune. Med over 600 medlemsvirksomheder udgør organisationen en væsentlig aktør i det lokale erhvervsøkosystem – både som netværksfacilitator, vidensformidler og politisk interessevaretager. Ifølge chefkonsulent Michael Freundlich er ambitionen at nå 700 medlemmer og en omsætning på 2,9 mio. kr. i 2025.

Men når virksomheder forlader organisationen, reduceres ikke blot indtægtsgrundlaget – også Business Viborgs netværkskapital og politiske legitimitet svækkes. Derfor er det afgørende at få indsigt i, hvilke faktorer der øger risikoen for udmeldelse, og hvordan man kan arbejde proaktivt med medlemsfastholdelse.

I dette projekt udvikles der en datadrevet løsning, der kombinerer teknisk analyse med brugervenlig indsigt og som respekterer både juridiske og etiske rammer. Målet er at styrke medlemskonsulenternes beslutningsgrundlag og understøtte en mere effektiv og målrettet medlemspleje.

### **3 Problemstilling**

Business Viborg er registreret under branchekoden 26104793, og arbejder målrettet for at skabe optimale rammer for erhvervslivet i Viborg Kommune. Som en medlemsorganisation med over 600 virksomheder i ryggen, er relationerne til medlemskredsen helt afgørende, både for at dele viden, styrke netværk og skabe lokal vækst. I forbindelse med præsentationen af Business Viborg udtalte chefkonsulent Michael Freundlich: "Vores mål for 2025 er at nå 700 medlemmer og en omsætning på 2,9 mio. kr."

Men når virksomheder melder sig ud, mister Business Viborg ikke kun en indtægt, men også værdifulde forbindelser, politisk legitimitet og mulighed for at gøre en forskel for erhvervslivet i området. For at handle proaktivt ønsker Business Viborg at få bedre indsigt i, hvad der driver churn og hvem der er i risikozonen.

Derfor skal der udvikles et datadrevet værktøj, som kombinerer teknisk analyse med brugervenlig indsigt. Et værktøj, der gør det muligt for både medlemskonsulenter og ledelse at træffe kloge beslutninger og handle i tide med respekt for både dataetik og jura.

### **4 Problemformulering**

Hvordan kan Business Viborg analysere og anvende medlemsdata til at udvikle et beslutningsunderstøttende dashboard, der forudsiger churn og forklarer centrale risikofaktorer – baseret på relevante maskinlæringsmetoder og med inddragelse af etiske og juridiske overvejelser?

#### **4.1 Underspørgsmål**

##### **Eksplorativ analyse (EDA)**

Beskriv hvilke mønstre og karakteristika kendetegner de virksomheder, der forlader Business Viborg?

## **Modelvalg og performance**

Hvordan kan forskellige machine learning-modeller anvendes til at forudsige churn i Business Viborgs kontekst, og hvilke modeller er mest velegnede?

## **Datavisualisering**

Hvordan kan resultater og churn-indsigter formidles via et brugervenligt dashboard, som understøtter daglig opsøgende indsats for medlemskonsulenter og ledelse?

## **Etik og jura**

Hvilke juridiske krav (fx GDPR) og etiske overvejelser bør indgå i udviklingen og brugen af et churn-forudsigelsesværktøj baseret på medlemsdata?

# **5 Afgrænsning**

I udviklingen af en datadrevet churn-model for Business Viborg er det nødvendigt at foretage en række metodiske og praktiske afgrænsninger for at sikre projektets gennemførlighed og fokus. Følgende underafsnit præciserer, hvordan projektets omfang er afgrænset i forhold til teknologisk anvendelse, datagrundlag, modeller, systemintegration og juridiske vurderinger.

## **5.1 AI-chatbots og anvendelse af ChatGPT**

ChatGPT 4.0 har været anvendt som et understøttende værktøj i forbindelse med idéudvikling, sproglig formulering og grammatisk korrektur. Modellen har alene fungeret som et supplement i arbejdet med tekstbaserede opgaver og har ikke erstattet selvstændig analyse, faglig vurdering eller besvarelse af projektets problemformulering. Chatbotten er således ikke anvendt til at generere indhold i den analytiske eller metodiske del af projektet.

## 5.2 Datagrundlag

Projektet baserer sig udelukkende på det datasæt, der er stillet til rådighed af Business Viborg. Datasættet indeholder oplysninger om medlemskab, virksomhedsdemografi, branchetilknytning, kontaktaktivitet, eventdeltagelse samt ydet rådgivning. Alle data er pseudonymiserede og begrænset til et afgrænset tidsrum. Dette kan påvirke modellens generaliserbarhed over tid og dens evne til at indfange nyere tendenser i medlemsadfærd.

## 5.3 Modellens omfang og valg af algoritmer

Formålet med projektet er at udvikle en forklarlig og anvendelig prototype frem for en produktionsklar løsning. Der er derfor ikke foretaget omfattende hyperparameter-tuning for alle modeller. Seks modeller er testet – herunder Support Vector Machine, Random Forest og XGBoost – og performance er evalueret på baggrund af F1-score og AUC som de primære metrikker. Fokus har været på at finde en balance mellem prædiktiv nøjagtighed og forklaringskraft.

## 5.4 Systemintegration

Den udviklede løsning er implementeret som en webbaseret prototype i R og er ikke integreret med Business Viborgs interne systemer, såsom CRM- eller medlemsdatabaser. Modellen kan tilgås og anvendes lokalt gennem RStudio Cloud eller ved afvikling på en dedikeret server, men kræver manuel opdatering af data. Fremtidig integration og automatisering er oplagte skridt i en potentiel videreudvikling.

## 5.5 Juridiske og etiske vurderinger

Projektet indeholder en overordnet vurdering af de juridiske og etiske rammer med fokus på dataminimering, transparens og behandlingsgrundlag i henhold til GDPR. Der er ikke foretaget en fuld juridisk gennemgang, og tekniske løsninger som adgangsstyring, kryptering og samtykkehåndtering er ikke implementeret i prototypen. Disse aspekter betragtes som en integreret del af en eventuel implementeringsfase og bør afklares i samarbejde med relevante juridiske rådgivere og systemansvarlige.

## 6 Definitioner

I dette afsnit defineres centrale begreber og forkortelser anvendt gennem rapporten:

**Churn:** Når en virksomhed ophører med sit medlemskab i Business Viborg. I datasættet angives dette som en binær variabel, hvor 1 betyder churn og 0 betyder fortsat medlemskab.

**Churn-model:** En prædiktiv model, der estimerer sandsynligheden for, at en virksomhed cherner. Den er baseret på historiske medlemsdata og konstruerede forklaringsvariable.

**Feature Engineering:** Fremstilling af nye forklarende variable fra eksisterende data, som styrker modellens evne til at forudsige churn. Eksempler inkluderer medlemsanciennitet, kontaktaktivitet og deltagelse i arrangementer.

**MeetingLength:** Længden af det seneste dokumenterede møde med en virksomhed, målt i minutter. Bruges som indikator for relationens styrke.

**har\_haft\_kontakt:** En binær indikator for, om virksomheden har haft kontakt med Business Viborg (f.eks. møder, telefonopkald eller rådgivning).

**deltaget\_i\_event:** Binær variabel der angiver, om virksomheden har deltaget i mindst ét event i analyseperioden.

**hjælp\_kategori:** En kategorisk variabel, der angiver typen af erhvervsfaglig støtte virksomheden har modtaget. Kategorierne er fx Strategi Udvikling, Organisation og Ledelse, Jura og Struktur m.fl.

**medlem\_antal\_år:** Antal år virksomheden har været medlem, beregnet som forskellen mellem analysedato og oprettelsesdato.

**Machine Learning (ML):** En metode til at bygge modeller, der kan lære mønstre i data og forudsige fremtidige hændelser. I projektet er ML anvendt til churn-forudsigelse.

**Random Forest:** En ML-algoritme, der kombinerer mange beslutningstræer for at skabe en robust og forklarlig model. Valgt som slutmodel i projektet.

**ROC AUC:** Et mål for modellens evne til at adskille churnere og ikke-churnere. En værdi tæt på 1 indikerer høj prædiktiv nøjagtighed.

**F\_meas (F1-score):** Et samlet præstationsmål, som balancerer præcision og recall – særligt velegnet ved skæve datasæt.

**Dashboard:** Et interaktivt visualiseringsværktøj, der præsenterer churn-risici og medlemsindsigter på en overskuelig måde til brug i den daglige medlemspleje.

**GDPR:** EU's databeskyttelsesforordning. Projektet tager højde for centrale principper som dataminimering, transparens og legitimt behandlingsgrundlag.

## 7 Analyse

### 7.1 Dataforståelse og fordeling

Analysen tager udgangspunkt i et datasæt bestående af 2.966 medlemsvirksomheder tilknyttet Business Viborg, som har et selvstændigt P-nummer. Datasættet afspejler en betydelig variation med hensyn til virksomhedsstørrelse, branchetilhørsforhold og interaktionsniveau med organisationen.

En indledende fordeling afslører, at virksomheder uden dokumenteret kontakt eller deltagelse i arrangementer har markant højere churn-rate. Denne observation antyder, at fraværet af kontakt og engagement kan være centrale indikatorer for medlemsophør.

### 7.2 Egenskaber ved virksomheder med høj churn

For bedre at forstå hvordan forskellige former for engagement påvirker medlemsstatus, har vi kombineret to centrale variabler: kontakt med Business Viborg og deltagelse i events. Dette giver fire grupper, som varierer i deres relation til organisationen. Figuren nedenfor viser tydeligt, hvordan virksomheder med både kontakt og eventdeltagelse i langt højere grad fastholdes som medlemmer, mens fravær af begge faktorer er stærkt forbundet med churn.



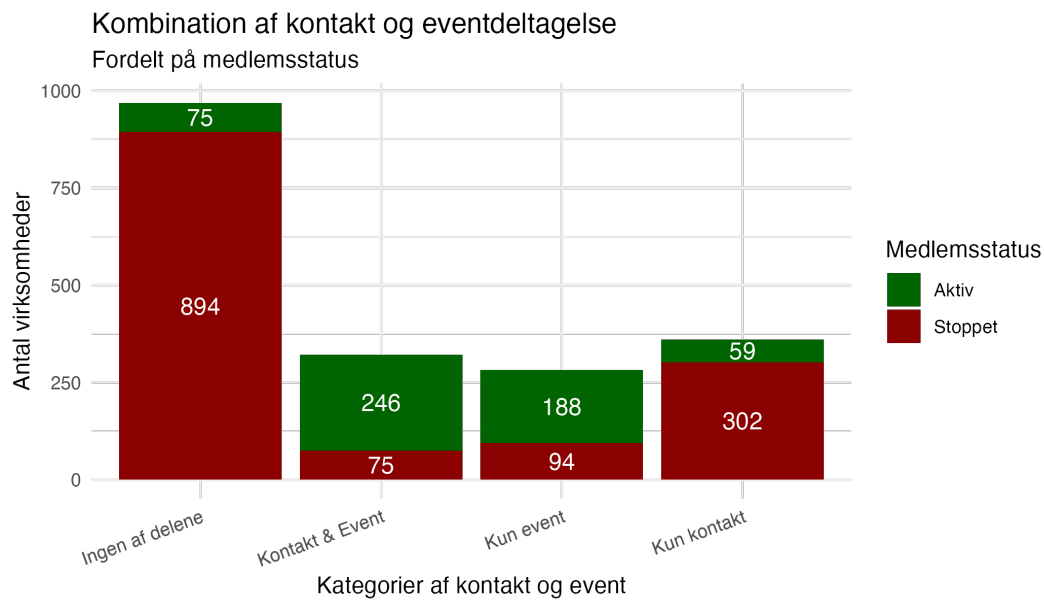


Figure 1: Fravær af begge faktorer er tæt forbundet med udmeldelse, mens dobbelt engagement viser stærk fastholdelse.

Virksomheder med begrænset kontakt til organisationen, lav deltagelse i arrangementer og uden dokumenteret interaktion har generelt en markant højere risiko for at opsige deres medlemskab. Dette underbygges af figuren nedenfor, der viser de fem postnumre med den højeste gennemsnitlige churn-risiko. Her ses, at geografiske områder med lav tilknytning til det centrale område (8800 Viborg) udviser særlig høj churn-sandsynlighed.

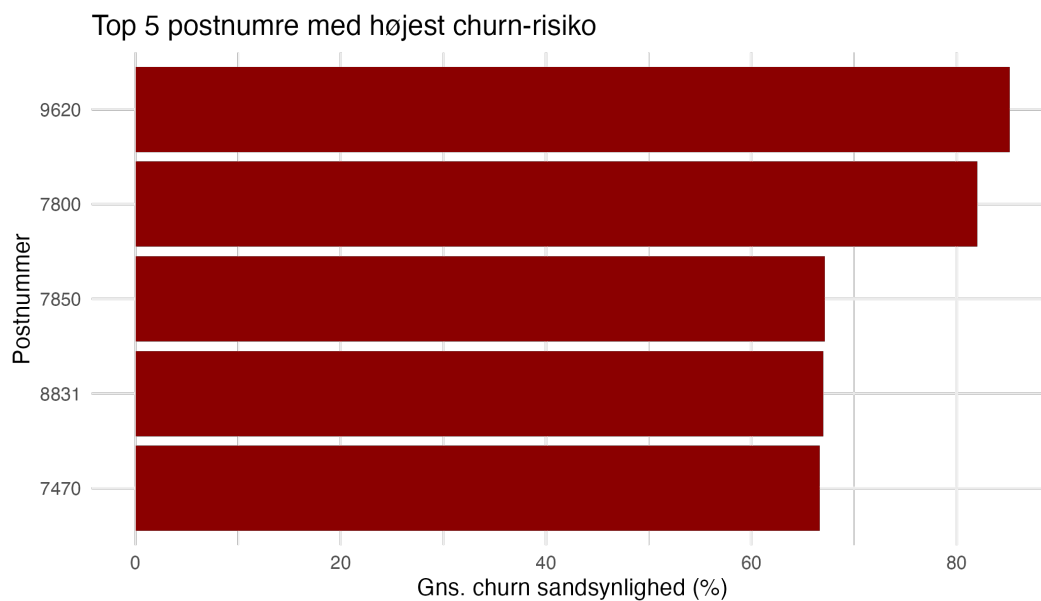


Figure 2: Geografisk afstand fra Viborgs centrum ser ud til at spille en rolle i udmeldelsestendens. 9620 Aalestrup, 7800 Skive, 7850 Stoholm, 7470 Karup, 8831 Løgstrup

Også på brancheniveau er der tydelige forskelle. Nogle brancher er kendetegnet ved begrænset interaktion med organisationen og har derfor en højere sandsynlighed for medlemsophør. Figuren herunder visualiserer de fem brancher med den højeste gennemsnitlige churn-risiko, hvilket bekræfter tendensen.

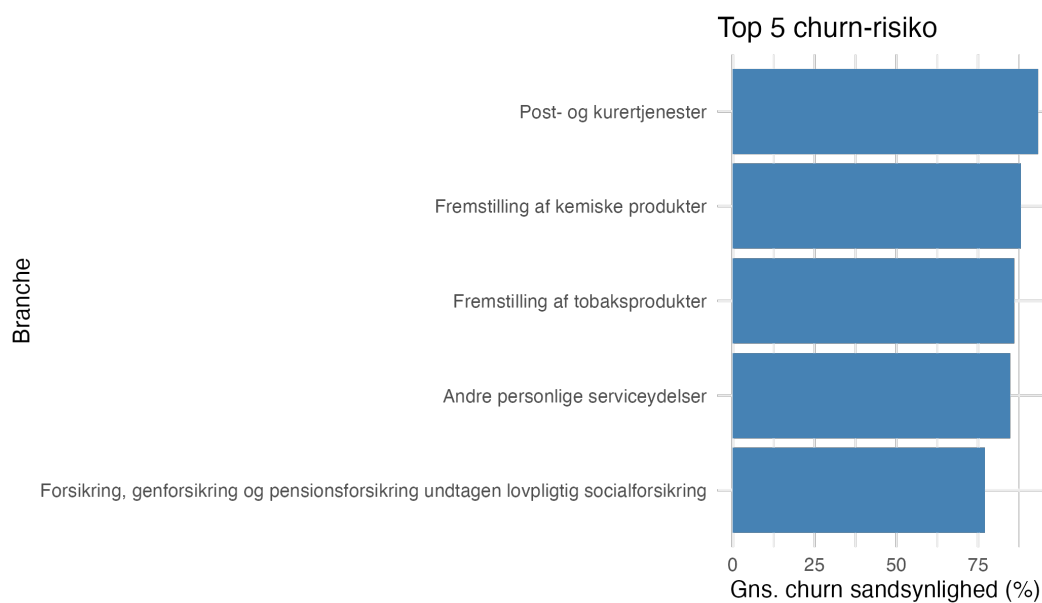


Figure 3: Brancher med lav netværksværdi og specialiserede ydelser udviser generelt højere risiko for medlemsophør.

Omvendt findes der brancher, hvor medlemmerne i langt højere grad fastholdes. Disse brancher har ofte en mere stabil tilknytning og deltager aktivt i organisationens tilbud. Den følgende visualisering viser de fem brancher, hvor medlemmerne i størst omfang forbliver tilknyttet – hvilket indikerer, at der her eksisterer en stærkere relation og et større udbytte af medlemskabet.

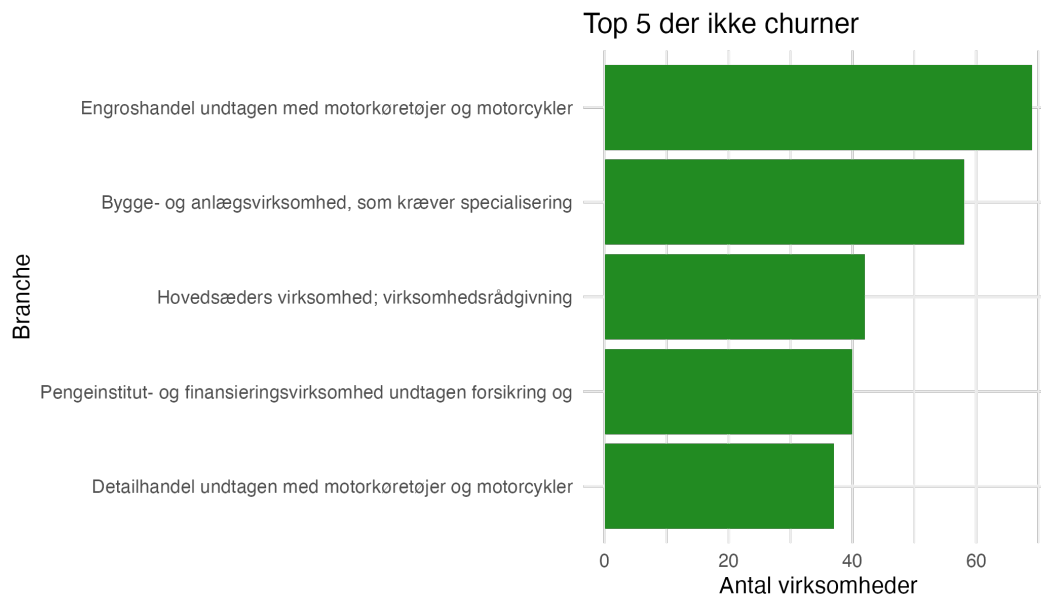


Figure 4: Branchenes høje fastholdelse kan skyldes stærkere relationer og oplevet værdi af netværket.

I de fem brancher hvor churnrisikoen er lille, er antallet af virksomheder højt. Dette kunne understøtte teorien om at netværkswærdi har høj prioritet, blandt dem der er medlemmer af Business Viborg.

### 7.3 Feature engineering

På baggrund af ovenstående mønstre blev der konstrueret nye forklarende variabler for at styrke modellernes prædiktioner. De mest centrale inkluderer:

- medlem\_antal\_år: længden af medlemskab målt i år
- har\_haft\_kontakt: binær indikator for, om der har været nogen form for kontakt
- deltaget\_i\_event: binær indikator for eventdeltagelse

Disse variabler blev udledt på baggrund af domæneviden og eksplorativ analyse, og bidrog væsentligt til forbedret modelperformance.

## 7.4 Modelperformance

Seks machine learning modeller blev afprøvet: Support Vector Machine (SVM), XGBoost, Random Forest, logistisk regression, K-nearest Neighbors (KNN) og Naive Bayes.

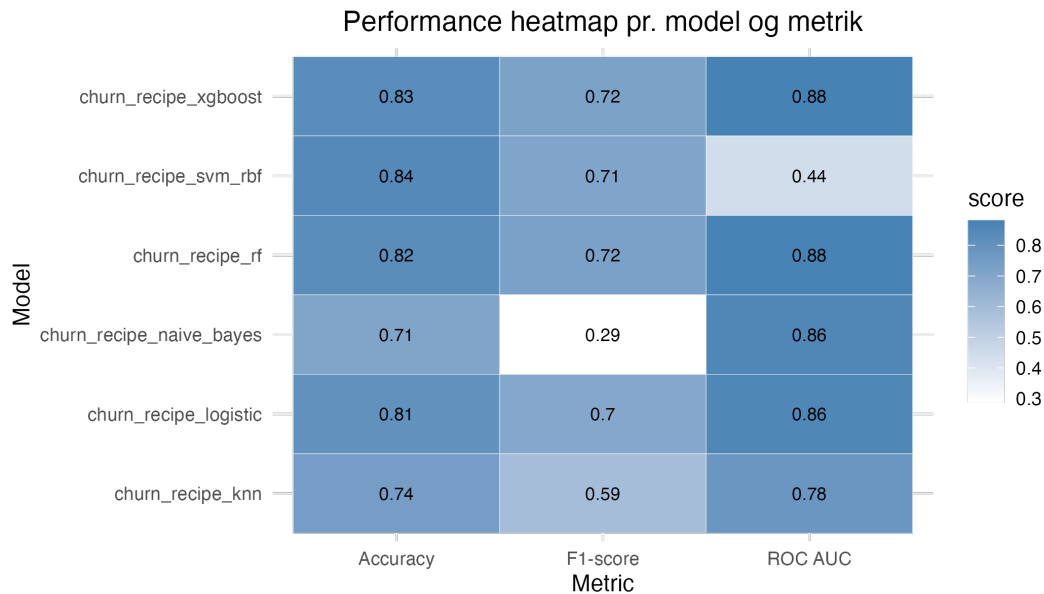


Figure 5: Farveintensitet viser performance – mørkere felter angiver højere score. Random Forest og XGBoost scorer generelt højt, mens Naive Bayes udviser lav F1-score.

På trods af XGBoost gode resultater blev Random Forest valgt som slutmodel. Dette skyldes modellens kombination af prædiktiv styrke og modelgennemsigtighed, hvilket gør den mere anvendelig i en praktisk kontekst. XGBoost blev fravalgt grundet behovet for yderligere parameteroptimering, som ikke var formålstjenligt inden for projektets rammer. Vi valgte den mest simple af de to modeller, da deres metrikker lå meget tæt.

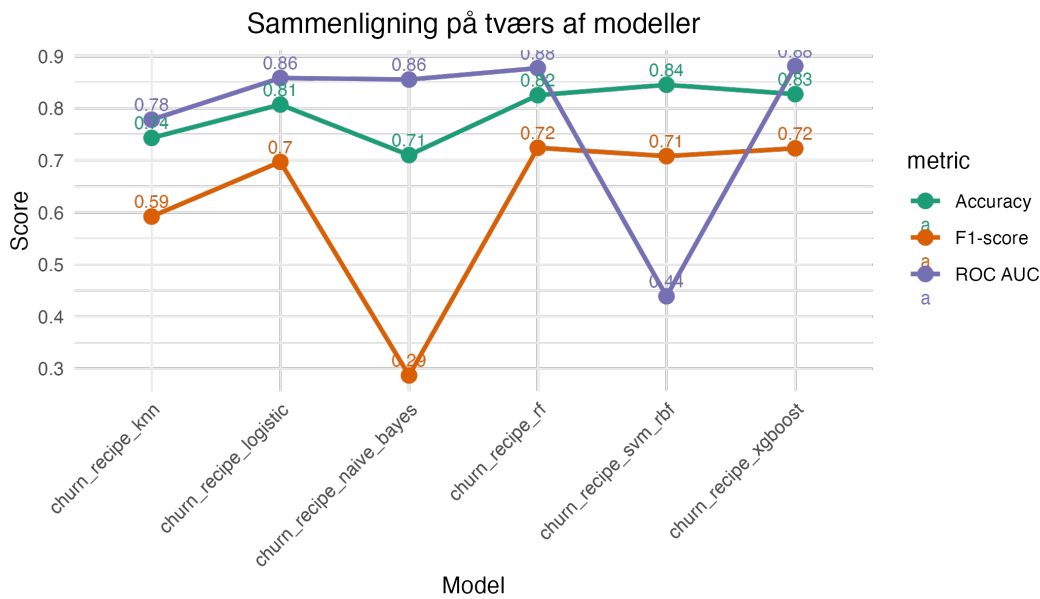


Figure 6: Lineplot over modelperformance. Random Forest og XGBoost opnår høj score på alle tre metrikker, hvilket underbygger deres styrke som robuste og præcise modeller.

## 7.5 Perspektiv: Modellering af churn over tid

Da vores datasæt ikke indeholdt tidsstempler for, hvornår virksomheder præcist meldte sig ud, har vi arbejdet med churn som en binær klassifikation. Hvis disse oplysninger forelå, ville det være oplagt at anvende time-to-event-modeller som fx Cox proportional hazards-model. Disse modeller gør det muligt at estimere, ikke blot om churn indtræffer, men også hvornår – hvilket kan styrke den opsøgende indsats og skabe bedre timing i medlemspleje. Det vil samtidig muliggøre en dynamisk forståelse af churn-risiko over tid og dermed forbedre prioriteringen af indsatser.

## 7.6 Variable importance og indsigt

Ved hjælp af vip()-pakken blev de mest betydningsfulde variable i Random Forest-modellen identificeret. De fem vigtigste prædiktorer var:

- deltaget\_i\_event\_Ja
- deltaget\_i\_event\_Nej
- MeetinLenght

- Employees
- medlem\_antal\_år

Disse variable udgør tilsammen et stærkt grundlag for at forstå churn-mekanismer i Business Viborgs medlemsbase og bekræfter den eksplorative analyses fund.

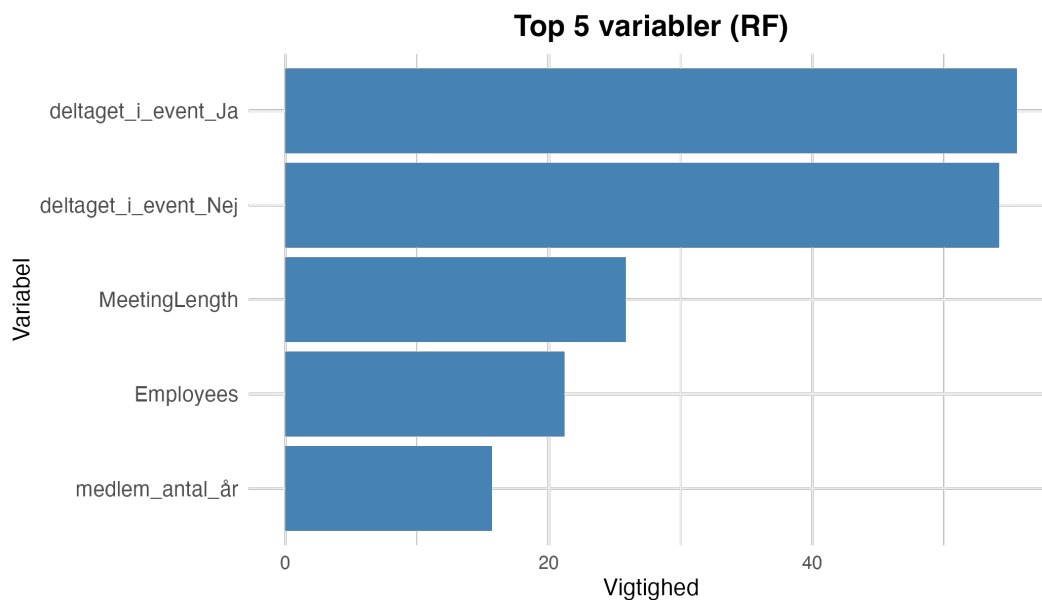


Figure 7: Eventdeltagelse, mødelængde, antal ansatte og medlem\_antal\_år er blandt de mest betydningsfulde prædiktorer for churn.

## 7.7 Fordeling af churn-risiko

For at operationalisere modellen i medlemsarbejdet er alle aktive virksomheder blevet klassificeret i fire risikokategorier baseret på deres sandsynlighed for churn. Som det fremgår af figuren nedenfor, har hovedparten af medlemsbasen en lav eller minimal risiko for udmeldelse, mens en mindre andel er vurderet til at være i moderat eller høj risiko. Denne fordeling giver medlemskonsulenterne et konkret udgangspunkt for at prioritere deres indsats og målrette dialogen mod virksomheder i risikozonen.

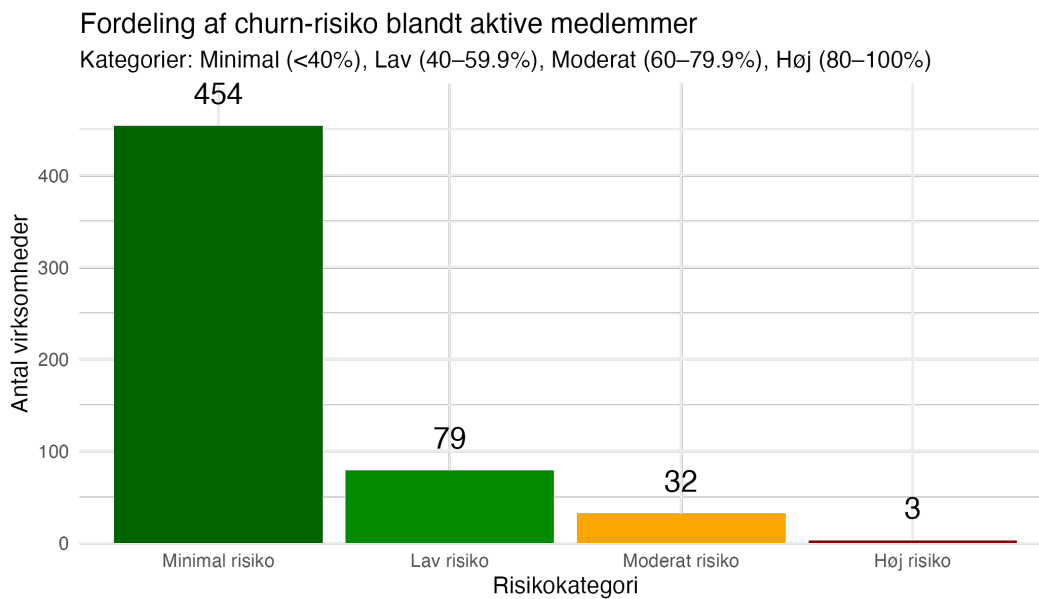


Figure 8: Fordeling af churn-risiko blandt aktive medlemmer baseret på modelprediktioner. Kategorierne “høj risiko” og “moderat risiko” rummer 38 virksomheder. De udgør et vigtigt proaktivt fokuspunkt.

## 8 Anvendelse af dashboard i medlemsarbejdet

Som led i samarbejdet med Business Viborg har vi udviklet et interaktivt dashboard, der omsætter churn-modellens output til konkret medlemsindsigt. Dashboardet er målrettet konsulenterne og ledelsen og gør det muligt at identificere medlemmer med høj risiko for at opsiges deres medlemskab. Visualiseringen er designet til at passe ind i konsulenternes og ledelsens arbejdsgange og skabe et klart beslutningsgrundlag i hverdagen.

### Adgangsuplysninger:

- **Link:** <https://maria-cristiana-maxim.shinyapps.io/churnapp/>
- **Brugernavn:** admin@businessviborg.com
- **Adgangskode:** Test1234



*Dashboardet er en prototype og bør anvendes til demonstrationsformål. Ved implementering i praksis anbefales det at justere adgangskontrol, datasikkerhed og systemintegration.*



Figure 9: QR-kode til churn-dashboard

## 9 Juridiske og etiske overvejelser

### 9.1 Juridiske og etiske forhold i churn-projektet for Business Viborg

I forbindelse med udviklingen af en datadrevet løsning til forudsigelse af medlems-churn hos Business Viborg er det essentielt, at både juridiske og etiske hensyn bliver nøje overvejet og integreret i hele udviklingsprocessen. Den teknologiske og analytiske dimension af projektet må ikke stå alene, men skal suppleres af en bevidsthed om datasikkerhed, individets rettigheder og ansvarlig brug af data. Da løsningen tager udgangspunkt i oplysninger om organisationens medlemmer, der potentielt kan knyttes til identificerbare personer, er databehandlingen omfattet af EU's Databeskyttelsesforordning (GDPR). Dette gælder, uanset om data fremstår pseudonymiserede for os som analytikere, da Business Viborg internt kan koble data til konkrete virksomheder eller kontaktpersoner. Formålet med dette afsnit er derfor at belyse de centrale juridiske forpligtelser samt de etiske retningslinjer, som Business Viborg bør overholde i forbindelse med udviklingen og implementeringen af churn-modellen.

### 9.2 Behandlingsgrundlag

Et fundamentalt krav i GDPR er, at al behandling af personoplysninger skal have et lovligt behandlingsgrundlag. Business Viborgs databehandling relaterer sig til almindelige personoplysninger såsom virksomhedsnavne, kontaktoplysninger, mødedeltagelse og interaktionshistorik. Det er derfor nødvendigt at vurdere, hvilket hjemmelsgrundlag der gør det lovligt at anvende disse data i en churn-model. Ifølge artikel 6 i GDPR kan behandlingen enten baseres på samtykke fra medlemmerne (art. 6, stk. 1, litra a) eller på organisationens legitime interesser (art. 6, stk. 1, litra f). I denne kontekst vurderes det, at Business

Viborg primært bør benytte den legitime interesse som behandlingsgrundlag, da organisationens formål – at forbedre medlemsservice, understøtte fastholdelse og sikre et bæredygtigt erhvervsfællesskab – er både sagligt, proportionalt og foreneligt med medlemmernes forventninger. Det er dog vigtigt at bemærke, at hvis data senere anvendes til andre formål, f.eks. målrettet markedsføring eller automatiseret profilering, kan det være nødvendigt at genoverveje behandlingsgrundlaget, herunder at indhente eksplicit samtykke.

### **9.3 Personoplysninger og dataminimering**

Analyserne tager udgangspunkt i almindelige personoplysninger, som ikke i sig selv er følsomme, men som i kombination med andre oplysninger stadig udgør personoplysninger i GDPR-forstand. Eksempler kan være data om deltagelse i arrangementer, mødeaktivitet eller medlemskabets varighed. For at overholde principperne om dataminimering og formålsbegrænsning (jf. GDPR art. 5), skal Business Viborg sikre, at der kun behandles data, som er nødvendige og relevante i forhold til det definerede formål – nemlig at kunne identificere virksomheder i risiko for udmeldelse. Alle unødvendige oplysninger bør fjernes eller anonymiseres, og det samlede datasæt skal reduceres til det minimum, der kræves for at modellere churn med høj præcision. Derudover skal der være dokumentation for, hvordan og hvorfor de valgte variabler indgår i modellen. Denne dokumentation skal kunne anvendes til intern kontrol og overfor Datatilsynet, hvis der føres tilsyn.

### **9.4 Oplysningspligt og transparens**

Business Viborg har i henhold til GDPR artikel 13 og 14 en klar oplysningspligt over for de registrerede medlemmer. Det betyder, at medlemmerne skal informeres om, at deres data indgår i analyser, hvad formålet med analysen er, hvilke rettigheder de har, og hvordan de kan kontakte organisationen for spørgsmål eller indsigelser. Oplysningen bør være let tilgængelig og formuleret i et sprog, der er forståeligt for ikke-specialister. Ideelt set bør informationen formidles både via organisationens privatlivspolitik og i forbindelse med medlemskommunikation – f.eks. i velkomstmateriale eller nyhedsbreve. Transparens er i denne sammenhæng ikke blot et juridisk krav, men også et middel til at styrke medlemmernes tillid til Business Viborgs databrug.

## 9.5 Pseudonymisering og identifikationsrisici

Selvom datasættet, som analysen baseres på, er pseudonymiseret for dataanalytikerne, betyder det ikke, at oplysningerne er anonyme i GDPR-forstand. Business Viborgs medarbejdere har adgang til nøgleoplysninger såsom medlemsnummer, virksomhedsnavn eller kontaktpersoner, som gør det muligt at identificere de registrerede. Dette understreger, at databehandlingen fortsat er omfattet af alle GDPR-krav. Der skal derfor træffes passende forholdsregler for at sikre, at oplysningerne ikke anvendes til andre formål uden nyt behandlingsgrundlag og ikke utilsigtet afslører følsom information om specifikke virksomheder. Datasikkerhed og organisatorisk ansvar Business Viborg er som dataansvarlig forpligtet til at beskytte personoplysninger mod uautoriseret adgang, tab, ændring eller misbrug. I henhold til artikel 24 og 32 i GDPR skal organisationen iværksætte både tekniske og organisatoriske sikkerhedsforanstaltninger. Det inkluderer brug af adgangsstyring, kryptering, logning af datatilgange, opdaterede sikkerhedspolitikker og uddannelse af personale i datasikkerhed. Derudover skal der foreligge databehandleraftaler, hvis eksterne samarbejdspartnere inddrages i projektet. Manglende sikkerhed kan ikke alene få juridiske konsekvenser, men også underminere den tillid, som projektet skal baseres på.

## 9.6 Etiske overvejelser og ansvarlig dataanvendelse

Udover den juridiske ramme bør Business Viborg også forholde sig aktivt til dataetik og samfundsansvar. Ifølge anbefalinger fra bl.a. Dataetisk Råd handler ansvarlig dataanvendelse om at sætte mennesket i centrum, sikre gennemsigtighed og undgå skævvridning eller diskrimination. Anvendelsen af en churn-model må ikke resultere i, at bestemte virksomheder eller grupper automatisk vurderes som mindre værdifulde baseret på statistiske mønstre, som ikke er sagligt begrundede. Der skal derfor tages højde for fairness og risikoen for bias – både i datasættets sammensætning og i de variable, der anvendes i modellen. Transparens er også afgørende her: medarbejdere skal kunne forstå og forklare modellens logik og resultater. En “black box”-model, som ikke kan forklares, kan føre til uforståelige eller urimelige beslutninger, hvilket ikke er foreneligt med en ansvarlig og tillidsvækkende medlemsorganisation. Ledelsen i Business Viborg har desuden et særligt ansvar for at sikre, at modellen anvendes i overensstemmelse med organisationens værdier. Det anbefales derfor, at der formuleres en dataetik-politik, som dækker ansvar, kontrol og kommunikation i relation til anvendelsen af churn-modellen. Politikken bør revideres løbende i takt med, at modellen og databrug udvikles.

## 9.7 Må Business Viborg beholde data på udmeldte medlemmer

Ifølge GDPR artikel 5, stk. 1, litra e, må personoplysninger ikke opbevares længere end nødvendigt til det formål, de er indsamlet til. Når et medlem melder sig ud, er det normale formål – f.eks. kommunikation, arrangementer eller medlemsservice – ikke længere aktuelt.

Men der kan være lovlige grunde til at opbevare dem i en periode Business Viborg må godt gemme data i en vis periode efter udmeldelse, hvis de har et sagligt og dokumenteret formål, f.eks.:

- Bogføring og regnskab – fx fakturaer, betalinger osv. må gerne opbevares i op til 5 år (jf. Bogføringsloven).
- Statistisk analyse – men kun hvis data anonymiseres eller pseudonymiseres, så det ikke længere er muligt at identificere personen uden ekstra oplysninger.
- Retligt forsvar – hvis der er risiko for tvister eller krav, kan man argumentere for at opbevare data i en periode, typisk op til forældelsesfristen (som ofte er 3 år i civile sager).

De må ikke fortsætte med at bruge tidligere medlemmers data til markedsføring og ikke opbevare data “bare for en sikkerheds skyld” uden formål og dokumentation.

CVR-numre tilknyttet enkeltmandsvirksomheder betragtes som personoplysninger, da de kan føres direkte tilbage til en fysisk person – virksomhedens ejer. Selvom disse oplysninger identificerer en person, klassificeres de ikke som følsomme oplysninger i henhold til GDPR’s artikel 9, da de ikke afslører forhold som helbred, politisk overbevisning eller religiøs tilhørsforhold. CVR-oplysninger er desuden offentligt tilgængelige og indgår derfor under almindelige personoplysninger, som stadig skal behandles i overensstemmelse med GDPR’s generelle principper.

## 9.8 Delonklusion

Sammenfattende vurderes det, at Business Viborgs churn-projekt kan gennemføres i overensstemmelse med GDPR og etiske retningslinjer, forudsat at visse betingelser overholdes. Der skal foreligge et klart og dokumenteret behandlingsgrundlag, medlemmerne skal informeres tydeligt og rettidigt, og datasættet skal reduceres til relevante oplysninger i henhold til formålet. Desuden skal organisationen sikre

passende tekniske og organisatoriske sikkerhedsforanstaltninger og være opmærksom på, at data fortsat er personhenførbare, selvom de fremstår pseudonymiserede. Afslutningsvis bør Business Viborg betragte dette projekt som et skridt mod mere datadrevet medlemservice – men også som en mulighed for at styrke sin rolle som en ansvarlig og gennemsigtig aktør i det lokale erhvervsliv. Det forudsætter, at juridiske forpligtelser og dataetiske principper ikke ses som forhindringer, men som en integreret del af en moderne og tillidsbaseret organisation.

## 10 anbefaling

På baggrund af analysen anbefales det, at Business Viborg anvender den udviklede churn-model som et beslutningsunderstøttende værktøj i det opsøgende medlemsarbejde. Modellen kan identificere virksomheder med høj risiko for udmeldelse og derved muliggøre en mere målrettet og proaktiv indsats fra medlemskonsulenterne. Det foreslås, at den udviklede Shiny app indgår som fast element i konsulenternes arbejdsrutiner og prioritering af medlemmer.

- Medlemspleje prioriteres over for virksomheder uden kontakt, eventdeltagelse eller modtaget hjælp, da disse faktorer er stærkt associeret med churn. Særligt events er relevante for at bibeholde medlemmerne. De har allerede mange events, men det kunne anbefales at lave flere skræddersyede events, så man sænker risikoen for churn på samtlige medlemmer.
- Løbende opdatering af modellen sikres ved at integrere churn-værktøjet i Business Viborgs CRM eller medlemsdatabase, så nye data automatisk indgår i fremtidige analyser.
- Etisk og transparent kommunikation om brugen af data indgår i medlemsdialogen for at styrke tilliden og sikre overholdelse af GDPR.

## 11 Konklusion

Projektet har vist, hvordan Business Viborg med et afgrænset datagrundlag og begrænsede ressourcer kan anvende machine learning som et konkret værktøj til medlemsfastholdelse. Gennem analyse af 2.966 medlemsvirksomheder og systematisk afprøvning af seks modeller blev der udviklet en forklarlig Random

Forest-model med stærke prædiktive egenskaber ( $F1 = 0,79$ ,  $AUC = 0,93$ ). Modellen er operationaliseret i et interaktivt dashboard, som giver medlemskonsulenterne og ledelse et datadrevet grundlag for at prioritere og målrette deres opsøgende indsats.

Resultaterne peger entydigt på, at fravær af kontakt, manglende eventdeltagelse og manglende erhvervshjælp er tæt knyttet til udmeldelse. Samtidig viser analysen, at en langvarig relation og relationel dybde – målt gennem fx mødelængde – hænger tæt sammen med fastholdelse. Denne indsigt skaber en direkte kobling mellem data og daglig praksis i medlemsarbejdet. Den største faktor er deltagelse i events og der kan med fordel udvikles skræddersyede events, der skaber værdi og dermed fastholdelse af flere medlemmer.

Modellen er udviklet med udgangspunkt i GDPR og principper om ansvarlig dataanvendelse. Det understreger, at datadrevne løsninger godt kan gå hånd i hånd med transparens, etik og tillid. Der er taget højde for både behandlingsgrundlag, oplysningspligt og identifikationsrisici – og løsningen kan derfor fungere som en ansvarlig prototype, der med de rette tekniske og organisatoriske rammer kan bringes i anvendelse.

Et naturligt næste skridt er at inddrage tidsstempler for udmeldelser. Det vil gøre det muligt ikke kun at vurdere risikoen for churn, men også at estimere, hvornår en udmeldelse sandsynligvis vil finde sted. Dermed får Business Viborg et endnu bedre grundlag for at igangsætte målrettede og rettidige indsatser over for medlemmer i risiko.

Samlet set besvarer projektet sin problemformulering ved at kombinere teknisk analyse, brugervenlig formidling og etisk ansvarlighed. Det peger samtidig frem mod en bredere anvendelse af datadrevne beslutningsværktøjer i medlemsorganisationer, der ønsker at arbejde strategisk og relationsbaseret – uden at give køb på tillid, menneskelig dømmekraft og lokal forankring.

## 12 Litteraturliste

### AI

OpenAI. (2025). ChatGPT (version 4.0). Hentet fra <https://chatgpt.com/>

### Bøger

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). R for Data Science (2. udg.). O'Reilly Media. Hentet fra <https://r4ds.hadley.nz/>

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (2. udg.). Springer. Hentet fra <https://ggplot2-book.org/>

Kuhn, M., & Silge, J. (2022). Tidy Modeling with R. O'Reilly Media. Hentet fra <https://www.tmwr.org/>

### WWW-dokumenter

Europa-Parlamentet og Rådet. (2016). Forordning (EU) 2016/679 – Generel forordning om databeskyttelse (GDPR). Hentet fra <https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=CELEX%3A32016R0679>

Erhvervsstyrelsen. (2025). Vejledning til bogføringsloven. Hentet fra <https://erhvervsstyrelsen.dk/vejledning-bogfoeringsloven>

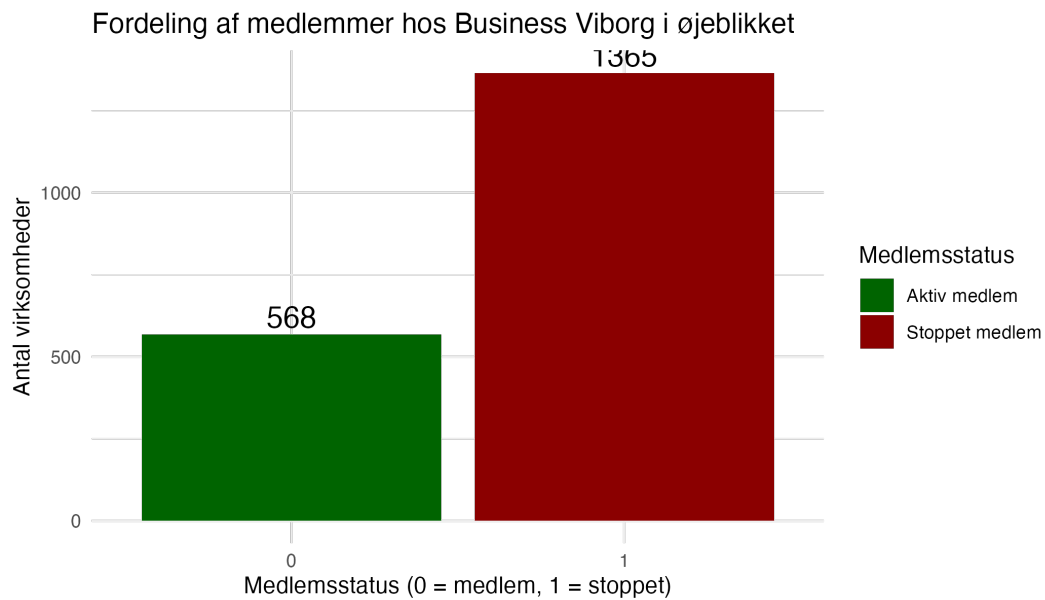
Dataetisk Råd. (2025). Officielt websted for Dataetisk Råd. Hentet fra <https://dataetiskraad.dk/>

## 13 Bilagsoversigt

### 13.1 Bilag 1–6: Eksplorativ dataanalyse

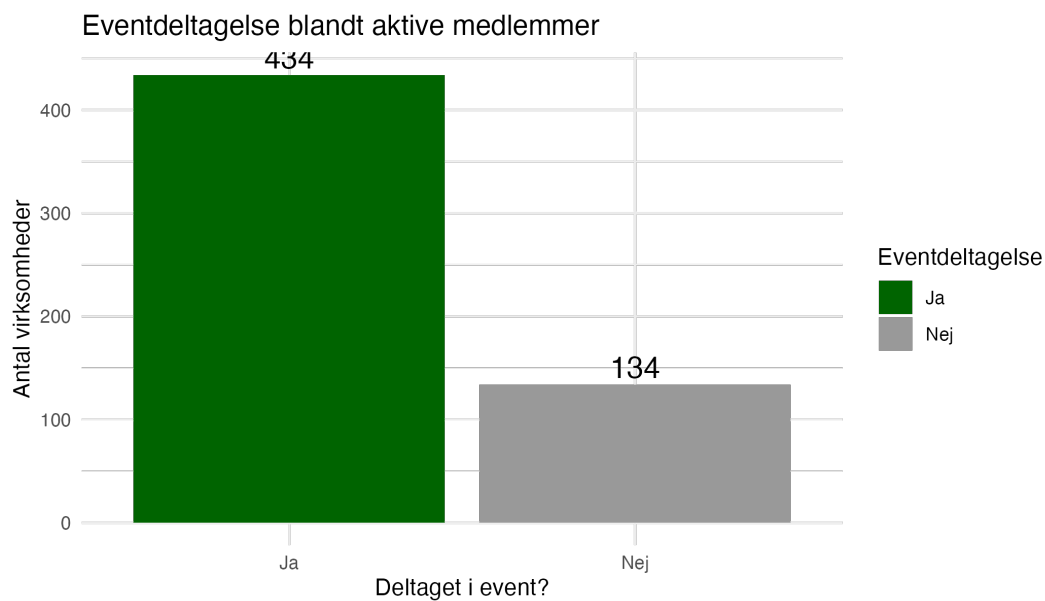
Anvendt i forbindelse med udarbejdelse og forståelse af datasættet:

- **Bilag 1:** Fordeling af medlemmer hos Business Viborg

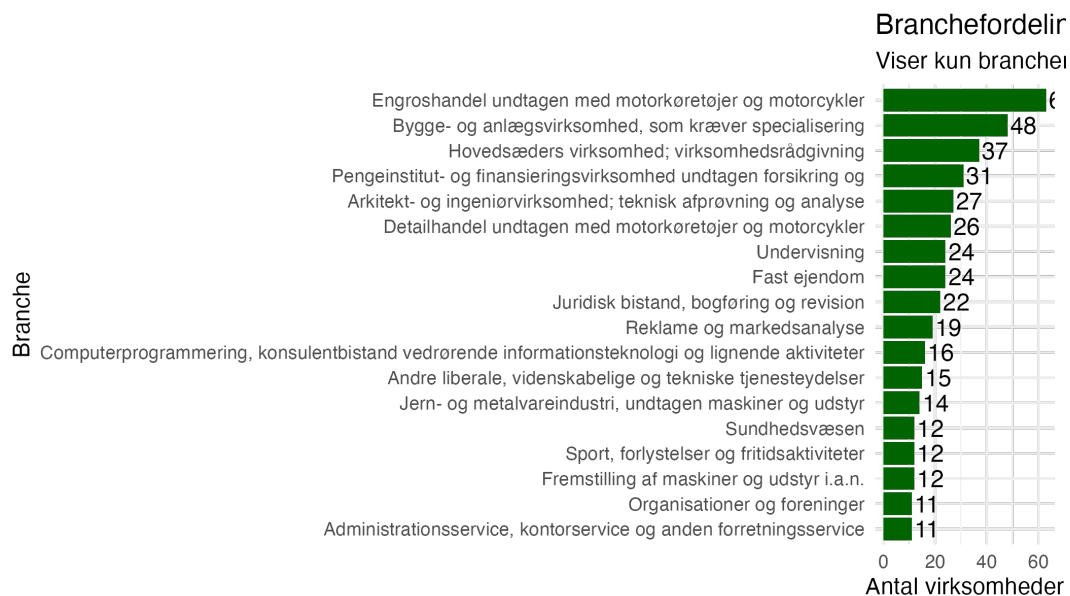


- **Bilag 2:** Eventdeltagelse blandt aktive medlemmer

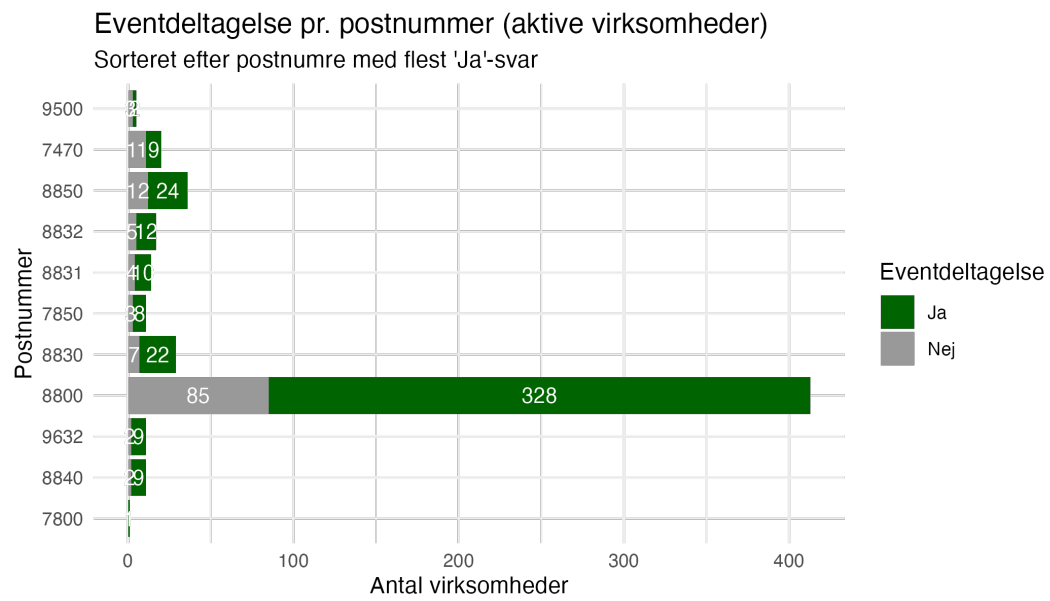




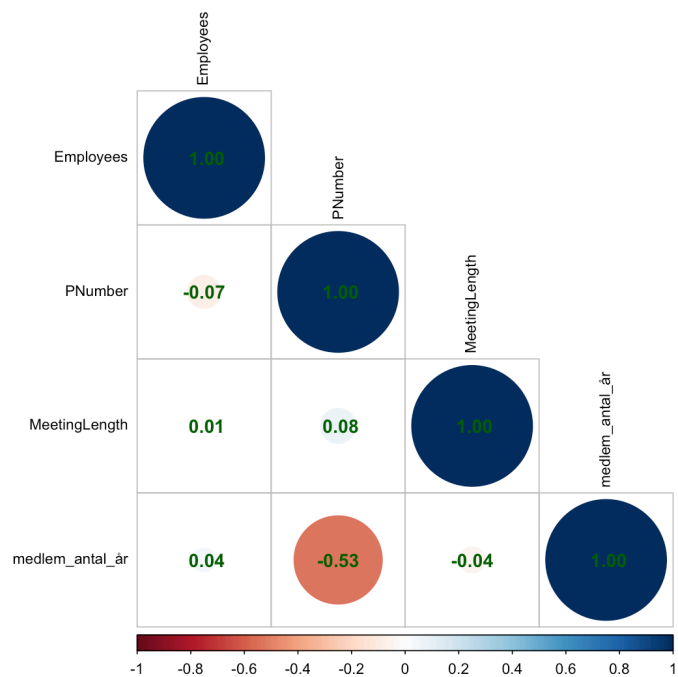
• **Bilag 3:** Branchefordeling blandt aktive medlemmer



• **Bilag 4:** Eventdeltagelse pr. postnummer



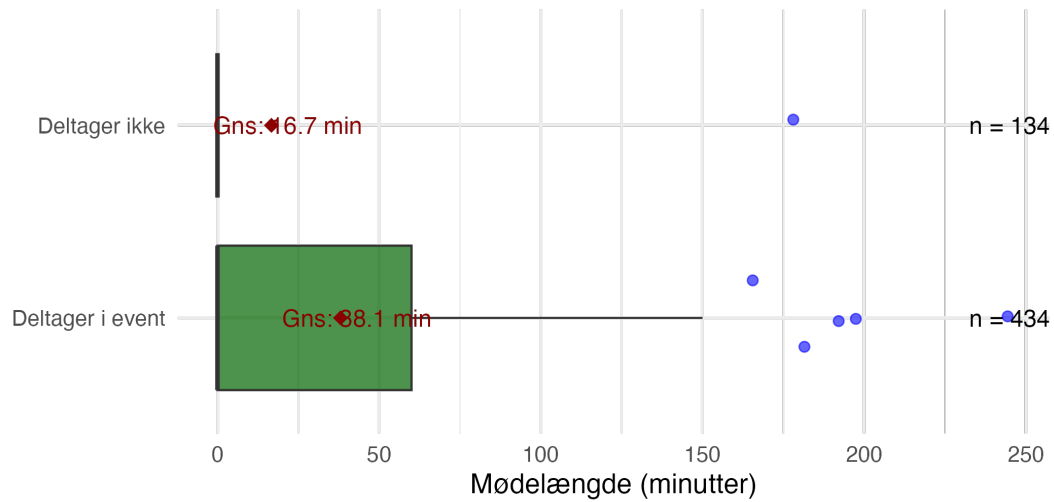
- **Bilag 5:** Korrelationsmatrix for numeriske variabler



- **Bilag 6:** Mødelængde vs. eventdeltagelse

## Mødelængde vs. eventdeltagelse (aktive medlemmer)

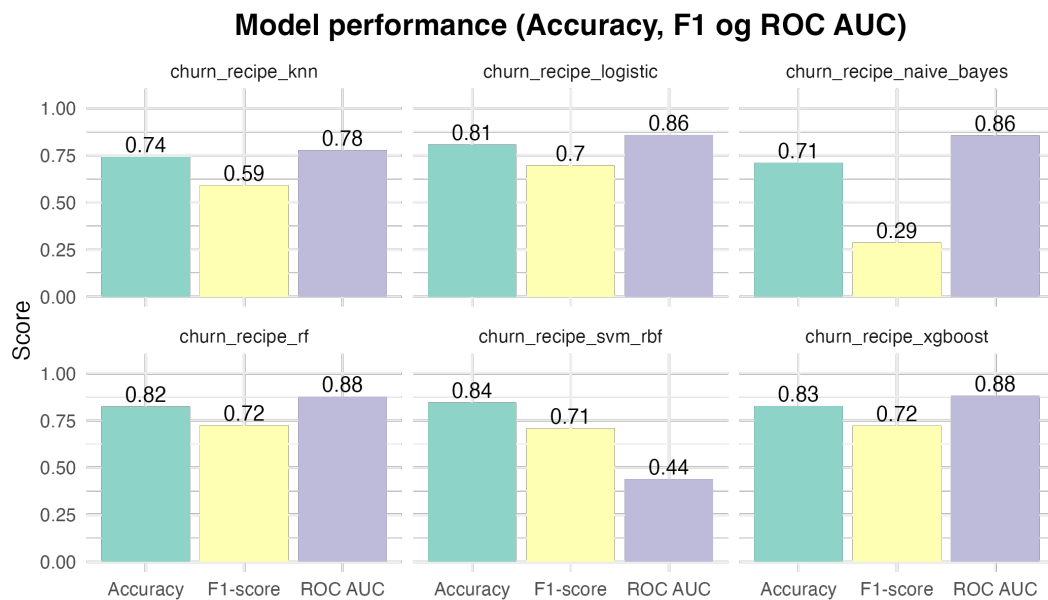
Rød prik = gennemsnit • Blå prikker = outliers



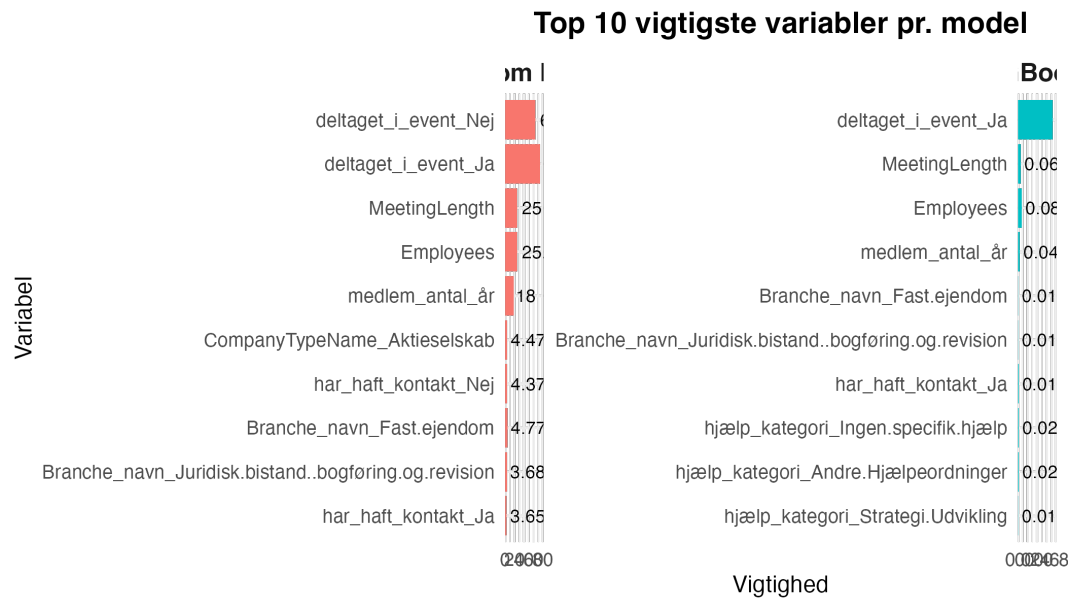
## 13.2 Bilag 7–15: Modeludvikling og evaluering

Anvendt i forbindelse med udvikling og evaluering af ML-modeller:

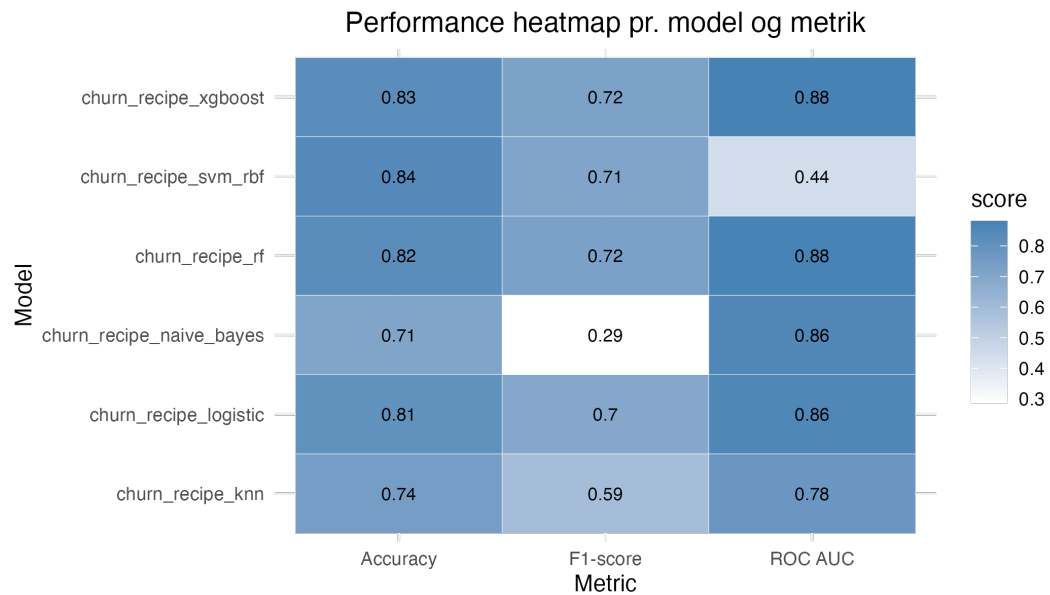
- **Bilag 7:** Model performance (Accuracy, F1 og ROC AUC)



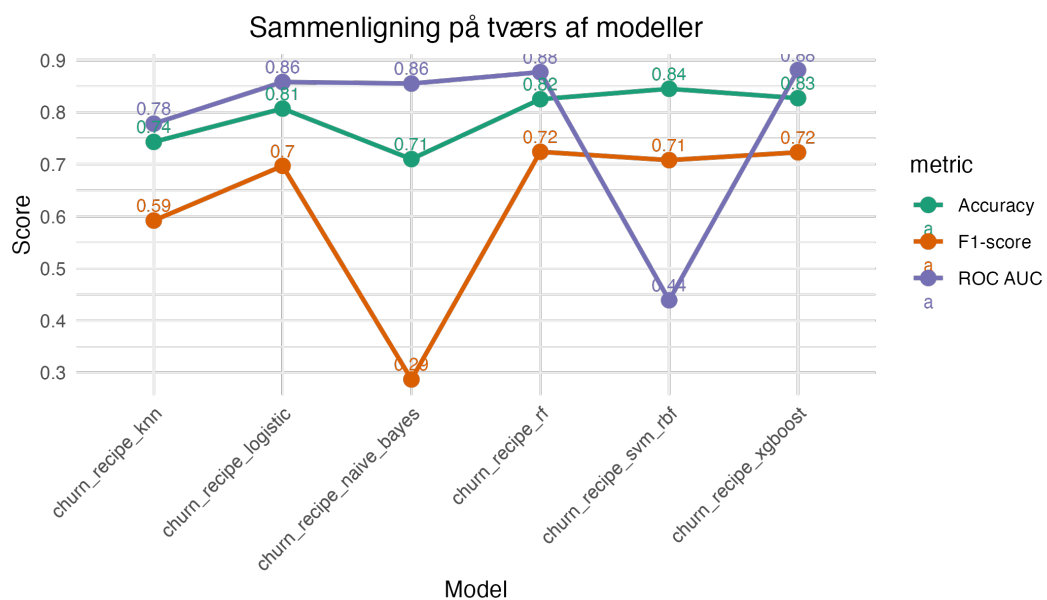
• **Bilag 8:** Top 10 vigtigste variabler pr. model



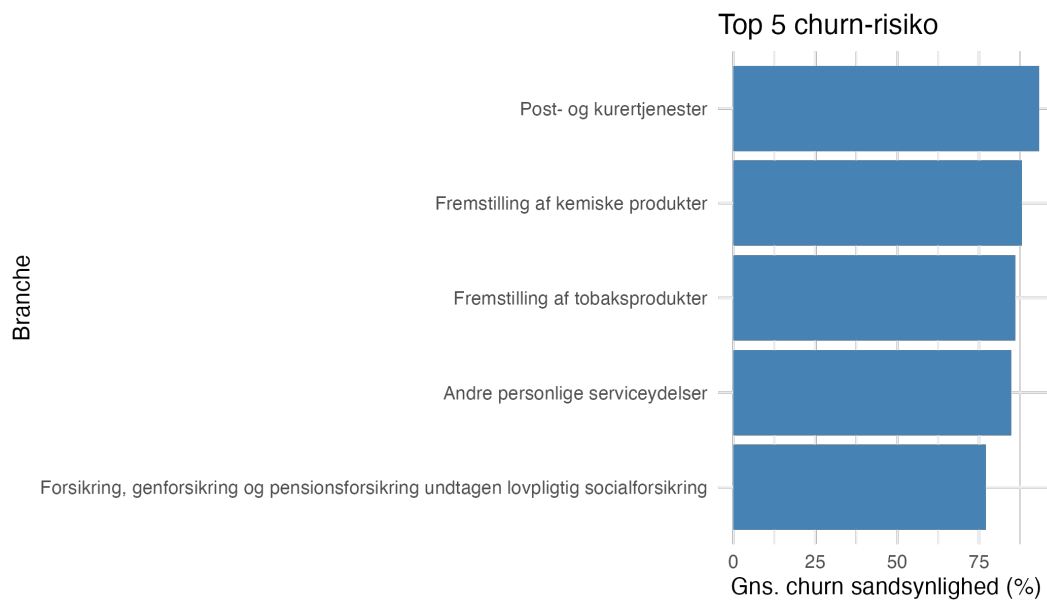
• **Bilag 9:** Sammenlignende heatmap over modeller



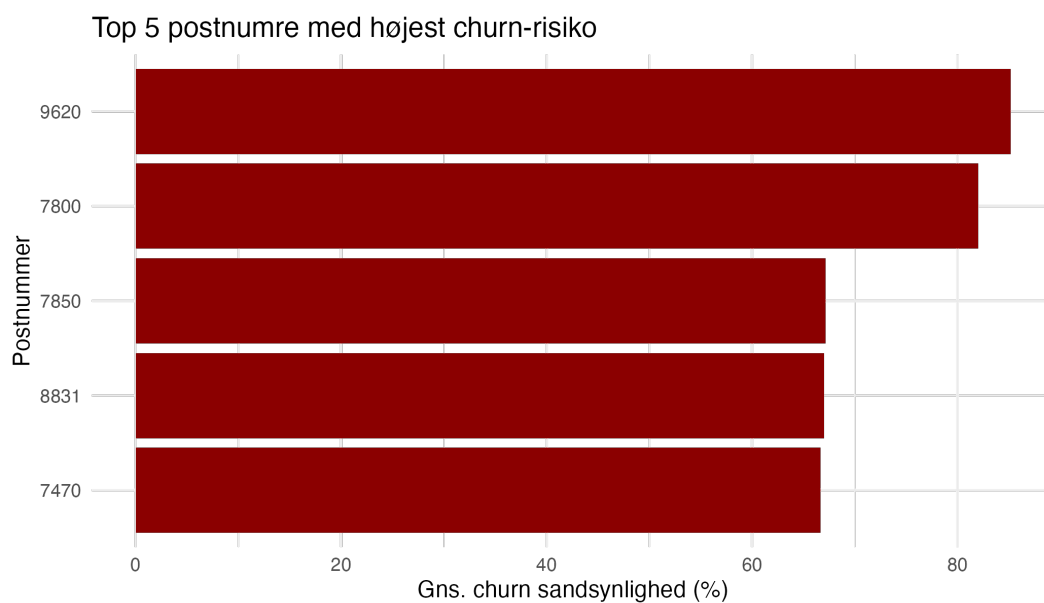
• **Bilag 10:** Udvikling over tid – churn-risiko



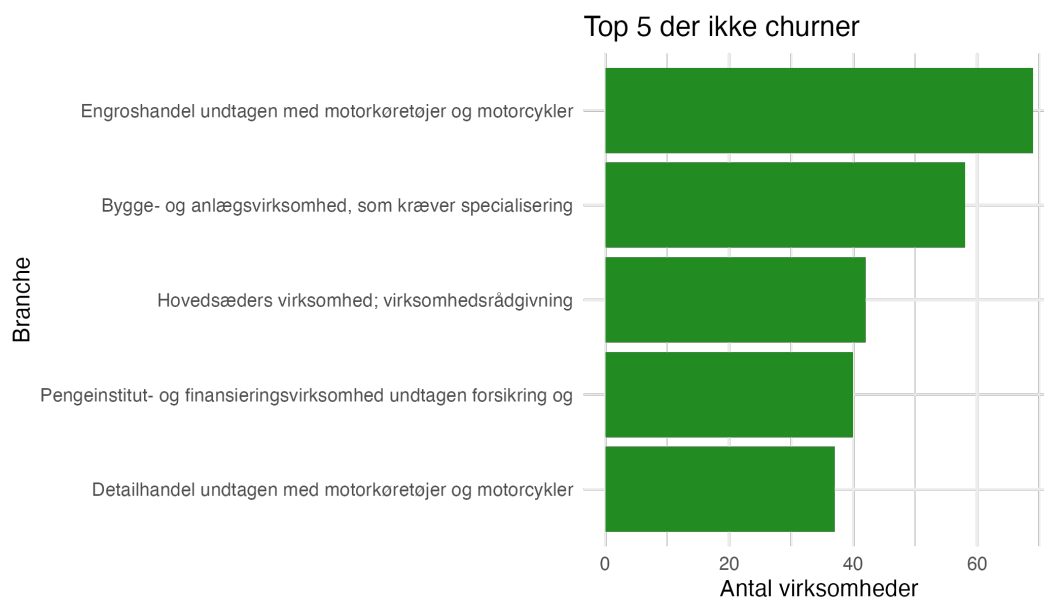
- **Bilag 11:** Brancher med højest churn-rate



- **Bilag 12:** Postnumre med højest churn-rate



- **Bilag 13:** Top 5 segmenter som ikke cherner



- **Bilag 14:** Churn-risikokategorier for aktive kunder