



# Forudsigelse af tilskuertal ved Viborg FF's hjemmekampe

– En dataanalyse af faktorer med betydning for billetsalg –

*Internt eksamensprojekt*

**Forfattere:**

Jeppé  
Lucas  
Konrad

**Dato:**

06/01/2026

Anslag: xxxx

# Indholdsfortegnelse

<b>1</b>	<b>Indledning</b>	<b>1</b>
<b>2</b>	<b>Problemstilling</b>	<b>1</b>
<b>3</b>	<b>Problemformulering</b>	<b>1</b>
<b>4</b>	<b>Afgrænsning</b>	<b>1</b>
<b>5</b>	<b>Videnskabsteori og metode</b>	<b>1</b>
5.1	Paradigmevalg . . . . .	1
5.2	Epistemologi – Hvordan viden opnås . . . . .	1
5.3	Metodologi – Hvordan undersøgelsen gennemføres . . . . .	1
5.4	CRISP-DM . . . . .	2
5.5	Empiri . . . . .	2
5.6	Validitet og reliabilitet . . . . .	2
<b>6</b>	<b>Analyser</b>	<b>2</b>
6.1	Alexandra-modellen . . . . .	2
6.2	Data Governance . . . . .	2
6.3	Datamodenhed . . . . .	3
6.4	Delkonklusion . . . . .	3
<b>7</b>	<b>Vurdering af modeller</b>	<b>3</b>
7.1	Indledende dataanalyse . . . . .	3
7.2	Anvendte modeller . . . . .	3
7.3	Sammenligning af modeller . . . . .	4
7.4	Variabeludvælgelse (Lasso) . . . . .	5
7.5	Fortolkning af modelresultater i praksis . . . . .	5
<b>8</b>	<b>Konklusion</b>	
<b>9</b>	<b>Anbefaling</b>	
<b>10</b>	<b>Kilder</b>	
<b>11</b>	<b>Bilag</b>	
11.1	Bilag Data Modenhed / Alexandramodel . . . . .	
11.2	Bilag Data Governance framework . . . . .	
11.3	Bilag Interview “Vff del 3 - Palle” . . . . .	
11.4	Bilag Interview “Vff del 1 - Daniel” . . . . .	
11.5	Bilag Interview “Vff del 2 - Praktikanter” . . . . .	
11.6	Bilag Kode . . . . .	

**Kode og repository:** <https://github.com/jepp3646/1semesterVFF>

## 1 Indledning

Tilskuertallet ved Viborg FF's hjemmekampe har stor betydning for klubbens økonomi, bemandingsplanlægning og stadiondrift. Med op til 10.000 tilskuere pr. kamp udgør billetsalget en central del af klubbens forretningsgrundlag. ("Energi Viborg Arena" 2024).

## 2 Problemstilling

Viborg FF råder over omfattende data fra billetsystemer og eksterne kilder, men data anvendes i begrænset omfang til at forudsige fremtidigt stadionfremmøde. Projektet har derfor til formål at undersøge, hvordan analyse af eksisterende data og maskinlæring kan identificere de væsentligste faktorer for billetsalg og understøtte mere databaserede beslutninger i klubben.

## 3 Problemformulering

Hvordan kan Viborg FF gennem indsamling og analyse af relevante data identificere de faktorer, der har størst indflydelse på billetsalget og anvende denne viden til at optimere salget af billetter til hjemmekampe?

## 4 Afgrænsning

Marketing og eksponeringsdata fra sociale medier er ikke inkluderet, da Viborg FF ikke har udleveret denne datatype. Projektet fokuserer alene på antallet af tilskuere.

## 5 Videnskabsteori og metode

### 5.1 Paradigmevalg

Projektet tager udgangspunkt i pragmatismen, som kan beskrives som et handlingsorienteret, praktisk og problemløsende paradigme. Pragmatismen adskiller sig fra andre paradigmer ved, at sandheden vurderes ud fra, hvad der fungerer i praksis. Dette passer direkte til projektets problemstilling, hvor der undersøges, hvordan Viborg FF kan anvende data i praksis til at forudsige tilskuertal og dermed optimere billetsalget.

### 5.2 Epistemologi – Hvordan viden opnås

Viden opnås gennem abduction, hvor observationer danner grundlag for forklaringer, som efterfølgende afprøves empirisk. Eksempelvis indikerer historiske data, at kampe mod attraktive modstandere er forbundet med højere tilskuertal, hvilket undersøges gennem kvantitativ modellering. Kvalitative input fra interviews anvendes til at understøtte hypotesedannelsen, hvorefter resultaterne føres tilbage til praksis. (Egholm 2014)

### 5.3 Metodologi – Hvordan undersøgelsen gennemføres

Der anvendes metodetriangulering:

- Kvalitative interviews: Interviews med nøglepersoner fra drift og dataafdeling anvendes til at indsigte i datamodenhed og behov.
- Kvantitative modeller i R: Kvantitative modeller i R til prædiktions af tilskuertal.
- Desktop research: Relevante offentlige datakilder som Superstats, DMI, Nager kobles med billetsalg via SQL.

## 5.4 CRISP-DM

I den kvantitative del af projektet anvendes CRISP-DM som metodisk ramme for arbejdet med dataanalyse og modellering. Modellen bidrager med en struktureret tilgang, hvor data bearbejdes med udgangspunkt i et forretningsmæssigt formål og videreudvikles frem mod anvendelige modeller og resultater. (“What is CRISP DM?” 2024)

## 5.5 Empiri

Empirien består af interviews/præsentationer fra klubben samt kvantitative data om billetsalg (3/7/10 dage før kamp) suppleret med kamp, vejr og kalenderdata.

## 5.6 Validitet og reliabilitet

Validitet understøttes ved at joine flere datakilder i ét fælles datasæt, samt ved at sammenligne flere modeltyper (lineær regression, Ridge, Lasso og Best Subset) på samme opgave. Reliabilitet sikres gennem reproducerbare data splits og trænings/test-evaluering og crossvalidation ved valg af lambda for Ridge/Lasso, hvilket reducerer risikoen for overfitting.

# 6 Analyser

## 6.1 Alexandra-modellen

Alexandra-modellen anvendes til at analysere samspillet mellem forretning, teknologi og mennesker i Viborg FF. Analysen viser, at klubben i stigende grad anvender data til at understøtte driftsmæssige beslutninger, særligt inden for rapportering og billetsalg. Teknologisk er et grundlæggende datagrundlag etableret, men manglende systemintegration begrænser en mere sammenhængende og automatiseret anvendelse af data. Samtidig er dataanvendelsen primært integreret i dataafdelingen, mens resten af organisationen i begrænset omfang inddrages i det analytiske arbejde. Viborg FF’s væsentligste udfordringer i forhold til dataanvendelse er ikke teknologiske, men organisatoriske. En mere moden anvendelse af data forudsætter derfor en bredere integration af dataanvendelsen i hele forretningen. (Se bilag [11.1](#)).

## 6.2 Data Governance

Analysen af data governance i Viborg FF viser, at ansvar og ejerskab for data i høj grad er centraliseret i dataafdelingen. Der mangler fortsat fælles og formaliserede retningslinjer for datadeling, datakvalitet og beslutningsrettigheder på tværs af organisationen. Datadeling og kvalitetssikring sker primært gennem manuelle processer, hvilket skaber usikkerhed om datakonsistens og begrænser data som en strategisk ressource. (se bilag [11.2](#)).

## 6.3 Datamodenhed

Viborg FF's datamodenhed vurderes ved at sammenholde interviewdata med Alexandra Instituttets dataforståelsesmodel. Analysen viser, at data primært anvendes til rapportering og understøttelse af driftsbeslutninger. Processer er endnu ikke fuldt automatiserede, og databrug er fortsat centreret omkring deres driftsafdeling.

## 6.4 Delkonklusion

Samlet set viser analysen, at Viborg FF har opbygget et solidt fundament for databrug, men at data fortsat anvendes primært operationelt og afgrænset organisatorisk. For at udnytte data mere strategisk på tværs af organisationen peger analysen på et behov for tydeligere governance-strukturer og en bredere organisatorisk integration af databrug. Viborg FF's datamodenhed vurderes samlet set som middel.

# 7 Vurdering af modeller

Modeludvikling og evaluering følger CRISP-DM-faserne 'Modeling' og 'Evaluation', hvor modellerne sammenlignes på test-RMSE på tværs af tidshorisonter.

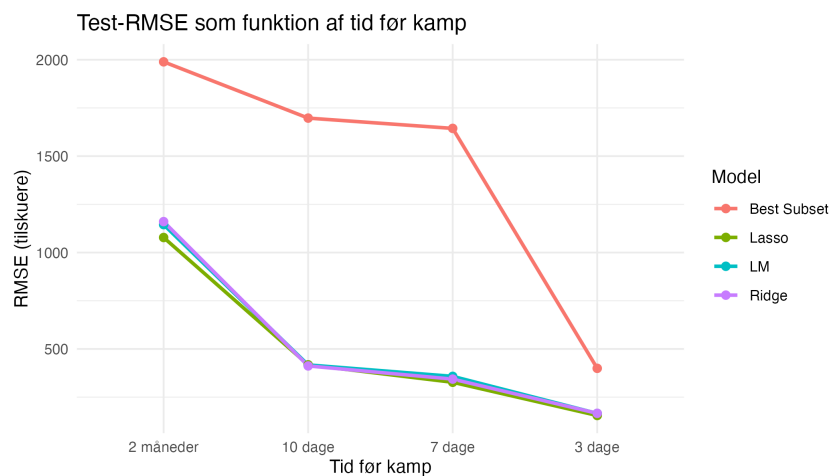
## 7.1 Indledende dataanalyse

Data er indsamlet, rensat og joinet til et samlet datasæt via SQL (Superstats, DMI, Nager og interne billetsalgsdata). Der er håndteret missing values og dubletter samt standardiseret variabler. Target er tilskuertal, og centrale features omfatter blandt andet historiske tilskuergennemsnit, modstander, kampafvikling (ugedag/kickoff), sæson/runde samt vejr og kalendereffekter.

## 7.2 Anvendte modeller

Datasættet opdeles i trænings- og testdata ved et 80/20-split for at sikre en konsistent vurdering af modellernes prædiktionssevne. Træningsdata anvendes til estimering af modellerne, mens testdata benyttes til evaluering. Der anvendes fire regressionsmodeller: lineær regression (LM), Ridge, Lasso og Best Subset Selection. Den lineære regression fungerer som reference for modellernes præstation. Ridge og Lasso anvendes til at håndtere multikollinearitet og reducere overfitting gennem regularisering; her anvendes Lasso desuden som et eksplorativt værktøj til variabeludvælgelse med fokus på prædiktionssevne. Modellerne sammenlignes på baggrund af test-RMSE.

### 7.3 Sammenligning af modeller

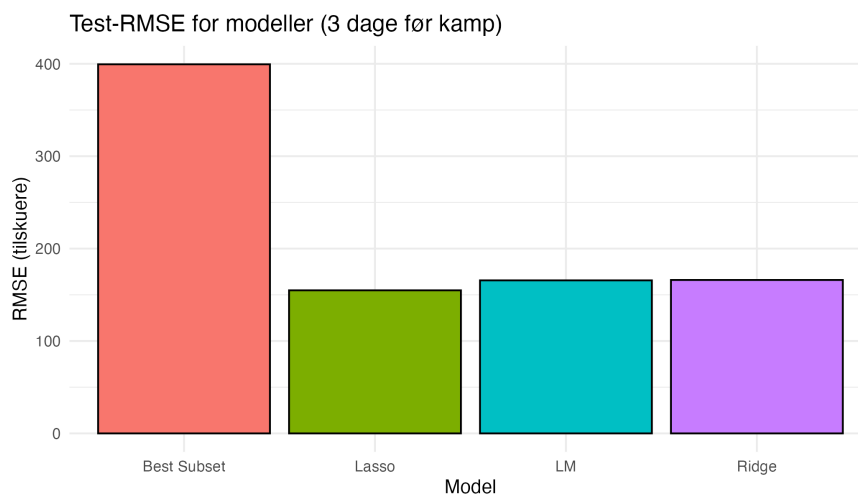


Figur 1: Test\_RSME for alle modeller 2m, 10d, 7d, 3d (Egen tilvirkning)

Tabel 1: Test-RMSE for de anvendte modeller ved forskellige tidspunkter før kamp.

Tid	LM	Ridge	Lasso	BestSubset
2 måneder	1144.7	1160.3	1077.9	1989.2
10 dage	417.1	411.9	417.4	1697.3
7 dage	357.8	343.6	327.1	1643.9
3 dage	165.6	166.0	154.8	399.4

Overordnet ses et klart fald i RMSE for alle modeller, jo tættere analysen foretages på kampdagen. Dette indikerer, at information om billetsalg tættere på kampafvikling har stor betydning for modellernes præcision. Den klassiske lineære regressionsmodel samt de regulariserede modeller Ridge og Lasso præsterer relativt ens på tværs af tidspunkter. Lasso-modellen opnår konsekvent den laveste eller næstlaveste RMSE, særligt 3 og 7 dage før kamp, hvilket indikerer en mere stabil og præcis model i denne fase.



Figur 2: Test-RMSE for modeller 3 dage før kamp (Egen tilvirkning)

Figuren viser test-RMSE tre dage før kamp. Lasso-modellen opnår den laveste test-RMSE og udviser den højeste prædiktionssevne i denne fase. LM og Ridge præsterer på et tilsvarende niveau. Best Subset Selection har markant højere RMSE, hvilket indikerer svag generalisering, hvilket kan være tegn på overfitting.

## 7.4 Variabeludvælgelse (Lasso)

Lasso kan samtidig bruges til variabeludvælgelse ved at reducere koefficienterne eller udelade mindre relevante variable.

Tabel 2: Koefficienter fra Lasso-model anvendt til variabeludvælgelse (3D).

Variabel	Koefficient
Intercept	145.987
season2006/2007	-1.359
season2023/2024	120.261
modstanderBIF	50.212
modstanderFCN	-47.236
kategoriB	-5.117
mean_tilskuere	0.004
weekday <sup>4</sup>	-36.542
kickoff_hour	0.737
temp_kamp	-0.664
regn_dag	-9.710
d10_tilskuere	0.234
d7_tilskuere	0.248
d3_tilskuere	0.828

Lasso-modellen viser, at salgsdata tæt på kampdagen har størst betydning for forudsigelsen af tilskuertallet, mens historiske gennemsnit og vejrforhold bidrager i begrænset omfang. Modellen reducerer kompleksiteten og identificerer de mest centrale faktorer for billetsalg, hvilket giver et anvendeligt beslutningsgrundlag.

## 7.5 Fortolkning af modelresultater i praksis

Tabel 3: Best- og worst-case estimer for tilskuertal til VFF – FCN (2 måneder før kamp).

kamp	scenario	predicted_tilskuere
Viborg vs FCN (1/3 kl. 17)	Best-case (2m)	5545
Viborg vs FCN (1/3 kl. 17)	Worst-case (2m)	4203

Tabellen viser best og worstcase-scenarier for det forventede tilskuertal til kampen mod FCN lørdag den 1. marts. Scenarierne illustrerer forskellen i det forventede fremmøde og kan anvendes som beslutningsgrundlag for planlægning af markedsføring og ressourceallokering. Resultaterne viser, at Viborg FF kan udarbejde brugbare estimer for billetsalget flere dage før kampstart, og at prognosernes præcision øges markant, jo tættere man kommer på kampdagen. Tidlige estimer kan hjælpe med overordnet planlægning, mens mere præcise prognoser tættere på kampafvikling kan anvendes til beslutninger i forhold til bemanning og kapacitetsstyring.

## 8 Konklusion

Projektet har haft til formål at undersøge, hvordan Viborg FF kan anvende data til at identificere de faktorer, der har størst betydning for billetsalget og understøtte mere databaserede beslutninger. Analysen viser, at klubben råder over et solidt datagrundlag, som muliggør opbygning af statistiske og maskinlæringsbaserede modeller til estimering af tilskuertal. Resultaterne dokumenterer, at modellernes præcision øges markant, jo tættere man kommer på kampdagen, hvilket gør det muligt at anvende forskellige prognoser afhængigt af beslutningshorisonten.

Sammenligningen af modeller viser, at regulariserede modeller som Ridge og Lasso håndterer datasættets kompleksitet bedre end den klassiske lineære regression og dermed er mere stabile i praksis. De organisatoriske analyser indikerer samtidig, at Viborg FF befinder sig på et middel niveau af datamodenhed, hvor data primært anvendes operationelt. Samlet set har klubben et solidt fundament for datadrevet billetsalg, men et yderligere potentiale kan realiseres gennem tydeligere governance-strukturer og bredere organisatorisk forankring af databrug.

## 9 anbefaling

På baggrund af projektets analyser anbefales det, at Viborg FF i højere grad anvender de udviklede datamodeller som et fast beslutningsstøtteværktøj i forbindelse med billetsalg og planlægning af hjemmekampe. Modellerne 5 kan anvendes forskelligt afhængigt af tidshorisonten, hvor tidlige estimater understøtter overordnet planlægning, mens mere præcise prognoser tættere på kampdagen kan anvendes til beslutninger såsom bemanning og indkøb.

Derudover anbefales det, at Viborg FF arbejder mod en tydeligere struktur for data governance, herunder klarere retningslinjer for dataejerskab, datadeling og anvendelse på tværs af organisationen. En bredere organisatorisk forankring af databrug kan bidrage til at øge værdien af de eksisterende data og styrke klubbens datamodenhed.

Afslutningsvis anbefales det, at klubben fortsætter udviklingen af datakompetencer og anvendelsen af data i praksis, så data i stigende grad anvendes strategisk og ikke udelukkende

## 10 Kilder

Camilla Kølsen and Laura Lynggard Nielsen and Rasmus Bækby. 2017. “Find vej i din dataindsats”. Alexandra Instituttet. <https://alexandra.dk/wp-content/uploads/2020/09/Alexandra-Instituttet-BDBA-Find-vej-i-din-dataindsats.pdf>.

Dataafdelingen. 2025. “Dania x Viborg FF”. Dataafdelingen.

Egholm, Liv. 2014. *Videnskabsteori*. 1. udg. Frederiksberg: Hans Reitzels Forlag.

“Energi Viborg Arena”. 2024. 2024. <https://www.vff.dk/viborg-f-f/stadionoversigt>.

“The DGI Data Governance Framework”. 2024. 2024. <https://datagovernance.com/the-dgi-data-governance-framework/dgi-data-governance-framework-components/>.

“What is CRISP DM?”. 2024. 2024. <https://www.datascience-pm.com/crisp-dm-2/>.

## 11 Bilag

### 11.1 Bilag Data Modenhed / Alexandramodel

For at vurdere Viborg FF’s datamodenhed anvendes interviews med centrale nøglepersoner i organisationen, herunder repræsentanter fra ledelsen, dataafdelingen samt praktikanter. Interviewene analyseres og sammenholdes med Alexandra Instituttets dataforståelsesmodel.



Alexandra modellen anvendes til at analysere samspillet mellem forretning, teknologi og mennesker og dermed identificere, hvor Viborg FF befinder sig i sin datamodenhed, samt hvilke barrierer og potentialer der eksisterer for en mere datadrevet organisation. (Camilla Kølsen and Laura Lynggard Nielsen and Rasmus Bækby 2017)

### 11.1.1 Forretning

Viborg FF anvender i stigende grad data som understøttelse af forretningsmæssige og driftsmæssige beslutninger, særligt inden for billetsalg, kampafvikling og ressourceplanlægning. Data anvendes blandt andet til at analysere tilskuertal, udnyttelsesgrader, billetsalg samt fordeling af publikum på stadionafsnit.

Interviewene viser, at data primært anvendes til optimering af eksisterende processer, frem for til grundlæggende at gentænke forretningen. Beslutninger om eksempelvis bemanding, indkøb til boder samt kapacitetsstyring understøttes af historiske data og løbende opfølgning på nøgletal. Dette understøttes af Palle, som beskriver, hvordan data anvendes direkte i den daglige drift: “Det er helt rigtigt. Det er data, jeg arbejder med, så det er meget faktabaseret.” (se bilag 11.3)

### 11.1.2 Teknologi

Den teknologiske infrastruktur i Viborg FF er etableret, men fremstår fortsat fragmenteret. Organisationen anvender flere systemer til håndtering af billetter, partnerdata og rapportering, som endnu ikke er fuldt integrerede. Dette medfører, at data ofte trækkes manuelt og deles via Excel. Dette bekræftes af praktikanterne, som beskriver, hvordan data i høj grad modtages i forskellige formater og efterfølgende struktureres manuelt: “Jeg har fået udleveret en hel masse data fra hovedsystemerne i alle mulige former – både PDF og Excel-filer – som jeg så skulle finde ud af, hvordan man kunne strukturere.” (11.5)

Selvom Viborg FF er begyndt at anvende Power BI og dashboards, fremgår det, at data ofte kræver rensning og manuel bearbejdning, før de kan anvendes analytisk. Daniel fra dataafdelingen understreger, at organisationen stadig befinder sig i en overgang fra manuelle til mere automatiserede løsninger: “Tidligere blev vores KPI’er lavet manuelt i PowerPoint, og i dag har vi gjort det sådan, at analytikerne bare kan bede om dataen.” (se bilag 11.4)

Der indhentes både data fra driftafdeling, salgsafdelingen og Markom. Driftsafdelingen tager sig af data på boder og billetter. Salgsafdelingen indsamler data fra deres BtB kunder. Markom afdelingen tager sig af data på sociale medier, data på fans igennem appen, og data på salg af merchandise. (Dataafdelingen 2025)

### 11.1.3 Mennesker

Interviewene viser, at der generelt er en positiv og stigende interesse for data blandt medarbejderne i Viborg FF. Dataafdelingen fungerer som en central drivkraft, mens øvrige afdelinger primært anvender data som beslutningsstøtte frem for selvstændigt analytisk værktøj. Daniel peger på, at der er stor nysgerrighed omkring data, men også en manglende forståelse for, hvad arbejdet med data kræver: “Jeg oplever generelt en rigtig stor nysgerrighed på, hvad vi kan få ud af vores data. Udfordringen er forståelsen for, hvor lang tid forskellige dele af arbejdet med data kan tage.” (11.4)

### 11.1.4 Samlet vurdering af Alexandra fase

På baggrund af analysen vurderes Viborg FF at befinde sig på niveau 3 i Alexandra-modellen. Organisationen har etableret et fælles datagrundlag og anvender data systematisk til rapportering og driftsmæssige beslutninger. Samtidig er der igangsat strukturelle tiltag for at styrke databrug, hvilket Daniel beskriver således: “I dag kan vi give analyserne videre meget hurtigere, og det gør, at data bliver brugt mere i hverdagen.” (se bilag 11.4).

Datamodenheden vurderes derfor som middel, idet Viborg FF har taget væsentlige skridt mod en mere datadrevet organisation, men fortsat er præget af manuelle processer, silo-opdelt datadeling og en beslutningskultur, hvor data endnu ikke er det primære grundlag på tværs af hele organisationen.

## 11.2 Bilag Data Governance framework

DGI Data Governance Framework anvendes i projektet som et analytisk redskab til at vurdere, hvordan Viborg FF aktuelt arbejder med data, samt hvordan data governance kan understøtte en mere systematisk, ensartet og værdiskabende anvendelse af data. Frameworket giver en struktureret tilgang til at analysere både organisatoriske, tekniske og ledelsesmæssige forhold omkring data og anvendes her i relation til billetsalg og beslutningsstøtte. (“The DGI Data Governance Framework” 2024)

### 11.2.1 Mission and value

Formålet med data governance i Viborg FF er at skabe et fælles og pålideligt datagrundlag, som kan understøtte bedre beslutninger på tværs af organisationen. Særligt i forhold til billetsalg er målet at sikre, at data kan anvendes til både planlægning, prognoser og løbende opfølgning frem mod kampdag. Værdien af data governance ligger dermed i at reducere usikkerhed, skabe overblik og gøre data mere anvendelige i praksis frem for udelukkende at fungere som historisk rapportering.

### 11.2.2 Beneficiaries

En styrket data governance-struktur skaber værdi for flere aktører i Viborg FF. Ledelsen får et mere solidt beslutningsgrundlag, som kan anvendes strategisk i forhold til kapacitet, prisstrategi og prioritering af ressourcer. Salgs- og marketingfunktionerne kan i højere grad anvende data til målrettede initiativer, mens dataafdelingen får en tydeligere rolle i forhold til struktur, kvalitet og sammenhæng i data. Samlet set understøtter dette en mere datadrevet organisation.

### 11.2.3 Data Products

De centrale dataprodukter i Viborg FF er i dag primært rapporter og dashboards, som samler information om billetsalg, kampforhold og historiske mønstre. Set i lyset af organisationens nuværende datamodenhed er fokus på overskuelige og operationelle dataprodukter, som let kan anvendes af beslutningstagere. Disse dataprodukter fungerer som bindeled mellem data og forretning og gør det muligt at omsætte analyser til konkrete handlinger.

### 11.2.4 Controls

Kontrolmekanismer omkring data er nødvendige for at sikre konsistens og troværdighed. I Viborg FF sker dette primært gennem faste arbejdsgange, manuelle kvalitetstjek og fælles definitioner af centrale variabler. Selvom processerne endnu ikke er fuldt automatiserede, bidrager disse kontroller til at reducere fejl og sikre, at data anvendes på et ensartet grundlag.

### 11.2.5 Accountabilities

Ansvarsfordelingen omkring data er et centralt element i data governance. Dataafdelingen har det overordnede ansvar for dataindsamling, struktur og kvalitet, mens de enkelte forretningsområder har ansvar for indhold og fortolkning af deres data. Denne fordeling bidrager til klarhed omkring ejerskab og understøtter samarbejdet mellem data og forretning.

### 11.2.6 Decision Rights

Beslutningsrettigheder vedrørende data er overvejende centraliserede, hvilket sikrer konsistens i anvendelsen af data. Samtidig anvendes data decentralt i organisationen, hvor de enkelte afdelinger bruger rapporter og analyser i deres daglige arbejde. Denne balance mellem central styring og lokal anvendelse er hensigtsmæssig i en organisation, der fortsat er under udvikling på dataområdet.

### 11.2.7 Policy and Rules

Viborg FF arbejder med overordnede retningslinjer for databrug, herunder adgang til data, datakvalitet og overholdelse af gældende lovgivning. Politikkerne har til formål at sikre ansvarlig og korrekt håndtering af data, samtidig med at de understøtter en fælles forståelse af centrale begreber og målepunkter.

### 11.2.8 Processes, Tools and communication

Data governance understøttes af eksisterende værktøjer og processer, som muliggør indsamling, lagring og formidling af data. Kommunikation spiller en vigtig rolle, da data først skaber værdi, når den forstås og anvendes. Dette sker blandt andet gennem fælles rapportering og dialog mellem dataafdelingen og forretningen.

### 11.2.9 DG Work Program og 10. Participants

Arbejdet med data governance i Viborg FF foregår gradvist gennem konkrete initiativer, der tager udgangspunkt i aktuelle behov og organisatoriske ressourcer. Dataafdelingen fungerer som drivkraft i dette arbejde, mens ledelse og nøglemedarbejdere inddrages i både prioritering og anvendelse af data. Samlet set understøtter data governance en mere struktureret og fremadrettet brug af data, som kan bidrage til optimering af billetsalget til hjemmekampe.

## 11.3 Bilag Interview “Vff del 3 - Pelle”

Vff del 3 - Pelle

Transskriberet af TurboScribe.ai. Opgrader til Ubegrænset for at fjerne denne meddelelse.

Har du noget af det at sige? Undskyld? Nej, men ikke svor for meget på noget af det, jeg siger. Jamen det er okay, vi skal nok tage det med et græns salt, men var det Pelle eller Pelle? Pelle. Pelle, der her er med på, og vi er godt med at optage.

Vi har fået lov til at låne Pelle. Pelle, han er på kontoret nu. Hej, mit navn er Pelle.

Jeg har været så heldig at være i VFF næsten 20 år. Så det har været en lang rejse med både op- og nedture. Og heldigvis er der været et par gode opture, som man husker meget tydeligt.

Min rolle omkring klubben har været mange. Da vi var på et tidspunkt tre og en halv ansat, så havde man rigtig mange roller. Og i dag, der synes jeg, de sagde i går, at vi var rundt 100.

Så jeg skal lige sige, at de tre sagde alt det var administration, så i dag er vi 25 stykker. Så det er nogle helt andre roller. Så i dag går min rolle 100 procent på kampeopviklingen.

Så når vi har kampe inde på stadion, så er det mit overordnede ansvar, at hele rammen omkring kampen fungerer.

At folk møder ind til de forskellige jobs, der er i boede og i sikkerhed. Hvor mange vil lige skyde på, at når vi åbner dørene, hvor mange vil skyde på, at vi har i gang på stadion under tribuneafsnittet til at servicere folk?

Hvis vi er 6.000, hvor mange har så en sving nede på stadion? Det er faktisk rigtigt.

Det er helt rigtigt. Det er data, jeg arbejder med, så det er meget fakta baseret. Hvad siger vi? 100. Det er der i hvert fald. 400. Ja, det er for mig. Hvad siger du? Under 200. Under 200? Nej. Vi runder 250 mennesker til sådan en mellemkamp.

Så der er rigtig mange i sving for at servicere i de områder, der er. Og egentlig, det er mit ansvar, at det kører nu ikke sådan, at jeg er på vej med alle dem, som vi har en masse, som er ledere i de enkelte områder. Men det

er sådan for at sige lidt om, hvilken kampramme den er i dag.

9Er der nogen af jer, der kommer fast på stadion? Nej. Men det er jo også det der er sjovt ved det job, der er blevet her. Da jeg for mange år siden blev ansat, der blev man ansat i klubben, fordi det var hjerteblodet.

Og fordi, at man havde en eller anden indgang til fodbolden, enten som frivillig på en eller anden måde, eller som fodboldtræner, eller som noget med ungdomseliten. Men i dag, der er det jo meget mere specialister, som bliver taget ind på enkelte områder. Og vi begynder jo stille og roligt at have nogle kolleger, som måske ikke har den store indgang i fodbold, men bare fordi de er pisse dygtige til de områder, de har.

Og så må vi så håbe på, at vi kan få fodboldet med stille og roligt. Men det er jo den udvikling, der er blevet inde. Og så sent som i går, der vælger vi en ny mand ind i bestyrelsen, som eftersigende på boldet, der er Brøndby-fag.

Men bare har en hel masse, og kan bidrage med i vores bestyrelsearbejde, vores strategiarbejde, fordi han har siddet i Arbejdernes Landsbank, og forhandlet en stor erhvervlighed og sponsoraftaler med mange klubber. Så han kommer helt sikkert til at bidrage en masse i vores bestyrelse. Så det er en helt ny verden.

Det vi skal snakke lidt om i dag, det er specielt kampafviklingen, og specielt rammen omkring kampafvikling, og nogle af de ting, som vi går ind og arbejder med. Det øverste billede her, det var sådan, vi vores stater så ud for, lidt før corona og nogle år før der, hvor vi sådan åbnede rammen for 2.000 mennesker, hvis det var en god dag, og skulle klare at gøre lidt i det. Ned til en ramme i dag, hvor vi spiller mod Westham, og har stort set udsolgt på tribunerne.

Og vi har en kapacitet på 10.000. Så det er egentlig den udvikling, der er sket på få år, hvor rammen omkring staterne er blevet en helt anden. Og det har vi brugt ekstremt meget tid på, i forhold til vores brug af data. Og udviklet hele elektroniske platform, udviklet brugen af de data, vi opsamler og køber på kampene.

Brugsgrader og omsætningstallet i broderne, på mange forskellige fronter. Så der forsøger vi at tage nogen skridt, og har brugt det til at komme videre med os. Hvis vi kigger på tilskuereudviklingen, så kan I se her, 18-19.

I første edition var vi egentlig oppe at have et tilskuertal på 3.000. Så dykker det selvfølgelig lige her lidt med corona. Men ellers kan I se, hvor stejlt en kurve det har gået opad. De her to år her, de tre år her er rimelig sammenlignelige, fordi der spiller vi i det, man kalder nedrykningsspillet, eller kvalificeringsspillet, synes vi lige. Det vil sige, at i de sidste ti kampe møder man de seks nederste hold, hvorimod den her ene runde, der var vi heldige at komme i medaljeslutspillet, samtidig med at vi er ude at spille europæisk. Men det der er vildt nok at se, det er at 24-25 sætter vi faktisk en tilskuerrekord på 6.400 for Superliga-kampen og det. Hvorimod det her over, der er vi selvfølgelig over, men det er med et slutspil, og det er med et europæisk spil.

10Så hvis man tog det europæiske fodboldudpræt, så går den faktisk under. Og vi opnår faktisk et højere tilskuertal

i sidste sæson. Det vil sige, at vi kæmper en lille smule med det nu.

Vi ligger på omkring 6.200. Det er ikke helt piget, som vi gerne vil. Så der er nogle ting, som vi begynder at arbejde på, og kigger på nogle af de her tal bagved, og prøver at dykke lidt ned i, hvad er det for nogle områder, hvor vi lige taber lidt terræn. Så det er nogle ting, vi prøver at bruge det på.

Det samme med sæsonkort. I 18-19, der var sæsonkort den traditionelle gamle ting, med nærmest et plastikkort,

som man fik udleveret, og man købte på et årskort. Ret hurtigt, så begynder der at ske noget med, at man får indført abonnementer, på samme måde som Netflix og alt muligt andet.

Så det introducerer vi lige der i begyndelsen, hvor corona er det, der starter. Vi har jo ikke helt en stor udvikling på det i starten, fordi corona kom. Men med det samme efter, hvor corona blev snydt op igen, og vi såndet ret kraftige markedsføder på sæsonkort og abonnenter, der eksploderer det for os.

I dag er vi 4.100 abonnenter på stadionet. I 18-19 var vi 504 årlige sæsonkortholdere. Så en helt eksklusiv udvikling.

Jeg tror bare, at det brænder ikke. Abonnement? Du køber en plads. Akkurat på samme måde, som du køber et overskort, hvor du har en plads.

Men her betaler du bare om måneden. Fordelen fra overskortene, de skal fornyes hver år. Det her er selvfølgelig ikke noget, jeg gør, men generelt er det en del.

For eksempel at få købt et fitnesskort, og det kører bare hele tiden. Egentlig er det på samme måde her. Når man først har fået købt abonnent, så kører det bare ind til, indtil man siger det er op.

Har I sat tal på, hvor mange der så bruger det? Ja. Fordi det er jo så det næste. Det kommer lidt senere.

Det er hele vores udnyttelsesgrad. Vi ønsker heller ikke bare at sælge et abonnement, og så bliver det stolt. Så det kommer lidt senere.

Men her i, hvad det hedder, 2023, så begynder vi egentlig at være i en situation, at vi har fået så mange abonnenter, eller så er overskortet solgt ud, at vi er nødt til at begynde at kigge på nogle andre løsninger. Fordi 11som sagt, når du køber et overskort eller et sæsonkort om morgenen, så optager du en plads. Og vi har 10.000 pladser.

Nogle af pladserne er sat af i 1420, og der er 13.400 til bespisende gæster, og 1.800 til partnerbilletter. Så det begynder at stille og roligt at plænge ned i de steder, man kan sælge. Når vi så har solgt omkring 3.500 sæsonkort, inden et par abonnementer lårs, så begynder der at være færre pladser at sælge.

Så vi introducerer et billetabonnement, der kommer herind til. Billetterabonnement er på den måde, at du stadigvæk har adgang til kampene, men du skal aktivt ind og hente billetten, når du ønsker at være der. Så ud af de her, vi er på omkring 780 stykker, jamen de har adgang til det, og der er en garantiperiode en uge før kampen, der er de helt 100% sikre på at få en billet.

Men hvis du kører ind i den sidste del af garantiperioden, så kan der være udsolgt. Så det er her, at jeg håber, at vi kommer til at udvikle os, fordi dem, der ligger i den her del med sæsonkortabonnenterne, hvis deres udnyttelsesgrader er for lave, så forsøger vi at få dem over på billetterabonnementen, fordi så har vi ikke nogen tomme pladser. Og dem her, de bliver stille og roligt i fasen, altså vi sælger ikke mere, men dem, der har været med igennem mange år, har har et års kort og sådan nogle ting, så de skal selvfølgelig have lov til at gøre det, men det er ikke den vej, vi ønsker at gå.

Vi forsøger at arbejde meget med målsætninger, i forhold til, hvor vi ligger henne på de enkelte kampe. Nu har vi nu her Vejle, den 23. 11., og vi har Midtjylland, den 7. 12.

Hvis man kigger på tallene hernede, med hvad vi ønsker at opnå, så er det faktisk sådan, i forhold til pladserne på stadion, når vi siger 75 procent, og I er så skarpe, så 10.075 procent, det kan vi godt regne ud alle sammen, hvad det er. Så det er egentlig næsten udsolgt, når vi siger 7.200, der skal fremmede, fordi vi er med på, at vi

har kun en fremmedeprocent på 75. Hvorimod til Midtjylland, der går vi op og regner med en fremmedeprocent på 85, så der vi kan få 8.500, og så har vi nok lidt overkostning den dag, der gør, at vi kan komme lidt højere op.

Men her kan I se, hvordan vi forsøger at målrette, ude i de forskellige afsnit på stadion, med 100 stykkes i vores Skybox, 100 stykkes i vores Gandusklaven, 720 stykkes i vores BDO, samlet 1.150 i VIP-afsnittet, 2.000 partnerbilletter, 2.500 på Østtribyen, 1.500 på Sydtribyen, og 300 i Udvendt-afsnittet. Så det sætter vi op målsætningsmæssigt, og det følger vi. Vi følger de tal, vi følger op på, hvordan ser det ud i de enkelte afsnit, og vi prøver at kigge lidt på, hvad er det for nogle mennesker, der møder ind specielt herovre på vores Østtribyen og Sydtribyen, og i Udvendt-afsnittet.

Så det forsøger vi at følge ret godt. Så var der det her med brugsgrader. Her har vi vores forskellige produkter. Vi har vores billetsal. Det ligger rimelig pænt med 92% benyttelse. Det er jo selvfølgelig fordi, det er her, du går ind og køber en billet.

12Så selvfølgelig møder du frem, når du har betalt 150 kroner eller hvad det er. Hvorfor er den så ikke højere end 92? Det er fordi, vi også laver nogle aktiveringer. For eksempel, så laver vi en aflæreraktivering til alle folkeskolerne.

At de kan have en gratis billetter til en af vores kampe, hvor vi ikke forventer så højt et tidsskud at sælge på. Og der er vi godt klar over, at når vi laver gratis billetter ved aflæringen, så ligger vi måske nede på 65% i fremmøde. Så det er derfor, at den ikke er højere.

Så kan I tage vores erhvervsklub. De har taget et kæmpe hop. Og det er meget tydeligt på, at vi konkret går ind og arbejder med tingene.

Hvor her i de første år, altså i den første tid, der var der ikke så stor fokus på, rent faktisk om de produkter, de havde, og om de blev benyttet eller ej. Og så kan man se, at så tager man fat i det og begynder at kommunikere med det, stiller krav om, og hvad gør jeg? Fordi for eksempel erhvervsklubben, de har alle sammen adgang til spisning. Det er sådan ret nederen.

Og regne med, at der kommer 800 mennesker, købt mad ind til 800 mennesker, og så kommer der kun 600. Så det er sådan ret tydeligt, at det der man tager fat i, at det dur ikke. Og der er jo masser af gode historier med madspil og alt muligt andet, som man kan putte på som kommunikation.

Og det gør vi sådan ret tydeligt. Det samme med sæsonkortholderne. Som I ser, altså vi havde lige det her forår 23, det er altså der, hvor vi spiller medaldeslutspliller.

Det er Brøndby, FCK, Midtjylland og alle dem der, der er hver eneste søndag. Så derfor er benyttelsen højere. Men I kan se, at ved at feste og arbejde med det, der kan vi altså flytte nogle ting.

Så samlet på eventen, der formår vi at rykke vores benyttelsesgrader rimelig positivt. Og igen, hvis vi så tager det her efterår, vi falder lidt, så det skal vi have kigget på, det skal vi have øget fokus på. Det prøver jeg at komme med nogle eksempler på lige om lidt.

Øhm Her der er egentlig bare nogle forskellige visninger af, hvordan tilskuertallen er enige at fordele på erhvervskort, puljebilletter, altså nu snakker vi i partnerdelen, hvordan de lægger fordel på det. Så her har vi partnere i Hospitality og herovre der har vi partnerbilletterne. Og forskellen det enige, altså vi har nogle her med menu på, vi har nogle almindelige sponsorbilletter og vi har nogle 3F og vi har nogle aktiveringsbilletter.

13Og der kan I sådan se lidt noget om, hvordan vores fordeling den er på det. Så for eksempel til en, hvad det hedder, BIF eller en Brøndbykamp, der har vi 397, som har med menu på, 268 med bare en almindelig sponsorbillet, 3F-billetter, 124 aktiveringsbilletter, det vil sige, det er nogle, som partnere er ude at aktivere. Det kunne være kløverfonden, det kunne være forskellige ting.

Og så har vi samarbejdsclubber til sidst. Så I kan se, vi er ret detaljeret med dataen på de enkelte områder og forsøger at kigge på, hvad er tendenserne på det. Hvad det hedder, herovre er det erhvervsgårdene og hvordan ligger det fra kamp til kamp og i forhold til de målsætninger vi har.

Yes. Det var meget med, hvad skal man sige, vores data inden omkring billetter og partnere. Er der nogen, der har spørgsmål til det? Hvor langt tilbage har I? Jamen, så konkret i dataet, og der skal du nok hjælpe mig lidt, men altså, hvad skal man sige, jeg har næsten lavet den samme måde at opgøre tilskuertallet på alle 20 år.

Men dybden i dataet er blevet en anden, efter at vores billettsystem er blevet så detaljeret, som det er nu. Så hvad har vi? 4 år? Hvor har vi den her detaljeringsgrad? Og du sagde, data må også gå ind? Ja, godt. I forhold til abonnenterne, så har det sådan, det har været en ret klar strategi, at vi gerne ville have flere tilskuere på stadion.

Vi går rigtig meget på fællesskabet i Viborg. Vi går til fodbold, fordi vi ønsker at gå til fodbold sammen med nogen. Og det er sjovere for alle parter, vores partnere, vores spillere, vores organisation, så er det sjovere at have en ramme, der hedder 5-6-7.000 mennesker, end at lægge rodet alt for langt ned.

Så det har været en helt klar strategi, i forhold til at få udviklet den her abonnementsdel. Og som I kan se, så har det været eksklusivt. Vi starter med at kigge på og sige, at vi har en klar målsætning om, hvad vi gerne vil låne.

Den er vi forbi, så vi har været nødt til at lave en ny. Der var jo ikke nogen af os, da vi satte målsætningen efter corona, der tænkte på, at vi får 4.100 abonnenter, når vi har 587 års kort. Det var da ikke nogen af os, der så. Men vi går ind og laver en målsætning om, at vi skal runde 3.000. Og vi går ind og kigger på krigsstrukturen omkring det. Så vi går klart ind og siger, okay, køber I abonnementen, så kan I spare nogle penge. Så det første, vi går ind og gør, det er, at vi går ind og rører ved prisen.

Og så går vi ind benhårdt efter corona og siger, hvis du vil med til fodbold, så køber du dit abonnement, ellers kan du ikke komme ind på stadion. Fordi vi har begrænsning på pladserne. Det er ikke for min side, så jeg er billetmand, jeg skal ud og sælge stadion eller sælge stadion ud, der er det ikke dårligere kapacitetsmangel.

14Altså, det er ikke dårligere at have en lavere kapacitet. Fordi med en lavere kapacitet, så stiger efterspørgselen efter de pladser, der er. Og det er noget nemmere at sætte krav på det.

AGF, de er i gang med at bygge et helt nyt stadion. De er gået fra at have et stadion på hvert 24, eller sådan noget, til at have, jeg ved ikke helt, hvor de er endnu, 12.000 eller sådan noget. Og de havde flere kampe, hvor de var på ringe på de 24, så de er lige pludselig gået ned til de her 12.000. De har kunne tillade sig at bruge nogle år her på, hvor de mega meget opdrager deres, hvad det hedder, tilskud af deres fans.

Altså, det er kun billetterabonnementen dernede, og hvis du ikke bruger dine ting, så bliver du bare skubbet en gang tilbage i køen. Partner, helt firkantet. Henter du seks partnerbilletter, og du ikke bruger dem, så får du bare ikke adgang til partnerbilletter næste gang.

Og det er jo fordi, at de har jo lige nu en langt større efterspørgsel i det antal pladser, de har. Så det vi gjorde i

coronatiden, det var, du kan ikke købe en billet. Du kan købe et abonnement. Færdigbum. Og det var selvfølgelig med til, fordi hele vores opstil, den startede kraftigt efter corona. Og det var selvfølgelig med til at sparke det hele ind i gang.

Vi gik også meget ind på kundesegmentet, og kiggede på, okay, vi har vores sydtribune. Hvad er det for nogle grupper, vi vil have dernede? Vi har vores østtribune. Hvad er det for nogle grupper, vi har dernede? Og vi har vores østtribune, hvor vi har forskellige kapriskategorier.

Altså helt almindeligt, det er med bløde sæder, og det er noget dyrere. Hvor hvis du så går ud i siderne, så er det noget andet. Vi gik meget ind og kiggede på, okay, hvor skal vi have faste pladser hen? Hvor skal vi have frie pladser hen? I starten var det ret vigtigt, at vi havde nogle frie pladser.

Fordi at man kan gruppere sig, som man har lyst til. Men jo højere krav vi stiller, jo flere steder fik vi faste pladser. I dag er hele vores østtribune faste pladser.

Fordi det vil folk bare gerne have, og de vil gerne betale for det. Så vi har kunnet skrue prisen op, på hvad skal man sige, når man tilbyder mere og mere. Så i starten var frie pladser, det var at lave, hvad det hedder, pris.

Og så kunne vi skrue på prisen, stille og roligt, i forhold til hvordan vi udviklede de kundesegmenter, der er der. Og i dag, noget af det vi forsøger på, det er jo at putte menyer ned oven i, oven i billetprisen. Så det vil sige, at når du køber en sæsonkortapparat, så har du betalt for billetten, plus du har betalt for stadionpladsen.

Og det er der, hvor vi forsøger at ændre på det nu. Og så kan vi ikke gøre det her, uden et ekstremt godt samarbejde med vores kommunikationsafdeling. Det er jo dem, der bestemmer budskaberne, og det har virkelig givet både.

15I forhold til hele den her udvikling på datasiden, i forhold til de elektroniske hvad det hedder, elektroniske muligheder, så var vi det første stadion i Danmark, der gik over og havde NFC-skaling. Så det vil sige, at alle vores kort, billetter, har i dag NFC indarbejdet, og det betyder jo, at når man går ind og skal ind på stadion, så

kan man bare lægge sin telefon op på møllen, og så er der kommunikation med det. Jeg siger ikke, at den normale

skanning med QR-kode og alt sådan noget ikke fungerer, men det er et kæmpe fordel med NFC-skanningen, fordi mange har nedtonet med lys på sin skærm, eller der er nok også et par i jeres skærm, der er smadret, og sådan nogle ting, men der er der bare ikke noget i forhold til NFC-skanningen.

Og vi kan se det ekstremt meget på vores indgangsdata, hvad det hedder, vi har aldrig købt, og det er sådan noget, jeg aldrig siger, fordi det er bare der, hvor man virkelig falder i, hvis det er sådan, at det er en gang, men altså kvarter før tid, da der er vores indgangen, der er det ryddet, for det går bare stærkt, man kommer hurtigt ind. Så det her, det har været et rigtig fedt. Plus, det var da også fedt at være de første, ikke, altså det plejer at være FCK og Brodby og dem der, men så er du sgu lige pludselig lige, hvor det står frem.

Og det var også sådan, at det er jo sådan noget, der skal godkendes, det er jo, det her NFC, det er jo et Apple-produkt, så det skulle godkendes af Apple til jo tre uger, altså fordi vi skulle huske at sætte skille-app det ene og det andet sted og frem og tilbage, fordi vi var de første. Så det har været ret fedt. Så har vi fået lanceret en ny app nu her, selvfølgelig hvor det her samarbejde med billetsystemet også er, således at du kan hente dine korte og dine billetter ned i din app, nede i wallet'en i app'en der.



Og specielt den del der med at kunne håndtere sin kort deri, så man kan sende det som e-mail eller man kan frigive pladsen, det forventer vi os helt vildt meget af og skal have dyrket meget mere nu her, fordi jeg siger ikke, at den profil man har, hvor sin sæsonkort ligger inde på eventidelen, jeg siger ikke, den ikke fungerer, men det er bare en barriere, du skal via en browser ind på en side og så først finde det. Nu ligger det bare direkte her i app'en og håndterer kort, så kan man sende det som billet, eller man kan frigive. Og det der frigivet er, hver eneste gang vi kan få folk, der ikke kommer til kampen til at frigive, så kan vi sælge pladsen igen.

Og det er bare vigtigt. Fordi vi er så højt af i abonnementer og det der, at vi har altså nogle gange kapacitetsproblemer i forhold til at sælge og sælge, og derfor så er det ærgerligt, hvis det er sådan, at der er nogen, der ikke får frigivet sine pladser. Så vi bruger belønningen til nogle kampe, belønningen for at få frigivet og for eksempel her til Vejle, der kom nu her, så er alle, der har en brugskred over 65, får lov at tage en melding til kampen, fordi de mener, der er plads der.

Og så den her send billet via e-mail, det er jo sådan, at vores kort, de er ikke personlige. Så I må godt være to herinde, der deler kort. Der er kun en, der kommer ind, men man må gerne dele sin billet, man må gerne sende sin kort videre til en, vi ønsker bare, at pladserne bliver optaget, hvis de ikke er frigivet.

Det her, altså både den del og det her med brugskreder, det kræver jo også bare, at vi har styr på vores kortdata, altså vi har styr på datan, vi har styr på scanningerne og fordi helt tydeligt nu her inde på deres profiler, så ligger deres brugskreder. Så Bent her, han er en af dem, der stadigvæk er tilbage med sæsonkort. Han har 100%. 16 Så seks kampe her i efteråret, fremmede seks gange, og ti kampe i foråret, så han ligger på 100%. Så holder han den her, de gør han, frem til hvad det hedder, jul, så leger vi have sådan en kort, så de kommer ned i krus eller et eller andet, kommer ned i vores kløver, sørger for hinder. Og det er sådan en stor ting, men det er en tilløsning, det er noget folk giver lige i poster bagefter.

Så er hun også en krusetjeneste. Så det er, ja, samtidig med, så kan man, hvad det hedder, optjene kløverpoinge. Det vi bruger kløverpoingene til indtil videre, det er for adgang til for eksempel kick-off events, vi har, eller udebærede ture, så der skal man optjene nok poinge.

Man får selvfølgelig point, som man kan se her, der får, han har fået 10 point for at benytte sit kort. Hvis man frigiver, får man også point. Hvis man deler sin kort, altså sender billet, til en ven, og vennen bruger kortet, så får man også point.

Så på den måde, så er der en hel masse opsamlings af data, og vi forsøger at gøre noget, hvad skal vi sige, synliggøre deres benyttelse, og også forsøge at give dem noget med, og sådan noget. Ja, lidt ejerskab, og lidt bonus for det. Ud i fremtiden, så kunne det være, at man kunne købe de bordene for det.

Der er vi ikke endnu, men det var så nogle steder, hvor vi kunne udvikle på det. Ja. Det var det, jeg havde tænkt mig at sige.

Jeg ved ikke engang, hvor lang tid jeg har brugt det her. Som altid perfekt på natten. Så jeg ved ikke, om der er nogen af jer, der har nogle spørgsmål på det her.

Hvordan personligt ser det ud, som en succes for en kamp? Hvad er dit mål for en største succes? Mit mål for en større succes? Vi har jo nogle kopimål. Jeg er selvfølgelig på tilskuertal. Vi har noget, vi skal omsætte billetmæssigt.

Vi har noget, vi skal omsætte i koordinerne. Det vil jeg selvfølgelig rigtig gerne opnå. Men en rigtig, rigtig god

kampeafvikling for mig, det er også, at der er ro på omkring banen.

At vi har styr på sikkerheden i specielle udebaneafsnitter. At vi ikke har nogen issues i forhold til boederne med det personale, der er og kontaktilskuerne. Og også, at vores boedepersonale har en god og en hyggelig dag.

Vi er egentlig i situation, at de folk, der arbejder hos os, de er som regel nødt til at have to, måske tre jobs.

Fordi vi har 16 events om året. Så når folk kommer på arbejde hos os, og du ikke lige er frivillig og du brænder på det den vej der, men du kommer som timelønne, skal du helst have det godt.

17Så det er sådan en blanding af de der ting. Men selvfølgelig vil jeg gerne opnå vores kopimål på det. Men det er også lige så vigtigt, at vi har ro på, at vi er godt forberedt, så det kører.

Det er sådan, det er sådan, at vi har betragtet sådan lidt vores kampeafvikling som en stor tandhjul. Vi har rigtig, rigtig mange små ting, som skal fungere. Og for at det her tandhjul ikke kan køre rundt, så burde det ikke, at vi har et eller to områder, hvor tænderne er knækkede.

Så vi skal hænge sammen hele vejen rundt. Og det er ret vigtigt for mig personligt. Og så er det ret vigtigt, at vores fællesskab er åbent.

Man kan lade op med en spørgsmål. Mål af sig for det her med, hvor meget brug har der været, eller har der været nogle uheld? Ja, hver år, der bliver der lavet i Superligaen, der bliver der lavet sådan en Superliga-survey, hvor klubberne bliver, altså hvor tilskruerne til klubberne bliver spurgt. Hvor man kan måle sig op mod de andre klubber.

Så benchmarken og dem der. Udover det her, så har vi også kørt nogle tilskruer-undersøgelser for to, halvanden sæson siden, hvor vi prøver at arbejde med det. De der tilskruerne, synes det er noget, vi gerne vil gøre lidt mere i. Og derudover, så har vi også undersøgelser i vores bodepersonale og sådan noget, der er med feedback tilbage på, hvordan det ligger.

Men der er ikke noget fast... Den der Superliga-survey? Ja, Superliga, men der er ikke noget fast skete med intern. Nej. Nej.

Men det kunne være et fedt punkt at få udviklet. Ja, for der kunne man jo højst synes også have mere at måle på. Ja.

Betyder det, at der er mere salg, hvis der er ro? Der er mere salg, hvis der bliver scoret en masse mål, og der er spænding og sådan noget. Nej, men grunden til det der med roen, det er bare, at for mig, hvis der er ro på stadion, forstår du, altså i min del, jeg snakker ikke, at der er stille på stadion, der skulle gerne være stille, men når jeg snakker ro, så er det, fordi vi er på forkant. Så løber vi ikke tør for ting nede i boderne, altså.

Så er vores boder pakket ordentligt, og når boderne åbner, eller billetsalget åbner, så er klargøringen i orden.

Så møder folk ind til noget, der fungerer. Det er jo egentlig meget sjovt, hvis du tager sådan en dårlig oplevelse i en bud, hvis man som kunde kommer ind der.

18Altså, der er jo rigtig mange berøringsflader. Du kan starte med at komme i fanzonen, så det vil sige, at den åbner tre timer før, så derudover,

Denne fil er længere end 30 minutter.

Opgrader til Ubegrænset på TurboScribe.ai for at transskribere filer på op til 10 timer.

## 11.4 Bilag Interview “Vff del 1 - Daniel”

Vff del 1 - Daniel

Transskriberet af TurboScribe.ai. Opgrader til Ubegrænset for at fjerne denne meddelelse.

Jeg forsøger at optage det, og jeg skal nok lægge det ud til alle, så der ikke er nogen grund til, at vi alle sammen sidder med det. Dennis, hvad er det, du hedder? Daniel. Daniel fra BFF.

Er det i orden, at jeg optager forløbet her? Selvfølgelig er det, og vi vil også gerne gemme det til Bo. Tak skal du have. Men måske skal der være en mere, der gør det? Rigtig juridisk, fordi jeg selv er med, som Bo er.

Jeg tror, der er to, der optager nu, men I er også velkomne til at optage lørdag. Vi tager den lige i planum. Er der nogen, der er meget imod, at der ligger en mikrofon heroppe til at optage lyd? Ingen indsigelser.

Jeg tænker, det er godt nok. Er det okay, at jeg også gør det? Ja, det må du også gerne. Godmorgen.

Nu har jeg lige fået en lille rundvisning af virksomheden. Mit navn er Dan, og jeg kommer til at holde lidt længere oplæg, som freder om, hvordan vi arbejder med data i en gruppeklub. Så kommer der nogle lidt mere specialister ind senere i dag.

I kommer bl.a. til at høre fra Palle et par gange. Frem til frokost, så vil der være meget også, der holder nogle oplæg på jer. Efter det, så vil der være noget interview over på stadion.

Jeg kommer mere ind på det lige om lidt. Men vi er sindssygt glade for, at I kommer her i dag. Og over på den bænk der, der sidder tre, som har taget den uddannelse, som I sidder og tager lige nu.

Og I skal bare vide, at som virksomheder sidder vi sindssygt glade for de kompetencer, I kommer med. Og nu er det jo tredje gang i dagen, at jeg kommer på besøg. Så jeg håber, at vi bliver bedre hver gang.

Men det må være op til de tre hårddommere at udvidere det. Men programmet ser sådan her ud for i dag. Vi er indkommet.

19Der er en kort rundvisning. Så laver jeg en præsentation på ca. 45 minutter.

Det er selvfølgelig helt i år, at vi optager og bruger det til vores opgaver. Selvfølgelig er det det. I får de slides, jeg har med i dag.

Og generelt alle de slides, I bliver præsenteret for. Er der noget, som er forretningssimpel. Jeg tror, at jeg har et enkelt slide, hvor der er noget andet til at fylde vægt.

Så tager jeg det ud. Ellers så får I slidesene i sin egen. Jeg har prøvet at lægge nogle pauser ind.

Både til at man kan tanke lidt op på kaffe osv. Har man nogle spørgsmål under oplæggen, så er man meget velkommen til at stille. Så går Olga, Svine og Anja på kl.

10.10. Så er der noget pauseværk igen. Så går Palle på i en halvtimes tid. Så har jeg lagt en forkostpause slads. Nogle gode ture til stadion. Det er fordi, vi skal over på stadion i eftermiddag. Jeg kommer til at vise det rundt på stadion.

Og fortælle lidt om nogle af de ting. Eller vise frem, hvad det egentlig betyder. Fordi jeg går ud fra, at der også er nogle, som ikke er helt ligeså meget enige i fodbold.

Og hvad betyder det her. Så kommer jeg til at snakke lidt om det. Så bliver Palle intervjuet.

Det bliver sammen med Olga også om at bringe noget billetdata. Så vil alle os fire fra dataafdelingen være til råde for noget intervju. Og så runderen i 14.30. Så skulle der måske lige være, at man sidder tilbage med nogle

spørgsmål efterfølgende.

Når der kommer noget næste uge, og hvordan var det lige med det der. Så er I meget velkommen til at række ud til en af jeres tre undervisere. I forhold til at sende spørgsmål til os.

Så er det meget vanvittigt. Vi har fået hjælp, hvor vi kommer til at være så transparente, som vi overhovedet kan. Der er ting, vi ikke må sige.

20Der er ting, vi ikke er rent ude i skoven. Men vi havde en generalforsamling i går. Hvor der er rigtig mange ting,

der bliver lagt på bordet.

Så kom egentlig også bare. Men det kan være, at vi henviser til nogle andre. Hvis det ikke lige er overhovedet.

Cool. Jeg vil ikke gøre det så langt. Men bare lige sådan, så I ved, hvem det er, der står og præsenterer.

Jeg er 29 år i Danmark. Jeg har tidligere arbejdet sammen med Simon op i AAB. Ikke i helt lignende funktioner, som jeg har nu.

Men mere som en analytiker. Hvor man filmer træninger og kampe. Og hjælper omkring førsteholdet.

Jeg har været her i tre år. I nogle forskellige roller, da jeg startede her. Så var jeg en, der rejste med førsteholdet ud.

Og filmede kampe. Filmede træninger. Jeg synes, at trænerne havde de bedst mulige forudsætninger.

For at bruge video live under kampe. Men også under træninger. I dag, så har jeg en tilladelse Head of Data & Technology.

Og det betyder, at jeg varer sig af rigtig mange forskellige opgaver. Men mit hovedopgave, det er. Hvordan vi indsamler, opbevarer og bruger data på tværs af funktionen.

Så jeg har en kandidatgrad i økonomi fra OpenSD. Yes. Og det er lidt af det, jeg kommer til at snakke om.

Jeg bliver tit mødt med det der med. Hvad er det egentlig, man arbejder med data i en fodboldklub? Og hvad er fodbold egentlig for noget? Så hvis der er nogen, der har nogle spørgsmål. Hvor jeg tager noget for givet.

Det kan være, hvad laver en skavlingafdeling? Ræk ind i hånden op. For der kommer helt sikkert til, at folk har de samme spørgsmål, som jeg. Yes.

21Det var, hvad jeg var lidt inde på på rundturen. Jeg deler det sådan lidt op i to sektorer. Altså vi har administrationen.

Det er alle dem, der varer sig af opgaver. Som er uden for fodboldringen. Det vil sige, at det er dem, der sørger for, at der er bakker til på stadion.

De har billetter. De sælger sponsorater. Når vi kigger op på stadion.

Her er et billede af stadion. Og I kan se, der er solgt nogle sponsorater til nogle virksomheder. Så skal vi have nogle sælgere ud og aktivere de her virksomheder.

Som kan sørge for, at der bliver lavet nogle partnerskaber. Det er cirka halvdelen af vores virksomheder, der gør det. Den anden halvdelen af virksomheden tager altså et kerneopgaver.

Det er det, der har noget at gøre med fodboldspillet. Det har noget at gøre med fodboldbanerne. Jeg kommer også lidt ind på, hvordan vi deler det op på data.

Men der er selvfølgelig nogle trænere. I dag er der nogle mentaltrænere. Altså hvordan trives spillerne.

Der er nogle fysioterapeut og læger. Som tager sig af skader og problemstillinger. Fysiske trænere, der måler,

hvor langt løber vi.

Og så er der analytikere, som laver mange forskellige ting. Men også er det nogle kovireporter på data. Og så rigtig meget video.

Altså hvordan understøtter vi beslutningstærnerne med den video, vi skal bruge. Administrationen først. Jeg opdeler det sådan i fire afdelinger.

Vi har en driftsafdeling. Det kommer I til at høre en masse fra i dag. Og palde.

Det er en trafik af mennesker, som sørger for, at der er mad i båden. Der bliver solgt billetter. Der er nok personale på stadion.

22Alle de ting, der skal til for at facilitere en fodboldkamp. Det vil sige, når vi på søndag har en hjemmekamp. Hvordan sørger vi så for, at der er det rigtige i båderne? Hvordan kan vi få aktiveret banes? Hvor mange billetter har vi ude? Hvor mange frivillige betyder det, vi skal? Alle de her ting, der skal til for at kunne facilitere en god ramme.

Og bringe fodboldkamp. Det står der i forhold til. Salg.

De står for aktiveringssponsorer. Det vil sige, at de taler med virksomheder. I forhold til, hvordan kan vi finde et eller andet partnerskab.

Hvor vi kan tilbyde noget. Ofte er det, at vi kan tilbyde en unik platform. I forhold til, at vi er den mest eksponerede virksomhed i kommunen.

Selvom vi ikke er nær den største omsætning. Så er der rigtig mange, der interesserer sig for os. Så når man hænger et skilt op på stadion, så bliver det set af rigtig mange mennesker. Så vi har en platform, vi kan sælge på. Og så tit er det også noget med, at man går på tændingserførelse.

Som en del af det. Altså den stemning, man kan skabe på stadion. Hvordan kan man tilbyde noget til partnerne. Som de ikke kan få ved at købe almindelige billetter. Det står salg for. Det er samarbejde.

Når der bliver solgt billetter. Til sponsoraterne. Sponsoraterne, hvordan bliver det så ført over drift.

Det tror jeg, at Ole kommer mere ind på senere. Så har vi en afdeling, som hedder. Marketing og kommunikation. Men under den paraply. Så er der marketing og kommunikation. Der sidder.

Nu glemte jeg at sige, hvor mange der er i salg. Der er tre fuldstændig sælgeansatte. I marketing og kommunikation.

Der sidder to content creators. Det vil sige, de er ude at filme og optage vores kampe. Så man kan lægge noget godt indhold på sociale medier.

23Så er der marketing. Så er der en ansvarlig for vores LinkedIn. Det vil sige vores virksomhedskommunikation. Og så er der en generel. Hvad kan man sige. Kommunikationsansvarlig.

Der er rigtig mange medier, der godt kunne tænke sig at høre. Hvordan går det med det. Hvordan sørger man så for, at man omligner sin kommunikation.

For at have et tydeligt billede. Så under det. Så er der også events og merchandise.

Events er jo sådan lidt en. Det er jo det, man selvfølgelig ikke skal gøre. Når man præsenterer.

Så er det i hvert fald langsomt. Events er jo. Kan man sige.

En del af det med at. Altså når der er noget, som en kund. Der tager på søndag efter kampen.

Hvor der bliver solgt brugt sportstøj. Så har vi faktisk en medarbejder. Der er ansat til at facilitere rammerne.

For at vi kan lave andre ting end blot. Hvad kan man sige fodboldkampe. Så når vi senere i dag går op og kigger på vores stadion.

Så har vi fået en ny lounge. Som kan hoste små 200 mennesker. Hvis der skal være konferencer og andet.

Så kan man komme derop hver dag. Vi har kun 16 hjemmekampe om året. Og på 365 dage, så er det ikke ret mange.

Når vi egentlig er fodboldkampe. Hvordan kan vi få aktiveret vores stadion. Og lave flere ting, som vi kan sælge. En af de muligheder er at have nogle personale. Som kan bruge de rammer vi egentlig har. Så når vi har nogle lounges.

24Skyboxe og så videre. Hvordan kan vi så på en eller anden måde. Skabe nogle rammer, der gør virksomheder. Kan se det som mødelokaler. Som konferencerokaler. Der skal vi også have nogle til dem.

Til merchandise, der har vi tornser. Som vi har anbrudt nede i byen. Som er om onsdag og torsdag.

Og så på kampdag. Det vil sige, hvordan får vi noget af de trøjer vi gerne vil have. Nu sørger jeg lige for at bevæge den her.

Så den ikke går i sort igen. Hvad hedder det? Hvordan får vi lavet nogle trøjer i samarbejde med vores trøjsponsor. Som er Capelli.

Hvordan sørger vi for at have noget udstyrt. For alle de fans, som går på tænk. Hvem er det, der køber handsker og huer osv.

Det er merchandise, der står for det. Økonomien. Der er en politismedarbejder, som hedder Pia.

Hun bor dernede. Og så vores CFO, vores økonomidirektør. Og jeg kommer lidt nærmere ind på, hvad de laver også.

Ja, så Palle. Palle kommer selvfølgelig til at gå lidt mere i dybden. Med præcis, hvad de laver senere.

Og så vil Holger også fortælle lidt om det. Men det, det handler om, det er Græften af broder og billetter. Sankt aktiviteter på stadion.

Det vil sige, når vi har en fanzone. Hvem er det så, der skal sørge for, at den har de rigtige ting i broderne.

Hvordan sørger vi for, at der er noget folk op, der kan sørge for at drikke.

Her er der data på broder. Det vil Subino måske afdække lidt mere bagefter. Og så er der nogle data på billetter, som Holger er ekspert på.

25Så det vil ikke være, som om jeg har udstyr til det. Selv, der er Det er lidt det her med, at vi har nogle sælger, der tager ud til virksomhederne og sørger for, at vi er stærkt repræsenteret af det hverkermæssige virksomhedsnetværk, der er her i Viborg. Vi kunne godt tænke os at være noget på byen.

Og det vi bedst muligt kan tilbyde, det er en platform, hvor man kan blive set og hørt. Og man kan skabe forbindelser på tværs af forskellige virksomheder. Det vil sige, at vi laver forskellige netværker.

Hvis man er en tømrervirksomhed, så kan man møde en mæglervirksomhed, så kan man møde en bankvirksomhed. Så vi kan sætte fund sammen. Det er noget af det, vi kan tilbyde.

På datasiden, så har vi et serialsystem. Hvor vi registrerer aftaler, hvor der styrer på produkter og indstillinger osv. Og en nøgle her er, at det her data tit bliver født hos os.

Det vil sige, at når der er en ny partner, som starter hos os, så bliver de lejet ind i serialsystemet. Men vi har et separat billetssystem, som hedder Eventi. Og hvordan sørger vi så for, at de to ting er godt integreret.

Men som regel, så bliver B2B-kunderne, de bliver lejet i vores serialsystem. Lagt til en kommission. Som jeg sagde, så handler det rigtig meget om, hvordan kan vi lave noget indhold, altså noget content, som vi kan eksponere på sociale medier.

Vi har, hvad hedder det, fra sidste måned af, eller starten af den her måned, har vi lanceret en ny app, hvor man kan få sin billetter igennem, men også hvor vi kan tilbyde noget videoindhold, som vi ikke kan tilbyde på vores sociale medier. Her, der har vi en masse data på, hvordan B2B er foldet sig rundt på medierne. Vi kan se på vores forskellige platforme, altså hjemmesider, hvad de trykker på, hvad de kigger på, så videre.

Vi har registreret, hvor mange af det, når vi laver et opslag på sociale medier, hvor mange af det, så der rent faktisk reagerer på det. Jeg kan afsløre, at når vi vinder, så er der rigtig mange, der reagerer, men når vi taber, så er der rigtig mange, som skal fortælle, hvor dårlig vores præstation var, hvor dårlig maden var i broderne. Et sjovt eksempel på det med fodbold og løvelser er, at vi har en lounge på kampdag, hvor folk kommer op og spiser og så videre.

Og så har vi nogle gange spurgt, ude i spørgeskemaet, hvad synes I så om maden? Og hvis vi har vundet, så er maden altid god. Altid god, fantastisk mad. Og hvis vi har tabt, så er det lige meget, hvad vi har lavet.

Så det er den dårligste mad, der nogensinde har været lavet, og ideelle klyder hælder op og alle sådanne ting. Så vi er også en virksomhed, som er relativt præget af, hvordan går det egentlig ned på fodboldbanen. Og det er ikke noget, vi kan styre, men det er bare sådan, det er.

26Jeg var lidt inde på den app der. Igennem den, så får vi også indsamlet en masse, kan man sige, stamdata på bands, som vi kan bruge i et marketing-solvek. Det vil sige, at vi bliver meget mere skarpe på, hvordan er det, vi lige præcis skal målrette vores kommunikation ind imod de enkelte bands.

På sigt, så er planen sådan set, at vi godt kunne tænke os at blive endnu skarpere på det, så når vi taler med sponsorer, at vi kan være meget tydelige omkring, okay, du er en bankvirksomhed, vi har lige nu rettet de her bands, som I kan kommunikere til igennem os. Det er et mål, vi har på, og så er der data på merchandise, altså hvad er det for noget, vi sælger, hvornår er det, vi sælger det, hvad er det, vi skal være opmærksomme på. Og I er meget velkomne til at stille spørgsmål om det også.

Yes. Yes, økonomi, det er en afdeling, der skal være der. Som I kan se, det er her, vi betaler løb.

Det er her, der bliver afregnet frakturer osv. Jeg vil ikke komme for meget ind i det, det er meget sådan klassisk, men nogle gange så glemmer man, at selv en fodboldklub skal også have en medarbejder, som sørger for, at vi holder styr på de her ting. Men det består primært af fire vores økonomimedarbejdere, som løser mange ting, og så vores CFO, Lars.

Ja, sporten. Og her kan jeg godt komme til at tage mange ting for givet, så inden I har nogle spørgsmål omkring, hvad er det, hvad er det, så spørg ind. Den første afdeling, jeg vil komme ind på, det er den, vi kalder performance, og i mange fodboldklubber, så er performance noget med fysisk data, men den kalder vi physical, hos os.

Ja, der er det, man kigger på, der har vi en analytiker på førsteåret, vi har en analytiker ansat på akademiet, som kigger på, hvad er det egentlig, for noget, der afgør fodboldkampe? Hvad er det for nogle kopier, vi skal have? Rettet op imod, hvordan er det, vi gerne kan tænke os at spille fodbold? Hvad er det for nogle ting, som afgør fodboldkampe? Så kigger vi på over tid, hvordan har vi klaret os? Fodbold er meget sådan en sport, at

du kan have en rigtig god kamp, du rammer soldaten fire gange, den går ikke ind, modstanderen løber op, og laver et heldigt mål, og så har du altså bare tabt 1-0. Du kan synes, du krydser dig tilbage og tænker, vi har virkelig gjort alt, hvad vi kunne for at vinde, men på dagen var vi der bare ikke. Og hvordan er det så, at vi sørger for, at vi holder øje med, hvordan vi præsterer? Godmorgen.

Hvordan har vi klaret os? Så kigger vi på lidt mere over tid. De ting, vi godt kunne tænke os at være gode til, er det så rigtigt faktisk de ting, vi er gode til over tid? Det kommer jeg mere ind på lidt mere. Benchmark vores præstation op med andre klubber.

Der snakker man meget om de ting, der afgør fodboldkampe. Hvor ude er vi til det? Når vi skal ned fra AC Horsens i dag, hvordan er vores styrkeforhold så op imod dem i forhold til præstation? Og hvad for noget data har vi til rådighed? Her har vi fuldstændig vanvittige mængder af data. Vi har nogle dataleverandører på Superligaen, som gør, at vi egentlig har mere data til rådighed, end vi har tid til at kigge på.

27Så vi har i dag 12-13 millioner rækker data, der kommer bare fra én leverandør. Udover det, så har vi 2-3 andre leverandører, der leverer data til os. Udover det, så har vi nogle ejere, som sidder i London, men som er danskere, som også producerer data til os.

Som ikke fodboldnørd, når du så siger data, hvad er det for noget, vi taler om? Er det antallet af skud på mål? Er det et hurtigt beløb? Det er et rigtig, rigtig godt spørgsmål. Jeg deler det op i to. Det, der sker, det kan for eksempel være skud på mål, måltakling og alle sådan nogle ting.

Indkast, målspar. Og så er der tracking data, det vil sige, du har XY-koordinaterne på alle spillerne samtidig. Det vil sige, du kan faktisk gøre stort set, hvad du vil.

Fordi du ved fuldstændig, hvor alle spillerne er, og du ved, hvor bolden er. Så der kan du sådan set alt efter dine kompetencer, kan du sådan set lave det data, du vil. Rigtig mange fodboldklubber bruger venddata, fordi det er lige til, at du har noget færdigt data, som er processeret.

Men i virkeligheden, så burde vi gå mere, hvad kan man sige, tracking data vejen. Problemet er bare, at det ikke handler om, at du har nogle kompetencer i virksomheden, som kan bruge det til lige nok, det du vil. Så vi har egentlig begge ting til rådighed i forhold til målbrugspræstation.

Vi bruger noget af det her tracking data, men er ikke kommet langt nok med det. Og det handler simpelthen om ren manpower. Og i dag har vi faktisk også skeletet data, som er, at vi får til alle fodboldkampe, så får vi 25 punkter på koppen på fodboldspillerne, som faktisk gør det i stand til os, så det egentlig ligner FIFA eller et spil.

Så vi kan se, hvordan står spillerne hele tiden i løbet af fodboldkampen. Vi kan se, hvordan de løber. Alting. Og det har vi sådan set data på. Før det, jeg skal nok komme til, så havde vi egentlig kun XY på spillerne, og så havde vi XY også sæt på bolden, så vi vidste, om bolden var op i luften. Men i dag har vi faktisk sådan, så vi også kan sige, og det har vi på hele sidste sæson også.

Det eneste, vi bruger det til i dag, helt lavet faktisk, det er, at vi kan se situationen på den vinkel, vi gerne vil. Altså, vi er... Hvis man forestiller sig, at skud på mål i en fodboldkamp, så tit tager det eneste, man har til rådighed, det er videobillederne. Men det kan godt være lidt svært at se.

28Kunne målmanden se skud, kunne han ikke se skud? I dag, der kan vi bare gå ind, bruge skilitedata, dreje det,



og så kan vi se, hvad kunne målmanden bare kan se i det øjeblik, bolden er sparket. Så det er brugbart, men vi har... Altså, der er et meget, meget stort potentiale, som vi bruger i dag. Rigtig fedt.

Jeg tror, det er herovre først. Jeg ved ikke... Hvordan bliver alt det der data indtastet og opdateret? Er der noget, eller er det noget, der sker automatisk? Ja, altså... På Second Spectrum, som er dem, der laver tracking, så er der otte kameraer på stationen, som egentlig indsamler det fuldt automatisk. Der er nogle forskellige tjeke og sådan nogle ting der, fordi hvis I forestiller jer, at I fandt efterårsvær, forårsvær, så kan der godt være noget toge, og nogle gange kan der også være noget pibe på stationen.

Det kan være lidt svært at se, hvor er alle spillerne samtidig. Så der er nogle forskellige tjeke, som man kan køre igennem. Og alle klubberne spiller ikke med samme trøjnummer, så der kan godt være nogle trøjnummer, kameraerne kan svære ved at tracke en eller andre.

Så er det nr. 1 eller nr. 7, eller sådan en ting.

Er det nr. 17 eller nr. 11? Altså, det er ret, ret svært.

Så det kører fuldt automatisk. Sådan udgangspunkt. Så har vi eventdataen, vi har lavet igennem, hvad kan man sige, taggingmanuals.

Så de firmaer, der har det, har en fast procedur for, de får nogle videobiler, og så tester de egentlig ind alt, hvad der sker. Så det sker manuelt hos dem, men vi får uddannet den, kan man sige, efter forlænet. Hvad laver vi? Ja, og så laver vi også selv noget data, men det får vores manuelle udviklere derom.

Næste folkeuddannelse gør jeg mig derom. Hvad er det for nogle mennesker, hvilken baggrund har dem, der sidder og kigger på alt herinde under kontrol? Rigtig gode spørgsmål. Øh, begge to har trænerbaggrund, så der er ting, og det ved de godt, fordi vi har internet med, men de ved godt, hvad det er, deres kompetence selv er, og hvad det er af vores dataerhjælp.

Så vores, det er datainfrastrukturen sørger for at sætte det rigtigt op og sådan nogle ting. Deres kompetence er, om data rent praktisk viser det, som de går på tændelse af. Så for eksempel, hvordan kunne vi godt tænke os at spille fodbold? Jamen det er dem, der har bedre tilsæt med øjnene, og så hvor vi hjælper dem her pænt i dag, så er det dem, der støtter det.

Ja, så det er de to, vi har, har trænerbaggrund i analyseforholdene. I forhold til det data, I får ind på jeres spillere,

især fordi det er et sted for dem. Hvor i det også for andre klubber? Ja, altså til vores kampe i Superligaen.

29Der spiller vi imod Fredericia i aften, og det er altid sådan lidt tricky, fordi de spiller også i Superligaen, vi spiller på et Superliga stadion. Hvorfor får vi det så ikke på den kamp? Fordi kameraet er jo faktisk på stadion. Det gør vi ikke, fordi aftalen er, at vi får det til alle vores Superliga kampe.

Til alle vores Superliga kampe, der har vi det på både os selv og modstanderen. Hvorfor? Og det gør, at vi kan sådan set gøre med data, hvad vi har lyst til. Vi ved også, hvor langt hvert hold har løbet, osv.

Ja. Fedt. Det her, det er, der er noget af det, jeg kommer til at tage ud i forhold til det, jeg sender til jer, men det er bare et eksempel på, at tidligere, så har vores KBI'er blevet lavet i Powerpoint og et program, der hedder Keynote, og det har været datainfrastrukturen.

Det var der, det er blevet kæmpet. Og så hver eneste gang vores analytikere, de der skulle lave en opsummering af, hvordan har det faktisk gået de sidste 10 gange, så havde de skulle tage Powerpoint-dokumenter og fremskrive

nede i hånden over i Excel. Alle sådan nogle ting.

I dag, der i samarbejde med Olga, der har vi egentlig gjort det sådan, at nu kan de bare komme og bede om, hvad de har lyst til, og så har vi fået datainfrastrukturen her, og så er det en ret tæt samarbejde med dem om, hvordan skal det her se ud, så det viser de ting, de skal bruge. Så nu kan de bare skrive til os, jeg skal bruge vores KPI'er på de sidste 10 kampe, så kan vi give det til dem. Hvor tidligere så var det en manuelt proces, der godt kunne tage en halv dag.

Så skulle han åbne alle filerne, måske printe dem ud, skrive over i, hvad hedder det, i Word måske, sætte over i Excel, få coderne til at virke i Excel. Det var et større arbejde for dem. Og det er helt færdigt, fordi det er forskellige kompetencer selv.

Men det ser sådan ud her i dag. Den grønne er os, den orange det er modstanderen, og forskellige KBI'er, som vi kigger på. Færdsruer bliver udbudt af vores ejere, så det er et mål for, hvem var egentlig bedst i den her kamp her.

Der kunne vi godt tænke os, at den grønne er højere end den orange. Det giver os en større sandsynlighed for at vinde fodboldkampen. Det betyder ikke, at vi har vundet, det betyder, at vi har skabt forudsætninger for at vinde, eller bedre forudsætninger for at vinde.

Det her, det er ting, som vi ved statistisk underbygger, at man øger sin sandsynlighed for at vinde. Det betyder ikke, at man vinder, men at være bedre på de her ting her, øger vores sandsynlighed for at vinde. De ting, der er her, de er mere bygget op på, hvordan kunne vi godt tænke os at spille.

30Og hvorfor er de to så ikke de samme, er et rigtig godt spørgsmål. Fordi, hvordan vi godt kunne tænke os at spille, skal jo gerne bygges op på, hvordan vinder man fodboldkampen. Det vil give meget god mening.

Der er bare nogle ting, som godt kan have afledte effekter. Det vil sige, hvordan vi godt kunne tænke os at spille fodboldkampen på førstehåret, kan gøre det nemmere at finde nye fodboldspillere. Det kan gøre det nemmere for hele vores akademifdeling at bygge et program op, hvordan man skal lige faktisk træne.

Så der er nogle afledte effekter af det her, som kan ses på tværs af den sportlige sektor. Og de vide her er mål for præstationen. Så er der fysiker, og jeg har lige taget billeder af en spiller, der i dag spiller AGF, Tobias Beck. Han har navnet med en KBS-vest der. Han har skruer. Det, man ser her, kan godt ligne netop en sportsbog. Men omme bagved, der sidder en KBS, som sådan set måler X, Y og alt, hvad han laver. Så det er ligesom, hvis man løber med et løb. Det, vi får ud her, det er over 200 datapunkter fra hvor langt han løbet, hvor stærkt han løbet, forskellige toner, hvor meget acceleration og deceleration han har lavet.

Og nu når vi jo det før, i forhold til at vi får data på modstanderen. Det gør vi jo ikke på vores KBS'er, fordi vi udleder ikke KBS-data med de andre klubber. Det vil sige, vi ved kun, hvad de andre har lavet til kamp. Det ved vi ikke. Vi ved ikke, hvad de har lavet til træning. Nu har vi to forskellige systemer lige pludselig, som begge introducerer fysisk data.

Hvordan sørger vi for, at de er sådan nogenlunde ens? Hvad kan vi bruge det her til? Jamen, vi kan tage KBS'erne med overalt. Så om vi træner herude, eller om vi træner på stadion, eller hvis vi skal på træningslejr i Spanien, så kan vi tage dem her med. Og det, vi bruger det til, det er, at vi kigger på, hvordan er den enkelte medarbejder, altså spiller, loaded.

Det I kan se her, det er taget fra en spiller på førsteholdet. Og det er bygget op på forskning. Så det er en mørk

grøn her.

Det er for, hvad har han lavet i total distance, altså hvor meget har man løbet i alt på den her uge. Så det er forskellige uger. I kan se, der er nogle uger, der er meget lave.

Det kan der være forskellige årsager til. Nogle uger er meget høje. Der har han løbet rigtig langt i løbet af ugen. 31Og det er den totale, det vil sige, med andre 10 stk. Den lysvegrønne er, den uge står i, plus de 3 sidste uger. Så det giver et mål for, har han lavet rigtig meget igennem de sidste 3 uger.

Forholdet mellem den mørkegrønne og den lysvegrønne er bevist i den gule. Så det er en ratio, som man skal ramme for at minimere sandsynligheden for, at spilleren bliver skadet. Og den skal helst ligge i det intervall her. Og nu kan I se, den hakker en del bare mere. Og vi er ikke så bekymrede for de her udfald vore. Altså det der, det er rigtig godt.

Dem der er vi ikke så bekymrede for, fordi de skal ikke så meget ud. Det der, det er som regel ikke så godt. Så er der en ret stor sandsynlighed for, at han går hen og bliver skadet.

Det er heller ikke super godt. Fordi så vil der helt naturligt ske, det er når man render baglæns, altså de sidste 3 uger sammen med den uge. Så kommer det til at se meget voldsomt ud den her uge.

Så at styre de her ting her, på tværs af nogle forskellige. Og det her, 204 og 205, det er nogle forskellige her steder, man løber ind. Yes.

Så det er det vigtigste, den fysiske sektor kan lave til os. Det er at kunne monitorere spillerne, så de ikke kan lave for meget. Fordi en af de dyreste ting for os overhovedet, det er vores spillere, de sidder ude.

Så vi skal sørge for, at spillerne ligger til røde. Det kan vi gøre på bedst mulig vis ved at sørge for, at de holder sig inden for den ratio der, i forhold til det, vi har sat op af den her fysiske koordinator. Ja.

Hvor længe har man bedt spillerne om at spille rundt i? Og har de udfordret dem i forhold til det? Har man bare

accepteret det? Det fede i fodbold med dataensamling på fodboldspillere, det er, at du kan få fodboldspillere til at gøre stort set bedst end altid. Fordi de vil bare spille. Så hvis træneren siger, vi begynder at spille med GPS fra i morgen, så står de klar til at tage den med GPS.

Fordi de skal ikke risikere, at der er noget, der risikerer, at de skal spille i weekenden. Så det der er aldrig en udfordring. Der kan godt være nogle udfordringer, helt separat med at spise og sådan noget.

Men sådan de ting der, det er der som aldrig udfordringer. Der kan være udfordringer i forhold til udstyr, så vi har haft andet udstyr end de her, hvor at de har siddet rundt om maven, og det har været lidt mere puls 32baseret. Men i en fodboldkamp, hvis man fortsætter op imod modstandere, så kan man rive lidt i en anden og rive lidt i toppen.

Så har vi oplevet, at spillerne har taget de der ned, og så bliver de irriteret, og så piller de med og smider dem ud. Så vi kan godt have nogle udstyrsudfordringer, men på det der har vi faktisk ikke haft udfordringer. Og vi har indsamlet det i sådan over tre år, men sådan struktureret på, at vi, altså igen, det der med datamodenhed og sådan nogle ting.

Vi har rigtig god data fra januar i år og frem. Der har vi en meget, meget fin datastruktur, og kan faktisk måle lige præcis på det, vi kan spille. Men vi har over tre års data på det.

Hvor tit følger I op på den dag til eneste dag? Det gør vi hver eneste dag. Altså på det her, det er, vi har to ansatte nærmest på fuld tid på første året, der følger det her hver eneste dag. Så når spiller X, om vi også har den her, når han skal ud at træne i morgen, hvad skal han så løbe, hvor han rammer det rigtige løb af ugen? Fordi han skal ikke for højt op, for eksempel.

Så det bliver fuldt hver eneste dag. Før træning, efter træning, når data kommer ind, okay, og vi kan bare så følge det live. Altså, uuh, han skulle kun ramme noget cirka 5.000 meter i dag.

Han har ramt 4.900 allerede, det går bare, at vi skal begynde at tage ham ud, så tager vi ham bare ud at træne. Fordi så har han noget af de fysiske mål, og så kan han godt deltage på lethold med noget let, og det kan være, at han får lov til at løbe 100 meter, bare for sig selv, så han rammer de mål, han skal nå. Men de bliver følt, altså fuldstændig slavisk, før, under og efter træning, Ja, jeg ved ikke, om der var først... Ja, nu ser vi da ikke så meget med fodboldspillerne.

Hvilke nogle problemer har I ellers oplevet i forhold til indtag med data? Har der været nogle modstød? Jeg oplever generelt en rigtig stor nysgerrighed på, hvad vi kan få ud af vores data. Jeg synes, der hvor vi oplever udfordringer, det er forståelsen, jeg kommer ind på det lige om lidt også, forståelsen af, hvor lang tid forskellige dele af arbejdet med data kan tage. Hvor hvis du ikke har en datamæssig baggrund, kan du godt tro, at det tager jo ikke meget lang tid.

Og så tager det bare fire uger, uden de oplever, at der kommer noget. Og så lige pludselig oplever de, at de ikke bare får det hele, fordi man har fået styr på datastudierne. Det vil jeg sige, det er nok sådan, nu skal I nok også over til at høre de andre, men det er sådan mine primære tanker, det er, at der har været rigtig meget hvad kan man sige, uddannelseafdelinger, uddannelser, hvad er der i forhold til, hvad er det, der tager tid, når vi går i gang, hvad er det så, I kommer til at opleve, hvornår er det det ikke virker og sådan nogle ting.

Det har der været en del af. Fordi vi er en virksomhed, som er meget sådan autodidakt og oplært. Du kan godt blive CEO'en i en fodboldklub, uden at have nogen lang uddannelse.

33Altså, du kan få nærmest virkningsmødesposition i en fodboldklub, uden at have en lang uddannelse. Det vil sige, tit til dem, der sidder rundt i ruller, i en fodboldklub kan godt have et behov for, at skulle lære det her data, og sætte det på grund. Nu har du også været i OB.

Hvordan ligger Rundborg i forhold til OB, i forhold til data? De hedder jo engang mig, Simon. Jeg synes, det er et godt spørgsmål. Altså, jeg kender ikke, jeg kender det ikke, og kender ikke OB, men det er godt sådan, at jeg kender Bilborg.

Så det er et godt sætning. Jeg oplevede den afdeling, men jeg understøtter det som minimum, ligeså nysgerrig, på grund af, at det er et sæt, som jeg er i. Men jeg kan ikke svare ret meget på det land.

Denne fil er længere end 30 minutter.

Opgrader til Ubegrænset på TurboScribe.ai for at transskribere filer på op til 10 timer.

## 11.5 Bilag Interview “Vff del 2 - Praktikanter”

Vff del 2 - Praktikanter Transskriberet af TurboScribe.ai. Opgrader til Ubegrænset for at fjerne denne meddelelse. Der bliver optaget igen nu her, så der bliver lige sagt nogle ting i forhold til, at man må gemme sig og sådan noget. Forhåbentlig afhjælpe, hvad man siger, nuklestigninger osv. Ja, som sagt, der bliver lige optaget lidt her, og hvis I lige gør mig den chance at lige nævne, hvem I er, og så forhåbentlig fortælle mig, at det er okay, at jeg optager.

Yes, altså jeg tænker, at det er okay. Der er ikke nogen indsigelser? Nej, der er ingen indsigelser. Jeg optager også.

Jeg hedder Josefine, og jeg kommer med en baggrund som serviceøkonom, og så har jeg arbejdet i tre år indenfor HR og løn, efterfølgende. Og imellem de her to ting har jeg faktisk også været en tur på universitetet, fordi jeg skulle finde ud af, hvad vil jeg egentlig gerne uddanne som år indenfor. Og jeg tror, at jeg havde en idé om, at det skulle være noget inden for tal, men jeg vidste ikke helt præcis, hvad det lige skulle være.

Og jeg vidste, at jeg skulle ikke læse erhvervsøkonomi i hvert fald. Og så havde jeg jo min serviceøkonom, og pludselig finder jeg ud af, at der er den her top-up uddannelse, man kan tage i dataanalyse. Men jeg opfylder så ikke adgangskravene til at komme ind, så jeg tænker, okay, det er sådan lidt et sats, så jeg søger ind og kommer så til optagelsesprøve ved Bjarne, og skal have taget sådan en matematikprøve først, for at vise, at jeg ikke godt finder ud af sådan noget matematik og sådan noget tal og sådan noget.

Og det lykkedes så, og jeg kommer ind på uddannelsen, og er i praktik her ved Viborg FF, og her har jeg været i tre måneder, den 9. til den 11. oktober. Og under praktikken, det har været sådan meget projektbaseret, der har jeg haft det her hoveddataprojekt, som jeg synes, jeg nævnte sidst, vi var ude ved jer og præsentere, hvor jeg har fået udleveret en hel masse data fra hovederne på stadion, der er fra deres hovedsystemer i alle mulige former, både PDF og Excel-filter og alt muligt sjovt, som jeg så skulle finde ud af, hvordan kan vi på en eller anden måde få struktureret den her data, så vi kan visualisere den på den ene eller anden måde.

Men sådan hovedformålet med det her projekt, det var egentlig at få den her struktureret data klargjort. Og så har jeg haft med et andet projekt at gøre, som jeg bare har kaldt for spillerstatistik, og det har handlet om, at man godt vil lave noget benchmarking på køb og salg og leje af spillere i Superliganden, som helt generelt tilbage fra 2021 og frem til nu. Og det er egentlig også stadig sådan et i gangværende projekt nu, og har primært involveret webscraping fra Wikipedia og fra, fordi jeg ikke ved noget om, hvilke Superliga hold, der har spillet i de forskellige år.

Og så har vi haft noget, hvad hedder det, Super... Transformer Transformer, ja, hvor man kan se, hvilke forskellige spillere, der har været i de forskellige hold i Superligaen. Og det første step i det, det er jo også at skabe en eller anden form for struktureret data, så man ud fra den data kan få det output, man ønsker til transportsvigelse.

Så har vi haft sådan noget med, hvad hedder det, arbejdsstruktur, organisering og dokumentation på den kode, man laver, og hvordan man bedst muligt kan gøre det, så man kan blive tilbage til, hvad er det egentlig, jeg har taget der kode, hvorfor har jeg gjort sådan og sådan, og lad os sige, at jeg skal overdrage min opgave, og når den praktisk slutter til en anden, så kan de også byde med i, hvordan har det her projekt forløb været.

Så helt til at starte med, blev vi også præsenteret for Power BI og Power Apps, og så har jeg mest brugt Power BI til, lige på søg med nogle visualiseringer af den her strukturerede data, jeg kom frem til fra brugerne, og så ved jeg, at Anja har brugt Power Apps rigtig meget. Ja, og opbygningen af praktikken, det har været en fire dages arbejdsuge, fordi jeg har et stupid job ved siden af, som jeg så har været på hver tirsdag, så ja, der har også været noget eksplikativt forhold til det. Så nu kommer der en QR-kode, tror jeg.

Det er altså først til senere. Jeg håber rigtig meget, at I gerne vil svare på nogle spørgsmål, vil jeg scanne den her QR-kode senere? Det er altid vores bachelorprojekt. Ja.

Øh, tillykke. Hej, jeg hedder Anja. Bare lige sådan kort om min baggrund.

Jeg bor i Viborg, fyrehjælpsarbejde. Efter en mange års tid har jeg taget en finanshygnoen. Har taget arbejdet, som en mani-medarbejder kan man kalde det.

Har taget både erhvervsøkonomi på Aarhus Universitet. Sidenløbende har jeg så også taget arbejde, som en fyrehjælpsassistent. Så har jeg arbejdet som revisorassistent.

Ja, og hvad kan man sige? Hvis jeg sådan skal tage mine tidligere uddannelser ind i den her, så vil jeg sige fra finanshygnoen, hvor jeg er rigtig glad for at have haft noget statistik og sådan nogle ting. På erhvervsøkonomi har jeg også haft noget operationsuddannelser. Så det synes jeg også er meget relevant.

Men jeg har så også hjemme. Jeg tror også, at jeg sagde det sidste gang, at jeg har ikke indeplads med mine børn i nogle år. Så jeg kommer også sådan lidt rustende ind til den her uddannelse.

Og jeg synes egentlig, at der er lidt grund på det. Men jeg synes det er mega interessant. Men det har så også drevet min indsats at lige forstå det hele der.

Ja, i min praktik, der har jeg som jeg skulle nævne, arbejdet i PowerApps-plattformen, hvor jeg har lavet en app. Og det vil jeg lige vise jer nu lidt. Så har jeg forsøgt at lave nogle visualiseringer i Power BI.

Det har jeg heller ikke arbejdet i før. Og i begge programmer helt nyt mig, har jeg også været på kursus en dag i hver her under min praktik. Så det er mega fedt, at jeg har plads til det.

Ja, det viser jeg lige om lidt. Nogle røde klods fra dashboardet. Og så er jeg begyndt med en datanalyse i R, som jeg kender.

Som, hvad skal man sige, i hørte jo Daniel, han nævnte, at der sidder nogen inde i den sportslige afdeling og laver nogle kopier. Eller beslutter, hvad for nogle kopier de skal kigge på. Hvad det er deres, så ved jeg sådan i forhold til om digitalt er det en eller anden kamp.

Der har jeg fået nogle GPS-data sammen med nogle andre indsamlede data. Og så har jeg selv fundet nogle data, som jeg har taget ind i det dataset. Og med vilje at ikke blive påvirket af, hvad de tror og mener.

Men kommer med mine nysgerrige øjne ud fra og mine perspektiver. Og er ikke sådan vildt meget ved fodbold, men ved da lidt mere efter at også have været her. Og synes det er mega spændende.

Så jeg kommer måske med nogle andre perspektiver. Men jeg har desværre ikke nået i mål med den endnu. Men det bliver spændende at se, om mine ansigter lige så rækker om deres teorier.

Eller om det er noget andet. Så det er sådan, ja, hvad jeg har lavet. Så har jeg taget, grunden til at jeg ikke viser PowerBI-reporten, det er fordi det er data fra fodboldspillerne.

Som jo ikke bare lige skal vises til alle. Så jeg har lige taget sådan nogle skærmbilleder, for ligesom at prøve at vise, at det her er en side, hvor jeg kan vælge den spil jeg gerne vil. De her visualiseringer bliver vist lige nu fra

hele hånden i den her periode.

Når jeg så vælger en spiller, kommer der data frem fra den spiller. Blandt andet navn, alder osv. Der kommer det her lidt på kropsomsætningen.

35Og det er jo i samspil med træneren. Hvad er det, jeg gerne vil se lige nu? Inden min app, kan man sige, der bliver data registreret i Excel-ark. Og faktisk også i yndtilføjede på fysisk papir.

Som bagefter kommer over i Excel. Og så bruger de i Excel, så laver de godt nok selv nogle målinger og beegninger. Hvad det er, de skal bruge og gerne vil se videre på.

Men det kommer jeg lidt mere ind på i app'en. I forhold til, hvordan data registreres og hvad de så kan bruge det. Jeg har jo så forsøgt med træneren at prøve, hvad er det egentlig, vi gerne vil se.

Og der kan man sige, lige i forhold til kropsomsætningen. På nuværende tidspunkt kigger træneren på, hvad er det, der har ændret sig fra denne her test til sidste test. Man vil gerne kunne zoome lidt ud og se udviklingen. Fordi det kan godt være, at der er den her ændring præcis, kan I se. Men hvis du kigger på det hele billede, så er det jo... Så det har jeg forsøgt at lege med. En anden måde, faktisk lidt det samme.

Bare fordi, at jeg er blevet inspireret af Jesfine, fordi hun er megateknisk og er så god. Så viste hun mig en måde, hvor jeg tænkte, det ville jeg også lige prøve. Så jeg har det lidt det samme, hvor jeg før havde drop down.

Til at vælge en spiller, så kan jeg trykke på den spiller, jeg gerne vil se data fra. Og så bliver det så vist herover. Så det var bare lige for, at den er flere måder at gøre det på.

Og sikkert også gøre det meget federe, hvis man bruger meget mere tid på det osv. Men ja, det har jeg også lavet at lege med i min praktik. Lidt det samme, bare med nogle testresultater og nogle hoppetest.

Så tror jeg faktisk, jeg vil gå videre til 18. Ja, så skal jeg bare lige have skiftet det her. Er du klar? Yes! Den her app, den er lavet til at registrere opdaget data fra både kropsomsætning og for nogle forskellige fysiske test. Så hvis man trykker start, så kan man starte med at se data fra kropsomsætningen først. Det er meget simpelt. Vælge en spiller.

Indtest. Så skriver vi bare noget. Gennem data.

Det kan man så gøre på. Det kan være, at jeg lige skal nævne, at jeg havde først lavet et bud på en app til dataregistreringen. Ud fra, hvad jeg tænkte havde mening.

I forhold til de her forskellige test. Så undervejs, så kom jeg med ud og observerede nogle test. Og kunne godt se, okay, den måde jeg først havde sæt min app op.

Meget simpelt. En person ad gangen bliver testet. Da jeg så kom ud og så det.

Så for eksempel for de her hoppetest. Så foregår det anderledes. Så foregår det med, der kommer en spiller her. Han hopper her. Kommer der en ny spiller, så du-du-du-du. Hvorfor det giver mening, at de bruger deres papir. Så jeg kan skrive forskellige forsøg. Jeg kan lige prøve at gå ind på en app, for det giver lidt mere mening. Her. Nu vælger jeg bare lige nogle. Så har første mand her for eksempel et forsøg. Okay.

Nå, men det er jo ikke. Det her, det er jo ikke. Nej, nej, nej.

Det var bare, at der var en idé der. Nå, ja. Okay.

Så kan det være, at det næste hopper. Så hopper de første spillere igen. Så skal man lige så.

Appen skulle ligesom have mulighed for. Det kunne ikke nytte noget, at jeg havde en side for hver spiller. Jeg

så skulle ind på hver gang.

Det ville tage alt for lang tid i praksis. Og træneren informerede mig så om, at de ville gerne lave det af hold. Når de har test med op til fem på hver gruppe.

Og derfor så kom jeg med en ny løsning. Hvor de så ligesom havde mulighed for at indsætte sig og fungere i praksis. Så kan jeg lige prøve at vise.

Det her er så et eksempel på en test. Hvor jeg faktisk, imens jeg lige har været her. Er der ikke nogen, der har lavet den her test.

Så den har jeg ikke kunne være med ude og observere. Hvilket jeg jo nu har fundet ud af, at det er ret vigtigt. Eller til gengæld har jeg fået gode beskrivelser af, hvordan den foregår.

36Og har taget udgangspunkt i det. Men det er også bare lige så meget for at sige. At vigtigheden i at forstå. Præcis hvad det er, man skal bruge det til. Og hvordan i praksis. For man kan godt handle sig nogle forestillinger om.

Det er da bare lige sådan og sådan. Men det er også ikke dem, der skal tilpasse sig min løsning. Det er jo mig, der skal komme med en løsning og fungere for dem.

Så det var i hvert fald noget, jeg fik bekræftet i min praktik. Så har jeg bare prøvet at lave. For igen.

Det her med de gamle test. Fem spiller ad gangen. Øhm.

Nu tager jeg jo spanden. Så kan det være, at der er en, der starter på. Ja.

Og så level man kan jo sætte den til at starte på samme tid. Men lad os så sige, at der skal stadigvæk mulighed for. Okay. Han kommer kun heroppe. Så skal der være en pause. Eller nej, han får en advarsel. Det kan være, at den her spiller kommer hertil og skal stoppe. Så igen med udgangspunkt i, at der skal være mulighed for at indtaste for fem ad gangen.

Hvad betyder det, at man får en advarsel? Jamen det er, fordi det er sådan en løbetest. Hvor spillerne skal nå at løbe ind for noget tid. Ja, præcis.

Og hvis de så ikke når det, så får de først en advarsel. Ja, jeg tror, at jeg har rigtig forstået. Nu de økonomiserer jo lidt med det.

Og så sælger de jo flere runder, når de kan. Så de prøver lige at trække den til den her bil. Og så de løber den anden vej.

Ja. Øhm. Øhm.

Og her, der bliver spilleren testet en ad gangen. Hvor de starter med en starthasthed. Øhm.

Jeg tager bare lige igen. Øhm. Så starter vi testen.

Så går timehånden i gang. Så når spilleren når det her punkt i vindtestet. Jeg ved ikke sådan helt specifikt, hvordan der var lavet med testen.

Men det er bare sådan for at vise, at så kan man registrere det. Det er stoptidspunktet. Og så hvornår man skal.

Det kan være, at der er en pause. Det er også stopt. Meget hurtigt gennemgang.

Men bare lige for at de visuelt kunne se noget af det output. Så det var egentlig det, jeg havde. Jeg kan bare lige fortælle, hvor gemmer man dataen hen? Ja, fra appen.

Øhm. De data, der bliver registreret i appen. De kommer så ind i en Excel-ark.



På nuværende tidspunkt. Øhm. Og det er jo så fra det Excel-ark, jeg har taget over i Power BI.

Der kan være noget datarensning inden Power BI. Men inden man kan gøre i Power BI. Der kan man gøre en del.

Men ellers kan man bruge R. I starten brugte jeg R. Fordi det følte jeg mig tryk i her på uddannelsen. Jo bedre jeg lærte Power BI at kende. Jo mere begyndte jeg også at bruge det.

Så det var hele rejsen fra dataindsamling, datarensning og output. Så nu skal vi snakke om data. Så mit navn er

Olga. Jeg har været her siden i sommer. Og så har jeg arbejdet som finansøkonom. Og så har jeg den forbindelse med at arbejde i brugtiv.

Det har jeg på stedet. Og så har jeg samfundet for at bruge den til hjælp. Og så har jeg efterfølgende været til at bruge den til uddannelsesstil.

Samfundet er det andre. Og så har jeg samfundet det samme til at bruge den til hjælp. Så nu får jeg samfundet for hjælp.

Mine præmier er at prøve at lide at nå hjælp. Jeg har prøvet at lide at nå hjælp. Det er det, som jeg gerne vil. Og så har jeg havet lyst til at prøve at tage hjælp. Implements plus ingangsdata Også full join resurrectioner plus ingangsdata Det er der, hvor jeg har prøvet at få det for billetter Også 50 words i næste gang Også skulle 37 jeg præsentere det på en eller anden måde Så det kunne sidde Vi har ikke lavet farger i vores forløp Så vi har lavet ShinyApp Derfor jeg har brugt ShinyApp bare for at vise dem Og jeg viser nu en screenshot Vi har tre forskellige ting Overview Her kan man vælge den partner, som man er interesseret i Også kan man se alle de der kort, som den enkelte partner har haft Og efter kan man vælge partner Her kan man se alle de der kort, som den enkelte partner har haft Vi har bestilt billetter Fremgjort Underskrevet Så kortene er opdelt i A, B og C Gruppen D, O, S, D, F og D Her viser man, hvilke trukker man måske skal have Så kan man for eksempel se, at det var med RANES Og næste gang var det med Fremgjort De forskellige farger er repræsenteret med de forskellige grupper I princippet kan man også gøre det samme med grafen Om partneren går til Fremgjort Om det er det, som er begyndt Om det er det, som er gået ned Eller hvordan fungerer det Jeg kan godt forstå, at det

måske ikke er så pænt ud over det der app Men vi har den app, hvor jeg skriver, hvilke ord Og det er det, som vi gerne vil have Så det vil være rigtig godt Og så den sidste, fordi det er os til mig Det vil være faktisk mega nice at se partneren ud fra kamp Så man kan vælge kamp For eksempel her har jeg væltet RANES Og så her kan man se alle de der partnere Her er der blevet filmet Hvor man har væltet Fremgjort Og så Ullus-ekranen Og jeg er træng Så det var det Så min praktik vil være sluttet Og så nu skriver jeg bare et tjelam Og så prøver jeg, hvad vi laver med dette Så her begynder det at snakke om eventet Og så tænker jeg, hvordan beslutter jeg det Okay, så tager vi den reservation Reservationer dette udtryk Sætter sammen med ingangsdata Og så nu giver det dataset kampdata Og det er også på den måde, som vi fungerer nu Og så når jeg kommer tilbage Så får jeg den mule Så det var både en god nyhed Og en lidt mindre god nyhed For mig For indtægterne, det er da super dejligt Nu er der lige meget mere data For den anden side, okay Så det betyder alt det, som jeg har lavet før Jeg kan ikke ligesom forsætte med at tvivle over det Men så fik jeg at vide At jeg skulle ikke blive uanset Så var jeg ikke sikker på det Okay, så her, vi starter på nyt Og så, ja, så Så blev det besluttet, at nu bliver det

Power BI Og hvorfor Power BI? Fordi det er nærmere at overtage Men det er også nærmere at arbejde med Men du kan jo sikkert skrive, at det er et langt kod Og så må det også være sådan sammenlignet Jeg har en spørgsmål Har I brugt Power BI på visualiseringen? Eller generelt visualiseringen? Ikke så meget Nogen måder, men ikke så meget Så Efter det, så startede jeg med at ligesom forfra Med hele processen i forhold til at Delette data Så jeg har fået data fra IVEX Så Tilmeldende dataset Stedigvæk indeholder tilmængder Og så kommer dataset Der hvor vi har visualizationer Plus indre data Så nu arbejder jeg kun med to set Så kan jeg udrejne Udrejne At sætte det samme tilmeldende Dataset Hvilket jeg mærker, fordi nu har vi vores kolonne Med den unikke fælles kolonne Så nu fungerer det bare Og så Jeg har fået to billetter, der kan filtrere det hele Uden at kæmpe det samme Sætter det sammen Og så Og så er vi lov Så vi er der Så nu er vi Denne fil er længere end 30 minutter. Opgrader til Ubegrænset på TurboScribe.ai for at transskribere filer på op til 10 timer.

## 11.6 Bilag Kode

### 11.6.1 A) Pakker og opsætning

```
# -----  
# Forbind til SQLite Database  
# -----  
  
# Liste over alle pakker (dokumentation: hvad projektet bruger)  
packages <- c(  
  "tidyverse", "janitor",  
  "RSQLite", "rvest", "xml2",  
  "dplyr", "stringr", "DBI", "lubridate", "leaps"  
)  
  
# Indlæs pakker vi bruger i dette script  
library(tidyverse)  
library(rvest)  
library(httr)  
library(dplyr)  
library(stringr)  
library(DBI)  
library(RSQLite)  
library(janitor)    # Bruges til at rense kolonnenavne (snake_case) og generel oprydning  
library(lubridate)  # Bruges til at arbejde med datoer/tid  
library(leaps)
```

### 11.6.2 B) Forbindelse til SQLite Database

```
# Opret forbindelse til SQLite databasen (filen ligger i dit projekt-folder)  
con <- dbConnect(  
  SQLite(),  
  dbname = "VffDataBaseNew.db"  
)  
  
# Sørger for at forbindelsen altid lukkes, selv hvis scriptet fejler undervejs
```

```
#on.exit(DBI::dbDisconnect(con), add = TRUE)
```

### 11.6.3 C) Web scraping af Superstats (udkommenteret)

```
# -----  
# WEB SCRAPING AF SUPERSTATS OG JOIN AF ANDRE TABELLER  
# -----  
  
#  
# # Vi har udkommenteret det, så den ikke scraber hver gang der loades.  
#  
# # Vi angiver URL til Superstats (bruges kun som udgangspunkt)  
# html <- "[https://superstats.dk/hold/sason?id=11&vis=hjemme&aar=2025%2F2026] (https://superstats.dk/ho  
#  
# # Vector med alle sæsoner vi vil hente data for (Efter nyt stadium)  
# seasons <- c(  
# "2025/2026", "2024/2025", "2023/2024", "2022/2023", "2022/2022",  
# "2016/2017", "2015/2016", "2013/2014", "2007/2008", "2006/2007",  
# "2005/2006", "2004/2005", "2003/2004", "2002/2003", "2001/2002",  
# "2000/2001", "1999/2000", "1998/1999", "1996/1997", "1995/1996",  
# "1993/1994"  
# )  
#  
# # Tomt dataframe som data fra alle sæsoner løbende samles i  
# st_webscrabe <- data.frame()  
#  
# # Loop der gennemgår hver sæson én ad gangen  
# for (s in seasons) {  
#  
# # Bygger korrekt URL for den pågældende sæson
```

```

# url <- paste0(

# "[https://superstats.dk/hold/sason?id=11&vis=hjemme&aar=](https://superstats.dk/hold/sason?id=11&vis=

# URLEncode(s, reserved = TRUE)

# )

#

# # Henter HTML-indholdet fra siden

# webpage <- read_html(url)

#

# # Finder alle tabeller på siden

# tables <- html_nodes(webpage, "table")

#

# # Tabel nr. 3 er den med kampdata

# st3_season <- html_table(tables[[3]], fill = TRUE, convert = FALSE)

#

# # Tilføjer sæson som variabel

# st3_season$Season <- s

#

# # Lægger sæsonens data oven i det samlede dataframe

# st_webscrabe <- bind_rows(st_webscrabe, st3_season)

# }

#

# # -----

# # DATARENS

# # -----

#

# st_webscrabe <- st_webscrabe %>%

# mutate(

# # Tilskuere omdannes fra tekst til numerisk værdi

```

```

# Tilskuere = as.character(Tilskuere),

# Tilskuere = str_trim(Tilskuere),

# Tilskuere = na_if(Tilskuere, "-"),

# Tilskuere = na_if(Tilskuere, ""),

# Tilskuere = str_replace_all(Tilskuere, "\\.", ""), # fjerner tusindtals-punktum

# Tilskuere = str_replace_all(Tilskuere, ",", ""),

# Tilskuere = as.numeric(Tilskuere),

#

# # Udtrækker modstanderhold fra kamp-teksten ("VFF - FCK" → "FCK")

# Modstander = str_replace(Kamp, "VFF - ", ""),

#

# # Konverterer dato fra tekst til Date-klasse

# Dato = as.Date(Dato, format = "%d.%m.%Y")

# )

#

# # -----

# # MODSTANDER-KATEGORIER

# # -----

#

# modstander_niveau <- st_webscrabe %>%

# group_by(Modstander) %>%

# summarise(mean_tilskuere = mean(Tilskuere, na.rm = TRUE), .groups = "drop")

#

# qs <- quantile(modstander_niveau$mean_tilskuere,

# probs = c(0.25, 0.5, 0.75),

# na.rm = TRUE)

#

# modstander_niveau <- modstander_niveau %>%

```

```

# mutate(

# kategori = case_when(

# mean_tilskuere > qs[3] ~ "A",
# mean_tilskuere > qs[2] ~ "B",
# mean_tilskuere > qs[1] ~ "C",
# TRUE ~ "D"

# )

# )

#

# st_webscrabe <- st_webscrabe %>%

# left_join(modstander_niveau, by = "Modstander")

#

# st_webscrabe <- st_webscrabe %>%

# group_by(Season) %>%

# mutate(mean_tilskuere_sæson = mean(Tilskuere, na.rm = TRUE)) %>%

# ungroup()

#

# st_webscrabe_db <- st_webscrabe %>%

# mutate(Dato = as.integer(Dato))

#

# DBI::dbWriteTable(con, "st_webscrabe", st_webscrabe_db, overwrite = TRUE)

```

#### 11.6.4 D) Indlæsning af data (.rds) og skriv til DB

```

# Load vores tabeller ind (fra .rds-filer)
vffkort01 <- readRDS("vffkort01.rds")
fcidk <- readRDS("fcidk.rds")

#Her loader vi df med dmi,nager og webscrabe fra superstats
vff_hjemmekampe_med_vejr <- readRDS("vff_hjemmekampe_med_vejr.rds")

# Skriv (eller overskriv) tabellerne i databasen, så DB altid er opdateret
dbWriteTable(con, "vffkort01", vffkort01, overwrite = TRUE)

```

```
dbWriteTable(con, "vff_hjemmekampe_med_vejr", vff_hjemmekampe_med_vejr, overwrite = TRUE)
#dbWriteTable(con, "fcidk", fcidk, overwrite = TRUE)
```

### 11.6.5 E) SQL-joins og kontroludtræk

```
# -----
# Join Tabeller
# -----

# Join st_webscrabe og DMI tabel (vff_hjemmekampe_med_vejr)
# OBS: st_webscrabe.Dato er numeric (dage siden 1970-01-01),
# mens vff_hjemmekampe_med_vejr.dato_index er tekst "dd/mm/yyyy".
# Derfor konverterer vi Dato -> dd/mm/yyyy i SQL før join.

joined_st_webscrabe_dmi <- DBI::dbGetQuery(con, "
SELECT *
FROM st_webscrabe s
INNER JOIN vff_hjemmekampe_med_vejr d
  ON strftime('%d/%m/%Y', date('1970-01-01', printf(' +%d days', s.Dato))) = d.dato_index
ORDER BY s.Dato DESC;
")

# Hurtig kontrol tjek at datoformater matcher i de to tabeller før join
DBI::dbGetQuery(con, "SELECT Dato FROM st_webscrabe ORDER BY Dato DESC LIMIT 10;")
DBI::dbGetQuery(con, "SELECT dato_index FROM vff_hjemmekampe_med_vejr ORDER BY dato_index DESC LIMIT 10;")

# Størrelse og kolonnenavne på joinet
nrow(joined_st_webscrabe_dmi)
ncol(joined_st_webscrabe_dmi)
names(joined_st_webscrabe_dmi)

# Da der laves udtræk fra SQL, så bliver "Dato" værdien retuneret som numeric,
# og det skal vi lave om til Date værdi igen i R
joined_st_webscrabe_dmi$Dato <- as.Date(joined_st_webscrabe_dmi$Dato, origin = "1970-01-01")

# Se hvilken datatype Dato har
class(joined_st_webscrabe_dmi$Dato)

# Join st_webscrabe og fcidk (giver modstanders gennemsnitlige tilskuere)
joined_st_webscrabe_fcidk <- dbGetQuery(con, "
SELECT
  s.*,
  f.tilskuere AS modstander_gns_tilskuere
FROM st_webscrabe AS s
LEFT JOIN fcidk AS f
  ON s.Modstander = f.kort
")

# Se hvilke tabeller der er i databasen (kontrol)
dbListTables(con)
```

```
# Join st_webscrabe med vffkort01 (sæson/runde join)
joined_st_webscrabe_vfkort01 <- dbGetQuery(con, "
SELECT
  st_webscrabe.*,
  vffkort01.*
FROM st_webscrabe
LEFT JOIN vffkort01
  ON st_webscrabe.Season = vffkort01.sæson
  AND st_webscrabe.Rnd    = vffkort01.runde
")
```

### 11.6.6 F) Opret samlet model-datasæt i databasen (model\_dataset)

```
# -----
# Byg samlet model-datasæt i databasen
# -----

# Nu vil vi lave en stor tabel, hvor vi joiner st_webscrabe med vffkort01, fcidk og vores DMI/Nager tab
# Vi dropper den først, så vi altid starter fra en "ren" tabel
DBI::dbExecute(con, "DROP TABLE IF EXISTS model_dataset;")

DBI::dbExecute(con, "
CREATE TABLE model_dataset AS
SELECT
  s.*,
  v.*,
  f.*,
  d.*
FROM st_webscrabe AS s
LEFT JOIN vffkort01 AS v
  ON s.Season = v.`sæson`
  AND s.Rnd    = v.`runde`
LEFT JOIN fcidk AS f
  ON s.Modstander = f.kort
LEFT JOIN vff_hjemmekampe_med_vejr AS d
  ON strftime('%d/%m/%Y', date('1970-01-01', printf('%d days', s.Dato))) = d.dato_index;
")

# Læs tabellen ind i R og kig på den
model_dataset <- DBI::dbReadTable(con, "model_dataset")
names(model_dataset)

# Vi bruger glimpse fra dplyr pakken, da den giver et glimrende overblik over rækker og kolonner
dplyr::glimpse(model_dataset)
```

### 11.6.7 G) Dataklargøring og feature engineering (tidy)

```
#Vi rydder op i dato, kolonnenavne, fjern junk + dubletter
# - clean_names() gør fx "VFF_mål" -> "vff_mal"
# - Vi laver dato om til Date
# - Vi gør udvalgte tekstfelter til factor (giver mening i modellering)
```



```

df <- model_dataset %>%
  janitor::clean_names() %>%
  distinct() %>% # undgå dubletter efter joins
  mutate(
    dato = as.Date(dato, origin = "1970-01-01"),
    season = factor(season),
    modstander = factor(modstander),
    kategori = factor(kategori)
  )

# Fjern "join-rester": kolonner der næsten kun er NA
# (det sker ofte når en join kun matcher for nogle rækker)
na_frac <- sapply(df, function(x) mean(is.na(x)))
df <- df[, na_frac < 0.95] # behold kun kolonner med <95% NA

# Overblik efter oprydning
dplyr::glimpse(df)

# Vi gemmer datasættet
# saveRDS(df, "df_model_ready.rds")

df <- df %>%
  mutate(
    # Hvilken ugedag spilles kampen?
    weekday = wday(dato, label = TRUE),

    # Er det weekend?
    is_weekend = weekday %in% c("Sat", "Sun"),

    # Måned (sæson-effekt)
    month = month(dato),
  )

```

### 11.6.8 H) Datasplits: 2 måneder, 10 dage, 7 dage, 3 dage

```

#Nu laver vi fire forskellige splits, da vi skal lave modeller for hhv. 3d,7d,10d og 2 måneder

#2 Måneder før - Dataframe: Vi laver et dataframe, hvor udvalgte variabler er med. De er udvalgt udfra
df_2m <- df %>%
  select(
    tilskuere, # TARGET (det vi forudsiger)
    season,
    modstander,
    kategori,
    mean_tilskuere,
    mean_tilskuere_saeson,
    weekday,
    is_holiday,
  ) %>%
  drop_na() #Bruger kun rækker, hvor der ikke er missing value (NA)

```

```
#10 dage før - Dataframe:
```

```
df_10d <- df %>%
```

```
  select(
    tilskuere,
    season,
    rnd,
    modstander,
    kategori,
    mean_tilskuere,
    mean_tilskuere_saeson,
    weekday,
    is_holiday,
    kamp_tid,
    kickoff_hour,
    d10_tilskuere
  ) %>%
  drop_na()
```

```
#7 dage før - Dataframe:
```

```
df_7d <- df %>%
```

```
  select(
    tilskuere,
    season,
    rnd,
    modstander,
    kategori,
    mean_tilskuere,
    mean_tilskuere_saeson,
    weekday,
    is_holiday,
    kamp_tid,
    kickoff_hour,
    regn_dag_kat,
    temp_kamp_kat,
    vind_1h_kat,
    d10_tilskuere,
    d7_tilskuere
  ) %>%
  drop_na()
```

```
#3 dage før - Dataframe:
```

```
df_3d <- df %>%
```

```
  select(
    tilskuere,
    season,
    rnd,
    modstander,
    kategori,
    mean_tilskuere,
    mean_tilskuere_saeson,
    weekday,
    is_holiday,
    kamp_tid,
    kickoff_hour,
```

```

temp_kamp,
regn_dag,
vind_1h,
d10_tilskuere,
d7_tilskuere,
d3_tilskuere
) %>%
drop_na()

```

### 11.6.9 I) Fælles setup til modellering 80/20 split)

```

# Vi Laver et fælles "setup" til 2m, 10d, 7d og 3d, så vi kan genanvende objekterne

# -----
# 2m fælles før setup
# -----

# X og y (bruges af ALLE modeller)
x_2m <- model.matrix(tilskuere ~ ., df_2m)[, -1]
y_2m <- df_2m$tilskuere

# 80/20 split
set.seed(4)
n <- nrow(x_2m)
train_2m <- sample(1:n, size = floor(0.8 * n))
test_2m <- setdiff(1:n, train_2m)

y_test_2m <- y_2m[test_2m]

# -----
# 10d fælles før setup
# -----

# X og y (bruges af ALLE modeller)
x_10d <- model.matrix(tilskuere ~ ., df_10d)[, -1]
y_10d <- df_10d$tilskuere

# 80/20 split
set.seed(4)
n <- nrow(x_10d)
train_10d <- sample(1:n, size = floor(0.8 * n))
test_10d <- setdiff(1:n, train_10d)

y_test_10d <- y_10d[test_10d]

# -----
# 7d fælles før setup
# -----

# X og y (bruges af ALLE modeller)
x_7d <- model.matrix(tilskuere ~ ., df_7d)[, -1]
y_7d <- df_7d$tilskuere

```

```

# 80/20 split
set.seed(4)
n <- nrow(x_7d)
train_7d <- sample(1:n, size = floor(0.8 * n))
test_7d <- setdiff(1:n, train_7d)

y_test_7d <- y_7d[test_7d]

# -----
# 3d fælles før setup
# -----

# X og y (bruges af ALLE modeller)
x_3d <- model.matrix(tilskuere ~ ., df_3d)[, -1]
y_3d <- df_3d$tilskuere

# 80/20 split
set.seed(4)
n <- nrow(x_3d)
train_3d <- sample(1:n, size = floor(0.8 * n))
test_3d <- setdiff(1:n, train_3d)

y_test_3d <- y_3d[test_3d]

```

#### 11.6.10 J) Lineær regression (LM) + test-RMSE

```

# -----
# Modellering Lineær Regression (STOR MODEL) - 2 måneder før
# -----

# STOR model: vi bruger alle variabler der findes i df_2m (undtagen tilskuere).
# Denne burde vi kunne sammenligne med Ridge/Lasso/subset senere.
lm_2m_big <- lm(tilskuere ~ ., data = df_2m[train_2m, ])

# summary viser koefficienter, p-værdier og forklaringsgrad R^2.
summary(lm_2m_big)

# -----
# Modellering Lineær Regression (STOR MODEL) - 10 dage før
# -----

# STOR model: bruger alle variabler som vi har i df_10d.
lm_10d_big <- lm(tilskuere ~ ., data = df_10d[train_10d, ])
summary(lm_10d_big)

# -----
# Modellering Lineær Regression (STOR MODEL) - 7 dage før
# -----

```

```

# STOR model: alle variabler i df_7d.
lm_7d_big <- lm(tilskuere ~ ., data = df_7d[train_7d, ])
summary(lm_7d_big)

# -----
# Modellerings Lineær Regression (STOR MODEL) - 3 dage før
# -----

# STOR model: alle variabler i df_3d.
lm_3d_big <- lm(tilskuere ~ ., data = df_3d[train_3d, ])
summary(lm_3d_big)

# -----
# Test-RMSE for stor lineær regression (2 måneder før)
# -----

# Vi kan få fejl hvis test indeholder modstandere, som ikke findes i træningsdata.
# Derfor laver vi et nyt testdatasæt og filtrerer "ukendte" modstandere fra.

# Lav testdata
test_df_2m <- df_2m[test_2m, ]

# Fjern modstandere som ikke findes i træning
test_df_2m <- test_df_2m %>%
  dplyr::filter(modstander %in% df_2m$modstander[train_2m])

# Forudsig på det filtrerede test-sæt
pred_lm_2m <- predict(lm_2m_big, newdata = test_df_2m)

# Beregn RMSE
rmse_lm_test_2m <- sqrt(mean((test_df_2m$tilskuere - pred_lm_2m)^2))

rmse_lm_test_2m

# -----
# Test-RMSE for stor lineær regression (10 dage før)
# -----

# Samme problem kan opstå her: test kan have nye factor-levels (modstander).
# Derfor filtrerer vi test, så den kun indeholder modstandere fra træning.

# Lav testdata
test_df_10d <- df_10d[test_10d, ]

# Fjern modstandere som ikke findes i træning
test_df_10d <- test_df_10d %>%
  dplyr::filter(modstander %in% df_10d$modstander[train_10d])

# Forudsig på det filtrerede test-sæt
pred_lm_10d <- predict(lm_10d_big, newdata = test_df_10d)

# Beregn RMSE
rmse_lm_test_10d <- sqrt(mean((test_df_10d$tilskuere - pred_lm_10d)^2))

```

```

rmse_lm_test_10d

# -----
# Test-RMSE for stor lineær regression (7 dage før)
# -----

# Igen: Hvis test indeholder en modstander som ikke var i træningsdata,
# kan predict() ikke danne de rigtige dummy-variable og giver fejl.
# Derfor filtrerer vi test-sættet.

# Lav testdata
test_df_7d <- df_7d[test_7d, ]

# Fjern modstandere som ikke findes i træning
test_df_7d <- test_df_7d %>%
  dplyr::filter(modstander %in% df_7d$modstander[train_7d])

# Forudsig på det filtrerede test-sæt
pred_lm_7d <- predict(lm_7d_big, newdata = test_df_7d)

# Beregn RMSE
rmse_lm_test_7d <- sqrt(mean((test_df_7d$tilskuere - pred_lm_7d)^2))

rmse_lm_test_7d

# -----
# Test-RMSE for stor lineær regression (3 dage før)
# -----

#Det giver en fejl, da AB (holdet) ikke er med i testdataen og kun i trænings,
#Så Vi skal lige manuelt fjerne dem / lave et nyt test datasæt, som er ens med det andet uden AB dog

# Lav testdata
test_df_3d <- df_3d[test_3d, ]

# Fjern modstandere som ikke findes i træning (AB)
test_df_3d <- test_df_3d %>%
  dplyr::filter(modstander %in% df_3d$modstander[train_3d])

# Forudsig antal tilskuere på testdatasættet
pred_lm_3d <- predict(lm_3d_big, newdata = test_df_3d)

# Beregn RMSE
rmse_lm_test_3d <- sqrt(mean((test_df_3d$tilskuere - pred_lm_3d)^2))

rmse_lm_test_3d

# Vi laver en opsummering af RSME for alle fire splits

rmse_lm_overview <- data.frame(
  Tid = c("2 måneder", "10 dage", "7 dage", "3 dage"),
  RMSE_test_lm = c(rmse_lm_test_2m, rmse_lm_test_10d, rmse_lm_test_7d, rmse_lm_test_3d)
)

```

## 11.6.11 K) Lasso (glmnet) + CV + test-RMSE

```

#=====
# Modellering Lasso
# =====
library(glmnet)

# Vi bruger samme lambda-grid til alle Lasso-modeller
grid <- 10^seq(10, -2, length.out = 100)

# -----
# Modellering Lasso - 2 måneder før
# -----

# Vi estimerer en Lasso-model på træningsdata.
# alpha = 1 betyder Lasso (L1-regularisering),
# som både reducerer overfitting og kan sætte koefficienter lig 0.
lasso_2m <- glmnet(
  x_2m[train_2m, ],
  y_2m[train_2m],
  alpha = 1,
  lambda = grid,
  thresh = 1e-12
)

# Cross-validation bruges til at finde den bedste lambda-værdi.
# Seed sikrer, at fold-opdelingen i CV er den samme hver gang.
set.seed(4)
lasso_cv_2m <- cv.glmnet(
  x_2m[train_2m, ],
  y_2m[train_2m],
  alpha = 1,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

# Den lambda-værdi som giver lavest CV-fejl
bestlam_lasso_2m <- lasso_cv_2m$lambda.min

# RMSE fra cross-validation (træningsdata)
rmse_lasso_cv_2m <- sqrt(
  lasso_cv_2m$cvm[lasso_cv_2m$lambda == bestlam_lasso_2m]
)

# Forudsigelser på testdatasættet
pred_lasso_2m <- as.numeric(
  predict(lasso_2m, s = bestlam_lasso_2m, newx = x_2m[test_2m, ])
)

# Test-RMSE: gennemsnitlig fejl i antal tilskuere

```

```

rmse_lasso_test_2m <- sqrt(mean((pred_lasso_2m - y_test_2m)^2))

bestlam_lasso_2m
rmse_lasso_cv_2m
rmse_lasso_test_2m

# De variabler som Lasso faktisk bruger (koefficient 0)
coef_lasso_2m <- coef(lasso_cv_2m, s = "lambda.min")
coef_lasso_2m[coef_lasso_2m[,1] != 0, , drop = FALSE]

# -----
# Modellerings Lasso - 10 dage før
# -----

# Lasso-model på træningsdata
lasso_10d <- glmnet(
  x_10d[train_10d, ],
  y_10d[train_10d],
  alpha = 1,
  lambda = grid,
  thresh = 1e-12
)

# Cross-validation til valg af lambda
set.seed(4)
lasso_cv_10d <- cv.glmnet(
  x_10d[train_10d, ],
  y_10d[train_10d],
  alpha = 1,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

bestlam_lasso_10d <- lasso_cv_10d$lambda.min

# CV-RMSE
rmse_lasso_cv_10d <- sqrt(
  lasso_cv_10d$cvm[lasso_cv_10d$lambda == bestlam_lasso_10d]
)

# Test-RMSE
pred_lasso_10d <- as.numeric(
  predict(lasso_10d, s = bestlam_lasso_10d, newx = x_10d[test_10d, ])
)

rmse_lasso_test_10d <- sqrt(mean((pred_lasso_10d - y_test_10d)^2))

bestlam_lasso_10d
rmse_lasso_cv_10d
rmse_lasso_test_10d

coef_lasso_10d <- coef(lasso_cv_10d, s = "lambda.min")
coef_lasso_10d[coef_lasso_10d[,1] != 0, , drop = FALSE]

```



```

# -----
# Modelling Lasso - 7 dage før
# -----

lasso_7d <- glmnet(
  x_7d[train_7d, ],
  y_7d[train_7d],
  alpha = 1,
  lambda = grid,
  thresh = 1e-12
)

set.seed(4)
lasso_cv_7d <- cv.glmnet(
  x_7d[train_7d, ],
  y_7d[train_7d],
  alpha = 1,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

bestlam_lasso_7d <- lasso_cv_7d$lambda.min

rmse_lasso_cv_7d <- sqrt(
  lasso_cv_7d$cvm[lasso_cv_7d$lambda == bestlam_lasso_7d]
)

pred_lasso_7d <- as.numeric(
  predict(lasso_7d, s = bestlam_lasso_7d, newx = x_7d[test_7d, ])
)

rmse_lasso_test_7d <- sqrt(mean((pred_lasso_7d - y_test_7d)^2))

bestlam_lasso_7d
rmse_lasso_cv_7d
rmse_lasso_test_7d

coef_lasso_7d <- coef(lasso_cv_7d, s = "lambda.min")
coef_lasso_7d[coef_lasso_7d[,1] != 0, , drop = FALSE]

# -----
# Modelling Lasso - 3 dage før
# -----

lasso_3d <- glmnet(
  x_3d[train_3d, ],
  y_3d[train_3d],
  alpha = 1,
  lambda = grid,
  thresh = 1e-12
)

set.seed(4)
lasso_cv_3d <- cv.glmnet(

```

```

x_3d[train_3d, ],
y_3d[train_3d],
alpha = 1,
lambda = grid,
nfolds = 5,
thresh = 1e-12
)

bestlam_lasso_3d <- lasso_cv_3d$lambda.min

rmse_lasso_cv_3d <- sqrt(
  lasso_cv_3d$cvm[lasso_cv_3d$lambda == bestlam_lasso_3d]
)

pred_lasso_3d <- as.numeric(
  predict(lasso_3d, s = bestlam_lasso_3d, newx = x_3d[test_3d, ])
)

rmse_lasso_test_3d <- sqrt(mean((pred_lasso_3d - y_test_3d)^2))

bestlam_lasso_3d
rmse_lasso_cv_3d
rmse_lasso_test_3d

coef_lasso_3d <- coef(lasso_cv_3d, s = "lambda.min")
coef_lasso_3d[coef_lasso_3d[,1] != 0, , drop = FALSE]

# Vi laver en opsummering af RSME for alle fire splits

rmse_lasso_overview <- data.frame(
  Tid = c("2 måneder", "10 dage", "7 dage", "3 dage"),
  RMSE_test_lasso = c(rmse_lasso_test_2m, rmse_lasso_test_10d, rmse_lasso_test_7d, rmse_lasso_test_3d)
)

# Her kan de forskellige RSME ses for 2m, 10d, 7d, 3d
rmse_lasso_overview

```

### 11.6.12 L) Ridge (glmnet) + CV + test-RMSE

```

library(glmnet)
set.seed(4)

# Vi laver et grid af mulige lambda-værdier (fra meget stor til lille)
grid <- 10^seq(10, -2, length.out = 100)

# -----
# Modellering Ridge - 2 Måneder Før
# -----

# Vi træner en Ridge-model på træningsdata.
ridge_2m <- glmnet(
  x_2m[train_2m, ],

```

```

y_2m[train_2m],
alpha = 0,
lambda = grid,
thresh = 1e-12
)

# Cross-validation: finder den lambda der giver mindst fejl på træningsdata.
set.seed(4)
ridge_cv_2m <- cv.glmnet(
  x_2m[train_2m, ],
  y_2m[train_2m],
  alpha = 0,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

# Den bedste lambda fra CV (den som giver lavest CV-fejl)
bestlam_ridge_2m <- ridge_cv_2m$lambda.min

# CV-RMSE (træningsdata via cross-validation)
rmse_ridge_cv_2m <- sqrt(
  ridge_cv_2m$cvm[ridge_cv_2m$lambda == bestlam_ridge_2m]
)

# Forudsigelser på testdata med den bedste lambda
pred_ridge_2m <- as.numeric(
  predict(ridge_2m, s = bestlam_ridge_2m, newx = x_2m[test_2m, ])
)

# Test-RMSE: gennemsnitlig fejl i antal tilskuere på testdatasættet
rmse_ridge_test_2m <- sqrt(mean((pred_ridge_2m - y_test_2m)^2))

bestlam_ridge_2m
rmse_ridge_cv_2m
rmse_ridge_test_2m

# =====
# Ridge - 10 dage før
# =====

ridge_10d <- glmnet(
  x_10d[train_10d, ],
  y_10d[train_10d],
  alpha = 0,
  lambda = grid,
  thresh = 1e-12
)

set.seed(4)
ridge_cv_10d <- cv.glmnet(
  x_10d[train_10d, ],
  y_10d[train_10d],
  alpha = 0,
  lambda = grid,

```

```

    nfolds = 5,
    thresh = 1e-12
)

bestlam_ridge_10d <- ridge_cv_10d$lambda.min

rmse_ridge_cv_10d <- sqrt(
  ridge_cv_10d$cvm[ridge_cv_10d$lambda == bestlam_ridge_10d]
)

pred_ridge_10d <- as.numeric(
  predict(ridge_10d, s = bestlam_ridge_10d, newx = x_10d[test_10d, ])
)

rmse_ridge_test_10d <- sqrt(mean((pred_ridge_10d - y_test_10d)^2))

bestlam_ridge_10d
rmse_ridge_cv_10d
rmse_ridge_test_10d

# =====
# Ridge - 7 dage før
# =====

ridge_7d <- glmnet(
  x_7d[train_7d, ],
  y_7d[train_7d],
  alpha = 0,
  lambda = grid,
  thresh = 1e-12
)

set.seed(4)
ridge_cv_7d <- cv.glmnet(
  x_7d[train_7d, ],
  y_7d[train_7d],
  alpha = 0,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

bestlam_ridge_7d <- ridge_cv_7d$lambda.min

rmse_ridge_cv_7d <- sqrt(
  ridge_cv_7d$cvm[ridge_cv_7d$lambda == bestlam_ridge_7d]
)

pred_ridge_7d <- as.numeric(
  predict(ridge_7d, s = bestlam_ridge_7d, newx = x_7d[test_7d, ])
)

rmse_ridge_test_7d <- sqrt(mean((pred_ridge_7d - y_test_7d)^2))

```

```

bestlam_ridge_7d
rmse_ridge_cv_7d
rmse_ridge_test_7d

# =====
# Ridge - 3 dage før
# =====

ridge_3d <- glmnet(
  x_3d[train_3d, ],
  y_3d[train_3d],
  alpha = 0,
  lambda = grid,
  thresh = 1e-12
)

set.seed(4)
ridge_cv_3d <- cv.glmnet(
  x_3d[train_3d, ],
  y_3d[train_3d],
  alpha = 0,
  lambda = grid,
  nfolds = 5,
  thresh = 1e-12
)

bestlam_ridge_3d <- ridge_cv_3d$lambda.min

rmse_ridge_cv_3d <- sqrt(
  ridge_cv_3d$cvm[ridge_cv_3d$lambda == bestlam_ridge_3d]
)

pred_ridge_3d <- as.numeric(
  predict(ridge_3d, s = bestlam_ridge_3d, newx = x_3d[test_3d, ])
)

rmse_ridge_test_3d <- sqrt(mean((pred_ridge_3d - y_test_3d)^2))

bestlam_ridge_3d
rmse_ridge_cv_3d
rmse_ridge_test_3d

# -----
# Samlet oversigt over Ridge (Test-RMSE)
# -----

rmse_ridge_overview <- data.frame(
  Tid = c("2 måneder", "10 dage", "7 dage", "3 dage"),
  RMSE_test_ridge = c(
    rmse_ridge_test_2m,
    rmse_ridge_test_10d,
    rmse_ridge_test_7d,
    rmse_ridge_test_3d
  )
)

```

```
)
)

rmse_ridge_overview
```

### 11.6.13 M) Best Subset Selection

```
# =====
# Best subset selection
# =====

# regsubsets() har ikke en predict()-funktion indbygget, så vi laver en selv.
# Ideen er:
# 1) Lav samme model-matrix (dummy-variabler osv.) som ved træning
# 2) Hent koefficienter for den modelstørrelse vi vil bruge (id)
# 3) Gang X-matrix med beta-vektoren -> så får vi prediction
predict.regsubsets <- function(object, newdata, id, ...) {
  form <- as.formula(object$call[[2]])
  mat <- model.matrix(form, newdata)
  coefi <- coef(object, id = id)
  xvars <- names(coefi)

  # Håndtering af manglende factor-levels: hvis en dummy-kolonne mangler i newdata,
  # tilføjer vi den som 0, så dimensionerne matcher modellen
  missing_vars <- setdiff(xvars, colnames(mat))
  if (length(missing_vars) > 0) {
    missing_mat <- matrix(0, nrow = nrow(mat), ncol = length(missing_vars))
    colnames(missing_mat) <- missing_vars
    mat <- cbind(mat, missing_mat)
  }

  # Til sidst tager vi kun de kolonner modellen forventer (i samme rækkefølge)
  mat <- mat[, xvars, drop = FALSE]

  # Prediction = X * beta
  mat %*% coefi
}

# -----
# Best Subset: 2 måneder før (bruger 80/20 fra setup)
# -----

# Vi tager træning og test fra de fælles setup-indeks (train_2m / test_2m)
vff_train <- df_2m[train_2m, ]
vff_test <- df_2m[test_2m, ]

# Cross-validation setup:
# k = antal folds, n = antal observationer i træningsdata
k <- 10
n <- nrow(vff_train)

set.seed(4) # styrer fold-inddelingen (så vi får samme CV hver gang)
folds <- sample(rep(1:k, length = n))
```

```

# Matrix til at gemme CV-fejl (MSE) for hver fold og hver modelstørrelse
# Antal modelstørrelser = antal variabler - 1 (fordi tilskuere er target)
cv.errors <- matrix(
  NA, k, dim(vff_train)[2] - 1,
  dimnames = list(NULL, paste(1:(dim(vff_train)[2] - 1)))
)

# Vi gennemløber alle folds:
# - træner på k-1 folds
# - validerer på fold j
for (j in 1:k) {

  # Forward selection: modellen bygger variabler på én ad gangen
  best.fit <- regsubsets(
    tilskuere ~ .,
    data = vff_train[folds != j, ],
    nvmax = dim(vff_train)[2] - 1,
    method = "forward"
  )

  # For hver mulig modelstørrelse (1..p) udregner vi MSE i fold j
  for (i in 1:(dim(vff_train)[2] - 1)) {
    pred <- predict(best.fit, vff_train[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

# Gennemsnitlig CV-MSE for hver modelstørrelse
mean.cv.errors <- apply(cv.errors, 2, mean)

# Vælg modelstørrelsen med lavest CV-MSE
best_size_2m <- which.min(mean.cv.errors)

# Fit endelig model på ALLE træningsdata (så koefficienter estimeres på mest data)
reg.best <- regsubsets(
  tilskuere ~ .,
  data = vff_train,
  nvmax = dim(vff_train)[2] - 1,
  method = "forward"
)

# Prediction på test-sættet med den valgte modelstørrelse
pred_best_subset <- predict(reg.best, vff_test, id = best_size_2m)

# Test-fejl (RMSE)
mse_best_subset <- mean((vff_test$tilskuere - pred_best_subset)^2)
rmse_bestsubset_test_2m <- sqrt(mse_best_subset)

# CV-RMSE (kvadratrods af den mindste CV-MSE)
rmse_bestsubset_cv_2m <- sqrt(min(mean.cv.errors))

# Output
best_size_2m
rmse_bestsubset_cv_2m
rmse_bestsubset_test_2m

```

```

coef(reg.best, best_size_2m)

# -----
# Best Subset: 10 dage før (bruger 80/20 fra setup)
# -----

vff_train <- df_10d[train_10d, ]
vff_test  <- df_10d[test_10d, ]

k <- 10
n <- nrow(vff_train)

set.seed(4)
folds <- sample(rep(1:k, length = n))

cv.errors <- matrix(
  NA, k, dim(vff_train)[2] - 1,
  dimnames = list(NULL, paste(1:(dim(vff_train)[2] - 1)))
)

for (j in 1:k) {

  best.fit <- regsubsets(
    tilskuere ~ .,
    data      = vff_train[folds != j, ],
    nvmax     = dim(vff_train)[2] - 1,
    method    = "forward"
  )

  for (i in 1:(dim(vff_train)[2] - 1)) {
    pred <- predict(best.fit, vff_train[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)
best_size_10d <- which.min(mean.cv.errors)

reg.best <- regsubsets(
  tilskuere ~ .,
  data      = vff_train,
  nvmax     = dim(vff_train)[2] - 1,
  method    = "forward"
)

pred_best_subset <- predict(reg.best, vff_test, id = best_size_10d)

mse_best_subset <- mean((vff_test$tilskuere - pred_best_subset)^2)
rmse_bestsubset_test_10d <- sqrt(mse_best_subset)
rmse_bestsubset_cv_10d <- sqrt(min(mean.cv.errors))

best_size_10d
rmse_bestsubset_cv_10d
rmse_bestsubset_test_10d

```



```

coef(reg.best, best_size_10d)

# -----
# Best Subset: 7 dage før (bruger 80/20 fra setup)
# -----

vff_train <- df_7d[train_7d, ]
vff_test  <- df_7d[test_7d, ]

k <- 10
n <- nrow(vff_train)

set.seed(4)
folds <- sample(rep(1:k, length = n))

cv.errors <- matrix(
  NA, k, dim(vff_train)[2] - 1,
  dimnames = list(NULL, paste(1:(dim(vff_train)[2] - 1)))
)

for (j in 1:k) {

  best.fit <- regsubsets(
    tilskuere ~ .,
    data    = vff_train[folds != j, ],
    nvmax   = dim(vff_train)[2] - 1,
    method  = "forward"
  )

  for (i in 1:(dim(vff_train)[2] - 1)) {
    pred <- predict(best.fit, vff_train[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)
best_size_7d <- which.min(mean.cv.errors)

reg.best <- regsubsets(
  tilskuere ~ .,
  data    = vff_train,
  nvmax   = dim(vff_train)[2] - 1,
  method  = "forward"
)

pred_best_subset <- predict(reg.best, vff_test, id = best_size_7d)

mse_best_subset <- mean((vff_test$tilskuere - pred_best_subset)^2)
rmse_bestsubset_test_7d <- sqrt(mse_best_subset)
rmse_bestsubset_cv_7d <- sqrt(min(mean.cv.errors))

best_size_7d
rmse_bestsubset_cv_7d
rmse_bestsubset_test_7d

```

```

coef(reg.best, best_size_7d)

# -----
# Best Subset: 3 dage før (bruger 80/20 fra setup)
# -----

vff_train <- df_3d[train_3d, ]
vff_test  <- df_3d[test_3d, ]

k <- 10
n <- nrow(vff_train)

set.seed(4)
folds <- sample(rep(1:k, length = n))

cv.errors <- matrix(
  NA, k, dim(vff_train)[2] - 1,
  dimnames = list(NULL, paste(1:(dim(vff_train)[2] - 1)))
)

for (j in 1:k) {

  best.fit <- regsubsets(
    tilskuere ~ .,
    data      = vff_train[folds != j, ],
    nvmax     = dim(vff_train)[2] - 1,
    method    = "forward"
  )

  for (i in 1:(dim(vff_train)[2] - 1)) {
    pred <- predict(best.fit, vff_train[folds == j, ], id = i)
    cv.errors[j, i] <- mean((vff_train$tilskuere[folds == j] - pred)^2)
  }
}

mean.cv.errors <- apply(cv.errors, 2, mean)
best_size_3d <- which.min(mean.cv.errors)

reg.best <- regsubsets(
  tilskuere ~ .,
  data      = vff_train,
  nvmax     = dim(vff_train)[2] - 1,
  method    = "forward"
)

pred_best_subset <- predict(reg.best, vff_test, id = best_size_3d)

mse_best_subset <- mean((vff_test$tilskuere - pred_best_subset)^2)
rmse_bestsubset_test_3d <- sqrt(mse_best_subset)
rmse_bestsubset_cv_3d <- sqrt(min(mean.cv.errors))

best_size_3d
rmse_bestsubset_cv_3d
rmse_bestsubset_test_3d

```

```
coef(reg.best, best_size_3d)
```

#### 11.6.14 N) Sammenligning af RSME

```
#=====
# Sammenligning af RMSE Test og CV
#=====

# Vi bruger RSME i stedet for MSE for at give et lettere fortolkende tal.
# Herefter kan vi sammenligne dem med hinanden for at finde den model med bedst prædiktionssevne.

#RSME TEST 2 måneder
rmse_bestsubset_test_2m
rmse_ridge_test_2m
rmse_lasso_test_2m

#RSME Træning 2 måneder
rmse_ridge_cv_2m
rmse_bestsubset_cv_2m
rmse_lasso_cv_2m

#RSME TEST 10 dage
rmse_bestsubset_test_10d
rmse_ridge_test_10d
rmse_lasso_test_10d

#RSME Træning 10 dage
rmse_ridge_cv_10d
rmse_bestsubset_cv_10d
rmse_lasso_cv_10d

#RSME TEST 7 dage
rmse_bestsubset_test_7d
rmse_ridge_test_7d
rmse_lasso_test_7d

#RSME Træning 7 dage
rmse_ridge_cv_7d
rmse_bestsubset_cv_7d
rmse_lasso_cv_7d

#RSME TEST 3 dage
rmse_bestsubset_test_3d
rmse_ridge_test_3d
rmse_lasso_test_3d

#RSME Træning 3 dage
rmse_ridge_cv_3d
rmse_bestsubset_cv_3d
```

```
rmse_lasso_cv_3d
```

### 11.6.15 O) GGplot af RSME for modellerne (Kun for 3dage før)

```
# -----  
# Vi laver et GGplot af RSME for modellerne (Kun for 3dage før)  
# -----  
  
library(ggplot2)  
  
rmse_data_3d <- data.frame(  
  Model = c("Best Subset", "Lasso", "Ridge", "LM"),  
  RMSE = c(  
    rmse_bestsubset_test_3d,  
    rmse_lasso_test_3d,  
    rmse_ridge_test_3d,  
    rmse_lm_test_3d  
  )  
)  
  
ggplot(rmse_data_3d, aes(x = Model, y = RMSE, fill = Model)) +  
  geom_col(color = "black") +  
  theme_minimal() +  
  labs(  
    title = "Test-RMSE for modeller (3 dage før kamp)",  
    x = "Model",  
    y = "RMSE (tilskuere)"  
  ) +  
  theme(legend.position = "none")
```

### 11.6.16 P) Samlet plot med de forskellige RMSE ift tid før kamp

```
library(ggplot2)  
  
rmse_over_time <- data.frame(  
  Tid = factor(  
    c("2 måneder", "10 dage", "7 dage", "3 dage"),  
    levels = c("2 måneder", "10 dage", "7 dage", "3 dage")  
  ),  
  LM = c(  
    rmse_lm_test_2m,  
    rmse_lm_test_10d,  
    rmse_lm_test_7d,  
    rmse_lm_test_3d  
  ),  
  BestSubset = c(  
    rmse_bestsubset_test_2m,  
    rmse_bestsubset_test_10d,  
    rmse_bestsubset_test_7d,  
    rmse_bestsubset_test_3d  
  ),  
)
```

```

Lasso = c(
  rmse_lasso_test_2m,
  rmse_lasso_test_10d,
  rmse_lasso_test_7d,
  rmse_lasso_test_3d
),
Ridge = c(
  rmse_ridge_test_2m,
  rmse_ridge_test_10d,
  rmse_ridge_test_7d,
  rmse_ridge_test_3d
)
)

# Gør data "long" (uden ekstra pakker)
rmse_over_time_long <- data.frame(
  Tid = rep(rmse_over_time$Tid, times = 4),
  Model = rep(c("LM", "Best Subset", "Lasso", "Ridge"), each = nrow(rmse_over_time)),
  RMSE = c(rmse_over_time$LM, rmse_over_time$BestSubset, rmse_over_time$Lasso, rmse_over_time$Ridge)
)

ggplot(rmse_over_time_long, aes(x = Tid, y = RMSE, group = Model, color = Model)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(
    title = "Test-RMSE som funktion af tid før kamp",
    x = "Tid før kamp",
    y = "RMSE (tilskuere)"
  )

```

### 11.6.17 Q) Best-Worst case (2m)

```

## =====
# Prædiktion på nye data 2 måneder før - Best/Worst case
# =====

# Vi bruger ridge, da den havde lavet test RSME 2 måneder

# Her laver vi to "scenarier": 1) Best-Case ift tilskuere 2) Worst Case ift tilskuere
# Og så bruger vi vores ridge model til at lave en prediction.

# Når vi laver nye data manuelt, kan R ellers finde på at lave factor-levels forkert.
# Derfor tvinger vi værdien til at bruge de samme levels som i df_2m.
as_level <- function(value, ref_factor) {
  factor(value, levels = levels(ref_factor))
}

# -----
# Vi vælger nogle "standard" valg til scenarierne
# -----

# Vi tager den seneste sæson i datasættet (tænkt som "seneste forår" eller nyeste sæson)

```

```

season_latest <- levels(df_2m$season)[length(levels(df_2m$season))]

# Vi vælger en modstander som "best" og "worst".
# Her bruger vi bare rækkefølgen i levels, men man kan selv justere det til noget mere realistisk.
modstander_best <- levels(df_2m$modstander)[1]
modstander_worst <- levels(df_2m$modstander)[10]

# Vi vælger en standard kategori (igen: kan justeres hvis I har bestemte kategorier)
kategori_default <- levels(df_2m$kategori)[1]

# -----
# Best-case (2 måneder før)
# -----
# Her prøver vi at "skrue op" for de ting der typisk giver flere tilskuere:
# - høje historiske gennemsnit (90%-percentilen)
# - weekendkamp (lørdag)
# - ikke helligdag (kan også sættes TRUE hvis det giver mening i jeres data)
new_best_2m <- data.frame(
  season = as_level(season_latest, df_2m$season),
  modstander = as_level(modstander_best, df_2m$modstander),
  kategori = as_level(kategori_default, df_2m$kategori),
  mean_tilskuere = quantile(df_2m$mean_tilskuere, 0.90, na.rm = TRUE),
  mean_tilskuere_saeson = quantile(df_2m$mean_tilskuere_saeson, 0.90, na.rm = TRUE),
  weekday = as_level("Sat", df_2m$weekday),
  is_holiday = FALSE
)

# -----
# Worst-case (2 måneder før)
# -----
# Her gør vi det modsatte:
# - lave historiske gennemsnit
# - hverdag (mandag)
new_worst_2m <- data.frame(
  season = as_level(season_latest, df_2m$season),
  modstander = as_level(modstander_worst, df_2m$modstander),
  kategori = as_level(kategori_default, df_2m$kategori),
  mean_tilskuere = quantile(df_2m$mean_tilskuere, 0.10, na.rm = TRUE),
  mean_tilskuere_saeson = quantile(df_2m$mean_tilskuere_saeson, 0.10, na.rm = TRUE),
  weekday = as_level("Mon", df_2m$weekday),
  is_holiday = FALSE
)

# -----
# Predict med Ridge (glmnet)
# -----
# glmnet kræver at nye data laves om til en model.matrix (uden intercept).
# Det er samme "format" som vi trænede modellen på.
x_new_best_2m <- model.matrix(~ ., new_best_2m)[, -1]
x_new_worst_2m <- model.matrix(~ ., new_worst_2m)[, -1]

# Vi bruger ridge_2m modellen og den bedste lambda (fra CV) til at lave prediction
pred_best_2m <- as.numeric(predict(ridge_2m, s = bestlam_ridge_2m, newx = x_new_best_2m))
pred_worst_2m <- as.numeric(predict(ridge_2m, s = bestlam_ridge_2m, newx = x_new_worst_2m))

```

```
# -----
# Vi samler resultaterne i en lille tabel så det er nemt at vise
# -----
scenario_preds_2m <- data.frame(
  scenario = c("Best-case (2m)", "Worst-case (2m)"),
  predicted_tilskuere = c(pred_best_2m, pred_worst_2m)
)

print(scenario_preds_2m)
```

### 11.6.18 R) Best-Worst case (2m) VFF vs FCN 1/3 2026

```
## =====
## Prediction (2m) - Viborg vs FCN, 1/3 kl. 17 (lørdag)
## Best/Worst case ved at variere historiske gennemsnit
## =====

# Sikrer samme factor-levels som i df_2m
as_level <- function(value, ref_factor) {
  factor(value, levels = levels(ref_factor))
}

# Faste kamp-info (kendte variabler)
season_latest <- levels(df_2m$season)[length(levels(df_2m$season))]
modstander_fixed <- "FCN"
weekday_fixed <- "Sat" # 1/3 er lørdag

# Vi kender FCN-kategorien i vores inddeling (B)
kategori_fcn <- as_level("B", df_2m$kategori)

# Best/Worst antagelser: variation i historiske gennemsnit (10% og 90% percentil)
best_mean <- quantile(df_2m$mean_tilskuere, 0.90, na.rm = TRUE)
worst_mean <- quantile(df_2m$mean_tilskuere, 0.10, na.rm = TRUE)

best_mean_season <- quantile(df_2m$mean_tilskuere_saeson, 0.90, na.rm = TRUE)
worst_mean_season <- quantile(df_2m$mean_tilskuere_saeson, 0.10, na.rm = TRUE)

# Best-case (FCN, lørdag)
new_best_fcn_2m <- data.frame(
  season = as_level(season_latest, df_2m$season),
  modstander = as_level(modstander_fixed, df_2m$modstander),
  kategori = kategori_fcn,
  mean_tilskuere = best_mean,
  mean_tilskuere_saeson = best_mean_season,
  weekday = as_level(weekday_fixed, df_2m$weekday),
  is_holiday = FALSE
)

# Worst-case (FCN, lørdag)
new_worst_fcn_2m <- data.frame(
  season = as_level(season_latest, df_2m$season),
  modstander = as_level(modstander_fixed, df_2m$modstander),
  kategori = kategori_fcn,
```

```

mean_tilskuere = worst_mean,
mean_tilskuere_saeson = worst_mean_season,
weekday = as_level(weekday_fixed, df_2m$weekday),
is_holiday = FALSE
)

# glmnet (i det her tilfælde bruger vi ridge) kræver model.matrix (intercept er fjernet)
x_best <- model.matrix(~ ., new_best_fcn_2m)[, -1]
x_worst <- model.matrix(~ ., new_worst_fcn_2m)[, -1]

pred_best <- as.numeric(predict(ridge_2m, s = bestlam_ridge_2m, newx = x_best))
pred_worst <- as.numeric(predict(ridge_2m, s = bestlam_ridge_2m, newx = x_worst))

scenario_preds_fcn_2m <- data.frame(
  kamp = "Viborg vs FCN (1/3 kl. 17)",
  scenario = c("Best-case (2m)", "Worst-case (2m)"),
  predicted_tilskuere = c(pred_best, pred_worst)
)

print(scenario_preds_fcn_2m)

```