

8. Sortering og filtrering

Sortering og filtrering

I daglig tale anvendes data, information og viden ofte i flæng uden nærmere angivelse af hvad der forstås ved de forskellige udtryk. Den sammenhæng udtrykkene anvendes i, vil som regel give den fornødne forståelse og være tilstrækkelig i den daglige kommunikation. I en mere professionel sammenhæng vil det dog være nødvendig med en mere præcis afgrænsning og forståelse af begreberne.

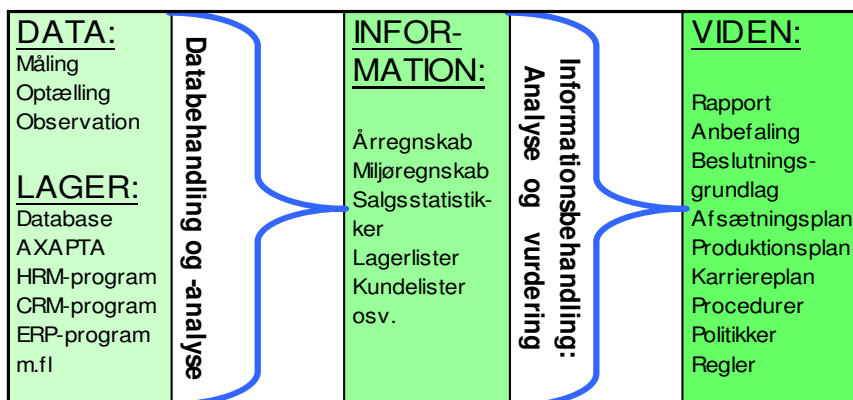
Basis er **data**. Data er måling, optælling eller observationer af forhold og begivenheder i den virkelige verden og er lagret fysisk. Bogstaverne i en bog er data, krydsene i fraværsprotokollen er data, en node er data og er fysisk lagret på papir ved hjælp af tryksværte eller blæk. I en computer gemmes data på harddiske, disketter eller chips, men her foretages den fysiske lagring i form af 0- og 1-taller ved hjælp af magnetisme eller strøm.

Næste skridt i udviklingen er **information**, som er sammenhænge mellem data eller mønstre i data. Ved at sammensætte bestemte bogstaver i bestemt mønster fås ord, som igen kan kobles sammen til sætninger. 108 km er data, som kombineret med "fra København til Korsør" bliver til information, om afstanden mellem de 2 byer. Her ses forskellen mellem data og information. Data er i sig selv meningsløs medens information giver mening, har en udsagnskraft. Information handler om noget og er opstået i en bestemt situation, hvor der var behov for denne information.

Viden fås når information sættes i system, sættes ind i en helhed. Viden er netop information, som har relevans i en bestemt sammenhæng. Kendetegnende for viden er at den muliggør handling og at den gør en forskel i valg af handling. Relevant viden giver bedre og mere hensigtsmæssige beslutninger end uden denne viden. Skal man fra København til Korsør, vil information om afstanden give en viden, som giver et bedre beslutningsgrundlag for valg af transportmiddel end uden denne viden. Kombineres informationen om afstanden mellem de 2 byer med informationen i DSB's køreplan om billetpriser og tidspunkter, vil det give forøget viden til en rejsende i København som skal til Korsør.

Det 4. trin - **kompetence** - er opnåelse af personlig erfaring i form af vellykkede udførelser.

Vejen fra uvidenhed til viden kan således opdeles i stadier eller trin. Afhængig af sammenhænge er der mange forskellige betegnelser for overgangene, men mere generelt benævnes overgangen fra 1. til 2. stadium for databehandling og fra 2. til 3. for informationsbehandling. Sammenhængen er vist i skematisk form i figuren herunder.



Bogføring er en lagring af transaktionsdata – f.eks. i et ERP-(økonomistyrings)program som AXAPTA. Mindst én gang om året skal disse data behandles: kontienes saldi skal beregnes, saldiene fra forskellige konti skal lægges sammen, og organiseres de sammenlagte tal i et bestemt mønster fås et årsregnskab. Gøres denne databehandling manuelt anvendes som regel et afslutningsskema, men de moderne datalagringsprogrammer er programmeret til at udføre mange standardiserede databehandlingsopgaver automatisk og dermed producere den ønskede information.

Meget ofte vil der dog ikke findes standardprogrammer til udtrækning af den ønskede information, men her kommer regnearket på banen som et meget suverænt databehandlingsmedium.

8. Sortering og filtrering

Arbejdsprocessen er sædvanligvis fra venstre mod højre medens søgeprocessen går den modsatte vej - et vidensbehov søges dækket vha. data. Indkøbschefen i BioGRAIN (kap. 7) havde et ønske om at spare omkostninger og med data fra bogholderiet og indkøbsafdelingen kunne han beregne og belyse omkostningsforløbet ved forskellige antal indkøb og herudfra danne sig et beslutningsgrundlag for fremtidige indkøb afhængig af forbrug, priser og omkostninger.

Dataanalyse kan på et mere overordnet niveau deles op i 3 forskellige former: kalkulatorisk - beregning af gennemsnit, spredning o. lign., grafisk - afbildning af søjlediagrammer, lagkagediagrammer mv. og strukturelt - visning af sammenhørende og -hængende data. Medens de 2 første er gamle kendinge, er den undersøgende datastrukturanalyse stort set et resultat af fremkomsten af regneark.

Næppe ret mange tænker over at den systematiske dataanalyse er det egentlige grundlag for verdens største EDB-firma, nemlig IBM. Et af IBM's moderselskaber udviklede i 1890'erne hulkortet og var derfor i stand til udføre den - efter datidens forhold - enorme dataanalyse af de amerikanske befolkningstællinger, der udføres hver 10. år. Befolkningstællingen i 1890 gav så mange data, at det skønnedes at den traditionelle manuelle databehandling ville tage ca. 12 år, men ved hjælp af hulkort og sorteringer af dem lykkedes det at halvere databehandlingstiden. Alle data lagredes nemlig som huller i et standardkort og ved hjælp af maskiner var det muligt at sortere kortene og dermed dataene efter de huller de indeholdt på en langt hurtigere og mere sikker måde end den manuelle databehandling. Helt indtil fremkomsten af de store EDB-anlæg i 1950'erne anvendtes denne stive og langsommelige mekaniske databehandling og i mange tilfælde anvendtes bogføringssystemer med hulkort helt op i 1970'erne. Først med fremkomsten af PC'en og regnearket i begyndelsen af 1980'erne fik man et så billigt, smidigt og fleksibelt værktøj, at sortering kan anvendes analytisk og af alle.

I dette og efterfølgende kapitel behandles Excels 3 indbyggede værktøjer til **interaktiv dataanalyse**, nemlig sortering, filter og i kapitel 9 pivottabeller, der hver især kan frembringe informationer om den virkelighed, som dataene beskriver, men som naturligvis også kan anvendes sammen med de øvrige dataanalyseformer. Dataformatet er i al væsentlighed det samme for alle 3 teknikker: en afgrænset datamængde uden tomme kolonner og rækker - dvs. de afgrænsende rækker og kolonner skal være tomme. Pivottabellen kræver desuden at kolonneoverskrifterne skal være entydige - dvs. de skal alle være forskellige.

Selvom de 3 teknikker har mange resultatmæssige lighedspunkter, er der dog en meget stor forskel på den datamanipulation som de udfører. Ved *sortering* flyttes dataene fysisk rundt på regnearket medens *filtrering* blot skjuler de uinteressante data - dvs. dataenes fysiske placering i regnearket forandres ikke. Ved pivottabeller anvendes en helt tredje metode idet dataene kopieres og organiseres i usynlige og utilgængelige hierarkisk ordnede tabeller i PC'ens RAM-hukommelse og som udgør grundlaget for den synlige pivottabel. Konsekvensen heraf er dels at kildedataene kan slettes uden at pivottabellen bryder sammen og dels at regnearkets størrelse forøges betydeligt hvis dataene gemmes sammen med regnearket.

Den mest enkle form for sortering består i at ordne en liste alfabetisk, efter en værdi eller en dato således at man hurtigt kan finde et enkelt element i listen. Ofte vil der dog være tale om sortering efter flere forskellige kriterier. Det er sådan telefonbogen er sorteret - efter abonnentens efternavn, fornavn, adresse og i så fald er der tale om en betinget eller afhængig sortering - dvs. alle sorteringer efter efternavnet er afhængig af alle forudgående sorteringer. Ved analytisk anvendelse af sortering forstås en anvendelse til løsning af en opgave. Det kunne f.eks. være i forbindelse med udvælgelse af nye biler til virksomhedens sælgere. Her er der en lang række forhold (kriterier) der skal tages i betragtning, som anskaffelsesprisen, brændstoføkonomi, grøn afgift, forsikring, værditab, reparations- og vedligeholdelsesudgifter, virksomhedens image (ikke for stor, ikke for lille), sælgernes ønsker mv. Det første skridt i analysen er at finde værdierne (dataene) for disse forhold for en række relevante forskellige bilmærker, men da tallene ikke kan lægges sammen til en total, må man prioritere de forskellige forhold og så vælge de bilmærker, der er bedst på de forskellige kriterier og her kommer sortering og filtrering ind som de relevante analyseredskaber.

8. Sortering og filtrering

📖 Autosortering - et kriterium

Den mest simple sortering vil her blive belyst ved hjælp af en A-B-C-analyse. Ideen i en A-B-C-analyse er at finde de bedste kunder, de bedste produkter, de største lagerbeholdninger osv., idet det kan give virksomheden god information om hvilke emner den primært bør koncentrere sig om. De opdeles ofte i 3 grupper: A, B og C hvor A er den vigtigste gruppe, B den næstvigtigste og C den tredjevigtigste.

Analysen hviler på den såkaldte 80/20-regel, der f.eks. kan formuleres sådan:

- 80 % af fortjenesten stammer fra 20 % af kunderne
- 80 % af leverancerne kommer fra 20% af leverandørerne
- 80 % af omsætninger stammer fra 20 % af produkterne, osv.

Virksomheden bør naturligvis prioritere behandlingen og opmærksomheden om de ca. 20 % af kunderne, leverandørerne, produkterne, der bidrager mest. Med en A-B-C-analyse opdeles kunderne, leverandørerne, produkterne i 3 grupper, hvor A-gruppen som regel omfatter de øverste 20 %, B-gruppen som regel de næste 30 % og C-gruppen de sidste 50 %.

Firmaet **Beslagdepotet** har udarbejdet nedenstående lagerliste over hvilke varenumre, der ligger på afdelingen i Esbjerg. Bemærk her er kun vist de 3 første og de 3 sidste varenumre på listen. Med udgangspunkt i en A-B-C-analyse ønskes en vurdering af de mest kapitalkrævende (værditunge) og dermed de mest belastende varenumre. Første skridt i analysen er en beregning af værdien i kolonne E med formelen som vist i formelfeltet. Med cursoren placeret på en værdi i kolonne E klikkes dernæst på Sorter faldende ikonet og resultatet ses nedenunder. Bemærk at alle sammenhængende kolonner / rækker helt automatisk følger med i sorteringen - dvs. det er rækkerne der sorteres efter indholdet i kolonne E. Det ses at katalog nr. 2812 er den mest værdifulde på lageret, medens katalog nr. 2030 har den laveste værdi.

E2		=		=C2*D2				
	A	B	C	D	E	F	G	H
1	Nr	Kat.nr.	Stk.	Kr.	Værdi	Akkum. værdi %	Akkum kat.nr. %	Akkum kat.nr. %
2	1	1003	1.146	6,00	6.876			
3	2	1017	24.725	4,50	111.263	Sorter faldende		
4	3	1018	24.195	5,00	120.975			
24	23	3802	850	25,00	21.250			
25	24	3807	416	12,00	4.992			
26	25	3812	612	15,00	9.180			
27	I alt							

	A	B	C	D	E
1	Nr	Kat.nr.	Stk.	Kr.	Værdi
2	12	2812	6.027	28,00	168.756
3	11	2805	5.702	25,00	142.550
4	13	3001	4.085	32,00	130.720
24	18	3504	2.270	2,00	4.540
25	19	3604	908	5,00	4.540
26	8	2030	1.833	1,00	1.833
27	I alt	25			1.096.741

Næste skridt i analysen er optælling af antal varenumre i B27 - gøres med TÆL-funktionen og en opsummering af lagerværdien i E27. Beregning af den akkumulerede værdi i procent af værdien i alt gøres med en akkumulerende sumformel (se nedenfor), der opsummerer værdien fra E2 og nedefter ved at dollar-fiksere den første celle. På samme måde opsummeres antal varenumre, men her anvendes TÆL-funktionen. Det ses at katalognr. 2812 udgør 15,39% af den samlede værdi, men kun 4% af det samlede antal katalognumre. Med en betinget formatering - mindre end eller lig 20 (%) sættes til rød og værdier mellem 20 og 50 sættes til blå - er det nemt at afgrænse de 3 kategorier A, B og C.

I efterfølgende diagram er vist den akkumulerede værdi i % ud af y-aksen og akkumuleret antal katalognr. i % ud af x-aksen - den mørkeblå kurvede linie - og sammenholdt med en fuldstændig ligelig fordeling af produkternes værdi - dvs. hvis hvert produkts værdi var 4 % af lagerværdien ligesom hvert produkt udgør 4% af det samlede antal produkter. Bemærk for at kurverne starter i 0-punktet må disse værdier indlægges manuelt i regnearket.

8. Sortering og filtrering

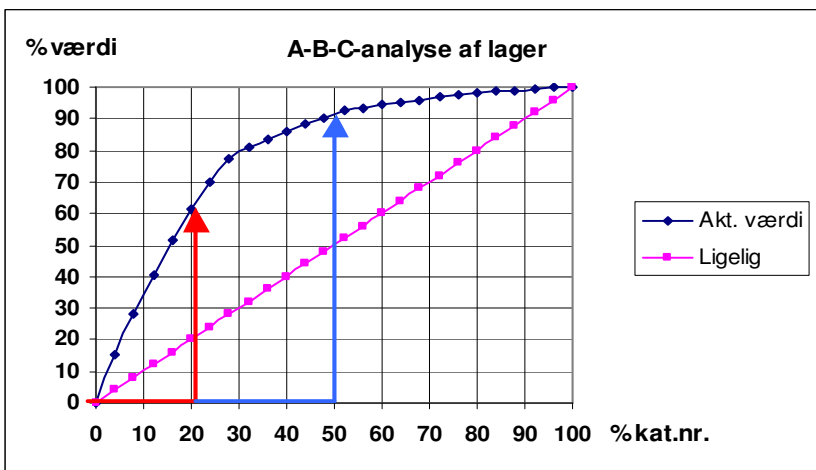
	E	F	G
1	Værdi	Akkum. værdi %	Akkum kat.nr. %
2	=C2*D2	=SUM(\$E\$2:E2)*100/\$E\$27	=TÆL(\$B\$2:B2)*100/\$B\$27
3	=C3*D3	=SUM(\$E\$2:E3)*100/\$E\$27	=TÆL(\$B\$2:B3)*100/\$B\$27
4	=C4*D4	=SUM(\$E\$2:E4)*100/\$E\$27	=TÆL(\$B\$2:B4)*100/\$B\$27

Y-værdi i den ligelige fordeling

Bemærk startværdien fastholdes med en absolut adressering

Indsatte værdier for at fuldende grafen

	A	B	C	D	E	F	G	H
1	Nr	Kat.nr.	Stk.	Kr.	Værdi	Akkum. værdi %	Akkum kat.nr. %	Akkum kat.nr. %
2					0	0	0	0
3	12	2812	6.027	28,00	168.756	15,39	4	4
4	11	2805	5.702	25,00	142.550	28,38	8	8
5	13	3001	4.085	32,00	130.720	40,30	12	12
6	3	1018	24.195	5,00	120.975	51,33	16	16
7	2	1017	24.725	4,50	111.263	61,48	20	20
8	9	2417	47.591	2,00	95.182	70,16	24	24
9	22	3701	4.430	18,00	79.740	77,43	28	28
25	18	3504	2.270	2,00	4.540	99,42	92	92
26	19	3604	908	5,00	4.540	99,83	96	96
27	8	2030	1.833	1,00	1.833	100,00	100	100
28	I alt	25			1.096.741			



Det ses at de 20 øverste % af varenumrene dækker 61% af lagerets værdi; (den røde pil) de næste 30% giver en lagerværdi på 30% (91-61) (den blå pil) og de sidste 50% af varenumrene udgør mindre end 10 % af den samlede lagerværdi. Bemærk i eksemplet kan 80% af lagerværdien henføres til 30 % af katalognumrene

Sammensat sortering - flere kriterier

Ved flerkriteriesortering er der 2 muligheder - en afhængig / betinget sortering, hvor de efterfølgende kriteriers sortering er betinget af de forudgående sorteringer eller en uafhængig sortering.

Det næste eksempel er hentet fra månedsmagasinet Penge og Privatøkonomi. Nedenstående regne-ark viser et uddrag af en oversigt over årlige forsikringspræmier (=betaling for forsikringen) for en elitebilist i forskellige forsikringselskaber - ordnet alfabetisk - og for forskellige bilmærker - ordnet efter salgspris. Den samlede oversigt fylder 3 sider, men her er blot vist et lille udsnit.

Sortering af en sådan omfattende tabel er langt mere kompleks end en simpel rangordning efter en enkelt værdi. Dels omfatter oversigten 8 forsikringselskaber = 8 kriterier, dels varierer præmierne fra selskab til selskab og fra mærke til mærke på en ganske uens måde. Sorteringsproceduren må derfor gennemføres mange gange.

Når man skal sortere en større oversigt er det altid vigtigt at kunne komme tilbage til udgangspunktet. Bilmærkerne er her rangeret efter salgspris, men den findes ikke tabellen. Derfor er det en god ide at starte med at nummerere rækkerne, så kan man altid sortere oversigten tilbage til udgangspunktet ved at anvende numrene som sorteringskriterium.

8. Sortering og filtrering

Betinget sortering

Sortering - Data Sorter - Alt+do

Med cursoren placeret på et tal i opstillingen og med Alt+do fås dialogboksen Sorter - se figuren nedenfor. Som det fremgår af dialogboksen er Excels sorteringskapacitet mærkværdigvis og helt unødvendigt begrænset til 3 kriterier - f.eks. kan en sortering af medarbejderne efter anciennitet, fornavn, efternavn, stilling og afdeling derfor ikke gøres med Excels sorteringsværktøj.

Bemærk i øvrigt at hvis der er tydelig forskel mellem kolonneoverskrift og tabelindhold vil Excel gå ud fra at den øverste række er kolonneoverskrift og ikke skal inddrages i sorteringen, men er der eksempelvis årstal i øverste række kan Excel tage fejl, men med klik i radioknappen Kolonne-overskrift udgår øverste række af sorteringen. Ved klik på boksens pile kan vælges hvilke kolonner, der skal indgå i sorteringen og sorteringsordenen - stigende eller faldende - bestemmes ved markering i radioknapperne.

I eksemplet er valgt de 3 første forsikringselskaber som sorteringskriterium. Sorteringen foregår på den måde at først sorteres efter Alm. Brand og denne sortering fastholdes medens der sorteres efter Baltica og for fastholdt Alm. Brand og Baltica sorteres efter Codans præmier. En sådan sortering vil resultere i at Honda Civic vil komme øverst og Mazda 323 vil komme ind på en fjerde plads. Vælges Codan som første kriterium og Alm. brand og Baltica som efterfølgende kriterier vil rækkefølgen derimod blive helt anderledes - der er således tale om en betinget eller afhængig sortering.

	A	B	C	D	E	F	G	H	I	J
1	nr	mærke	Alm. Brand	Baltica	Codan	Dial	GF	Kgl Brand	Tell	Top DK
2	1	Mazda 323	5360	4656	4752	4384	4000	4856	3952	4780
3	2	Nissan Primera	4524	4084	5516	4868	4000	5400	4116	4780
4	3	Honda Civic	4524	4084	4752	4876	3784	5400	4252	4780
5	4	Mitsubishi Lancer	4524	4084	5516					
6	5	VW Golf CL	5360	4656	5516					
7	6	Opel Astra	5360	4656	4752					
8	7	Peugot 405	5360	4656	6044					
9	8	Mazda 626	5360	4656	5516					
10	9	Citroen Xantia	5360	4656	5516					
11	10	Volvo 440	5360	5128	6044					
12	11	Audi A4	5360	5128	5516					

Sorter

Arrangerer oplysningerne i markerede rækker eller lister ordnet alfabetisk eller efter nummer eller dato.

Sorter

Sorter efter
Alm. Brand Stigende Faldende

Og derefter
Baltica Stigende Faldende

Og derefter
Codan Stigende Faldende

Listen har
 Kolonneoverskrift Ingen kolonneoverskrift

Indstillinger... OK Annuller

Ubetinget sortering

Præmierne i de øvrige 5 selskaber vil ikke påvirke sorteringen. En sortering efter alle 8 selskabers præmier må derfor tilrettelægges på en anden måde - nemlig med en sortering af hvert enkelt selskabs præmier og et andet fælles kriterium, der rangordner ens præmier i det enkelte selskab forskelligt og en fastholden af denne sortering ved at rangordne hver sortering for sig. Hvert bilmærkes samlede placering kan derfor beregnes som summen af de 8 placeringer og sorteres den, findes den samlet set billigste bil med hensyn til forsikringspræmie. Som andet kriterium vælges i hvert tilfælde bilens salgspris bestemt af dens nr. i kolonne A.

Nedenfor er vist indstillingen og sorteringen af bilmærkerne i henhold til Codans præmieberegning. Bemærk hvert forsikringselskabs rangordning fastholdes i en selvstændig kolonne ved at nummerere sorteringen. Rangordningen efter det enkelte selskab kan derfor sammenlignes med rangordningen efter de øvrige selskaber og efter de enkelte mærkers salgspris i kolonne A.

8. Sortering og filtrering

	A	B	E	F	G	H
1	nr	mærke	Baltica	R _E	Codan	R _G
2	1	Mazda 323	4656	4	4752	1
3	4	Honda Civic	4084	2	4752	2
4	5	Opel Astra	4656	6	4752	3
5	2	Nissan Primera	4084	1	5516	4
6	3	VW Golf CL	4656	5	5516	5
7	6	Mitsubishi Lancer	4084	3	5516	6
8	8	Mazda 626	4656	8	5516	7
9	9	Citroen Xantia	4656	9	5516	8
10	11	Audi A4	5128	11	5516	9
11	7	Peugot 405	4656	7	6044	10
12	10	Volvo 440	5128	10	6044	11

Sorter [?] [X]

Sorter efter: Codan Stigende Faldende

Og derefter: nr Stigende Faldende

Og derefter: Stigende Faldende

Det bemærkes at der er betydelig forskelle i Balticas og Codans rangordning af de forskellige mærker: Codan betragter Mazda 323, Honda Civic og Opel Astra de billigste at forsikre medens Baltica giver de laveste præmier til Nissan Primera, Honda Civic og Mitsubishi Lancer - se kolonne F og H - og at den også er forskellig fra salgspriserne, jfr kolonne A. Denne sorteringsmåde er i modsætning til tidligere viste flerkriteriesortering helt uafhængig af hvilket kriterium, der vælges først eller sidst.

☞☞☞ Indsættelse af tomme kolonner efter hver udfyldt kolonne gøres sådan: klik på den første kolonnes bogstav (her C), hold Ctrl nede og klik på de øvrige kolonnens bogstaver og med Ctrl++ (plustegnet ved tastaturet og kun 1+) fås 1 tom kolonne efter hver markeret kolonne.

Hvis rangordningstallene skal anvendes til yderligere beregning - f.eks. en vægtning af de forskellige kriterier - kan de med fordel anbringes samlet og umiddelbart til højre for tabellen, idet de skal indgå i alle sorteringer.

Det endelige resultat af sorteringen ses til højre. Efter rangordningen efter alle forsikringselskaber er summeret hvert bilmærkes rang i kolonne S og sorteringen af bilerne efter denne, giver rangordningen vist i kolonne U. I kolonne T er beregnet den gennemsnitlige rang.

Resultatet viser at de dyrere mindre tyske biler af forsikringselskaberne vurderes til at være de billigste at forsikre medens de billigere mindre japaniske biler pålægges højere forsikringspræmier.

	A	B	S	T	U
1	nr	mærke	Tot	Gns.	R _r
2	5	Opel Astra	28	3,5	1
3	3	VW Golf CL	30	3,8	2
4	1	Mazda 323	31	3,9	3
5	2	Nissan Primera	33	4,1	4
6	4	Honda Civic	38	4,8	5
7	6	Mitsubishi Lancer	39	4,9	6
8	7	Peugot 405	45	5,6	7
9	8	Mazda 626	58	7,3	8
10	9	Citroen Xantia	60	7,5	9
11	10	Volvo 440	80	10,0	10
12	11	Audi A4	86	10,8	11

Vandret Sortering

Som nævnt er forsikringselskaberne ordnet alfabetisk, men der er tydeligvis forskelle i præmieni-veauet i de forskellige selskaber. Sorteringen ovenfor ordnede bilmærkerne efter forsikringsudgifterne for en elitebilist. Næste skridt er en ordning af selskaberne efter laveste præmier.

Som sorteringskriterium anvendes den opsummerede præmie for alle 11 bilmærker i række 13. Ved sortering fra venstre mod højre skal man være opmærksom på at de 2 første kolonner ikke skal indgå i sorteringen og derfor skal hele området afmærkes først og med Alt+do fås dialogboksen Sorter. Her klikkes på Indstillinger nederst hvilket giver dialogboksen Sorteringsindstillinger. Her markeres radioknappen Sorter fra venstre mod højre, OK, og i feltet Sorter efter vælges række 13. Den færdige sortering med en betinget formatering - <=4300 rød typografi og mellem 4300 og 5000 blå typografi - ses herunder. En sådan sortering giver et langt bedre beslutningsgrundlag for valg af såvel forsikrings-selskab som bilmærke end den oprindelige tabel.

8. Sortering og filtrering

	A	B	C	D
1	nr	mærke	Alm. Brand	Baltica
2	5	Opel Astra	5360	4656
3	3	VW Golf CL	5360	4656
4	1	Mazda 323	5360	4656
5	2	Nissan Primera	4524	4084
6	4	Honda Civic	4524	4084
7	6	Mitsubishi Lancer	4524	4084
8	7	Peugot 405	5360	4656
9	8	Mazda 626	5360	4656
10	9	Citroen Xantia	5360	4656
11	10	Volvo 440	5360	5128
12	11	Audi A4	5360	5128
13			56452	50444
14				
15				
16				
17				
18				

Sorter [?] [X]

Sorter efter _____

Række 13 [v] Stigende Faldende

Sorteringsindstillinger [?] [X]

Sorteringsrækkefølge for 1. nøgle

Normal [v] [OK]

Forskel på store og små bogstaver [Annuller]

Retning _____

Sorter fra oven og ned Sorter fra venstre mod højre

[Indstillinger...] [OK] [Annuller]

	A	B	C	D	E	F	G	H	I	J
1	nr	mærke	GF	Tell	Baltica	Dial	Top DK	Alm. Brand	Kgl Brand	Codan
2	5	Opel Astra	4000	4120	4656	4000	4424	5360	4432	4752
3	3	VW Golf CL	4000	4252	4656	4000	4434	5360	4432	5516
4	1	Mazda 323	4000	3952	4656	4384	4780	5360	4856	4752
5	2	Nissan Primera	4000	4116	4084	4868	4780	4524	5400	5516
6	4	Honda Civic	3784	4252	4084	4876	4780	4524	5400	4752
7	6	Mitsubishi Lancer	4000	4260	4084	4000	5372	4524	4432	5516
8	7	Peugot 405	4000	4132	4656	4000	4424	5360	4432	6044
9	8	Mazda 626	4000	4268	4656	4000	5372	5360	4432	5516
10	9	Citroen Xantia	4212	4272	4656	4000	4424	5360	4432	5516
11	10	Volvo 440	4000	4284	5128	4936	5372	5360	5472	6044
12	11	Audi A4	4212	5828	5128	8308	5372	5360	9200	5516
13			44208	47736	50444	51372	53534	56452	56920	59440

Brugerdefineret sortering

Excel har yderligere en sorteringsmulighed, nemlig efter *brugerdefineret liste*. Et supermarked, der ønsker at sortere omsætningen efter ugedage vil få fredag øverst og torsdag nederst hvis der sorteres alfabetisk og ikke i den naturlige rækkefølge mandag til lørdag. I dialogboksen Sorteringsindstillinger kan dog vælges en brugerdefineret sorteringsorden - se standardmulighederne til højre. Kun 1. nøgle kan dog sortere efter en sådan liste.

Sorteringsindstillinger [?] [X]

Sorteringsrækkefølge for 1. nøgle

Normal [v] [OK]

Normal [v] [Annuller]

Ma, Ti, On, To, Fr, Lø, Sø

Mandag, Tirsdag, Onsdag, Torsdag

Jan, Feb, Mar, Apr, Maj, Jun, Jul

Januar, Februar, Marts, April, Maj

Marker den ønskede brugerdefinerede sorteringsrækkefølge.

Brugerdefinerede lister kan anvendes i mange tilfælde: afdelinger, der sorteres efter hvordan de hører sammen - salgsafdelinger, produktionsafdelinger osv., elever i en hhx-klasse sorteret efter hvilket gruppearbejde de deltager i, markeder sorteret efter geografisk områder: Skandinavien, Europa, osv.

Opret lister: Funktioner, Indstillinger, Lister - Alt+kl

I feltet Listeelementer kan en liste indskrives direkte - afslut med enter efter hver listeelement og afslut listen med Tilføj. Hvis listen allerede er indskrevet i et regneark i den ønskede rækkefølge kan en sådan liste importeres: klik på den røde pil i feltet og udpeg listen, enter, afslut med importer.

8. Sortering og filtrering

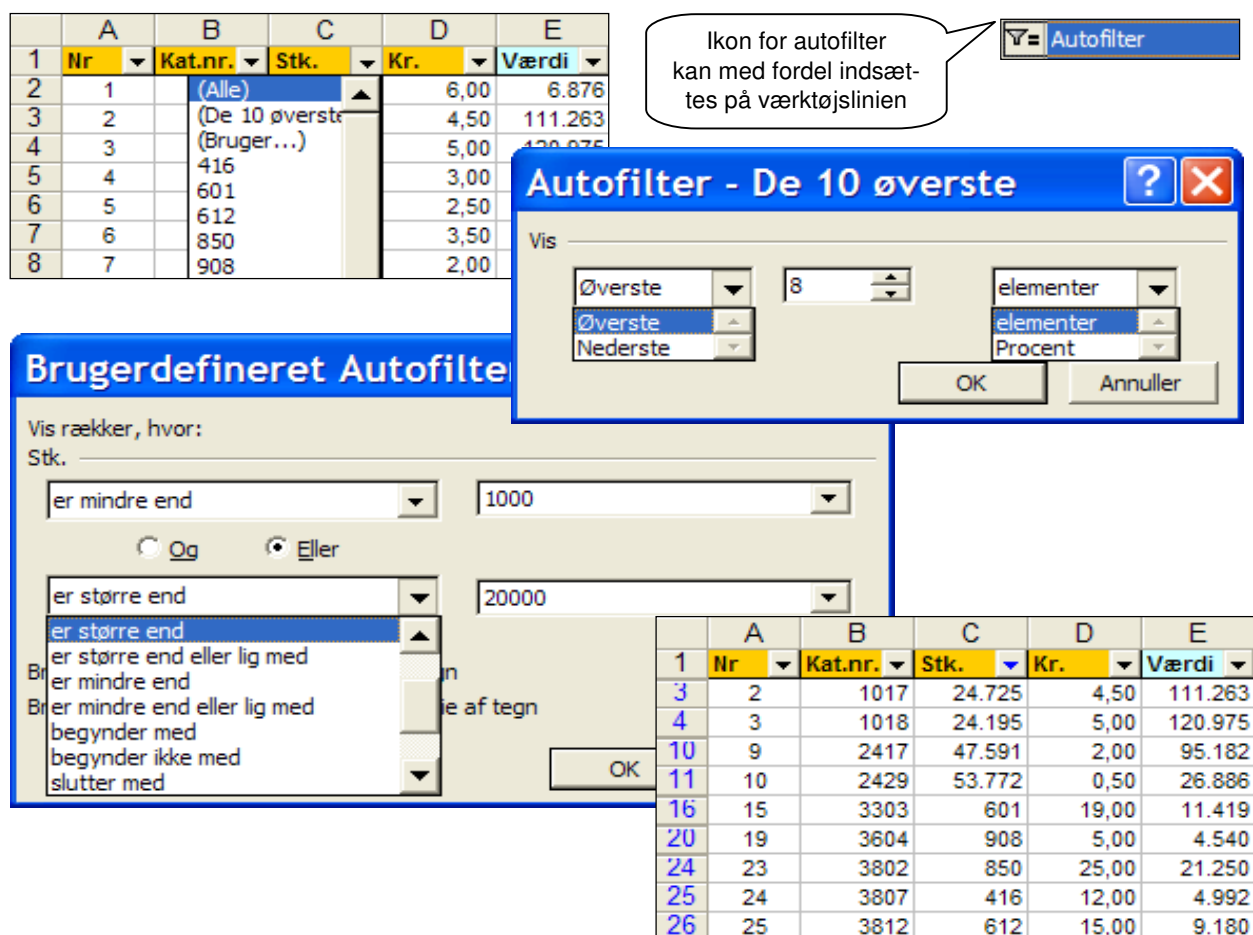
Filter - Data, Filter, Auto - Alt+dfa

Filter er en anden måde til en hurtig og systematisk omorganisering af en større datamængde opstillet i et sammenhængende tabelformat. Teknikken sigter egentlig mod udvælgelsen af rækker efter forskellige kriterier, men den har også betydelige analytiske anvendelsesmuligheder.

Til illustration af filterteknikken anvendes **Beslagdepotets** lagerliste. Med cursoren placeret i listen vælges Alt+dfa. Herved indsættes en lille grå pileboks ved alle kolonneoverskrifter. Ved klik på pilen åbnes en rullemenu med 3 standardpunkter og en størrelsessorteret oversigt over elementer i kolonnen - se efterfølgende figur. Herfra kan så vælges det eller de elementer der ønskes. Punktet (Alle) anvendes til at fjerne en filtrering medens de 2 næste giver muligheder for en meget specifik udvælgelse. Det andet punkt er benævnt (De 10 øverste), men det er kun 1 af mange muligheder - se dialogboksen nedenfor. Som det fremgår af de udslåede rullemenuer kan denne give de øverste eller nederste for et vilkårligt antal - her 8 - og udvælgelsen kan gøres ud fra antal elementer eller en procentdel. Filtreres listen efter Værdi i kolonne E og vælges 'øverste 20 procent' fås A-gruppens elementer i en A-B-C-analyse. De vil dog ikke være sorteret efter størrelse, men det kan gøres efterfølgende med klik på sorteringsknapperne.

Med (Bruger...) fås dialogboksen Brugerdefineret Autofilter, som muliggør en betinget udvælgelse af rækkerne. Der kan filtreres efter 2 kriterier, der knyttes sammen med et *og* eller et *eller*. Der er mulighed for at vælge blandt mange afgrænsningsmuligheder - større end, mindre end, lig med osv. - se rullelisten. I feltet til højre angives afgrænsningsværdien - her kan også angives ord der forekommer i kolonnen evt. med jokertegnene ? og * - se teksten på dialogboksen - eksempelvis vil *mælk finde sødmælk, letmælk, kærnemælk osv. I eksemplet herunder ønskes de mere problematiske lagerbeholdninger - for små (under 1000 stk.) *eller* for store (over 20000 stk.). Resultatet af denne filtrering ses til højre. Bemærk at rulle-ud-pilen bliver blå efter filtreringen - på den måde kan man let finde den eller de overskrifter, der er filteret efter. Farven på række-numrene ændres også.

Filtreringen fjernes ved at klikke på (Alle) og Autofilter ved at vælge kommandoen Alt+dfa igen.



The image shows a sequence of operations in Excel. It starts with a data table with columns A-E: Nr, Kat.nr., Stk., Kr., and Værdi. A callout box points to the filter icon in the toolbar, stating: "Ikon for autofilter kan med fordel indsættes på værktøjslinien".

Next, the 'Autofilter - De 10 øverste' dialog box is shown. It has a 'Vis' section with two dropdowns: 'Øverste' (selected) and 'Nederste'. A numeric input field shows '8'. To the right, another dropdown is set to 'elementer'. 'OK' and 'Annuller' buttons are at the bottom.

Then, the 'Brugerdefineret Autofilter' dialog box is shown. It asks 'Vis rækker, hvor: Stk.' and has two criteria: 'er mindre end 1000' and 'er større end 20000'. The 'Og' radio button is unselected, and the 'Eller' radio button is selected. A list of comparison operators is shown on the left, with 'er større end' selected. An 'OK' button is at the bottom.

Finally, the resulting filtered data table is shown, with rows 3, 4, 10, 11, 16, 20, 24, 25, and 26 highlighted in blue. The 'Nr' column values are 2, 3, 9, 10, 15, 19, 23, 24, and 25 respectively.

	A	B	C	D	E
1	Nr	Kat.nr.	Stk.	Kr.	Værdi
2	1	(Alle)		6,00	6.876
3	2	(De 10 øverste)		4,50	111.263
4	3	(Bruger...)		5,00	120.975
5	4	416		3,00	
6	5	601		2,50	
7	6	612		3,50	
8	7	850		2,00	
9	8	908			

	A	B	C	D	E
1	Nr	Kat.nr.	Stk.	Kr.	Værdi
3	2	1017	24.725	4,50	111.263
4	3	1018	24.195	5,00	120.975
10	9	2417	47.591	2,00	95.182
11	10	2429	53.772	0,50	26.886
16	15	3303	601	19,00	11.419
20	19	3604	908	5,00	4.540
24	23	3802	850	25,00	21.250
25	24	3807	416	12,00	4.992
26	25	3812	612	15,00	9.180

8. Sortering og filtrering

Filtrering og kalkulationer

Som tidligere nævnt er basisfunktionen i filtrering alene at vise eller skjule visse udvalgte rækker. Det betyder bl.a. at sumformler m.fl. også omfatter de skjulte rækker når dataene filtreres. Ofte er det dog kun de synlige tal i en filtrering, der ønskes sammenfattet i en sum, et gennemsnit, et antal eller lignende. Til sådanne beregninger kommer Excel med 2 forskellige subtotaler (egentlig databasefunktioner), men det er også muligt selv at udvikle formler, der kun omfatter de synlige rækker.

SUBTOTAL - Funktionskategori mat & trig

Denne funktion kan anvendes ved regneoperationer i kolonner, der skal filtreres. Den kræver 2 argumenter, hvor det første angiver hvilken regneoperation, der ønskes udført og det andet omfatter adressen på den liste, der skal indgå i beregningen. I alt kan SUBTOTAL udføre de 11 regneoperationer, der er specificeret til højre herfor.

Eksemplet nedenfor er taget fra Penge og Privatøkonomis oversigt over 3 års omkostninger ved en ny bil, som i de 3 år i alt kører 45.000 km. Pris og Værditab (i kolonne I) er angivet i t.kr. Benzin er angivet i liter, Vedligehold omfatter service og 1 sæt nye dæk medens Faste er vægtafgift og billigste forsikring. Værditabsprocenten i kolonne J er beregnet i forhold til bilens pris i kolonne E.

SUBTOTAL's første argument:


- 1 = MIDDEL
- 2 = TÆL (tal)
- 3 = TÆLV (alt indhold)
- 4 = MAKS
- 5 = MIN
- 6 = PRODUKT
- 7 = STDAFV
- 8 = STDAFVP
- 9 = SUM
- 10 = VARIANS
- 11 = VARIANSP

B33	F33	=	=SUBTOTAL(3;B2:B31)	=SUBTOTAL(1;F2:F31)						
	A	B	C	D	E	F	G	H	I	J
1		Nat	Moto	HK	Pris	Benzin	Vedl.	Faste	Vtab	Vtab%
3	Alfa Romeo	Ita	1,4	90	173	3537	7225	18084	68,0	39,3
11	Citroen Advantage	Fr	1,8	103	182	3209	7108	18084	66,2	36,3
17	Alfa Romeo 145	Ita	1,6	103	188	3510	7225	19437	76,5	40,7
22	Volvo 440	Sv	1,8	90	193	3403	7125	18084	63,0	32,6
24	Citroen Xantia	Fr	1,6	89	195	3353	6671	19569	71,1	36,5
27	Citroen Xantia 1,8	Fr	1,8	103	202	3467	6671	19569	72,9	36,1
32										
33	Viste elementer	6,0				3413	7004	18805	69,6	36,9
34	Alle antal / gns.	30,0				3356	10817	18350	66,2	35,2
35	Forskel					-57	3813	-455	-3,4	-1,7

Oversigten, der kun omfatter 30 af de i alt 240 mærker indeholdt i bladet, er her filtreret til at vise de 6 nederste vedligeholdelsesudgifter = 20% billigste mht. service o.lign.

De anvendte formler fremgår af formellinien. I B33 er anvendt en SUBTOTAL med første argument 3 - dvs. den tæller alle viste celler med et indhold og i F33:J33 er anvendt en subtotal med første argument 1, som beregner gennemsnittet af de viste tal. Tallene i række 34 tjener som referencepunkt og er en 'hård kopi' af værdierne i række 33 uden filtrering - altså gennemsnittet af alle 30 biler - og i række 35 er beregnet forskellen mellem referencepunkterne og den aktuelle filtrering.

Figuren viser at de 6 billigste biler mht. vedligeholdelse i gennemsnit er 3813 kr. billigere end gennemsnittet for alle 30 biler. Til gengæld er værditabet 3400 kr. større, faste afgifter 455 kr. større og benzinforbruget 57 l større end gennemsnittet - tendensen synes altså at være, at fordel på et punkt modsvares af ulemper på andre punkter. Det er netop én af filtreringens styrker: *at man systematisk kan udvælge positive egenskaber og eliminere negative egenskaber samtidig med at man kan se konsekvenserne af valg / fravalgene.*

 I en filtreret datamængde kan SUBTOTAL-formelen aktiveres ved at klikke på Autosum. Den vil altid have første parameter 9 (=sum), men den kan ændres på formellinien. Vær dog opmærksom på at Autosum starter sammentællingen fra den første synlige række og ikke fra første række i de ufiltrerede data og slutter sammentællingen ved sidste synlige række.

8. Sortering og filtrering

Data Subtotaler - Alt+du

Under kommandoen Data findes også en Subtotaler, der omfatter de samme regnefunktioner og parametre. Den fungerer dog på en lidt anden måde, idet den gennemfører de ønskede regneoperationer hver gang indholdet i en udvalgt og **sorteret** kolonne ændres - se figuren.

I en liste der omfatter flere års udvikling kan det være ønskeligt med en i alt værdi for hvert år og i så fald kan denne funktion anvendes, idet hver gang årstallet ændres indsættes en ny række med den ønskede sammenfatning (sum, gennemsnit mv.) med teksten 200X Total. Nederst afsluttes med en Hovedtotal omfattende alle subtotaler.



Et datamateriale, der indeholder formler på tværs af rækkerne bryder normalt fuldstændig sammen ved sortering og filtrering fordi formler og adresser ikke følges ad i sorteringen. Med en SUM.HVIS-formel kan en filtrering gøres endnu mere anvendelig.

Økonomichefen i **Kuffertimportøren** følger nøje med i hvor store rabatter sælgerne giver kunderne og han noterer derfor hver måned hver sælgers salg (salg dm) og den afgivne rabat (rabat dm) - se figuren herunder. I kolonne E beregner han rabatprocenten (bemærk - omfatter kun data i 1 række) og i kolonne F beregnes hver sælgers 'år til dato' (sum over flere rækker) - salg med den SUM.HVIS-formel, der er vist i formellinien. Den siger: blandt navnene i kolonne A skal vælges det navn, der står i A11 og at de tal, der skal summeres står i kolonne C. Formlen er kopieret til kolonne G for at beregne rabatter år til dato - bemærk fordelingen af relative og absolutte adresser i formelen.

Anvendelsen af SUM.HVIS-formelen bevirker at fokus kan ændres fra måned til sælger uden at formelen bryder sammen - se efterfølgende figur hvor sælgerne er filtreret og Bent Jensen valgt.

F11		=SUM.HVIS(\$A\$2:\$A11;\$A11;C\$2:C11)						
	A	B	C	D	E	F	G	H
1	Sælger:	Mdr	Salg dm	Rabat dr	Rabat%	Salg åtd.	Rabat åtd.	Rabat%
10	Anders Hansen	3	645.821	73.548	11,4	1.906.611	201.527	10,6
11	Bent Jensen	3	649.587	88.957	13,7	2.001.719	253.963	12,7
12	Camilla Nielsen	3	612.854	66.103	10,8	1.827.843	189.462	10,4
13	Dennis Olsen	3	602.541	57.185	9,5	1.738.036	173.205	10,0
14								
15	I alt		2.510.803	285.793	11,4			

F11		=SUM.HVIS(\$A\$2:\$A11;\$A11;C\$2:C11)						
	A	B	C	D	E	F	G	H
1	Sælger:	Mdr	Salg dm	Rabat dr	Rabat%	Salg åtd.	Rabat åtd.	Rabat%
3	Bent Jensen	1	762.145	92.458	12,1	762.145	92.458	12,1
7	Bent Jensen	2	589.987	72.548	12,3	1.352.132	165.006	12,2
11	Bent Jensen	3	649.587	88.957	13,7	2.001.719	253.963	12,7
14								
15	I alt		2.001.719	253.963	12,7			

I en pivottabel - se næste kapitel - vil Bent Jensen og de øvrige sælgere kun indgå med 1 række hver - nemlig med summen i række 15 for hver sælger. Specifikationen på de enkelte måneder kan dog også gøres i pivottabeller med funktionen "drill down".

8. Sortering og filtrering

Filtrering og diagrammer

I et datamateriale med mange kolonner kan det ofte være vanskeligt at overskue konsekvenserne af en filtrering på alle kolonner - også selvom man har sammenfattet tallene med én eller flere SUB-funktioner. Her kan en grafisk afbildning af kolonnerne være til stor nytte. En afbildning af et filtreret materiale omfatter nemlig kun de synlige data ligesom SUB-funktionerne og diagrammet vil derfor ændre sig i takt med filtreringen. Eksempelvis vil en XY-punkt diagram med trendformel og R^2 -koefficienten vise sammenhængen mellem 2 eller flere variabler (kolonner) og genberegne trendformelen og R^2 ved hver filtrering og dermed føje en yderligere dimension til dataanalysen ved hjælp af filtrering.

"Dataskrubning"

Dette malerisk udtryk dækker over det forhold at et datamateriale meget ofte skal præpareres før det kan anvendes i forskellige analyser og til det formål er filtrering særdeles velegnet.

Et datamateriale, der indeholder mellemtotaler - f.eks. opsummering pr. måned eller pr. produkt - kan ikke anvendes i en pivottabel eller i statistiske analyser. Med filtrering kan udvælges alle rækker med tekster som Total, I alt, Sum o. lign. i principielt alle kolonner, men som regel er det kun de første kolonner, der har en sådan tekst. Med markering af *rækkenumrene* for de sorterede data kan rækkerne slettes med Ctrl+- (Ctrl og minus på tastaturet). De skjulte rækker forbliver intakte - klik på (Alle) i rullemenuen - og da det er hele rækker, der slettes, vil der ikke være tomme rækker i datamaterialet efter denne "skrubning"

Et andet hyppigt forekommende problem i et datamateriale er tomme celler - f.eks. kan man have glemt sælger, kunde, varenummer el. lign. ved bogføringen og i et udtræk vil disse mangler optræde som tomme celler. Hvis der er tomme celler i en kolonne vil der nederst på rullemenuen for den pågældende kolonne være yderligere 2 punkter - nemlig (tomme) og (ikke tomme). Ved at filtrere efter punktet (tomme) fås et hurtigt overblik hvilke rækker der har mangler og herudfra kan bedømmes om de manglende informationer skal indskrives manuelt i de tomme celler.

Udtræk fra økonomi- og bogholderisystemer vil ofte indeholde hele rækker med 0 - eksempelvis ubenyttede konti. Sådanne kan slettes, men ofte vil der i budgetkolonnen eller i Åtd-kolonnen være et tal og så kan rækken naturligvis ikke slettes. Den mest effektive måde at finde sådanne 0-rækker er ved at lave en kolonne til en "nonsenssum" som summerer alle talkolonner på tværs som sidste kolonne i talmaterialet. Hvis nonsenssummen er 0 er som regel alle celler i rækken lig med 0. Med en filtrering af nonsenssummen og med valg af 0 på rullemenuen kan alle betydningsløse rækker findes. Hvis kolonnernes SUBTOTALER også er 0 kan rækkerne slettes.

Filtrering kan **ikke** anvendes til afgrænsning af et datamateriale, der ønskes analyseres med en pivottabel. Ligesom de almindelige formler vil en pivottabel også medtage de skjulte rækker. Med filtrering kan et datamateriale nemt afgrænses til 'salget i 4. kvartal i de sidste 3 år', men en pivottabel vil medtage alle rækker - skjulte såvel som viste. Ønskes en pivottabelanalyse af et sådant afgrænset datamateriale kan der tages en 'hård kopi' af de filtrerede data og denne kopi kan placeres på et andet regneark, som derefter kan analyseres med en pivottabel. Det samme gælder opslagsformler.

 En liste eller en filtreret liste kan markeres med genvejstasten Ctrl+*