Define customer segments in the Insurance sample

When you define customer segments, you profile the customers into groups that have similar demand characteristics. In the Insurance sample, customers are profiled based on their financial sophistication.

Customers are segmented into financially sophisticated, and novice categories. This means that insurers can target each segment with cross-selling insurance policies that are appropriate to increase the effectiveness of cross-sell campaigns.

The example stream for defining customer segments is named Segmentation.str.

The model used is a TwoStep cluster.

The inputs for the example Segmentation model are customer master data and customer policy data, specifically:

* Demographic data: age, gender, marital status, and employment status.
* Insurance policy related data: insurance lines, policies, premiums, tenure, and insurance score.
* Financial data: income, retirement plans, home ownership status, vehicle ownership.

These inputs are aggregated. As each record is read, based on a distance criterion, the TwoStep cluster algorithm decides whether it should be merged with an existing cluster, or used to start a new cluster.

If the number of clusters becomes too large (you can set the maximum number), then the distance criterion is increased, and clusters whose distance is now less than the modified distance criterion are merged. In this way, the records are clustered into a preliminary set of clusters in a single pass of the data.

**Segmentation model**

Customers are segmented based on their financial sophistication. This model enables the insurer to sell insurance policies that are appropriate to the customer.

# Data used in the Insurance sample

In the Insurance sample, the Insurance company runs a multi-line business.

The following types of data are used.

**Customer master data**

This includes customer’s demographic data, employment and income data, and information about the household as well. POLICYHOLDER and HOUSEHOLD tables capture most of this data. Typically, Master Data Management systems are the source of customer master data.

**Customer policy data**

This includes aggregated customer information, such as the number and types of policies owned by the customer, total premium being paid by the customer, average claim amount, tenure of the customer, number of complaints, number of claims, and customer sentiment data. POLICYHOLDER\_FACT and POLICY\_FACT tables capture most of this data.

**Customer transaction data**

This includes data about all the customer transactions such as the policies purchased, their inception and maturity/renewal date, data related to all of the complaints made by the customer in the past, and also data related to all of the claims made by the customer. POLICIES, CLAIMS, COMPLAINTS, COMPLAINT\_DETAILS tables contain this data.

**Customer Social Media Data**

Apart from the customer data that is available within the enterprise, insurance organizations may also want to get insights from external sources of data. For example, the social media channels where customers post comments about their experiences with their insurers, as well as about their needs and life-events that can potentially lead to an opportunity to sell appropriate insurance products. SMA\_DATA and SMA\_DATA\_ANALYSIS tables capture such external data, as well as the summarized analysis of this social media data.

**Note:** The fetching of social media data is outside the scope of this solution.

**Customer web browsing data**

Many insurance organizations today allow their customers to buy or explore their insurance products online through their websites. Technology makes it possible to track customers' activities on their websites, giving them vital insights about the customers' current interest in specific insurance products. Web Analytics tools can be used to analyze customers' website activities and use this information along with other customer data to make the right recommendations to the customers at the right time. ACTIVITY\_FEED\_DATA, ONLINE\_BROWSING\_HISTORY and ONLINE\_BROWSING\_SUMMARY tables contain customers web activity data.

# Group customers based on their current life stage

The Insurance sample uses the Lifestage Segment model to group customers based on their current life stage.

The example stream for grouping customers into their current life stage is named Lifestage Current Segment.str.

The model uses simple rules to get the current life stage segment of customer. Some examples of defined segments are:

* Newly married.
* Young family.
* Young and affluent.
* Single.
* Divorced.

# Insurance data model

The historic data that is used for predictive modeling in IBM® Predictive Customer Intelligence is stored in an IBM DB2® database.

A Predictive Enterprise View (PEV) of the data is created which passes data to the real time scoring service. Database Views (DB Views) are also created to be used in IBM SPSS® Modeler streams.

The following table describes some of the data columns that are part of the Predictive Enterprise View and the Database Views.

| **NAME** | **DESCRIPTION** |
| --- | --- |
| AGE | The age of the policy holder. |
| CLTV | Customer Lifetime Value. |
| EDUCATION | The education level of the policy holder. |
| EMPLOYMENT\_STATUS | The status of the person’s employment. |
| GENDER | The person’s sex or gender. |
| INCOME | The policy holder’s annual income. |
| MARITAL\_STATUS | The marital status of the person. |
| MAINTENANCE\_COST | The cost of maintaining this policy. |
| MONTHS\_SINCE\_POLICY\_INCEPTION | The number of months since the policy holder started the policy. |
| MONTHS\_SINCE\_LAST\_CLAIM | The number of months since the policy holder filed the last claim. |
| NUMBER\_OF\_CLAIMS\_DENIED | The number of claims that were denied. |
| NUMBER\_OF\_CLAIMS\_FILED | The number of claims that are filed. |
| CLAIM\_SETTLEMENT\_DURATION | The time, in days between the date when the claim opened and the date when the claim was closed, and the customer satisfaction confirmed, based on the status of the claim. |
| NUMBER\_OF\_COMPLAINTS | The number of complaints the policy holder has submitted. |
| NO\_OF\_CLOSED\_COMPLAINTS | The number of complaints that have been closed. |
| NUMBER\_OF\_OPEN\_COMPLAINTS | The number of complaints that are open. |
| LATEST\_NOTE\_ATTITUDE | The noted attitude of the last communication. |
| AVG\_NOTE\_ATTITUDE | Average communication note attitude. |
| NUMBER\_OF\_POLICIES | The number of policies that the policy holder has. |
| POLICYHOLDER\_ID | Any value without business meaning that uniquely distinguishes each occurrence of this entity. |
| POLICY\_ID | Any value without business meaning that uniquely distinguishes each occurrence of this entity. |
| POLICY\_TYPE | Indicates the policy type, for example, fixed term and flexible term. |
| VEHICLE\_OWNERSHIP | Indicates whether the policy holder owns a vehicle or not. |
| VEHICLE\_TYPE | The vehicle type. |
| VEHICLE\_SIZE | The vehicle size. |
| HOME\_OWNERSHIP\_STATUS | The tenancy status of residence. |
| INSURANCE\_LINES | The number of types of insurance products held by the policy holder. |
| INSURANCE\_SCORE | The insurance score, based on the credit score, as well as other factors such as claim filing history. |
| LIFE\_CUSTOMER | Indicates whether a customer owns a life insurance policy. |
| NON\_LIFE\_CUSTOMER | Indicates whether a person owns a non-life insurance policy. |
| NUMBER\_OF\_CHILDREN | The number of children of the policy holder. |
| NUMBER\_OF\_INSURED\_CARS | The number of insured cars insured with the insurance company. |
| POLICYHOLDER\_TENURE | The number of years the policy holder has been a customer with the insurance company. |
| TOTAL\_PREMIUM | The total premium on all the policies paid by the policyholder to the insurance company. |
| RETIREMENT\_PLAN | The name of the retirement plan owned by the policyholder |
| HOUSEHOLD\_NUMBER\_OF\_CHILDREN | The number of children in household |
| HOUSEHOLD\_NUMBER\_OF\_INSURED\_  CARS | The number of insured cars in the household that are insured with the insurance company. |
| NUMBER\_OF\_POLICIES\_IN\_HOUSEHOLD | The number of policies held by the household. |
| HOUSEHOLD\_TENURE | The number of years that a household has been classified within its customer status. |
| HOUSEHOLD\_PREMIUM | The total premium paid by the household to the insurance company. |
| HOUSEHOLD\_DISPOSABLE\_INCOME | The amount of money that the household can spend after having paid all the fixed expenses such as rent, mortgage repayment and so on. |
| *Table 1. Key data columns in the Insurance sample* | |