Understanding Churn

- Context: retail sales
Understanding Churn Dataset

- Real data describing customers and transactions
  - Several department stores
  - Purchases performed over several years
  - Includes product details, customer ID

- articolo.csv
  - Textual description of the products (in Italian)

- cliente.csv
  - Basic information about customers (in Italian)

- data.csv
  - Translation table for date coding

- marketing.csv
  - Marketing hierarchy of products (in Italian)

- venduto.csv
  - Transactions, a line for each product sold
Understanding Churn Dataset
Understanding Churn
Data understanding
Understanding Churn
Data understanding

![Bar chart showing total transactions per month.

<table>
<thead>
<tr>
<th>Month</th>
<th>Num Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DICEMBRE</td>
<td>138,112</td>
</tr>
<tr>
<td>NOVEMBRE</td>
<td>127,418</td>
</tr>
<tr>
<td>OTTOBRE</td>
<td>136,106</td>
</tr>
<tr>
<td>SETTEMBRE</td>
<td>137,995</td>
</tr>
<tr>
<td>AGOSTO</td>
<td>135,206</td>
</tr>
<tr>
<td>LUGLIO</td>
<td>136,934</td>
</tr>
<tr>
<td>GIUGNO</td>
<td>136,205</td>
</tr>
<tr>
<td>MAGGIO</td>
<td>132,753</td>
</tr>
<tr>
<td>APRILE</td>
<td>131,916</td>
</tr>
<tr>
<td>MARZO</td>
<td>136,833</td>
</tr>
<tr>
<td>FEBBRAIO</td>
<td>120,278</td>
</tr>
<tr>
<td>GENNAIO</td>
<td>121,245</td>
</tr>
</tbody>
</table>
Understanding Churn
Data understanding
Understanding Churn
Data understanding
Understanding Churn
General objective

• The company realized that a large number of loyal customers of a shop are churning
  – Question 1: Who is churning?
  – Question 2: Why?
  – Question 3: What can we do about that?

• Sketch a KDD project plan aimed to tackle the problem
  – Use CRISP-DM as guidelines
  – Make questions as you would to the company managers (your client)
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

• Question 1: Who is churning?
  – Implies creating some profile of churners
  – Descriptive task, not predictive

• Possible reformulation
  – What is the personal and behavioural profile of churners?
Understanding Churn
Sample solution

• Question 2: Why?
• Question 3: What can we do?
  – Implies understanding what triggered the churn
  – Might look for clues that precede the churn

• Possible reformulation
  – What (customer-based) events most often precede a churn?
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

• Question 1: What is the personal and behavioural profile of churners?

• Data mining objective:
  – Build **classification rules or trees** with target variable “churner”
  – Extract **distributions or patterns** on churners and non-churners and compare them
Understanding Churn
Sample solution

• Questions 2 and 3: What (customer-based) events most often precede a churn?

• Data mining objective:
  – Define customer events: buying some specific products, changing brand for a frequently bought item, etc.
  – Extract sets or sequences of events that occur frequently in churners’ history (and not in non-churners’) a short time before churning
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

- Select data sources that are relevant
  - Purchases
  - Demographic info on customers
  - Promotions? (Were not mentioned initially)
  - Product hierarchy

- Check data quality
  - Especially demographic data, which often are unreliable or incomplete
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

• Question 1: What is the personal and behavioural profile of churners?
  – Involves rules, classifiers and distributions
  – Requires to compute indicators and aggregates, such as:
    • Purchase volumes (total, over key categories, split into days of week or hours of the day, etc.)
    • Frequency of visit, Recency, Tenure, etc.
    • Trends? E.g., variations of volumes in the last 3 months
  – Customers should be divided into churners and non-churners
    • Use predefined division or invent one
Understanding Churn
Sample solution

• Questions 2 and 3: Why and What can we do?
  – Involves events
  – Requires to define sensible types of events:
    • Abnormal purchases (large expenditure, unusual products, etc.)
    • Changes of habits (usual product from unfamiliar brand, change usual day or hour of visit, etc.)
    • Others?
  – Customers should be divided into churners and non-churners
    • As previous case
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

• Classification
  – Something human-readable: decision tree

• Event patterns
  – Frequent itemsets
  – Sequential patterns (w/ or without gaps)
Understanding Churn

General objective: Who is churning? Why? What can we do?

CRISP-DM Summary
Understanding Churn
Sample solution

- Classification
  - Accuracy

- Event patterns
  - Frequency and frequency delta
  - Select patterns that are frequent in one set (churn / non-churn) and infrequent in the other
Understanding Churn

General objective: Who is churning? Why? What can we do?
Understanding Churn
Sample solution

• Simple case, since the business problem is about understanding, not creating a service
  – Packing all the results in a report
  – Submit to marketing dept.
  – Maybe re-iterate the process to go deeper on specific points