



GENERATION DATA

USING DATA FOR PROFIT

ARBEJDER MED DATA



Erasmus+

This project has been funded with support from the European Commission

Dette projekt er blevet finansieret med støtte fra Europa-Kommissionen. Denne publikation [meddelelse] afspejler kun autoritetens synspunkter, og Kommissionen kan ikke holdes ansvarlig for enhver brug, der kan gøres af oplysningerne deri.



START MED EN HISTORIE

EKSPONERING

I begyndelsen af 1990'erne var Tesco andenplads i det britiske supermarked til Sainsburys. Som enhver enorm forhandler havde de MASSE data fra transaktioner, men ingen måde at personalisere dem på. Bare information om, hvilke produkter der sælger for hvor meget og hvor. Dette var fantastisk til at forstå tendenser med tilbagevirkende kraft, men ikke så nyttigt til at dyrke forretningen, øge kundernes forbrug og loyalitet.

I 1995 lancerede de Tesco Clubcard. Clubcard-indehavere kunne samle point (1 point for hvert brugt pund) og modtage kuponer til deres hjem med magasin hver 3. måned. De troede, at en salgsforhøjelse på 1,6% ville dække omkostningerne på 10 millioner pund ved lancering af kortet og belønningskuponerne. Måned efter lanceringen overhalede de Sainsburys og forblev foran i årtier. Hævningen var omkring 4% ved højdepunktet og derefter afregnet til 2%.

Det ser ud til, at succesen var baseret på at øge kundernes motivation for at vende tilbage til butikken, fordi de ville tjene kuponer. Imidlertid genererer et program med millioner af medlemmer en enorm mængde data. Tesco arbejdede sammen med Dunnhumby, et marketinganalysefirma, for at behandle dataene. Efter den første prøvekørsel, da de præsenterede deres fund og begrundelse for yderligere udvulning til bestyrelsen, sagde Tescos formand, Lord Ian MacLaurin, berømt: "Nå, dette bekymrer mig virkelig, fordi du ser ud til at vide mere om min virksomhed på tre måneder, end jeg har lært i 30 år."

I september 2002 rullede Tesco ud målrettede elektroniske kuponer i kassen - automatisk udskrevet på kasserne på købstidspunktet, men ikke som konkurrenter, ikke udløst ved køb af en bestemt vare, men af kunderne, der ejer den seneste købshistorik. Ud af 100 - 200 tilbud valgte systemet de relevante for den kunde.

I november 2002 tilføjede et online klubkort - da belønningspoint konverteret til ecoupons, der skulle indløses online, blev alle de daværende 10 millioner kunder opfordret til at logge ind online. Virksomheden brugte dataene ved at målrette kunder med mailshots, der fremhæver forskellige tilbud og fordele for velsegmenterede grupper baseret på kendte udgifts- og shoppingvaner.

(Bemærk: Da Dunnhumby var så succesrig med Clubcard og dets tilknyttede kundeanalyser, købte Tesco faktisk virksomheden i processen med at skabe multimillionærer af sine grundlæggere.)

FORTOLKNING

Tesco betragtes som det primære eksempel på vellykket kundeloyalitetsordning ved hjælp af big data. Omend på grund af hård konkurrence har de set et fald i overskuddet i de sidste 8 år. Hvorfor har denne sag haft så stor indflydelse?

Som en "normal" forhandler havde de enorme mængder data fra transaktioner, men ingen måde at gøre det rigtigt på. Ved at indføre clubcard fik de straks enormt mere berigede kundedata: hvem du var, hvor du boede, hvor du handlede, når du handlede, hvor meget du brugte, og hvilke afdelinger du handlede i. Senere satte Clubcard købsvaner med et navn (og et ansigt) og en e-mail-adresse. Så hvad der startede som et håb om en loyalitetseffekt, åbnede døren til en direkte marketingkanal for forbrugerne, som gjorde det muligt for dem at få adgang til uovertrufne strømme af data og drive virksomheden bedre. "Det var første gang, at en masseforhandler kunne tale med individuelle kunder på et personligt plan." Tim Mason, der blev Tescos marketingdirektør på bestyrelsesniveau i 1995.



EKSPONERING

1. DATAVÆRDIKÆDEN

Nøglen til at forstå, hvordan man arbejder med data er datastrømmen eller dataværdikæden. Det ses generelt som sekventielt. Nogle mennesker foretrækker at bruge en livscyklusvisualisering, dette har nogle fordele. I begge tilfælde skal du tage et helhedsbillede: selvom du bevæger dig trin for trin langs strømmen, har du altid brug for en klar idé om, hvor du vil ende, inden du starter.

2. DATAINDSAMLING: DATAKILDER

Data kan komme fra interne eller eksterne kilder. Hvad er fordelene ved hver datakilde?

Interne datakilder, hvor data oprettes af virksomheden, er typisk lettere at indsamle og kan være mere relevante for virksomhedens egne formål og indsigt. 1 De er gratis for virksomheden. kan starte en række big data-initiativer uden nogensinde at kigge ud over deres egne mure

Eksterne data er de, der er erhvervet uden for virksomheden.

- Tredjepartsdata - kan være specifikke for dit marked eller kunder, men vanskelige at opnå og / eller dyre.
- Åbne data - dvs. statslige statistikker, data relateret til social påvirkning. Kan være af stor værdi, men kræver ofte flere investeringer i analyse for at give indsigt af relevans.
- Kollektivt genererede data. Mange flere virksomheder deltager i møder og hackathons.

Casestudie - er eksterne data omkostningerne værd?

- 7 små virksomheder i Nordirland, der sælger forarbejdede fødevarerprodukter, fik adgang til data fra en stor supermarkedskæde om områder som f.eks. Forbrugernes livsfase og livsstil, analyse af markedskurven og de bedst mulige butikker for de små virksomheders produkter.
- Ejere og ledere deltog i workshops for at lære at hente og analysere de mest relevante data fra databasen for at besvare spørgsmål som "Hvordan klarer min kategori?" "Hvad er den mest populære smag af brød?" "Hvilken type forbruger køber et produkt, der ligner mit? En yoghurtproducent lærte, at ældre voksne var et nøglemarked, så da de besøgte supermarkeder til smagsprøver, prøvede de ikke længere at lokke yngre kunder og fokuserede i stedet på ældre mennesker. Taktikken forbedrede begivenhedernes produktivitet.
- Andre virksomheder var i stand til at forestille sig langtrækkende innovationer i stedet for blot at se på andre produkter på markedet og forsøge at efterligne dem eller følge retningslinjerne fra de store detailkøbere.
- Dataene forstærkede virksomhedernes ibrønde iværksætter karakter. Arbejdspladser blev mere kollegiale: De fleste ejeradministratorer delte kortoplysningerne med deres firmaer og opfordrede medarbejderne til at blive involveret og tilbyde nye ideer. 2

Ejere af små virksomheder står over for hindringer for at investere i eksterne data. Barrierer inkluderer omkostninger, manglende ekspertise eller tro på deres egen markedskendskab.

Når de først fik adgang til dataene, var de fleste af virksomhederne hurtige til at formalisere deres tilgang til markedsføringsplanlægning.

I dette eksempel kunne de små virksomheder få adgang til de eksterne data gratis. Hvor meget kunne de have betalt for at det stadig var umagen værd?

¹ <http://www.aunalytics.com/understanding-analytics-part-1-top-internal-sources-of-big-data/>

² <https://www.conference-board.org/blog/post.cfm?post=2526>



DATAINDSAMLING: TYPER DATA

Data præsenteres som et DATA SET. EN **datasæt** er en gruppering af oplysninger, der er relateret til hinanden.

- Datasæt kan være kvantitative eller kvalitative. Forskellen virker skarp, men der kan være mange måder at få dem til at interagere. Kvalitative data kan være vanskelige at måle og analysere nøjagtigt. Dataene kan være i form af beskrivende ord, der kan undersøges for mønstre eller betydning, undertiden ved hjælp af kodning. Kodning giver forskeren mulighed for at kategorisere kvalitative data til at identificere temaer, der svarer til forskningsspørgsmålene og til at udføre kvantitativ analyse. Data kan være diskrete eller kontinuerlige.
- Data kan være:
 - Nominel: Nominel data er kvalitativ; der er ingen iboende skala eller værdi knyttet til dataene.
 - Ordinal: Data, der har en naturlig orden, som en Likert-skala, hvor 1 betyder "hader det!" og 5 betyder "elsker det!" Husk bare, at forskellen mellem ordinaler (f.eks. Mellem 1 og 2 eller mellem 3 og 4) ikke nødvendigvis er ens. Ofte er afstanden mellem ekstrem lignende og modvilje større end mellem at føle sig neutral eller lidt positiv eller negativ.
 - Interval: Data, der har en naturlig rækkefølge, og afstanden mellem hver værdi er lig, ligesom temperatur.
 - Forhold: Data, der har en naturlig rækkefølge, lige afstand mellem værdier og et naturligt nulpunkt. 3

3. DATA OPBEVARING

Det er vigtigt at tænke fremad, når du planlægger din dataindsamling. Tænk over, hvordan du vil bruge de data, du indsamler, og hvilke beslutninger du vil tage ud fra dem. Ønsker du at få oplysninger om en realtids situation? Vil du forstå kausalitet? Vil du sammenligne to variabler og vælge den bedste?

Opbevaring på stedet v fjernlager

- Traditionelt blev data gemt på stedet, men stigende datamængde og faldende omkostninger for udbydere af datalagrings-tjeneste gør fjernlager i skyen mere attraktivt.
- Det skybaserede lagermarked er domineret af Amazon Web Services, Google og Microsoft Azure.
- Alle lagringsskyer oprettes ikke ens. Nogle skyer er optimeret til at håndtere arkivering, andre har ydeevne og stabilitet til at fungere som bagenden for et primært datalagrings-system, og stadig andre er ikke risikoen værd til ethvert formål. Den laveste prissky kan ende med at blive meget dyrere i det lange løb, hvis data går tabt eller utilgængelige
- Mange løsninger giver fleksibiliteten til at vælge, hvor data lagres: på stedet og / eller i skyen.

4

Hvordan vælger jeg, hvordan dine data skal gemmes?

Når det kommer til lagring af data, er der ingen 'one-size-fits-all' løsning.

Virksomheder skal forstå mængden og typen af data, de har sammen med motivationen bag lagring af oplysningerne.

HVEM har brug for adgang til hvilken type information? Hvor ofte og hvor hurtigt har vi brug for at få adgang til dataene?

- Når du opretter processer, skal du identificere organisationens vigtigste data og prioritere lagerstyringsressourcer korrekt. e-mail kan være en virksomheds højeste prioritet, men lagring og arkivering af e-mail-data for en bestemt gruppe, siger ledere, kan være mere kritisk end andre grupper
- "Spar penge ved kun at bruge din hurtigste lagerplads, f.eks. SSD, til data, som du aktivt bruger, og brug billigere platforme som skyen til at gemme dine arkiverings- eller backupdata,"

HVILKEN type data skal vi gemme?

- Måske skal vi kombinere forskellige datasæt i fremtiden?

³ <https://communitymedicine4asses.com/2013/01/13/scales-of-measurement-nominal-ordinal-interval-ratio/>

⁴ <https://techcloudlink.com/14-things-you-need-to-know-about-data-storage-management/>



HVOR MYE data har vi, og hvor meget vil de vokse i fremtiden?

- Gem ikke overflødige data. Mange virksomheder indfanger og gemmer overflødige data. Dette plejede at være et mindre problem, når disken var dyr, men da kapaciteten var mere rimeligt prissat, er tendensen at gemme alt.

HVOR LANGT har vi brug for at gemme dataene?

- En datalagringspolitik er en nødvendighed for både intern datastyring og juridisk overholdelse. Nogle af dine data skal opbevares i mange år, mens andre data muligvis kun er nødvendige i flere dage. "

HVOR SIKKER skal dataene være? Hvilke lovgivningsmæssige krav skal overholdes?

- Har en plan for genopretning af katastrofer. Uanset hvilken metode eller hvilke sikkerhedskopier du bruger, skal du teste dem og være sikker på at du kan gendanne dine data.

3 DATAANALYSE

For at besvare dine strategiske vejledende spørgsmål og få en dybere forståelse af din virksomhed skal du tage dine rådata og omdanne dem ved hjælp af en kombination af grundlæggende aritmetiske og beskrivende statistikker for at afsløre mønstre - både dem, du forventer at se, og dem, der kan være overraskende. 5

Hvad man skal se efter: fakta, statistik og tendenser er alle nyttige, men af forskellige grunde.

Bemærk: Når vi taler om datavisualisering, kan det ses som en del af den analytiske proces og separat som et middel til at præsentere resultater. De er iboende forbundet, og den ene kan normalt støtte den anden.

Eksempler på salgs- og rentabilitetsdiagram. Dette er ofte et typisk resultat af dataanalyse. (Kilde: Forbes) Nedenstående salgs- og rentabilitetsdashboard af Ann Jackson er et visuelt overbevisende, rent designet resume af dataene, der viser ændringer over tid, geografiske forskelle, tab for visse produktkategorier og opsummerer nøglepræstationsindikatorer som tal. Ann afleverer dette instrumentbræt til sine interessenter og kalder det gjort? Nej, for nu starter de virkelige diskussioner. Ann kan sidde med sit publikum og undersøge yderligere detaljer for at undersøge, hvorfor visse resultater er opstået, og identificere muligheder for at forbedre forretningsresultaterne. Yderligere undersøgelse af dataene, udforskning med dit publikum, det er den værdi, du tilføjer som analytiker ud over at producere dashboards. De giver dig et fremragende grundlag for disse diskussioner, men bør ikke være slutpunktet. 6

4. DATA TOLKNING OG BRUG

At have data er ikke nok. Det er ikke nok at analysere data. Det er vigtigt at oversætte indsigt fra data til handlinger, der understøtter forretningsvækst.

Som med enhver god strategi er det meget vigtigt at vide, hvornår og hvordan man anvender den, hvornår man skal ændre den, og hvordan man måler dens succes, for at organisationer går fra at være dem, der udfører data, til at være virkelig datadrevne.

Cyklussen med brug af data er i gang - det stopper ikke bare, når indsigten præsenteres, men kræver dybdegående diskussion. Analysering af data er et udgangspunkt, ikke et slutpunkt.

I hvilket diagram er der en større forskel?

Databrug stammer fra videnskab, men det er også en kunst. Data er objektive, men de kan præsenteres subjektivt. I betragtning af at datavisualisering er så vigtig for at give mening om store data og for at køre beslutninger, er det vigtigt at forstå, hvordan man gør det.

Slide 1: I det første diagram ser det ud til at der er stor forskel.

Slide 2: I dette diagram ser vi, at begge er en forskel mellem 35,5% og 41,5% - den samme forskel. Dette

⁵ <https://www.schusterman.org/playbooks/data/making-meaning/find-patterns-stats/>

⁶ <https://www.forbes.com/sites/evamurray/2019/03/29/what-is-the-difference-between-data-analysis-anddata-visualization/>



viser det

diagrammer kan være vildledende

det er vigtigt at kommunikere data i formater, der er nøjagtige og i rimelig skala. Når oplysninger tilføjes i skala, bliver det meget klarere.

Hvilken måde at vise data på er mest effektiv?

Du er nødt til at tænke over, hvordan du kommunikerer data til dine formål, OG hvordan du kommunikerer det tydeligt, så din besked bliver forstået.

I dette eksempel er det farvede diagram nemmest at forstå. Hvad ændrede sig?

Det originale blå diagram blev "decluttered" - aksens og etiketterne blev fjernet og det tomme rum reduceret. Derefter blev "fremhævede" fremtrædende detaljer - farverne blev tilføjet for at afspejle et kontinuum af svar, og tallene blev tilføjet for at vise relative værdier.

Konklusion:

Hver gang du bruger data til at formidle noget, fortæller du i det væsentlige en historie. Det kan være i en kvartalspræsentation for dit bestyrelse i en formel programevalueringsrapport i en årlig rapport til vælgere.

Historiefortælling med data er stærk, og det har et etisk imperativ. Data kan let manipuleres til at fortælle en historie, der ikke er der, eller for at minimere en historie, der er.⁷ Det er vigtigt at starte med dataene og lade dem fortælle historien i stedet for at lave din historie på forhånd og kun vælge de data, der understøtter den. For eksempel, hvis du præsenterer et resumé af en programevaluering, der kun fremhæver de succeser og aspekter af din service, som folk kunne lide, og ikke inkluderer de komponenter, der kom kort og afslørede forbedringsområder, fortæller du ikke hele historien.

⁷ <http://www.bdbanalytics.ir/media/1123/storytelling-with-data-cole-nussbaumer-knaflic.pdf>