

# TRAVAIL PRATIQUE

APPRENTISSAGE STATISTIQUE EN ACTUARIAT  
ACT-4114

ÉQUIPE 07

---

Premier rapport  
**Analyse exploratoire**

---

*Par*

Justin Poulin  
Cédric Provencher  
Philippe Gélinas  
Randy Lefebvre  
Alex Maggioni

*Numéro d'identification*

536 778 899  
YZX YZX YZX  
536 782 326  
XYD XYD XYD  
TRO LOL LOL

*Travail présenté à*

*Monsieur*

**OLIVIER CÔTÉ**

13 MARS 2023



## Table des Matières

<b>Introduction</b>	<b>2</b>
<b>Analyse Exploratoire</b>	<b>3</b>
Nettoyage des données . . . . .	3
Somme des réclamations . . . . .	3
Genre . . . . .	4
Âge du conducteur . . . . .	5
Année du véhicule . . . . .	6
Modèle du véhicule . . . . .	7
Area . . . . .	8
State . . . . .	9
Exposition . . . . .	10
Prime Totale . . . . .	11
SumInsAvg . . . . .	12
<b>Traitement des valeurs manquantes</b>	<b>13</b>
<b>Conclusion</b>	<b>14</b>

## **Introduction**

Pour ce travail, l'objectif est de construire un modèle de prime pure pour la somme des réclamations (tous types confondus).

## Analyse Exploratoire

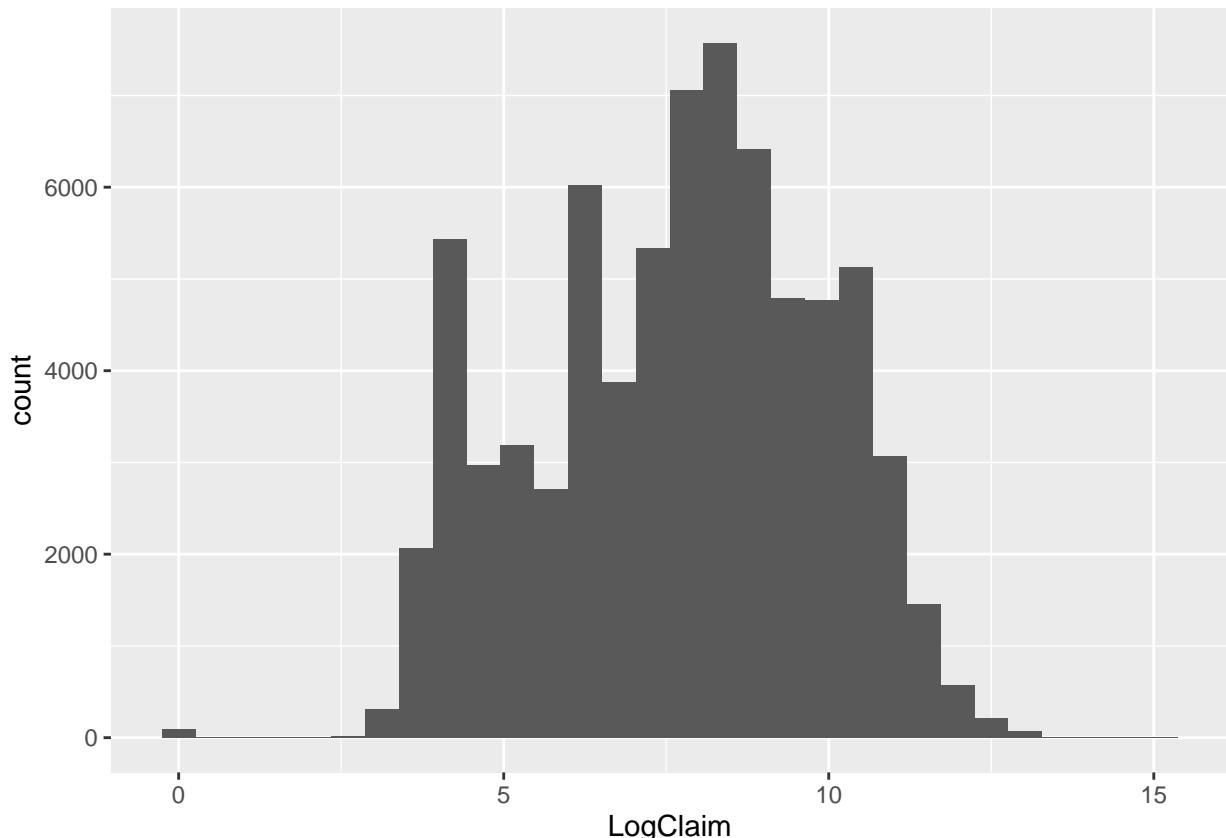
### Nettoyage des données

#### Somme des réclamations

La somme des réclamations est la variable réponse pour ce projet. Cette variable prend fréquemment la valeur de 0 319910 fois sur 393069. Cela revient à dire qu'un assuré ne fera pas de réclamation 81.4% du temps.

Moyenne	Médiane	Écart-type	Minimum	Maximum
2362.0805762	0	$1.5842131 \times 10^4$	0	$3.634933 \times 10^6$

Puisqu'un montant de réclamation peut prendre des valeurs assez élevées, les graphiques peuvent devenir plutôt difficiles à lire. Pour remédier à ce problème, le logarithme du montant total des réclamations est utilisé dans les graphiques. Cela a cependant un coût : toutes les réclamations de 0 doivent être retirées du jeu de données. Des informations supplémentaires par rapport à ces réclamations seront ajoutées au long de l'analyse exploratoire pour pallier ce défaut.



Comme on peut voir dans la Figure 1, il

## Genre

La variable **Gender** a quatre niveaux : *Male*, *Female*, *Corporate* et *NA*.

	Male	Female	Corporate	NA
Fréquence	171087	129166	75067	17750

Lorsqu'on regarde la quantité de ces personnes qui n'ont pas fait de réclamation, on remarque qu'il n'y a absolument aucune réclamation pour le genre *Corporate* et que *NA* semble avoir la plus petite proportion de personnes n'ayant fait aucune réclamation.

	Male	Female	Corporate	NA
Aucune réclamation	135245	101846	75067	7752

En terme de sévérité, les moyennes des réclamations ne semblent pas changer d'un genre à un autre, mais il est possible d'observer un peu plus de réclamations extrêmes lorsque le genre est inconnu.

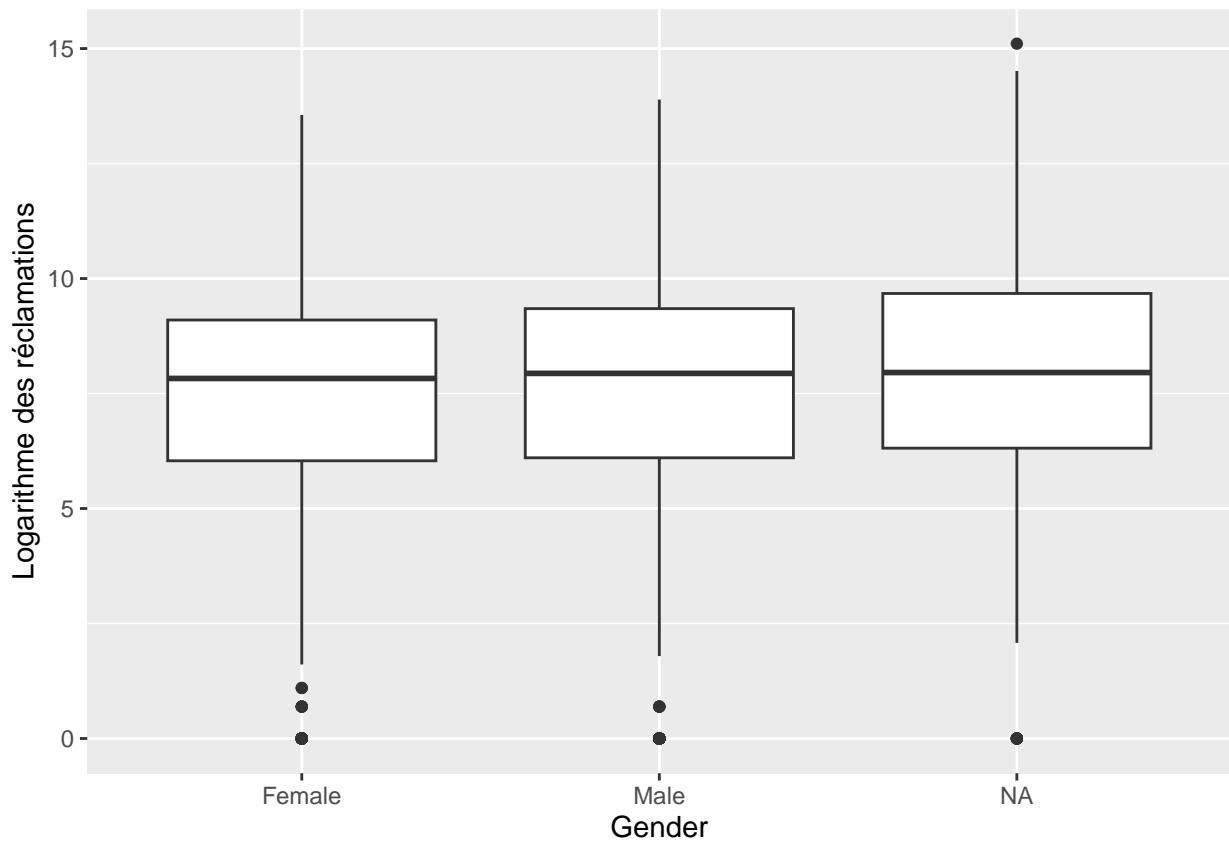


Figure 2: Logarithme des réclamations par genre

### **Âge du conducteur**

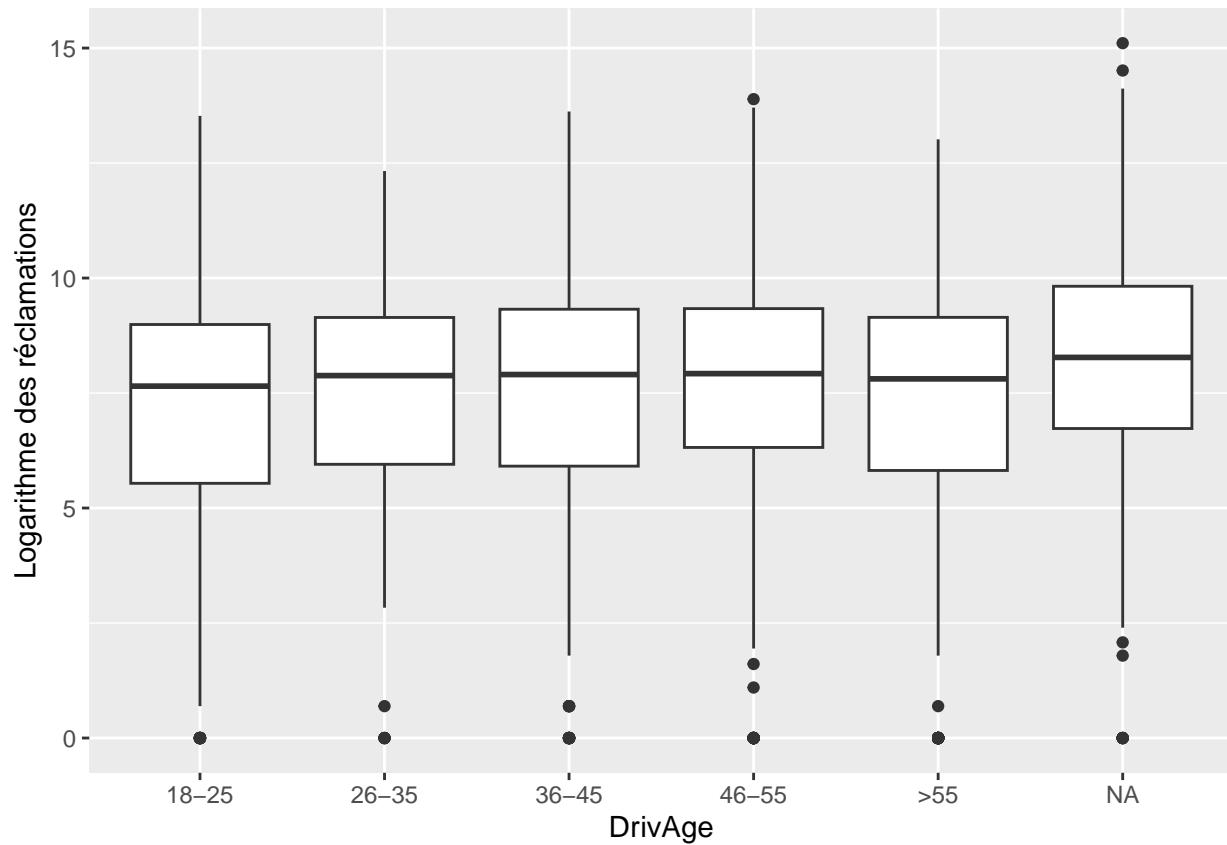


Figure 3: Logarithme des réclamations par Âge de conducteur

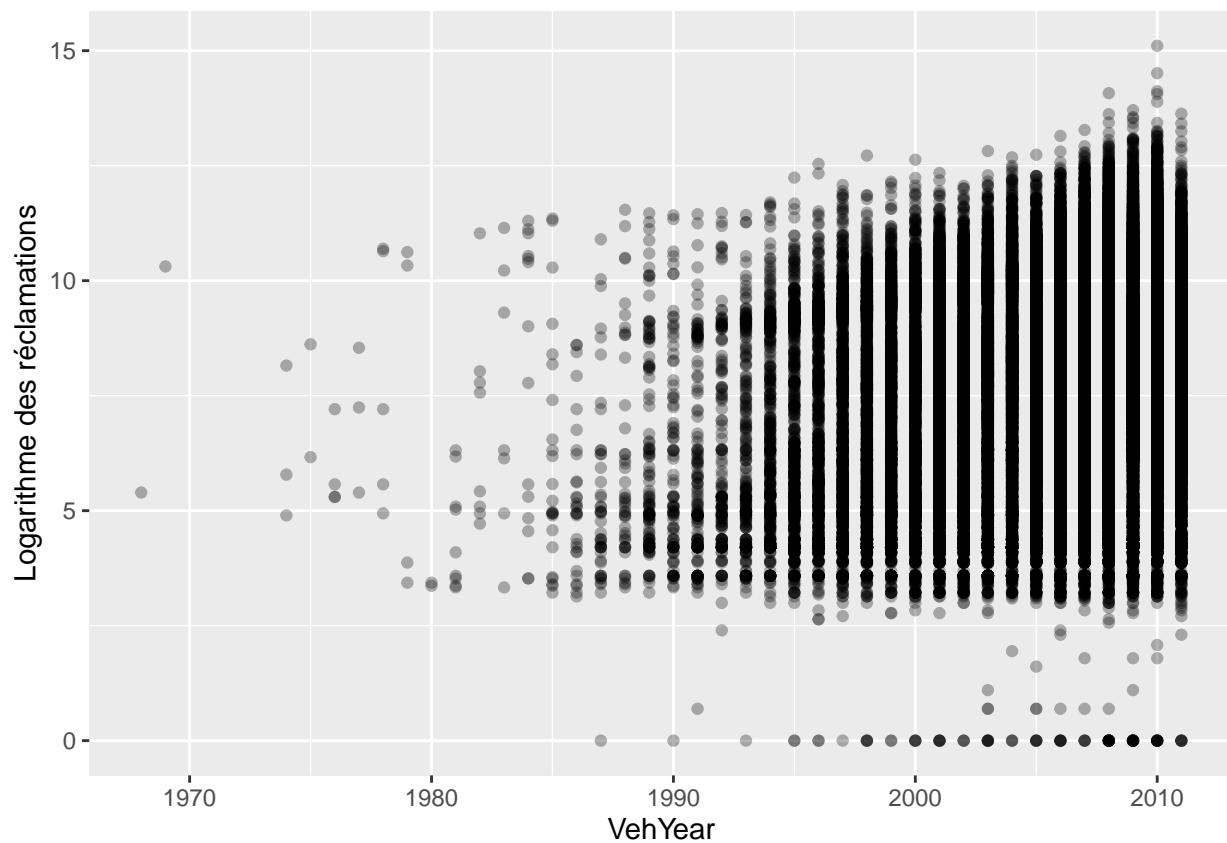
**Année du véhicule**

Figure 4: Logarithme des réclamations en fonction de l'année du véhicule

## **Modèle du véhicule**

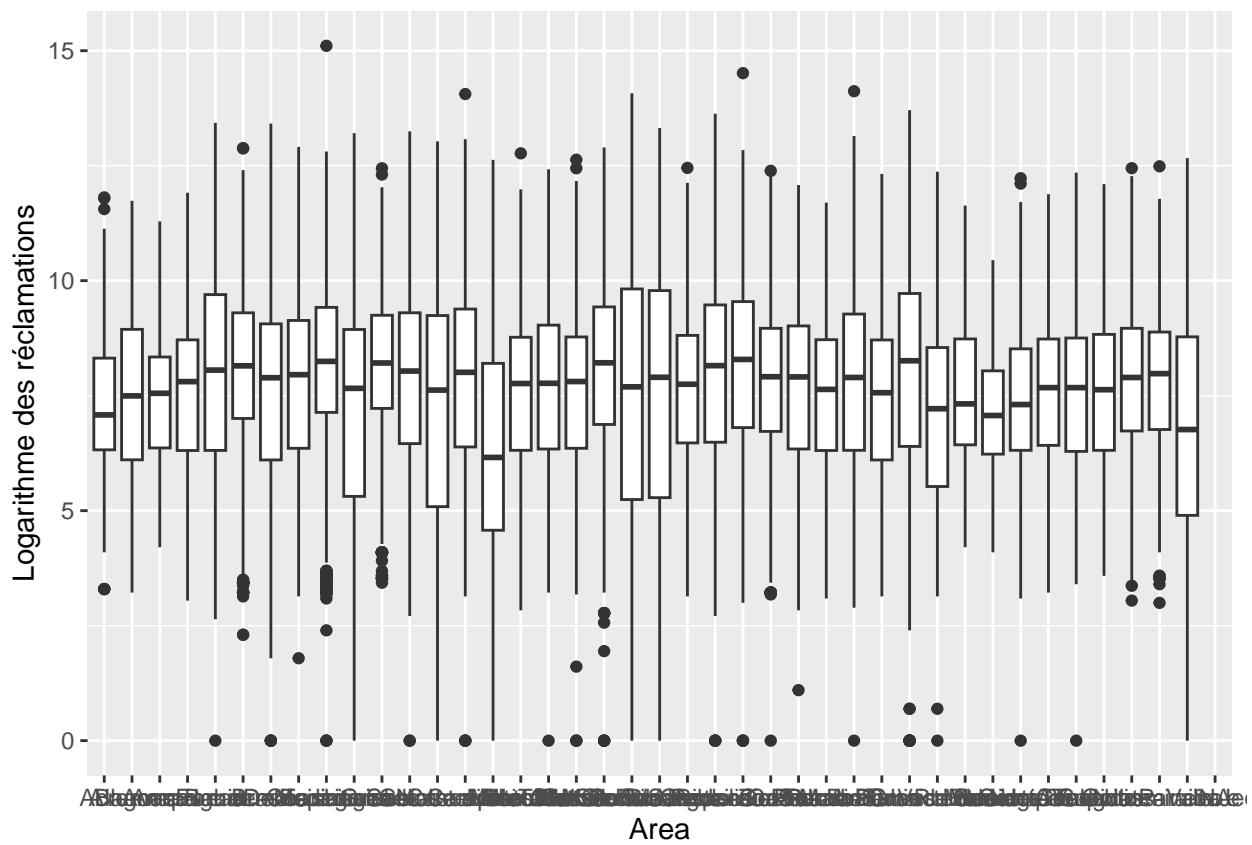
**Area**

Figure 5: Logarithme des réclamations par région

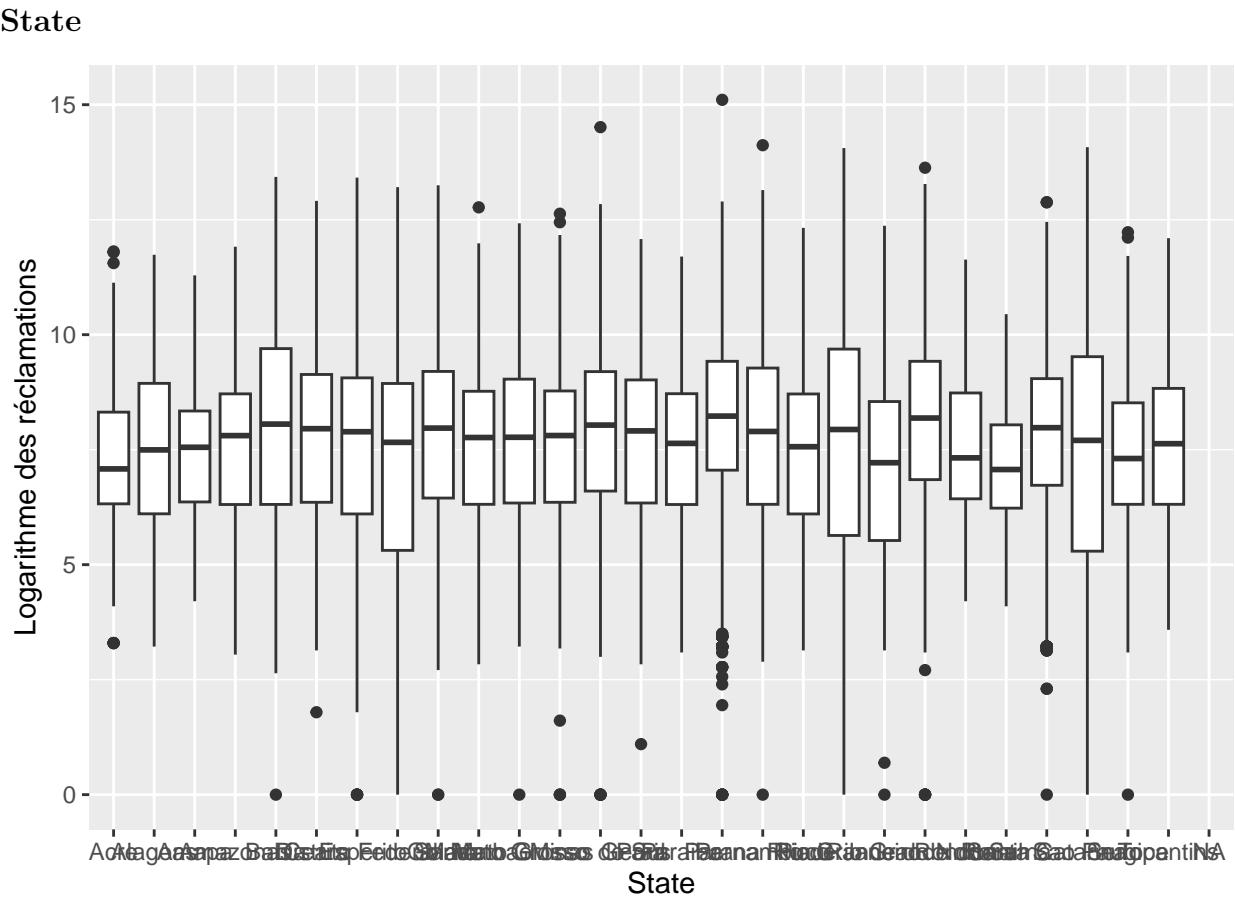


Figure 6: Logarithme des réclamations par état

## Exposition

Moyenne	Médiane	Écart-type	Minimum	Maximum
NA	NA	NA	NA	NA

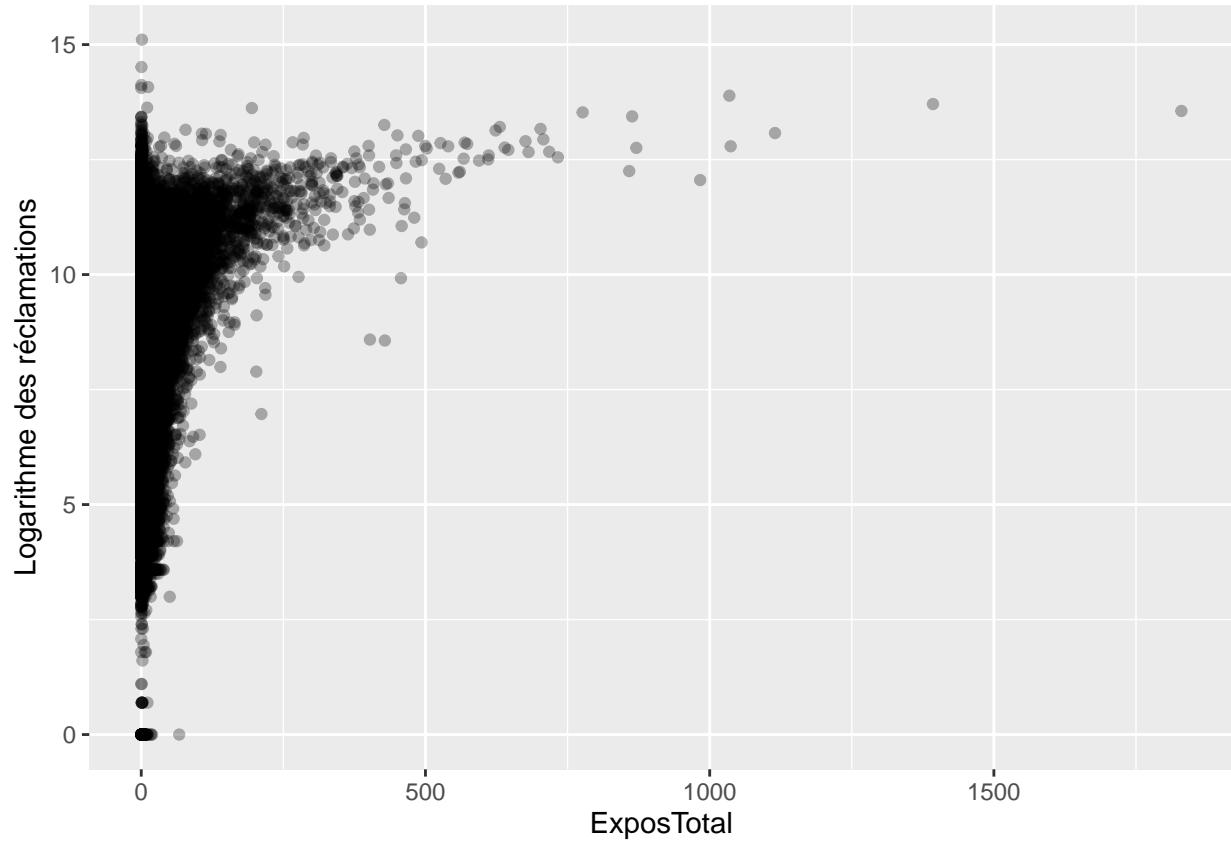


Figure 7: Logarithme des réclamations en fonction de l'exposition

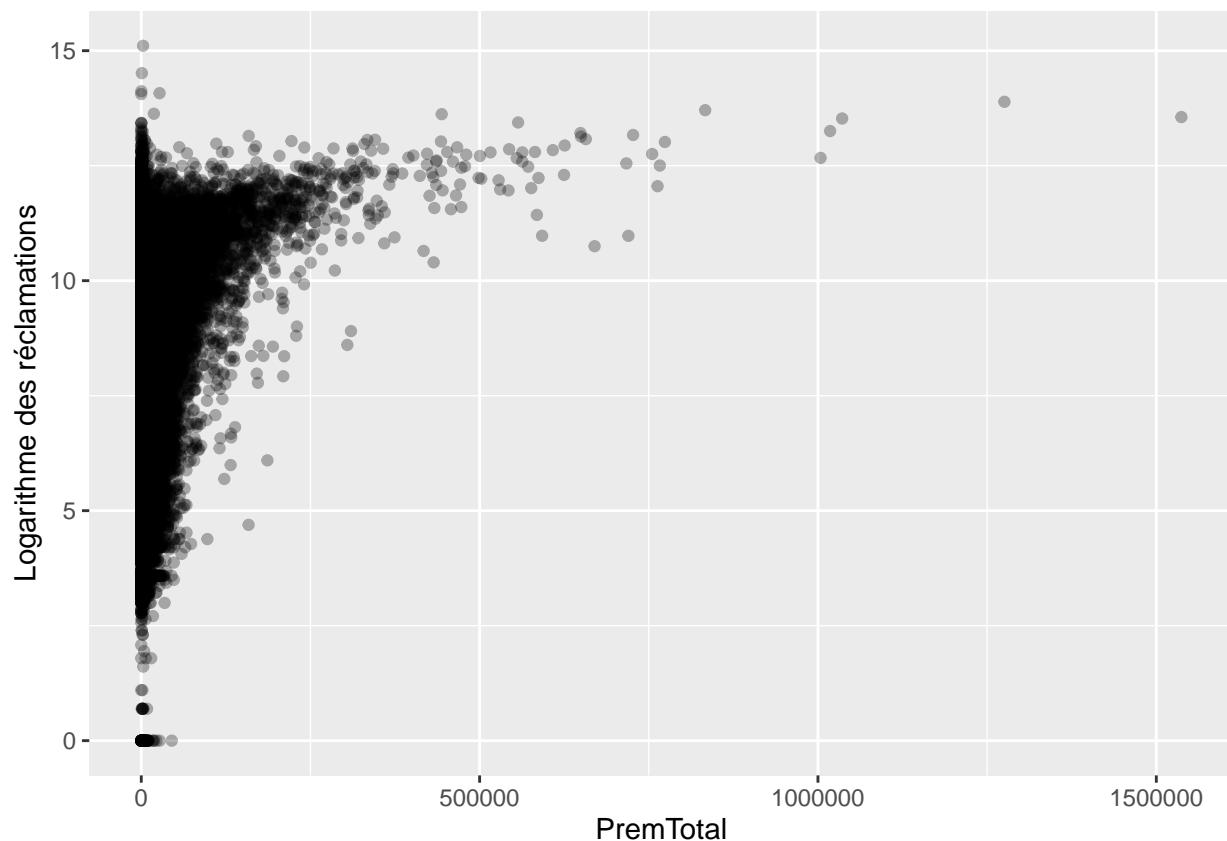
**Prime Totale**

Figure 8: Logarithme des réclamations en fonction de la prime totale

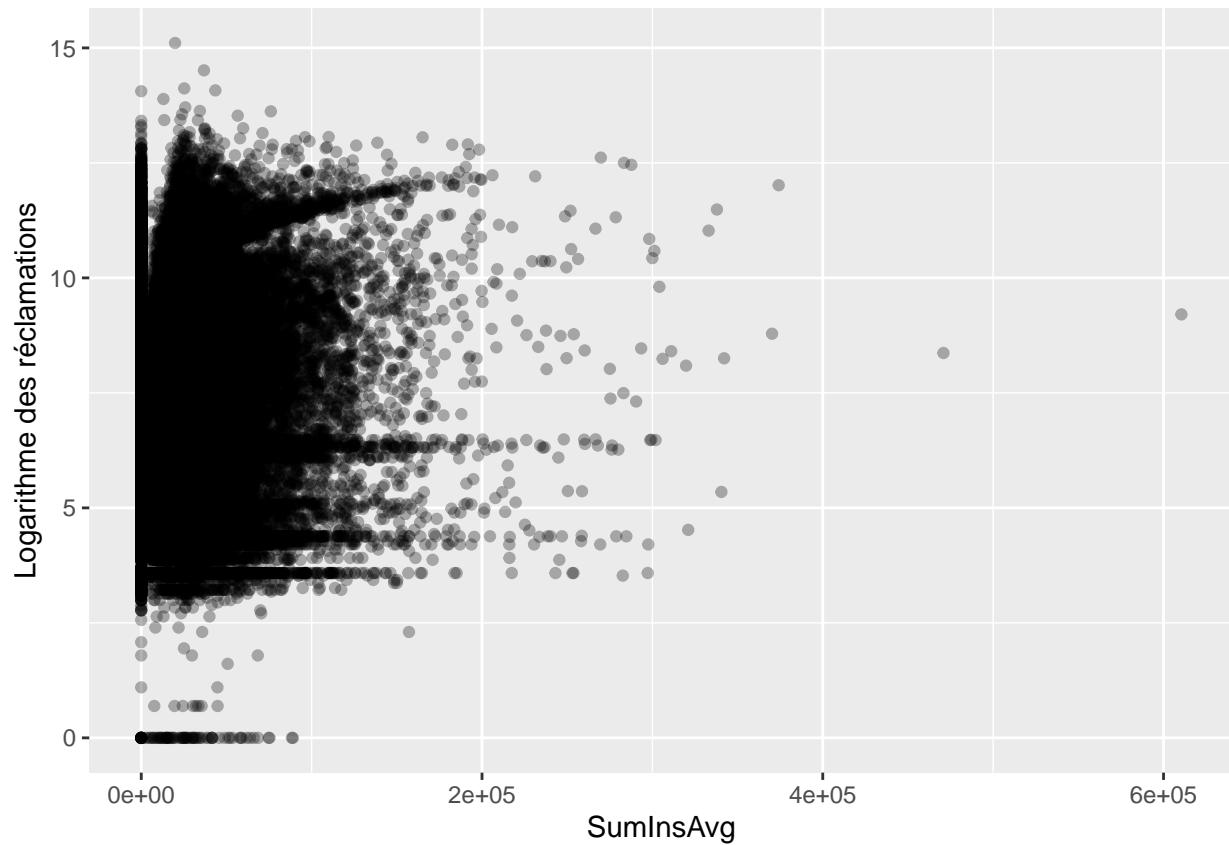
**SumInsAvg**

Figure 9: Logarithme des réclamations en fonction de SumInsAvg

## **Traitement des valeurs manquantes**

## Conclusion