

Appendix 1

Dataset Explanation

Table 1. Explanation of Variables in the Dataset

	Column Name	Variable explanation
Claimant-specific information	PolicyholderNumber	Unique policy number for each policyholder
	FirstPartyVehicleNumber	Vehicle number
	PolicyholderOccupation	Occupation of the policyholder
	FirstPolicySubscriptionDate	Subscription date
	FirstPartyVehicleType	Type of vehicle (car, motorcycle, ...)
	PolicyholderPostCode	Postal code of the policyholder
	PolicyWasSubscribedOnInternet	1 if policy subscribed online; 0 otherwise
	NumberOfPoliciesOfPolicyholder	Number of subscribed policies
	FpVehicleAgeMonths	Month age of the car at time of incident
	PolicyHolderAge	Age of the policyholder
	FirstPartyLiability	Percentage of first party liability
Incident-specific information	Referenceld	Unique claim number identifier
	ThirdPartyVehicleNumber	Vehicle number of third party
	InsurerNotes	Insurer notes about the incident (free text)
	LossDate	Date of the covered loss incident
	ClaimCause	Cause of the incident
	ClaimInvolvedCovers	Policy covers used by the claimant
	DamagesImportance	Importance of damages (assessed by an expert)
	ConnectionBetweenParties	Potential connection between parties
	LossPostCode	Address where the incident took place



	EasinessToStage	Indicator of easiness to stage (computed by Shift)
	ClaimWihoutIdentifiedThirdParty	1 if no other party involved; 0 otherwise
	ClaimAmount	The amount of money claimed
	LossHour	Time of the incident
	NumberOfBodilyInjuries	Number of injured persons in the incident
	Fraud	1 in the case of known frauds; 0 otherwise