

# Pricing Game # 2

100 % Data Science

Juillet-Novembre 2016

## 1 Description de la Base

Trois bases de données sont disponibles<sup>1</sup>

[http://freakonometrics.free.fr/pricinggame/train\\_contrats.csv](http://freakonometrics.free.fr/pricinggame/train_contrats.csv)  
[http://freakonometrics.free.fr/pricinggame/train\\_sinistres.csv](http://freakonometrics.free.fr/pricinggame/train_sinistres.csv)  
[http://freakonometrics.free.fr/pricinggame/test\\_contrats.csv](http://freakonometrics.free.fr/pricinggame/test_contrats.csv)

ou directement, en R, via

<http://freakonometrics.free.fr/pricinggame/pricinggame.RData>

qui contient les trois bases.

La base **train\_contrats** contient 87 228 polices d'assurance, avec les variables suivantes

**Num\_contrat**: Numéro d'affaire

**IMMAT**: Immatriculation

**Date\_deb\_situ**: Date de début de situation

**ANNEE**: Année de situation

**Date\_fin\_situ**: Date de fin de situation

**Classe\_age\_situ\_cont**: Classe age du contrat

**Type\_apporteur**: Type d'apporteur (agents ou courtiers ou autres)

**creation\_entr**: Indicateur de création entreprise (à 2 si création)

**Mode\_gestion**: Mode de gestion de la flotte

---

<sup>1</sup>Les bases peuvent être importées en R à l'aide de la commande `read.csv()`. Pour tout autre langage de programmation, attention lors de la lecture: la première colonne - sans titre - est un indice

**activité:** Activité du client (niveau contrat)  
**zone :** Zone géographique  
**segment:** Taille de la flotte  
**Fractionnement:** Fractionnement de paiement de la prime  
**Exposition\_au\_risque:** Véhicules année (exposition au risque)  
**Age\_du\_véhicule:** Classe d'âge du véhicule  
**Valeur\_puissance:** Combiné puissance valeur du véhicule  
**franchise:** Classe de franchise de la garantie DTA  
**Nombre\_véhicule:** Nombre de véhicule  
**valeur\_assurée:** Valeur assurée  
**formule:** Formule souscrite  
**Catégorie\_ensemble:** Catégorie du véhicule  
**nombre\_de\_sinistre:** Nombre de sinistres

La base **test\_contrats** contient les mêmes variables, mais sans informations relatives au nombre de sinistres. La base comporte 32 772 polices d'assurance.

La base **train\_contrats** contient des informations relatives à 4 568 sinistres observés

**Num\_contrat:** Numéro d'affaire  
**CHARGE\_SINISTRE:** Charge du sinistre  
**IMMAT:** Immatriculation  
**Date\_Deb\_Situ:** Date de début de situation  
**Date\_Fin\_Situ:** Date de fin de situation  
**IDA:** 1 si sinistre Indemnisation Direct de l'Assuré  
**Responsabilité\_IDA:** Taux de responsabilité de l'IDA  
**ANNEE\_CLOTURE:** Année de cloture du sinistre si le sinistre est cloturé  
**charge sinistre:** Charge sinistre  
**ANNEE:** Année de situation

## 2 Instructions, partie 1 (1er Juillet - 15 Septembre)

Le but est de proposer une prime pure pour les 20 000 contrats de la base **pricing**. Sont attendus

- une base (un fichier **csv**) constituée de deux colonnes (seulement): la première **nu\_affa** sera le numéro de police et la seconde **Premium** contiendra la prime pure proposée. La base sera constituée de xxx lignes (il est impératif de proposer une prime pour *tous* les contrats).
- un rapide descriptif de la méthodologie utilisée, décrivant les variables retenues.

Merci d'indiquer un nom d'équipe pour chaque modèle proposé, et d'envoyer le fichier **csv** avec un rapide descriptif à **PricingGame@institutdesactuaires.com**, avant le **15 Septembre**.

## 3 Instructions, partie 2 (20 Septembre - 20 Octobre)

Le 20 Septembre, tous les participants recevront les primes de 2 de leurs compétiteurs, au moins.

Les participants auront alors la possibilité de réviser (ou pas) la prime qu'ils proposent. Un fichier **csv** révisé - et un rapide descriptif de la méthodologie adoptée - devront être envoyés à **PricingGame@institutdesactuaires.com** pour le **20 Octobre**, au plus tard.

## 4 Règles du Jeu

Une fois les primes collectées, nous fonctionnerons comme un agrégateur de prix: chacun des xxx assurés se verra affecter une compagnie selon une méthode qui sera basée sur une sélection aléatoire parmi les primes les moins chères.

Les organisateurs normaliseront les prix pour éviter les stratégies de dumping. En l'occurrence, la somme des primes proposées par un assureur sera égalisée à xxxxx (par un facteur d'inflation appliqué uniformément).

Les résultats seront présentés lors de la journée 100% Data Science organisée à Paris (voir également sur <http://freakonometrics.hypotheses.org>).

Pour plus d'information, **PricingGame@institutdesactuaires.com** ou **@freakonometrics** sur Twitter.