

Projekt nr. 1:

Et datasæt vedrørende kundeloyalitet.

Datamaterialet til denne opgave er stillet til rådighed af økonomidirektør Ralph Mathiesen, JohnsonDiversey.

JohnsonDiversey er et firma som handler med rengøringsmidler og redskaber til rengøring. Der satses på kvalitet frem for discount priser, og kunderne er typisk rengøringsfirmaer og virksomheder.

Blandt andet for at undersøge i hvor høj grad kunderne er stabile, har man af firmaets fakturadatabase udtrukket en fil, som over en given periode for hver enkelt transaktion (dvs. faktura, eller leverance, det kommer stort set ud på et) fortæller hvilken kunde det drejer sig om, hvor stort et beløb det drejer sig om, samt datoen for leveringen. Detaljer følger nedenfor.

Opgavens formål er at undersøge i hvor høj grad kunder falder fra. Firmaet har et naturligt ønske om at registrere frafald så tidligt som muligt, for eventuelt at gribe ind på passende måde. Desuden ønsker man i almindelighed at vide noget om i hvor høj grad kundemassen er stabil, eller hvor stor udskiftning der er. Samtidig ønskes en generel vurdering af hvordan firmaets omsætning (såvel i kroner og øre som i antal transaktioner) varierer over tiden.

Data foreligger i form af filen 03-1-dat.txt, som kan hentes fra kursets hjemmeside.

www.mes.cbs.dk/~sttt/MPAS.html

Filen er på 35172 linier plus en overskriftslinie. Overskriftslinien og de første otte datalinier er gengivet her:

KONTO1	KONTO2	AAR	MD	DAG	TID	KR
1	1	2002	10	17	17	3124.20
1	1	2002	11	12	43	250.00
1	1	2002	12	10	71	4623.38
1	1	2003	1	28	120	762.87
1	1	2003	2	27	150	6911.74
2	2	2002	12	6	67	1047.08
2	2	2003	1	13	105	1522.86
2	2	2003	1	20	112	139.61
...						

Data omfatter samtlige transaktioner med kunder i perioden 1/10-02 til 1/4-03. Dog er kunder med under 5 transaktioner sorteret fra.

Hver linie svarer til en transaktion (faktura, levering). De syv oplysninger i hver linie er

1. KONTO1. Kontonummer, bestemt ved fakturaen.
2. KONTO2. Kontonummer, bestemt (formodentlig) ved leveringsadressen.
- 3-5. AAR, MD og DAG. Datoen på fakturaen (formodentlig tæt på leveringsdatoen).
6. TID. Datoen på fakturaen omsat til et tal mellem 1 og 183 (hele dage, ingen korrektion for weekender, helligdage etc.).
7. KR. Beløbet på fakturaen.

Variablen TID er udregnet på forhånd for at undgå for stor fokusering på den besværlige dataindlæsning. Af samme grund er KONTO1 og KONTO2 omkodet fra deres oprindelige form (et ID med adskillige cifre og undertiden bogstaver) til hele tal, på en sådan måde at sammenfald mellem KONTO1 og KONTO2 er bevaret. Oftest er KONTO1 og KONTO2 sammenfaldende, og der foreligger ikke nogen retningslinier for hvilken af de to det er mest relevant at bruge.

Huskeseddel vedr. specielle dage:

DAGNR	UGEDAG	DATO	
1	tir	1/10-2002	Første dag
2	ons	2/10-2002	Næsten ingen omsætning (???)
9	ons	9/10-2002	Meget lavt salg (???)
49	man	18/11-2002	Ingen omsætning (???)
85	tir	24/12-2002	Juleaftensdag
86	ons	25/12-2002	Første juledag
87	tor	26/12-2002	Anden juledag
92	tir	31/12-2002	Nytårsaftensdag
93	ons	1/1-2003	Nytårsdag
183	tir	1/4-2003	Sidste dag

Desuden foregår der (med en enkelt undtagelse) intet salg på lørdage og søndage.

Projektbesvarelsen afleveres mandag d. 20. oktober.