

Projekt nr. 2:

Analyse af kundedata fra Gyldendals Bogklub.

Data til dette projekt er stillet til rådighed af Malene Haxholdt, SAS Institute. Det drejer sig om et udtræk fra medlemsdatabasen for Gyldendals bogklub fra 2002. Formålet med at trække disse data ud var at undersøge årsagerne til “churn”, her defineret som udmeldelse af klubben i perioden 14/8–12/9 2002. Fra bogklubbens side er man naturligt interesseret i at forudsige hvilke medlemmer der står for at melde sig ud, så man evt. kan forhindre dette v.h.a. diverse tiltag.

Data er dokumenteret lidt mere udførligt i Word-dokumentet “Databeskrivelse” som findes på kursets hjemmeside. På datafilen 06-2.txt findes de 24 variable af længde 20058 i samme rækkefølge som i beskrivelsen. Her skal blot tilføjes et par ting:

Den første variabel, medlemsnummeret (som vi næppe skal bruge til noget), er stort set en fortløbende nummerering, men den har et enkelt “hul” (medlem nr. 6609 mangler af en eller anden grund).

I databeskrivelsen nævnes at man har foretaget en “oversampling” af data. Det betyder ifølge pålidelig kilde at nogle af “ikke-churnerne” er udeladt for at få et mindre datasæt. Man må gå ud fra, at ikke-churnere er udeladt ved passende randomisering. Sammenligningen af churnere med ikke-churnere bliver ikke ret meget dårligere af den grund, fordi ikke-churnerne stadig er langt de fleste (13.4 procent churnere, hvor der åbenbart var 4 procent oprindeligt, svarende til at ca. 73% af ikke-churnerne er udeladt).

Churn. Det er ikke helt indlysende hvorfor ordet “churn” i dokumentet bruges til at beskrive den handling at et medlem siger sit medlemskab op. Bortset fra at selve ordet på engelsk betyder “kærne” kan følgende opslag på <http://en.wiktionary.org/wiki/Churn> måske hjælpe lidt:

To churn is to agitate rapidly and repetitively, or to stir with a rowing or rocking motion.

Used literally, it generally applies to liquids, notably cream, which is churned to produce butter.

It can also be used figuratively, with the meaning of producing excessive and sometimes undesirable or unproductive activity or motion. The figurative meaning has been applied to stock transactions related to mutual funds. It describes a situation in which the fund managers move

the money of their investors rapidly from one stock to another, taking a fee for each transaction out of the fund, but not necessarily increasing the value of the investment portfolio substantially. This is called churning the fund portfolio.

Data mart. I databeskrivelsen bruges også ordet “data mart” om datasættet. Man bliver ikke meget klogere af at slå “mart” op i en engelsk–dansk ordbog, men forklaringen får man i følgende udskrift fra <http://www.webopedia.com/>

Data mart. A database, or collection of databases, designed to help managers make strategic decisions about their business. Whereas a data warehouse combines databases across an entire enterprise, data marts are usually smaller and focus on a particular subject or department. Some data marts, called dependent data marts, are subsets of larger data warehouses.

Korte variabelnavne. Følgende variabelnavne vil blive brugt ved forelæsningerne (idet selve beskrivelserne gentages her, for overskuelighedens skyld):

MEDLNR	Medlemsnummer
CHURN	Churn i target periode
STIND	Indmeldt på Storkampagne
TOTALMD	Måneder siden indmeldelse
ANDKLUB	Medlem af anden klub
CHURNA	Churn fra anden klub
TIDFRI	Måneder siden forpligtigelse klaret
GLMEDL	Gammelt medlem
RETUR	Antal retur fuldpris artikler
MDBOG	Antal månedsbøger
FPRIS	Antal fuldpris artikler
PSFPRIS	Antal perioder siden sidste køb af fuldprisartikel
KOEB	Samlet værdi af køb
VRETUR	Samlet værdi af retur
NETTO	Værdi af nettokøb
MD_01	Køb af månedsbog for 1 periode siden
MD_02	Køb af månedsbog for 2 periode siden
FP_01	Køb af fuldpris artikel for 1 periode siden
FP_02	Køb af fuldpris artikel for 2 periode siden
MRETUR1	Retur af månedsbog for 1 periode siden
MRETUR2	Retur af månedsbog for 2 periode siden
LTVTOT	Samlet livstidsværdi af kunden
LTV6M	Værdi af kunden de sidste 6 måneder
KOEN	Køn

Projektbesvarelsen afleveres mandag d. 15. maj.