Description of the Travel Insurance dataset.

A third-party travel insurance servicing company that is based in Singapore.

The attributes:

1. Target: Claim Status (Claim)
2. Name of agency (Agency)
3. Type of travel insurance agencies (Agency Type)
4. Distribution channel of travel insurance agencies (Distribution Channel)
5. Name of the travel insurance products (Product Name)
6. Duration of travel (Duration)
7. Destination of travel (Destination)
8. Amount of sales of travel insurance policies (Net Sales)
9. Commission received for travel insurance agency (Commission)
10. Gender of insured (Gender)
11. Age of insured (Age)

Data preparation used in studies [(Rawat et al. 2021)](https://www.sciencedirect.com/science/article/pii/S2667096821000057) of the Travel Insurance dataset.

The data consists of 63,326 rows and 11 columns: 10 features and 1 target variable.

First, the data is checked for any missing values. There are 45,107 missing values for Gender i.e. 71.2%. Hence, the column is dropped from the datasets before proceeding for further evaluation.

Next, the maximum age is 118 which is not a suitable age to travel for anybody, also most of the insurance companies do not provide insurance to people above 85 years of age, hence considering 100-118 as an outlier category which comprises of 1.55% (984 / 63326) of the total policyholders, it is replaced by 99 hence keeping 99 as the maximum age of a policyholder.

The minimum negative value of Net Sales is justified as net sales are calculated as the difference between the value for which the insurance was sold and the expenses incurred, or the claim amount paid by the insurance company to the policyholder / beneficiary. So the net sale may be negative if the claim amount paid or even if the claim amount is not paid it can be negative when the claim is rejected and the expenses incurred for doing the investigation is more than the actual policy amount paid by the policyholder.

The minimum value of duration is -2 is which not possible under any circumstances and the maximum value is 4881. Even if the unit for the duration is considered to be in days then also this value is not possible as travel insurance policies can be applied for a maximum duration of 1–2 years in the case of an Annual Plan. Considering a maximum duration of 731 days i.e. one year and one leap year, the values above 731 and below 1 are imputed as they constitute only 0.13% (81 / 63326) of the whole dataset. Values above 731 are imputed as 731 and values below 1 are imputed by the median value of duration.