



GENERATION DATA

USING DATA FOR PROFIT

M2: Arbejde med data

Fuldfør dette modul sammen med modul 2-noter

Dette program er finansieret med støtte fra
Europa-Kommissionen



Føl dig sikker på at koordinere hele datalogi-pipeline fra erhvervelse til offentliggørelse.

- en) Forstå stadierne i dataene rørledning.*
- b) Klassificer data efter kilde og forstå dets behandlingsmuligheder.*
- c) Vælg passende opbevaring og analyse muligheder.*
- d) Evaluer omkostninger og fordele ved opbygge forudsigelige modeller eller bruge algoritmer til at opdage struktur i data.*
- e) Præsentere data i en attraktiv og nøjagtig måde til forskellige formål.*



Oversigt

1 Dataværdikæden

2 Dataindsamling

Datakilder

Datatyper

3

Data opbevaring

Onsite v. Cloud opbevaring

4

Dataanalyse

Grundlæggende teknikker

Avancerede teknikker


5

Datatolkning og anvendelse

Datavisualisering

Implementering af indsigt

Dette program er finansieret med støtte fra Europa-Kommissionen. Forfatteren er udelukkende ansvarlig for denne publikation (meddelelse), og Kommissionen påtager sig intet ansvar for brug af informationen deri.



Indstil Scene Case Study
(se ressourcenotater 2)

Tesco Clubcard

*” Dette bekymrer mig virkelig,
fordi du ser ud til at vide mere om
min forretning i
tre måneder, end jeg har
lært på 30 år. ”*

Lord Ian MacLaurin, formand

*Du kan ikke styre det, du
ikke måler.*

W. Edwards Deming og Peter Drucker

””““

1. DATAVÆRDKÆDEN - et klart overblik over værdikæden



Data indsamles eller oprettes.

Sørg for, at det er nøjagtigt, konsekvent og forsættlig.

Bestem, hvad du skal vide, inden du begynder at samle.

Data opbevares sikkert.

Forstå den aktuelle brug og den mængde lager, du har brug for i fremtiden.

Overvej hastighed på adgang og lethed af sikkerhedskopier.

Data behandles for at udtrække betydning.

Vælg passende analysemetoder.

Data bruges til at informere om beslutningstagning.

Overvej hvordan resultater præsenteres for at muliggøre fortolkning.

Bestem, hvem der skal modtage og handle efter data.

Vores næste diagram viser en mere kompleks oversigt



GENERATIONSDATA | BRUG AF DATA TIL PROFIT



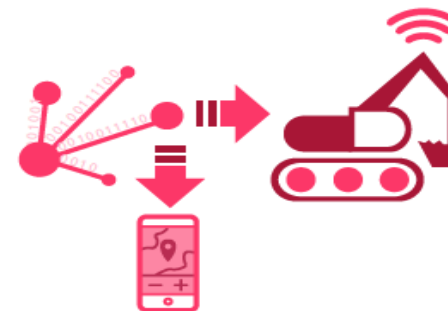
Frigør data

En person, proces eller system skaber og offentliggør / deler data



Manipulere data

En person, proces eller system transformerer, flytter eller analyserer data



Forbruge data

En person, en proces eller et system drager fordel af manipulerede data



Erhverve

Indtast data fra sensorer, systemer eller mennesker, der registrerer dens oprindelse og samtykke til brug hvor det er muligt



butik

Registrer data på et betroet sted, der er både sikkert og let tilgængeligt til yderligere manipulation



Samlet

Kombiner forskellige datasæt for at oprette et større datasæt, der er større end summen af dets dele.



Analysere

Undersøg og transformer data med formålet med at udtrække information og opdage nye indsigter.



Brug

Anvend den indsigt, der er opnået ved dataanalyse, til at tage beslutninger, påvirke ændringer eller levere et produkt eller en tjeneste.



Del / Sælg

Giv adgang til datasæt eller dataindsigt til nye sæt datamanipulatorer eller forbrugere.



Kassér

Fjern data fra servere for at forhindre fremtidig frigivelse eller brug

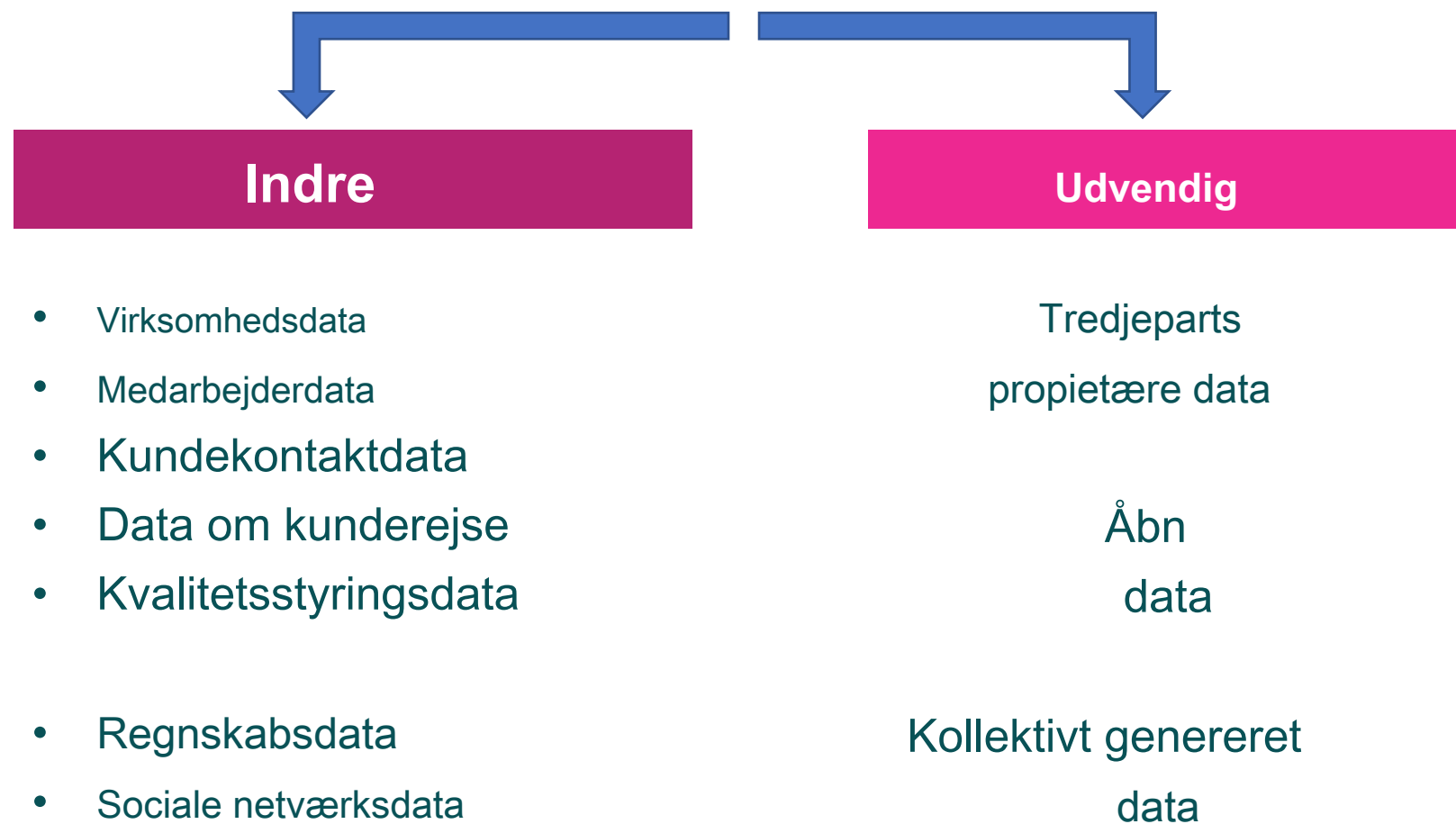
2.

Dataindsamling

Processen for samling information om variabler af interesse på en systematisk måde.

Hvilke fordele eller uheld er hver enkelt datakilde?

KILDER DATA



Er eksterne data værd at koste?

I et forskningsprojekt:

- 7 små virksomheder i Nordirland, der sælger forarbejdede fødevarer, fik adgang til en stor supermarkeds kædes data.
- Ejere og ledere deltog i workshops for at lære at hente og analysere de mest relevante data.



Er eksterne data værd at koste?

- Virksomheder kunne forestille sig innovationer i lang rækkevidde snarere end at forsøge at efterligne eksisterende produkter.
- Dataene forstærkede virksomhedernes iboende iværksætterånd. De fleste af lederne delte kortoplysninger med deres firmaer og opfordrede medarbejderne til at blive involveret og tilbyde nye ideer.



2.

Dataindsamling

Processen for samling information om variabler af interesse på en systematisk måde.

Hvilke fordele eller fordele er der ved hver type data?

TYPE DATA



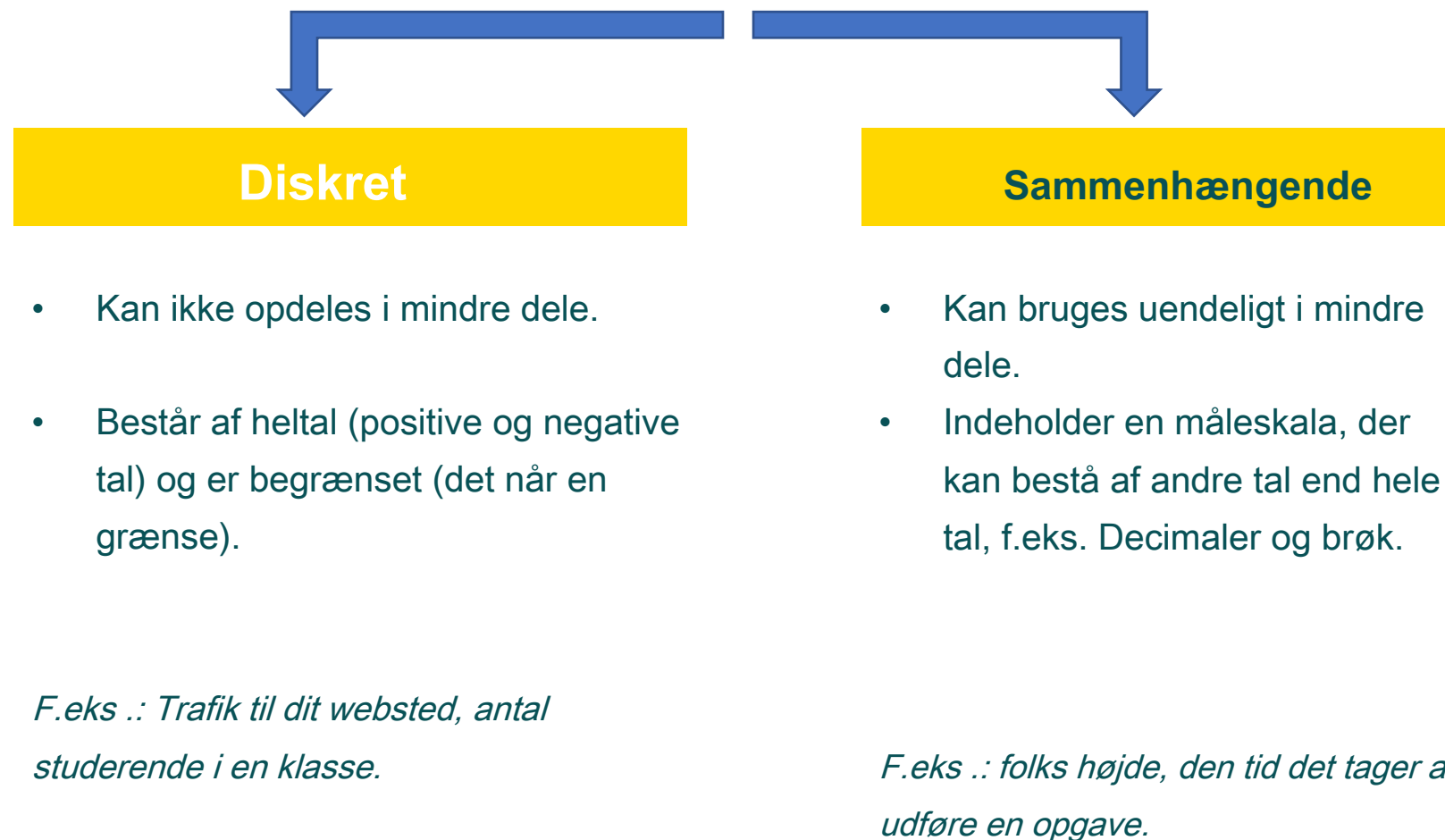
2.

Dataindsamling

Processen for samling information om variabler af interesse på en systematisk måde.

Hvilke fordele eller fordele er der ved hver type data?

TYPE DATA



TYPER KVANTITATIVE DATA



nominelt

Kvalitative data
uden iboende
skala eller værdi
knyttet.

F.eks. Type boliger.



ordenstal

Data organiseret på et
system eller skala.

*F.eks. Ja måske,
ingen.*



Interval

Data, der har en naturlig
rækkefølge, og afstanden
mellem hver værdi er ens.

*F.eks. Temperatur, uretider eller
IQ-scoringer.*



Forhold

Data, der har en naturlig
rækkefølge, ens afstand mellem
værdier, men nul på skalaen
betyder ikke
eksisterer.

*F.eks. Vægt, års uddannelse,
indtjent indkomst.*

*Hvis du ikke ved, hvor du skal, hvordan vil du
vide, hvordan man gør det
kom derhen?*

””““

3. Datalagring

Arkivering af data i digital form til senere adgang.

Hvad skal du tage højde for, når du beslutter, hvordan du gemmer dine data?

OPBEVARINGSMuligheder



2. Datalagring

Arkivering af data i digital form til senere adgang.

Hvad skal du tage højde for, når du beslutter, hvordan du gemmer dine data?

HVORDAN VÆLGER?

HVEM har brug for adgang til hvilken type information, og hvornår?

HVILK type data skal vi gemme?

HVORDAN MANGE data har vi, og hvor meget vil de vokse i fremtiden?

HVOR LANG skal vi gemme dataene?

HVORDAN SIKKER skal dataene være?



Bestem, hvilken lagringsindstilling der bedst er baseret på dine behov.



Opret en datalagringspolitik, som alt personale kender og overholder.

3. Dataanalyse

Evaluerer af data ved hjælp af analytiske og statistiske værktøjer til at finde nyttige oplysninger og hjælpe til beslutningstagning i virksomheden.

FORBEREDELSE

RENGØRING. Forkerte eller ufuldstændige data kan føre til skæve resultater. Analytikere renser datasæt ved f.eks. At fjerne ugyldige poster eller duplikatoplysninger.

SAMMENLÆGNING. Rå data kan indsamles og gemmes i flere forskellige formater. Datasamling er samlingen af oplysninger fra forskellige datasæt, så dataanalyseværktøjer kan importere den og arbejde med dem.

ANALYSE

STATISTISK ANALYSE. Den mest almindelige type analyse involverer den nøje at undersøge for det meste kvantitative data for at identificere tendenser. Standardanalyser inkluderer det aritmetiske middel, standardafvigelse, regression og hypotesetest.

TEKSTANALYSE. Semistrukturerede eller ustrukturerede data i form af fortællende tekster kan analyseres ved at "analysere" teksterne for at udtrække maskinlæsbare datamaterialer.

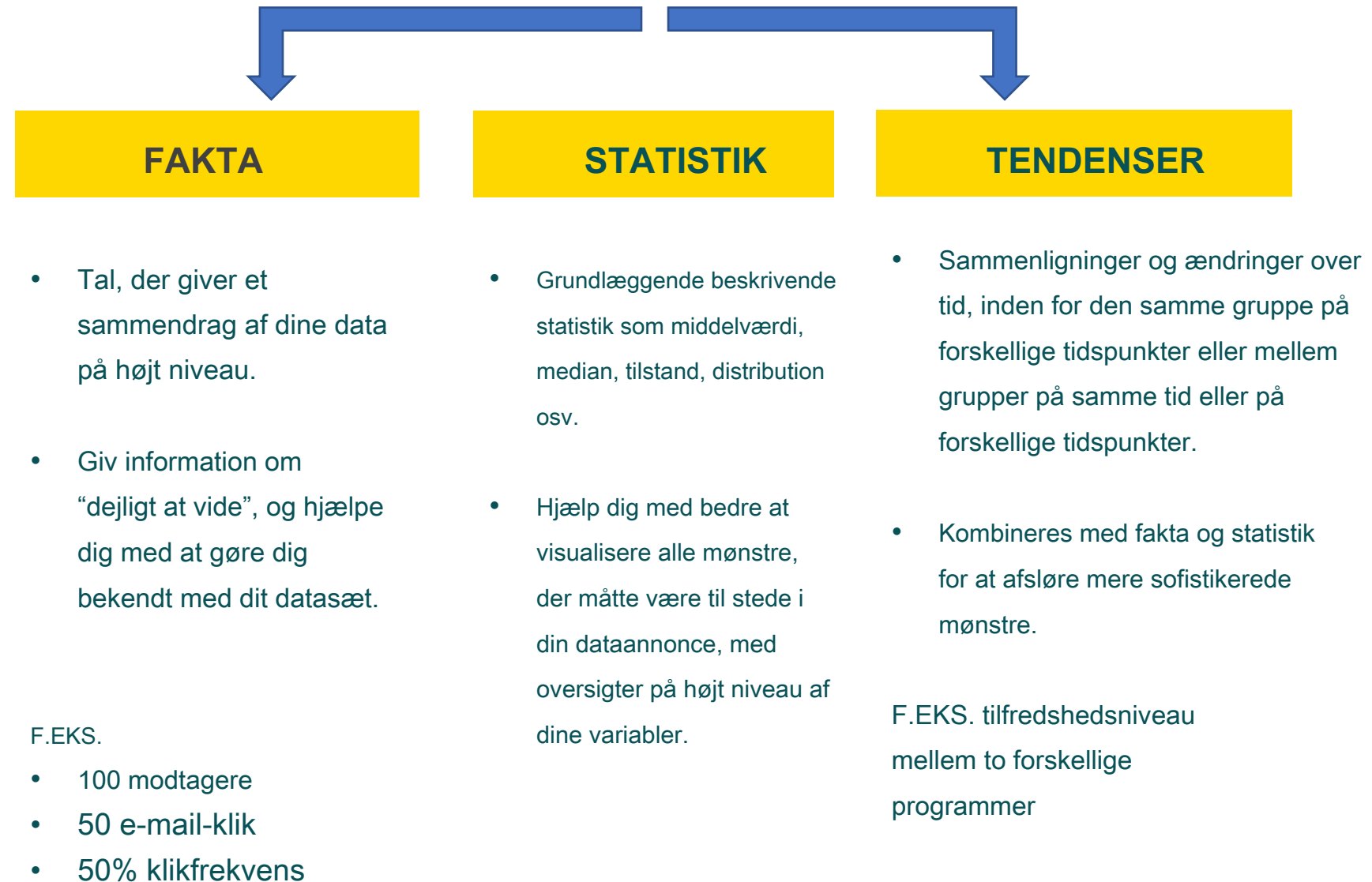
DATA MINING. Hvis du arbejder med meget store datasæt, bruger data mining maskinlæring, kunstig intelligens og statistik til at finde mønstre og korrelationer for at forudsige resultater.

VISUEL ANALYTIK. For at identificere outliers, huller, tendenser og interessante datapunkter bygger mange analytikere diagrammer og visuelle repræsentationer, der giver forskellige perspektiver og peger på områder, der berettiger yderligere undersøgelse.

3. Dataanalyse

Evaluerer af data ved hjælp af analytiske og statistiske værktøjer til at finde nyttige oplysninger og hjælp til beslutningstagning i virksomheden.

HVAD TIL Kigge efter



3. Dataanalyse

Evaluering af data ved hjælp af analytiske og statistiske værktøjer til at finde nyttige oplysninger og hjælp til beslutningstagning i virksomheden.

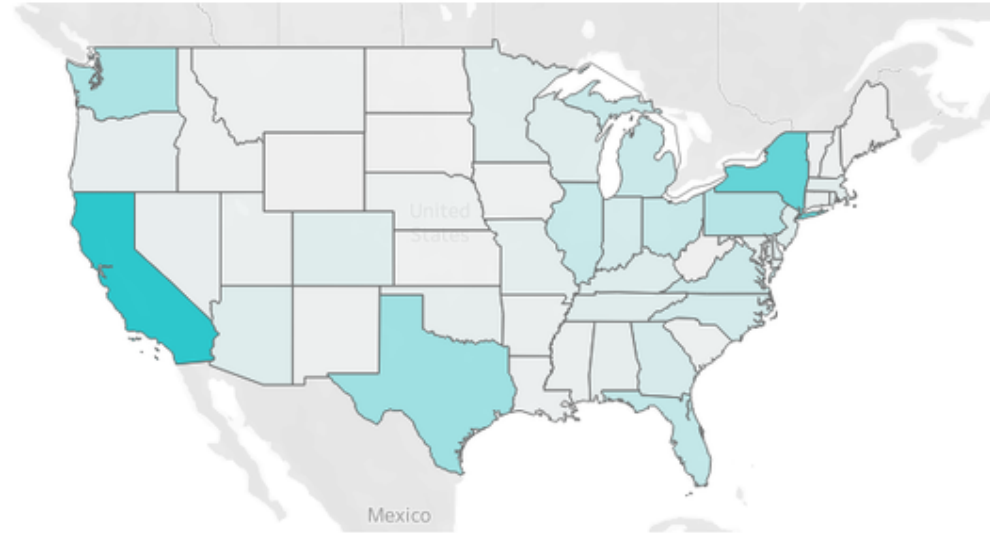
"Hvad er forskellen mellem dataanalyse og datavisualisering?"

KILDE Forbes, marts 2019 Jackson, Ann. "Salg og rentabilitet af amerikansk stat" i Murray, Eva,

SALES & PROFITABILITY



SALES BY STATE
CLICK TO FILTER TO STATE



\$2,297,201

Sales

12%

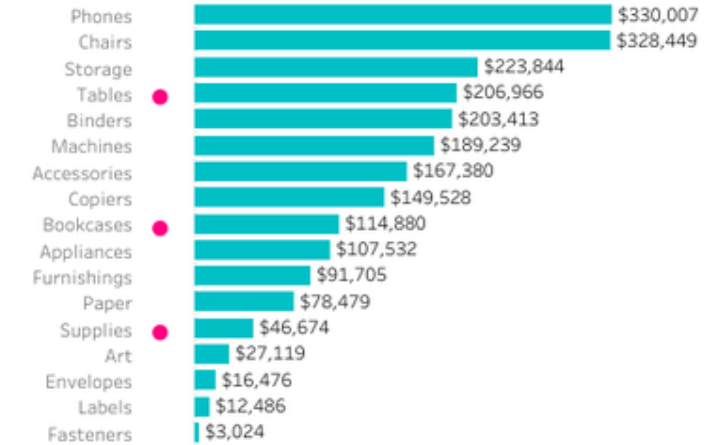
Profit Ratio

16%

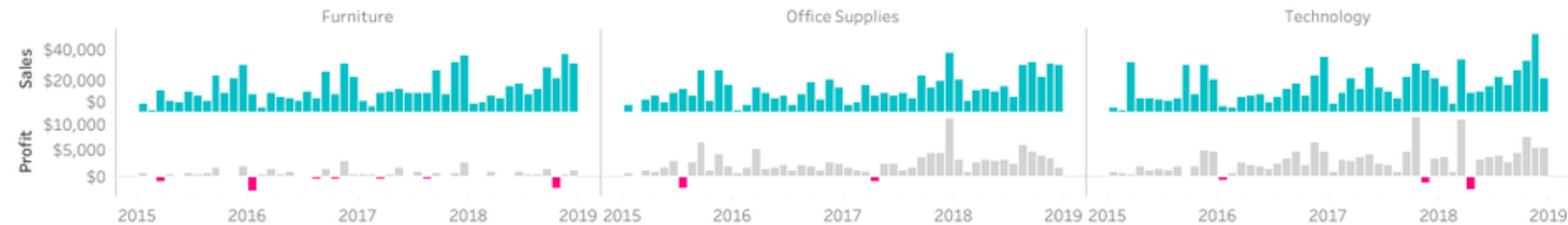
Discount

SALES BY SUBCATEGORY

DOT INDICATES NEGATIVE PROFIT



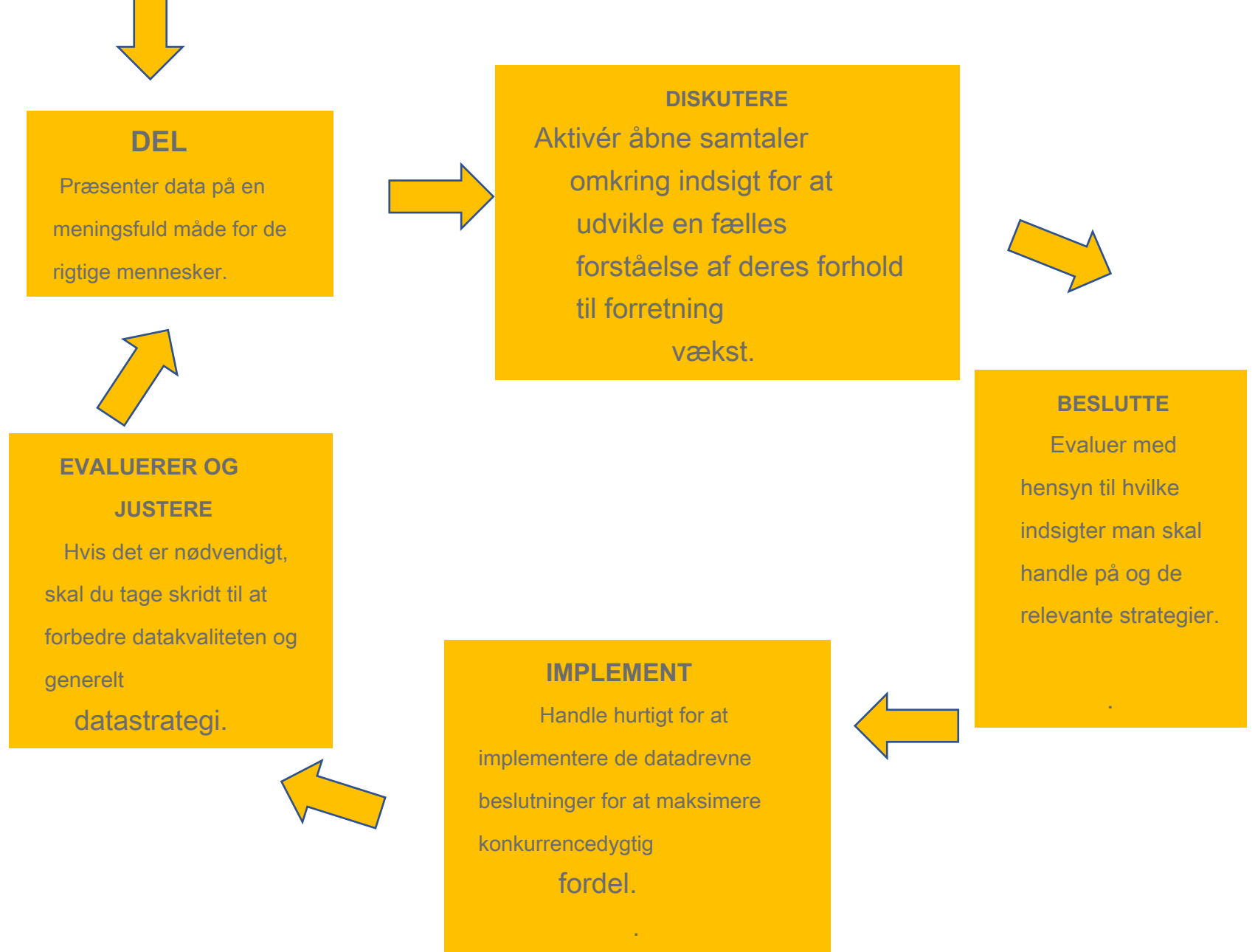
SALES & PROFIT BY CATEGORY



4. Data fortolkning og brug

Den effektive
oversættelse af dataindsigt i
handling, der understøtter
forretningsvækst.

*Hvad kræver det at være en
datadrevet virksomhed?*



4. Data fortolkning og brug

Den effektive
oversættelse af dataindsigt i
handling, der understøtter
forretningsvækst.

*Betydningen af
datavisualisering.*

Spørgsmål

I hvilket diagram er der en større forskel mellem datasæt?



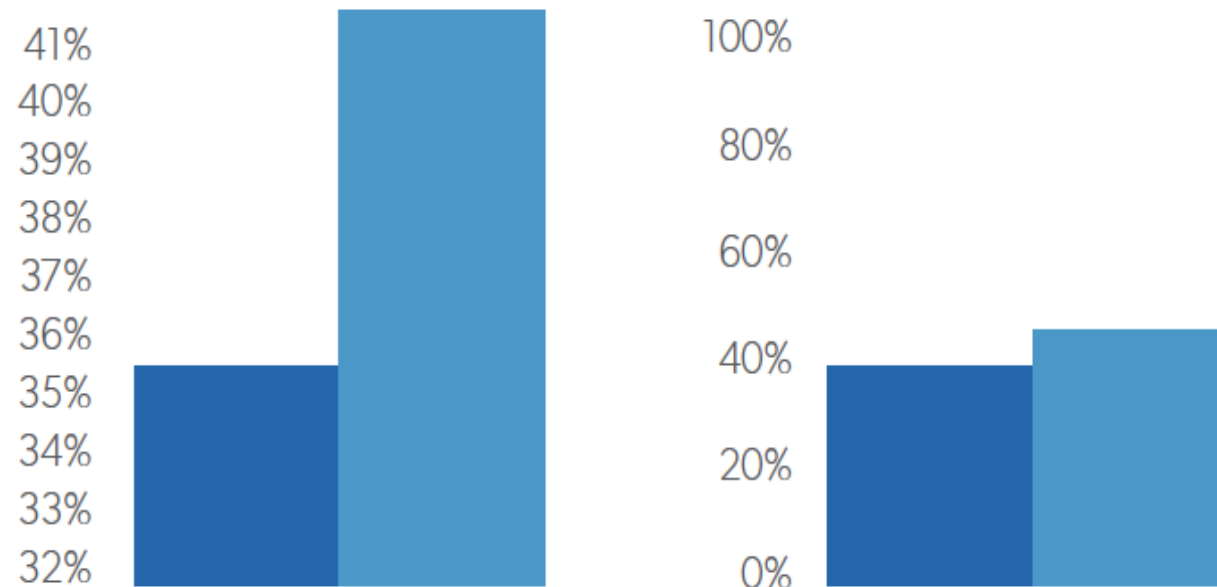
4. Data fortolkning og brug

Den effektive
oversættelse af dataindsigt i
handling, der understøtter
forretningsvækst.

*Betydningen af
datavisualisering.*

Spørgsmål

I hvilket diagram er der en større forskel mellem datasæt?



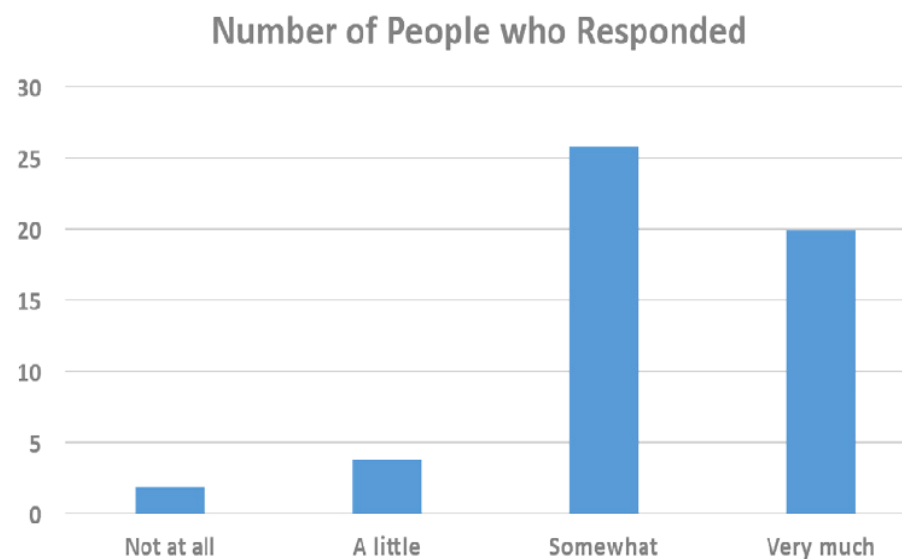
4. Data fortolkning og brug

Den effektive
oversættelse af dataindsigt i
handling, der understøtter
forretningsvækst.

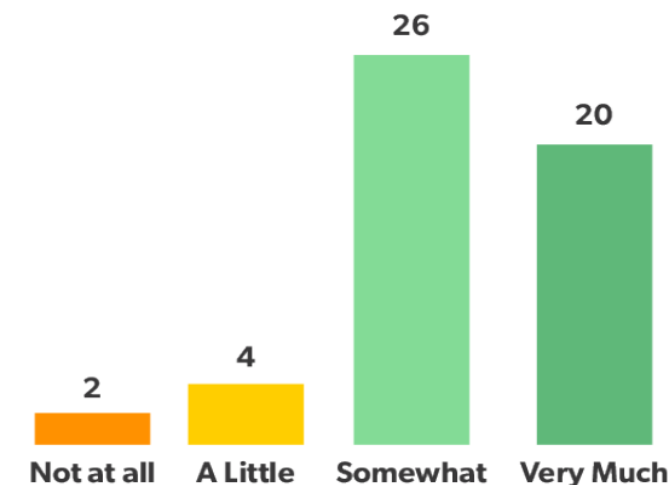
*Betydningen af
datavisualisering.*

SPØRGSMÅL Hvilken måde at vise dataene er mest effektiv på?

Kategori	Nej. Svar
4 Meget meget	20
3	26
2	4
1 slet ikke	2



How much did this program change your
opinion about the importance of service?



ORDLISTE

Datatilgang

Handlingen eller metoden til at se eller hente lagrede data.

Datasamling

Handlingen med at indsamle data fra flere kilder med henblik på rapportering eller analyse

Rengøring af data

Processen med at gennemgå og revidere data for at slette duplikater, rette fejl og give konsistens.

Dataindsamling

Enhver proces, der fanger enhver type data

Dataintegritet

Målet for tillid en organisation har i nøjagtigheden, fuldstændigheden, aktualiteten og gyldigheden af dataene.

Datamigrering

Processen med at flytte data mellem forskellige lagringstyper eller formater eller mellem forskellige computersystemer.

Data mining

Processen med at udlede mønstre eller viden fra store datasæt. Formålet er at finjustere data til et mere forståeligt og sammenhængende sæt information.

Datamodel, datamodellering

En datamodel definerer strukturen af dataene med det formål at kommunikere mellem funktionelle og tekniske personer for at vise data, der er nødvendige for forretningsprocesser, eller til at kommunikere en plan for at udvikle, hvordan data gemmes og fås adgang til blandt medlemmer af applikationsudviklingsholdet.



ORDLISTE

Datasikkerhed

Praksisen med at beskytte data mod ødelæggelse eller uautoriseret adgang.

Datasæt

En samling af data, typisk i tabelform.

Datakilde

Enhver udbyder af data - for eksempel en database eller en datastrøm.

Data steward

En person, der er ansvarlig for data, der er gemt i et datafelt .

Datastruktur

En bestemt måde at lagre og organisere data på.

Datavisualisering

En visuel abstraktion af data designet til det formål at udlede mening eller formidle information mere effektivt.

Data varehus

Et sted at gemme data med henblik på rapportering og analyse.

AKTIVITET 1

Data kategorisering

Forestil dig, at din virksomhed har et callcenter til kundeservice. Du har et VoIP-system, der opretter digitale logfiler og optagelser af opkald.

Skriv en liste over alle de data, du kunne indsamle og analysere, og dataene:

- a) Hvor gemmes disse data?
- b) Er det kvantitativt eller kvalitativt?
- c) Er det nominelt, ordinært, interval eller forhold? Diskret eller kontinuert?
- d) Hvilken type analyser ville være passende at køre på disse data for at udtrække betydning?

Eksempler:

- *Demografisk nummeropkald: navn og opkaldskode*
- *Angivet grund til opkaldet*
- *Opkaldets længde*
- *...*

AKTIVITET 2

Matching af data
med formålet.

Hvad er den rigtige datakilde til at hjælpe med følgende opgaver?

1. At forstå kundeadfærd og motivation
2. Identificering af tendenser
3. Tjek konkurrencen
4. Forbedring af operationer
5. Rekruttering og styring af talent
6. Tilpasning af din forretningsmodel