Churn Analysis: Fundraising

The data at hand is a part of the database of a fundraising organization.

As always, before we can start the modeling phase, we need to create the basetable. In order to do this, the information of different tables has to be combined (see Appendix).

The fundraising organization gave you a data dump as of 02/02/2007.

They will use the model you create to score their customer base in terms of their churn probability.

NOTE: All datasets attached are in the SAS data format (sas7bdat). You will need to figure out how to import this into pandas.

1. **Divide your data into different time windows.**

For example: If you have data for 5 years,

- Use the first 3 years to train models
- validate on the 4th year and pick your best model
- The 5th year's data will be your hold-out sample for determining out-of-sample accuracy

2. **Create the Basetable (for training models)**

- 1 observation per customer
- Only include information present during training period
- Churn should happen during validation period or later

3. **Follow the instructions below**

1) Subset the **extrel** dataset according to the appropriate **timewindow** and only select donors with a commitment.

2) Create the following independent variables :
    a. Frequency
    b. Recency
    c. Total and average donation per donor

d. Pay type per customer
   → Create new variables that signify whether a donor ever used sendout, order, own initiative and unknown
e. Preferred mailing language
f. Dummy whether the donor made a complaint
g. Dummy whether communication direction was ever incoming

3) Create your dependent variable [if relationship has ended, churn = 1]

4) Merge everything and indicate those rows that have a missing with a dummy

5) Impute missings, treat outliers

6) Create competing models

7) Asses the performance of your model

Appendix: Description of the tables

Extrel: All the donors of the organization

Variable	Description
Extrelno	Unique identifier of each donor
Exrelactcd	Activity code of the donor
Extrelstdt	Start date of the relationship
Exreldaten	End date of the relationship (Missing: not ended)

Extrelty: Description of the activity

Variable	Description
Exrelactcd	Activity code of the donor
Exrelactde	Description of the activity

Nameaddr: Sociodemographical information

Variable	Description
Extrelno	Unique identifier of each donor
Name1title	Title to address someone
Postcode	Postcode
Languagecode	Preferred mailing language

Payhistory: Paymenthistory of each donor

Variable	Description
Pid	Unique identifier for each payment
Pdate	Date of payment
Pamt	Amount of payment
Extrelno	Unique identifier of each donor
Paytypecd	Paytype *O Bank transfer* *D Permanent order* *E Own initiative* *X Unknown*
Status	Status of payment *OK Normal/Real payment* *CO Correction (internal)* *RF RF (Refund)* *RC Recall*

Communication: All possible communication between the donor and the organization

Variable	Description
Contid	Unique identifier for each contact
Mediumcode	Medium of the contact (CI is unknown)
Mntopcode	Main topic code of the contact
Classcode	Class of the contact
Extrelno	Unique identifier for each donor
Contdirec	Direction of the communication *I Incoming* *P Outgoing*
Contdate	Date of the contact

Commediu: Description of medium type

Variable	Description
Mediumcode	Code of the mediumtype
Mediumdesc	Description

Commaint: Description of the main topic code

Variable	Description
Mntopcode	Main topic code
Mntopdesc	Description

Comclas: Description of the contact class

Variable	Description
Clascode	Code of contact class
Clasdesc	Description