a.s.r.
de nederlandse
verzekerings
maatschappij
voor alle
verzekeringen

a.s.r./aegon car churn data description

Overview

Description

- The dataset provides information of car policies over time.
  - The variable of interest is churn
  - The independent variables contain information related to the
    - **Policy**: Length, initiation year, coverage, premium, discounts…
    - **Customer**: Age, postcode.
    - **Car**: brand, value, fuel type…

- The data was retrieved on December 15th, 2023.

Variable of interest: Churn

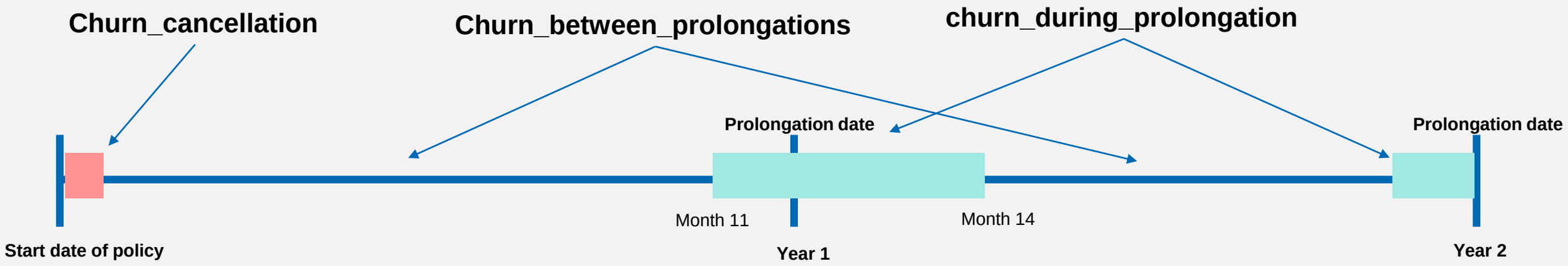
We recognize different types of churn

- **Churn cancellation**
  - Customers that have churned within the first 2 weeks of the start date of the policy.
  - Customers have the option of cancelling their policy in this time period without incurring into a “penalty”.
- **Churn during prolongation**
  - A car insurance policy is renewed every year. Prolongation refers to the period in which the customer has to decide whether he or she would like to prolong the policy for another year.
  - This period starts on the month before the policy reaches the end of each year*.
    - For example, if a policy starts on the 15th of January 2022, the prolongation date will be from the 1st of December 2022 to the 14th of January 2023.
- **Churn between prolongations**
  - This category refers to any churn that does not fall into the first two categories.
    - For example, a policy that starts on the 15th of January 2022 and ends on the 15th of April 2022.

** This is a bit more nuanced, see the slide about churn during prolongation for a whole explanation of how this type of churn is obtained.*

Variable of interest: Churn

We recognize different types of churn



Variable of interest: Churn

Note on churn during prolongation

- Prolongation occurs once a year.
- In the churn during prolongation variable we include not only the month before a policy ends but also the next two months (see previous slide for a visualization of this).
- We do this because we expect some customers to not churn at the time they receive the prolongation letter but after the invoice of the new (and possibly higher) premium is received.

Independent variables

Additional notes

- Each policy has at least one version per year (prolongation).
  - In addition, it can also have more mutations/changes throughout the year (e.g., customer changes car or payment details).
  - This information is included in the mutation_X variable which indicates what mutation/change has happened as well as the premium change that this change has incurred, if any.
- A policy consists of one or more main coverages and zero or more supplementary coverages.
  - This information can be found on the dataset together with its corresponding premium.
  - We include information about which main coverage the customer has bought, these are:
    - allrisk basis
    - allrisk complete
    - allrisk royal
    - wa-extra
    - wettelijke aansprakelijkheid

Independent variables

Additional notes

- There are several variables that indicate the date in which an event has happened:
  - year_initiation_policy: this is constant throughout the policy length
  - year_initiation_policy_version: this variable changes every year as long as the customer prolongs his/her policy.
  - year_end_policy: This variable is filled for policies that have churned, and hence have an end date; and it is empty if the policy is still active at the time this data was retrieved, 15th December, 2023 (i.e., data_collection_date variable).
- Some policies also had a welcome discount during the first year, which was up to 30% of the premium.
  - In the variable welcome_discount_control_group you can find information regarding whether a given policy on a given year had any welcome discount, list price adjustment, both or neither.

Closing remarks

- It is possible that you will find inconsistencies/errors in the data
  - In some cases, this is due to a wrong input from the user's side (e.g., the customer has inputted a wrong date of birth or age of car)
  - In other cases, this can be the consequence of our data preprocessing pipeline (e.g., there are a few cases in which the variable d_churn is 0 but if you go the the mutation_X column you'll find cancel as a reason for the mutation). In this specific case the customer actually churned.
- For these and other cases that you will for sure find, we'd recommend just to apply common sense.
- If there are wrong values in the data for which you cannot find a reasonable explanation you can of course ask us and we'll check it out.