



# Introduktion til PowerBI

## Introduktion

PowerBI er en Microsoft-plattform rettet mod business intelligence. Det er en kompetence vi ser efterspurgt blandt mange virksomheder. PowerBI bruges typisk til at formidle data og analyser, fx til ledere i en virksomhed, så man kan tage datadrevne beslutninger. Formålet med ugens opgave er derfor tredelt:

- Opbyg erfaring med de basale elementer i PowerBI
- Eksperimentér med forskellige typer af plots og grafer, og tænk over hvordan man bruger dem til at kommunikere teknisk information til modtagere med forskellig baggrund.
- Få praktisk erfaring med at kommunikere jeres analyser ud fra de visualiseringer I har valgt.

For at øve det sidste punkt, afsluttes ugen med en præsentation af jeres dashboard.

Der kommer en uge med mere avanceret PowerBI senere, hvor man kan dykke ned i den mere data-tekniske side af det, og eksperimentere med mere avancerede analyser. Så brug denne uge på at eksperimentere med forskellige typer af visualiseringer og fokusér på kommunikationsdelen.

Hvis ikke PowerBI allerede er installeret på din PC, kan du finde det her:

<https://www.microsoft.com/en-us/download/details.aspx?id=58494>

Se også teoridokumentet [PowerBI-intro](#), som er en kort introduktion til PowerBI. Deri er også links til dokumentation og andre steder I kan finde hjælp.

---

## Afleveringsformat

Afleveringen til denne uges opgave består af en individuel præsentation på fredag:

- Varighed: Omkring 10 minutter
- Jeres PowerBI-dashboard er det visuelle element, så I skal ikke lave en separat powerpoint.
- Hold jer til at præsentere 2-5 sider af jeres PowerBI dashboard.
- Det vi lægger vægt på, er de tanker I har om jeres valg af visualiseringer, og hvordan I kommunikerer jeres analyser. De specifikke resultater I er nået frem til, er mindre vigtige.



- **Husk at tænke over, at dem I præsenterer for ikke kender til jeres analyser, så vær OBS på at definere enheder og kommunikere hvad X og Y-akserne repræsenterer.**
- 

## Opgavebeskrivelse

Opgaven er at analysere et datasæt for at finde sammenhænge, der potentielt kan bruges til at tage datadrevne beslutninger i en virksomhed. Sammenhængene skal præsenteres vha. det PowerBI-dashboard I laver. Opgaven er stillet ret åbent, men hvis I har behov for det er der forslag til ting man kan kigge på i afsnittet *opgavevejledning* længere nede.

Start med at skabe jer et overblik over jeres datasæt.

- Hvilke sammenhænge vil give mening at kigge på?
- Hvilke udregninger er nødvendige for at undersøge de sammenhænge?
- Hvordan kommunikerer man potentielle fund, så personer der ikke har forudgående indsigt i datasættet, også kan forstå det?

Vi anbefaler at I tænker lidt over det selv og evt. diskuterer det med hinanden, før I kigger i opgavevejledningen. Skriv evt. lidt ned omkring jeres strategi. Overvej om I kan finde nogle sammenhænge man kan lave konkrete handlinger ud fra, for at optimere fx profit. Hvis I har behov for sparring, er I også velkomne til at henvende jer til en teamcoach.

---

## Datasæt

Der er lidt forskellige datasæt at vælge imellem. Evt. kan man starte med det lidt simple "video\_game\_sales\_data.csv", for at komme i gang. Så kan man gå videre med det lidt mere komplekse "super\_store\_sales\_data.csv", hvis man vil have mulighed for at lave mere avancerede analyser og visualiseringer. Man kan også arbejde med "DKHousingPricesSample100k.csv", hvis man hellere vil det. Det gør det måske nemmere at komme igang, hvis man allerede kender datasættet.

Det er helt frit hvilket af datasættene man vælger. Hvis man vil holde sig til det simple for at fokusere mere på PowerBI gør man det, hvis man vil gå direkte til det lidt mere komplekse, gør man det.

Nedenfor er der en beskrivelse af de to nye datasæt:



## Videogame sales

Datasættet er en enkelt .csv-fil med computerspil rangeret efter deres globale antal solgte enheder.

- Rank
- Name
- Platform
- Year
- Genre
- Publisher
- NA\_Sales - mio. enheder solgt i Nordamerika
- EU\_Sales - mio. enheder solgt i Europa
- JP\_Sales - mio. enheder solgt i Japan
- Other sales
- Global\_Sales - mio. enheder solgt globalt (brugt til rangering)

## Super store sales

Datasættet er en .XLSX-fil med 3 ark.

Orders:

- Order ID - A unique identifier for each order.
- Order Date - The date of the order placement.
- Ship Date - The date the order was shipped.
- Ship Mode - The shipping mode for the order (e.g. standard, same-day).
- Customer ID - A unique identifier for each customer.
- Customer name - Name of the customer
- Segment - The customer segment (e.g. Consumer, Corporate, Home Office).
- Postal Code
- City
- State - Region or state within a country



- Country
- Region - The global region where the customer is located (e.g. West, Central, East).
- Market - Another global region designation
- Product-ID
- Category - The category of the product purchased (e.g. Furniture, Technology, Office Supplies).
- Sub-Category - The sub-category of the product purchased (e.g. Chairs, Desktops, Paper).
- Product Name - The name of the product purchased.
- Sales - The sales revenue for the products purchased.
- Quantity - The number of units of the product purchased.
- Discount - The discount applied to the product purchased.
- Profit - The profit generated by the product purchased.
- Shipping cost
- Order priority

#### Returns:

- Returned - was the order returned
- Order ID
- Region

#### People:

- Person - name of the sales person responsible for a region
- Region

---

### Tips til fremgangsmåde

Her er nogle forslag til analyser I kan starte med. Det er ikke ment som en checkliste, hvor I skal udføre alle sammen. Brug dem som udgangspunkt, hvis I har brug for det, mens I lærer PowerBI at kende (Se også afsnittet i primeren om filters). De er løst arrangeret efter kompleksitet. I kan også vælge at udføre de samme analyser I lavede i pandas, hvis det er et nemmere udgangspunkt.



- Start med at lave et "sanity-check" af datasættet - er tallene korrekte? Er der fejl som skal renses. **Skriv noter sideløbende med at du laver opgaven, så du kan dokumentere din vej til resultatet.**
- Lav et "Column Chart" over summen af salg, grupperet efter en kolonne, fx:
  - Video Game Sales:
    - global sales per år.
  - Super Store Sales:
    - sales per region.
- Find en måde at vise salg grupperet efter to forskellige kolonner, fx:
  - Video Game Sales:
    - global sales af forskellige genrer, per år.
  - Super Store Sales:
    - sales af forskellige produkt-kategorier, per region.
- Eksperimentér med mere avancerede hierarkier af inddelinger, men tænk på hvordan og hvad I vil kommunikere. Super Store Sales-datasættet er nok bedst egnet her.
- Specifikt til Video Game Sales:
  - Nogle spil er udgivet til flere platforme, og kan derfor have flere "entries". Hvordan ser ranglisten ud, hvis du summerer salget af hver titel på tværs af de platforme, hvor de er udgivet?
- Specifikt til Super Store Sales:
  - Kan I komme med forslag til optimeringer man kunne lave for at øge indtjeningen? Er der fx bestemte produkter/lande der har bedre indtjening end andre, eller hvor der tabes penge?
  - Hvis man gerne vil identificere outliers (fx produkter med høj/lav indtjening), hvordan visualiserer man outliers, så de bliver tydelige?
  - Kan indtjeningen i fx et land forklares ud fra hvilke produkter/kategorier de sælger?

---

## Ressourcer

- Se dokumentet [PowerBI-Primer](#) & [visualiseringstyper](#)