

21^e Congrès des Actuaires – Éclairer l'avenir dans un monde en transition

Représentation moderne et dynamique de la data au service de la prise de décision

23 juin 2022 – 12h00-12h45

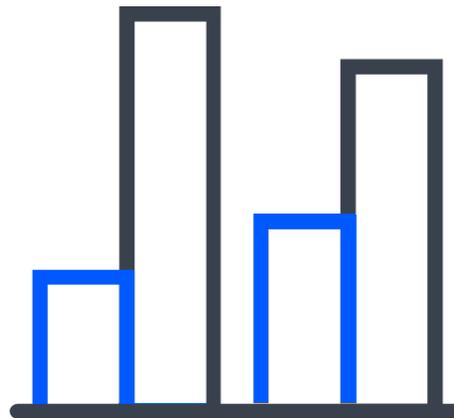


Rémi BELLINA
Chief Data Scientist & Manager Actuariat
Milliman



Guillaume BERAUD-SUDREAU
Responsable R&D & Actuariat
AKUR8

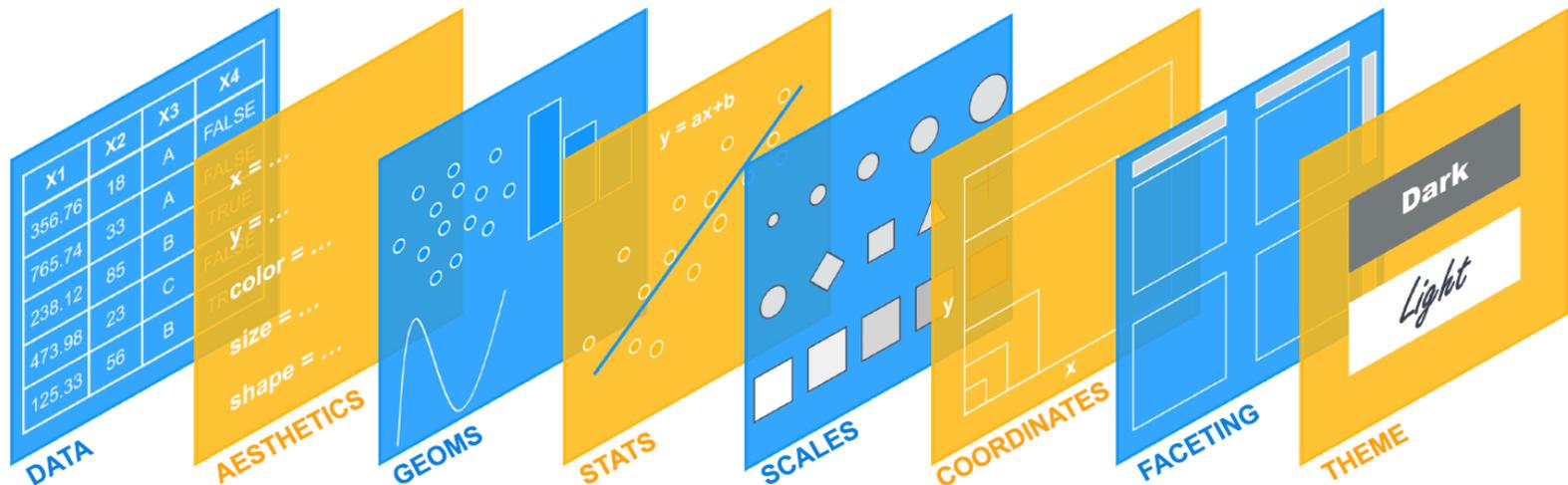
Les enjeux autour de la **data viz'**



Les enjeux autour de la data viz'

Les bases

- La **data viz'** est « l'art » ou plutôt « **la science** » de représenter les données pour la transformer en **information pertinente** et **explicite**.
- La source **data-to-viz.com** résume parfaitement sa richesse, tout comme l'illustration de la « **grammaire graphique** » (cf. ggplot2).

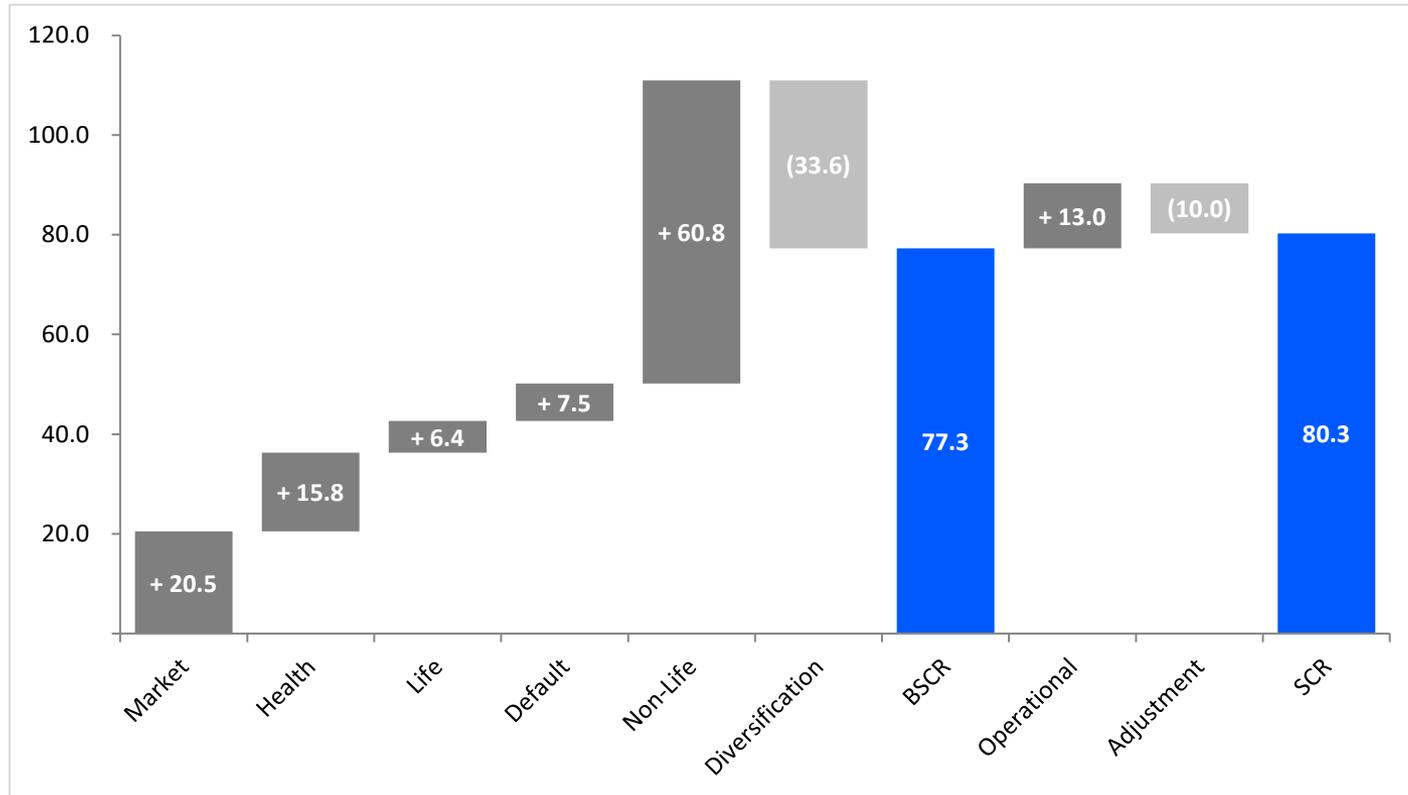


- La data viz' occupe une place de plus en plus importante, obéit à un ensemble de **bonnes pratiques** (choix du support, *data-ink ratio*, *highlight*, choix des couleurs, **éthique**, ...),
- et surtout **se modernise**, en faveur d'une **aide efficace à la décision**.

Les enjeux autour de la data viz'

Exemple introductif (1/2)

- Ci-dessous une représentation de type *waterfall* à partir d'Excel (volumes en formule standard Solvabilité 2).



→ Est-elle **simple à réaliser** ?

Les enjeux autour de la data viz'

Exemple introductif (2/2)

- La réponse dépend des outils à disposition

Version récente d'Excel ([rapide](#))

| Name | Market | Health | Life | Defa |
|-------------------------|--------|--------|-------|------|
| Actual Amount | 20.5 | 15.8 | 6.4 | 7.5 |
| "Top" Value | 20.5 | 36.3 | 42.6 | 50.: |
| Base Amount (invisible) | | 20.5 | 36.3 | 42.1 |
| Base Amount (blue) | | | | |
| Up (red) | + 20.5 | + 15.8 | + 6.4 | + 7. |
| Down (green) | | | | |

Anciennes versions d'Excel ([complexe](#))

| Name | Market | Health | Life | Defa |
|-------------------------|--------|--------|-------|------|
| Actual Amount | 20.5 | 15.8 | 6.4 | 7.5 |
| "Top" Value | 20.5 | 36.3 | 42.6 | 50.: |
| Base Amount (invisible) | | 20.5 | 36.3 | 42.1 |
| Base Amount (blue) | | | | |
| Up (red) | + 20.5 | + 15.8 | + 6.4 | + 7. |
| Down (green) | | | | |

=MAX(...,...)-
ABS(...-...)

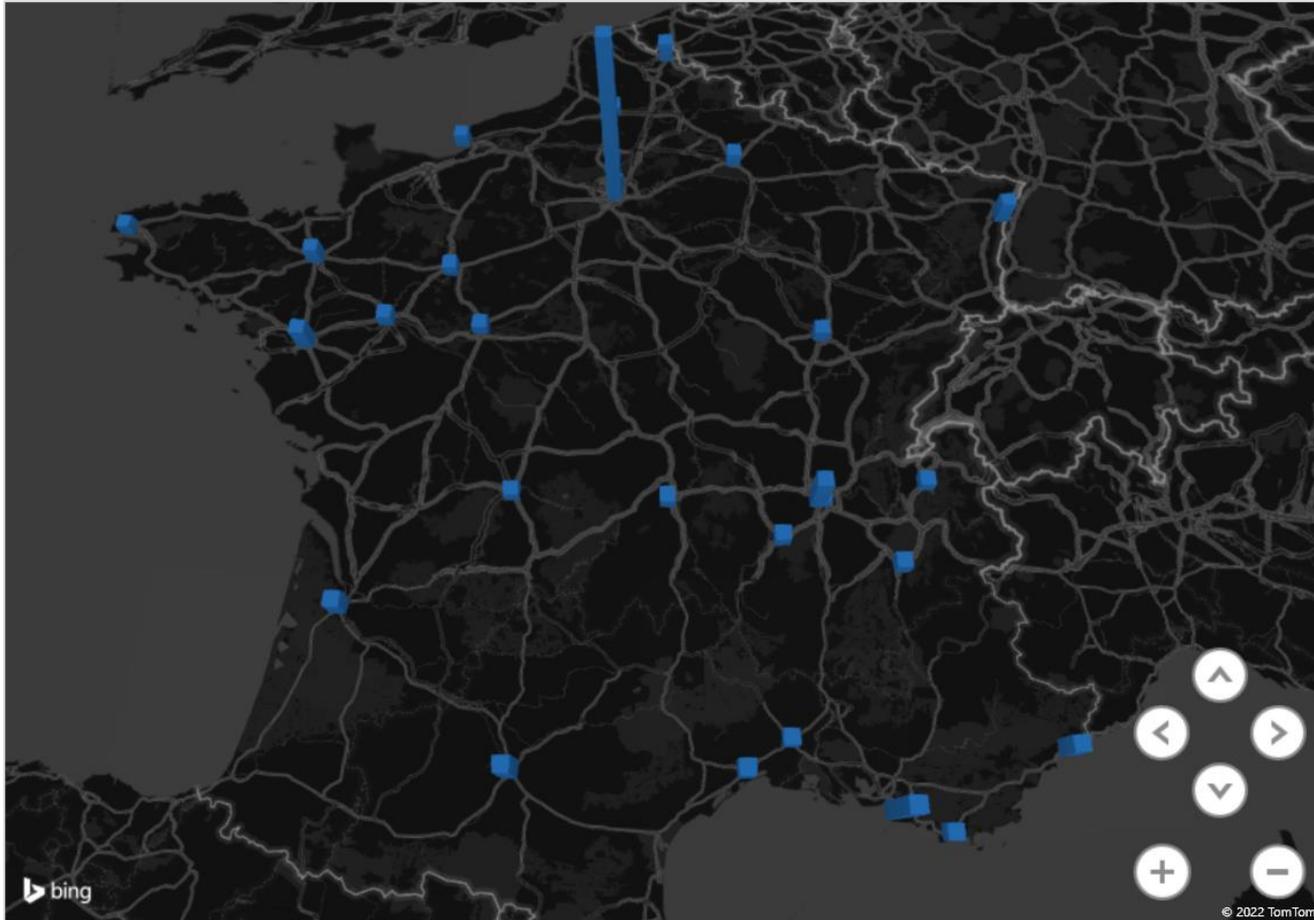
→ La data viz' est un **sujet profond**, qui peut être complexe avec des outils non adaptés.

→ Cela **évolue vite**, notamment avec l'Open Source ou les outils BI.

Les enjeux autour de la data viz'

Exemple bonus

- Un autre exemple : **quel outil** a généré cette représentation ?

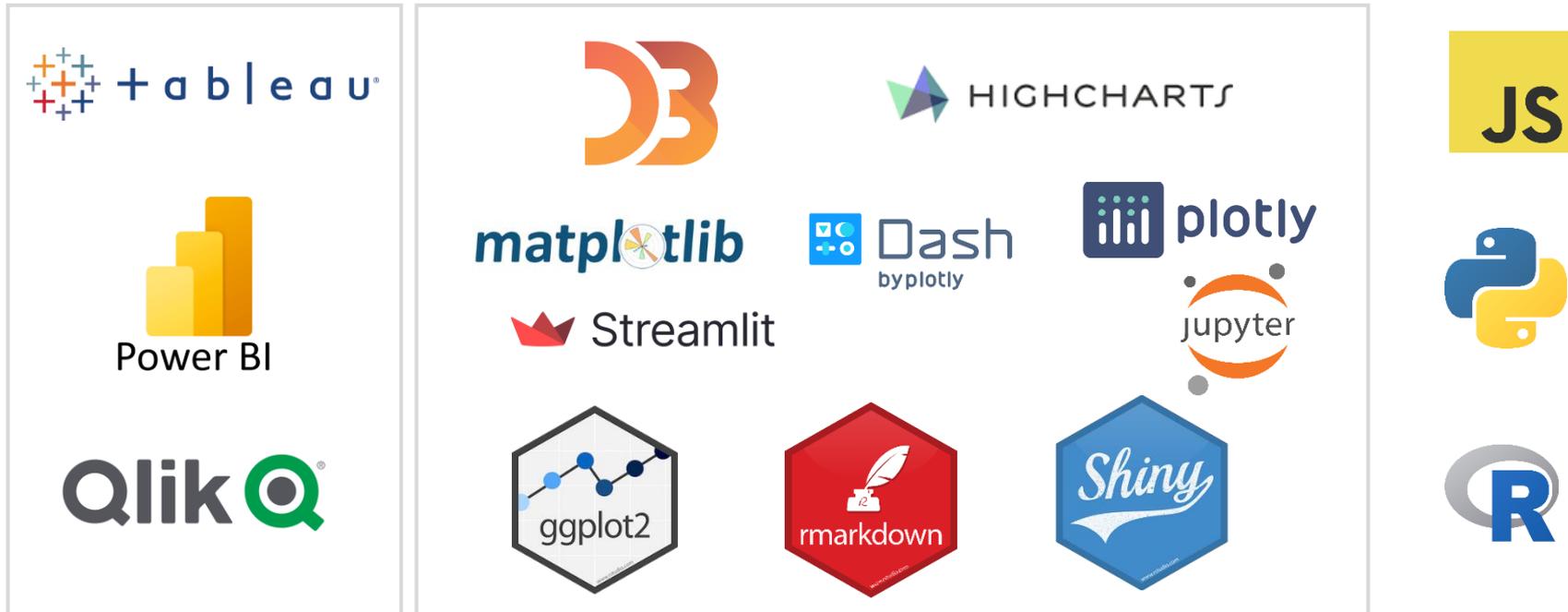


Top 30 des villes en France par nombre d'habitants

Les enjeux autour de la data viz'

Les outils

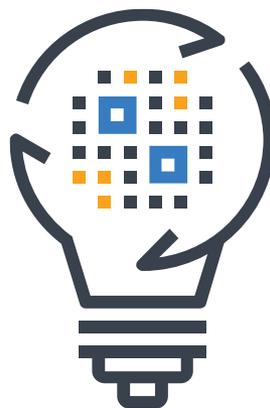
- De **nombreux outils** sont désormais à la disposition des actuaires et statisticiens, mais aussi des profils moins orientés « data ». (liste **non exhaustive**)



- La data viz' peut être ainsi échelonnée selon les besoins, et aller au-delà du *simple* graphique avec la mise en place de **rappports**, de **dashboards** et d'**applications web**.

La **data viz'** par l'exemple

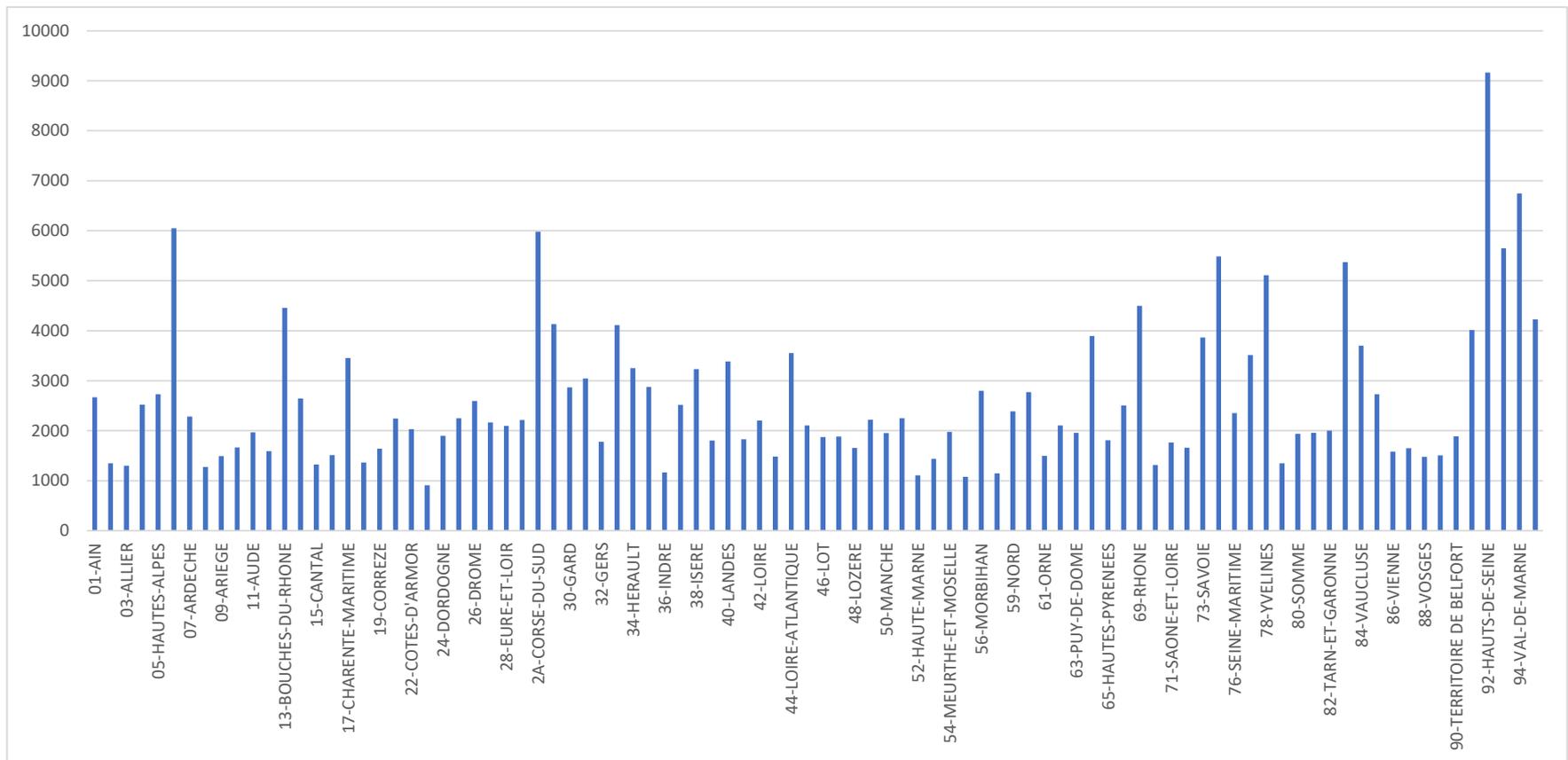
Cas d'étude : Exploration de données externes



La data viz' par l'exemple

Du simple graphique ...

- Voici une représentation du **prix au m² moyen des maisons** (exclusion de cas atypiques) en France métropolitaine hors Paris à partir des données du Ministère de l'économie, des finances et de la relance (Demande de Valeurs Foncières sur data.gouv.fr). Elle a été faite **rapidement** sous **Excel**.



Source DVF <https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/>

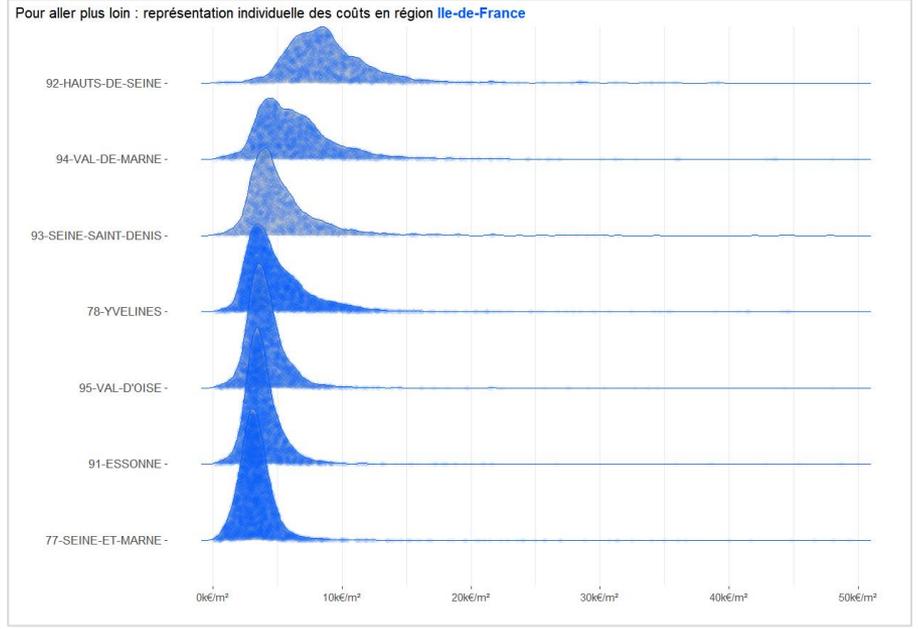
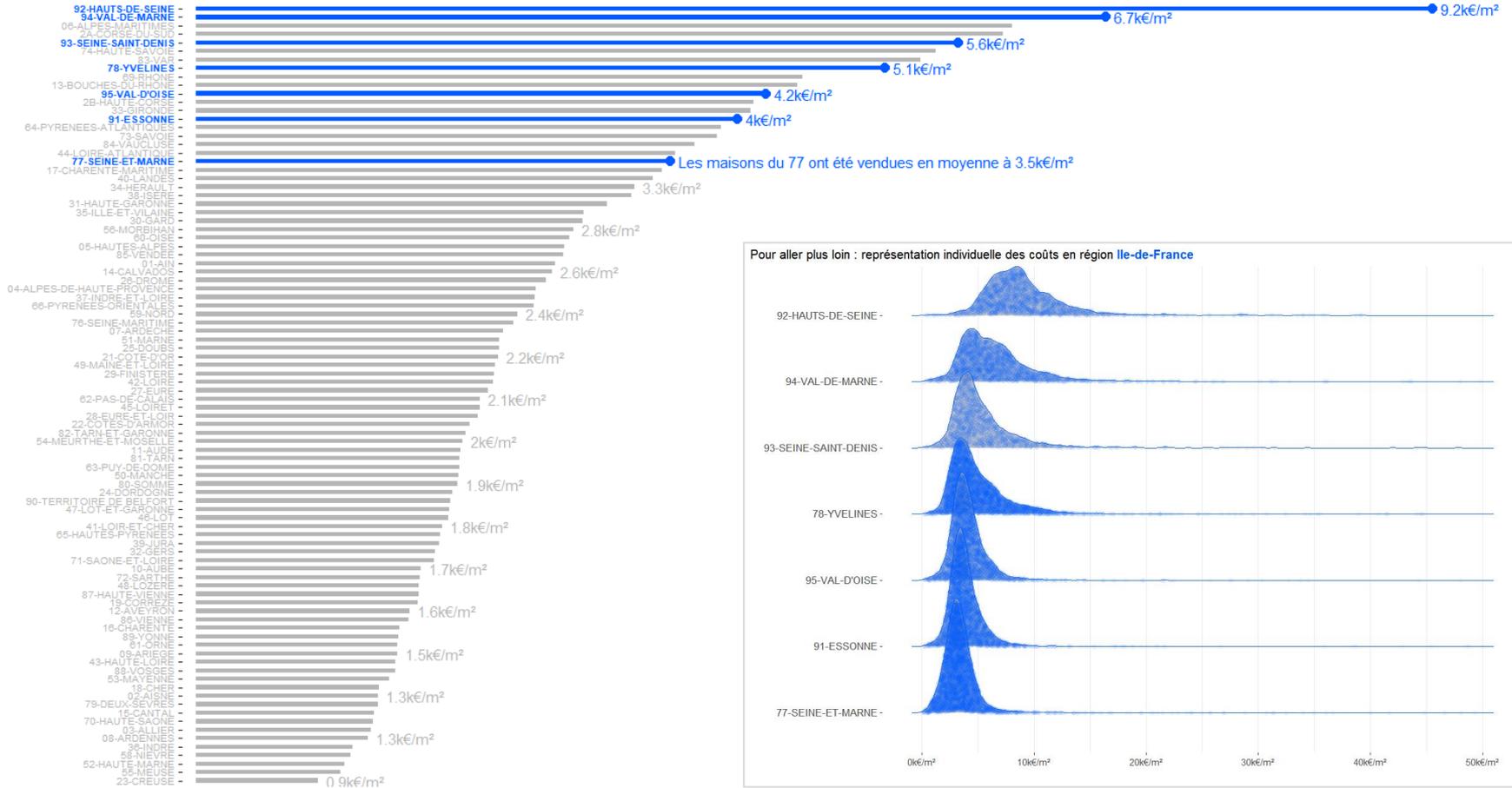
La data viz' par l'exemple

... au storytelling

- La même représentation avec une **valorisation** de l'information.

Prix de vente moyen (m²) des maisons (zoom sur la région Ile-de-France)

Périmètre : Année 2021, France métropolitaine, hors Paris, Maisons, exclusion de cas atypiques



Source DVF
<https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/>

La data viz' par l'exemple

Ajout d'une dimension temporelle

- Test N°1 : le graphique **animé** (*race chart* sous Power BI)
- Test N°2 : le passage à la **3D** (*rayshader* [<https://github.com/tylermorganwall/rayshader>])



La data viz' par l'exemple

Ajout d'une dimension temporelle

- Test N°3 : l'utilisation des **possibilités géométriques** (*ggplot2*)

Prix de vente moyen (m²) des maisons (zoom sur la région Ile-de-France)

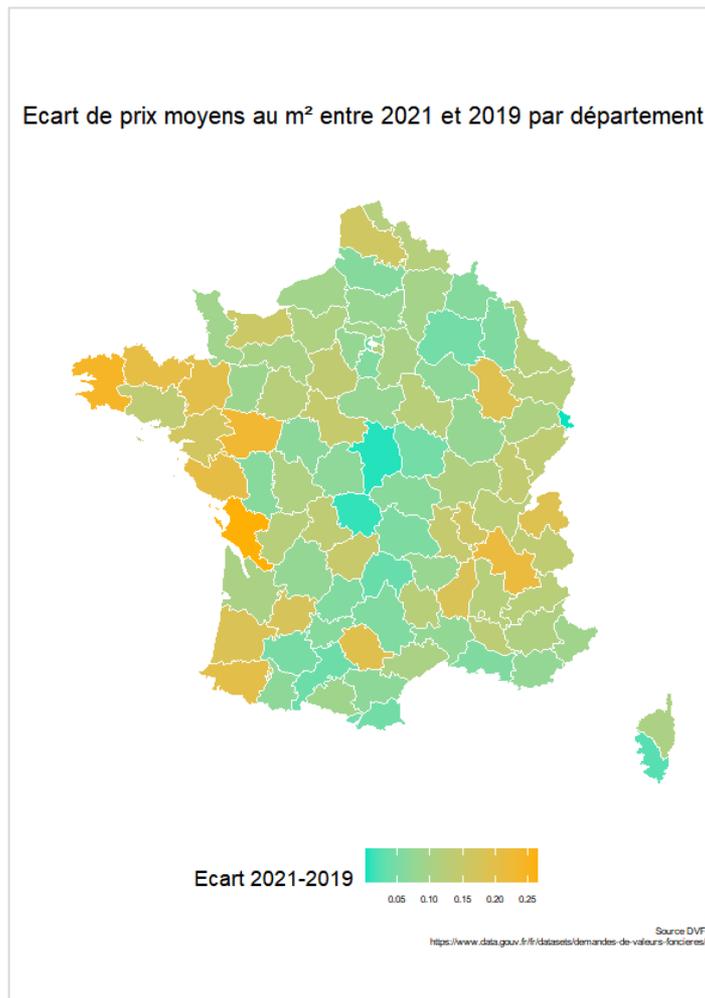
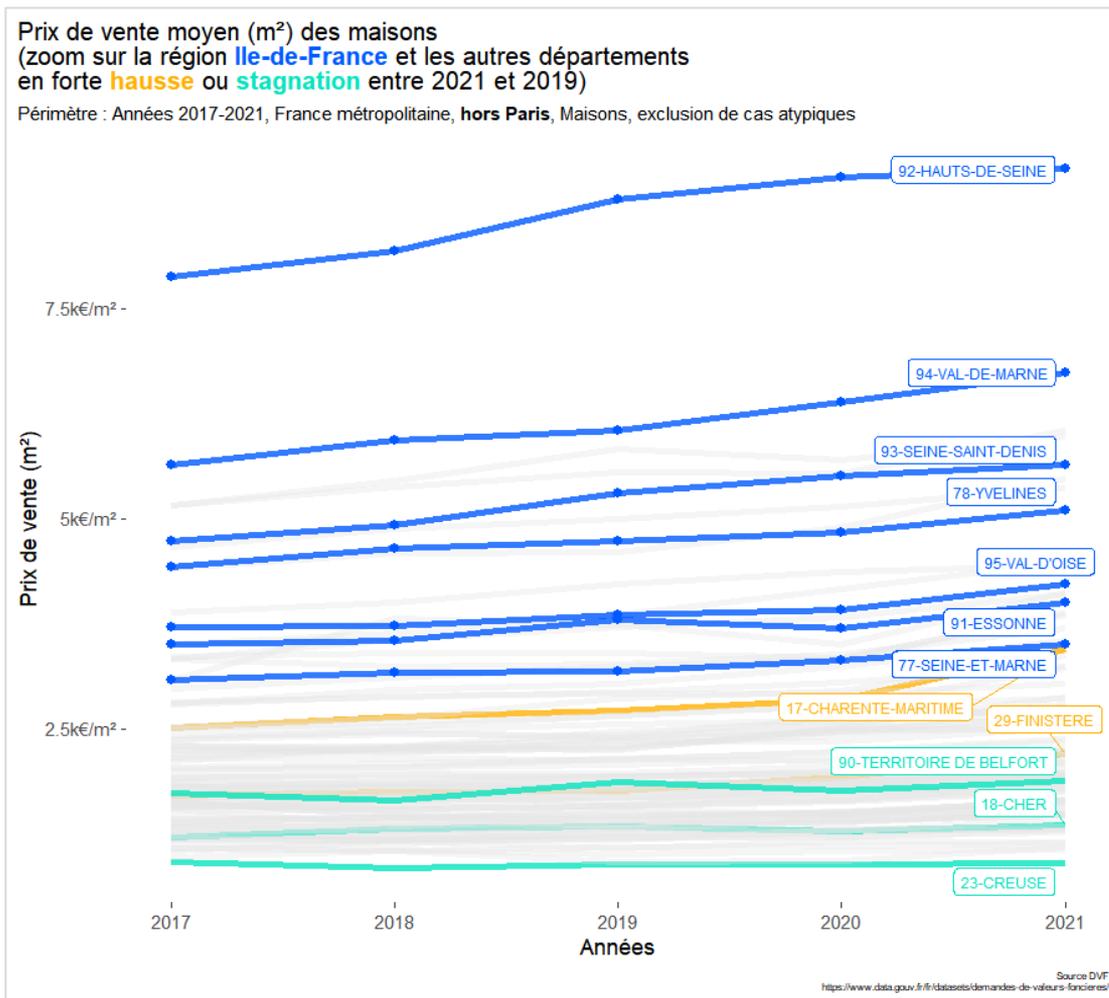
Périmètre : Années 2017-2021, France métropolitaine, hors Paris, Maisons, exclusion de cas atypiques



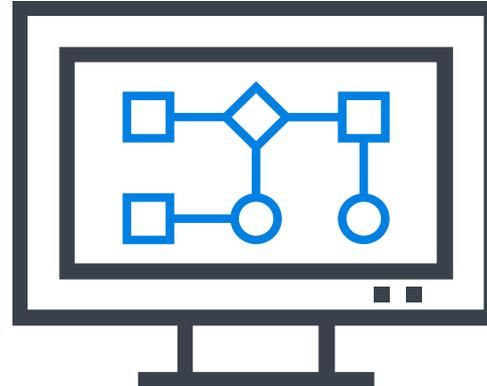
La data viz' par l'exemple

Ajout d'une dimension temporelle

- Proposition N°1 : retour à des concepts **simples** (*ggplot2*)



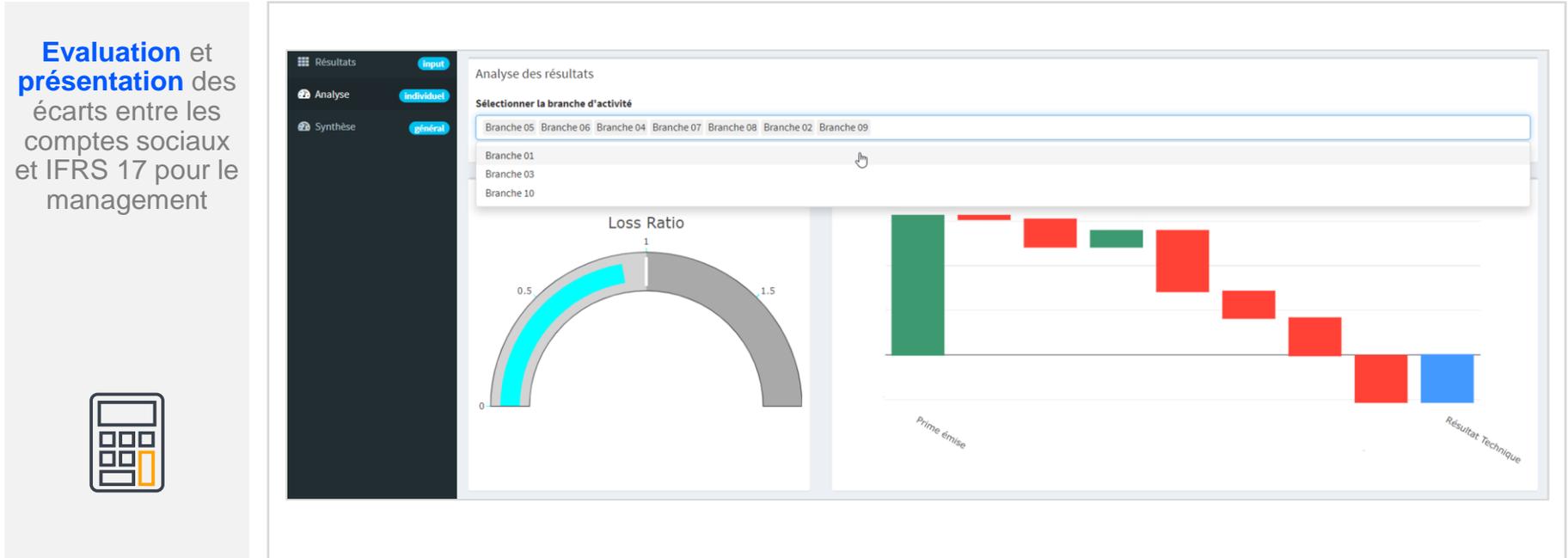
Quelques cas d'étude autour de la **data viz'**



Quelques cas d'étude autour de la data viz'

Illustrations

- Le **choix de l'outil** dépend de la cible et des utilisateurs.



Excel (Power)
R Shiny

Mise en œuvre d'une application R-Shiny permettant d'analyser les résultats des projections IFRS 17

Quelques cas d'étude autour de la data viz'

Illustrations

- Le **choix de l'outil** dépend de la cible et des utilisateurs.

Evaluation et **présentation** des écarts entre les comptes sociaux et IFRS 17 pour le management

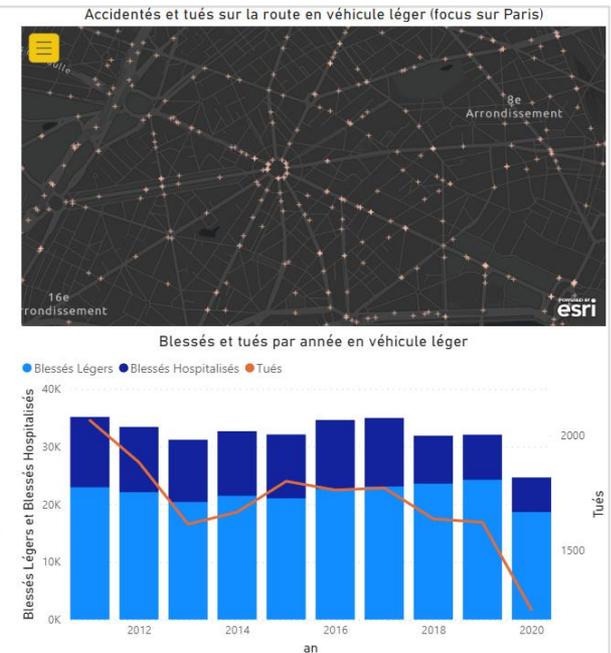
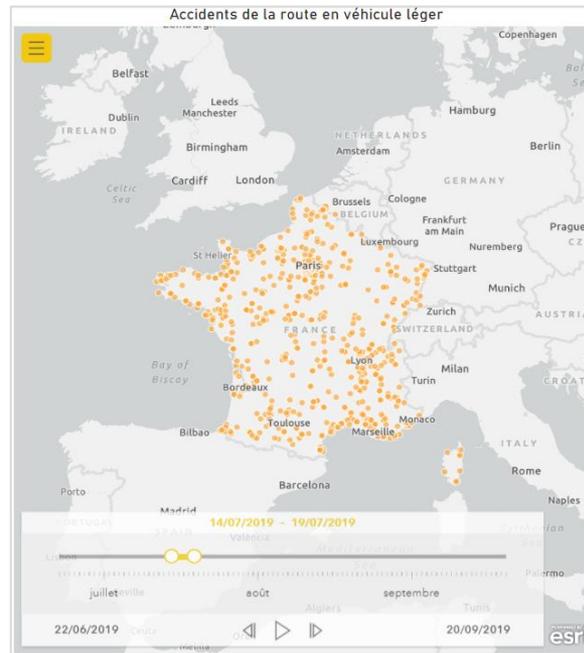


Excel (Power)
R Shiny

Exploration d'une **donnée** externe **nouvelle** pour échange au sein des équipes et évaluation de son potentiel



Ggplot2
Power BI

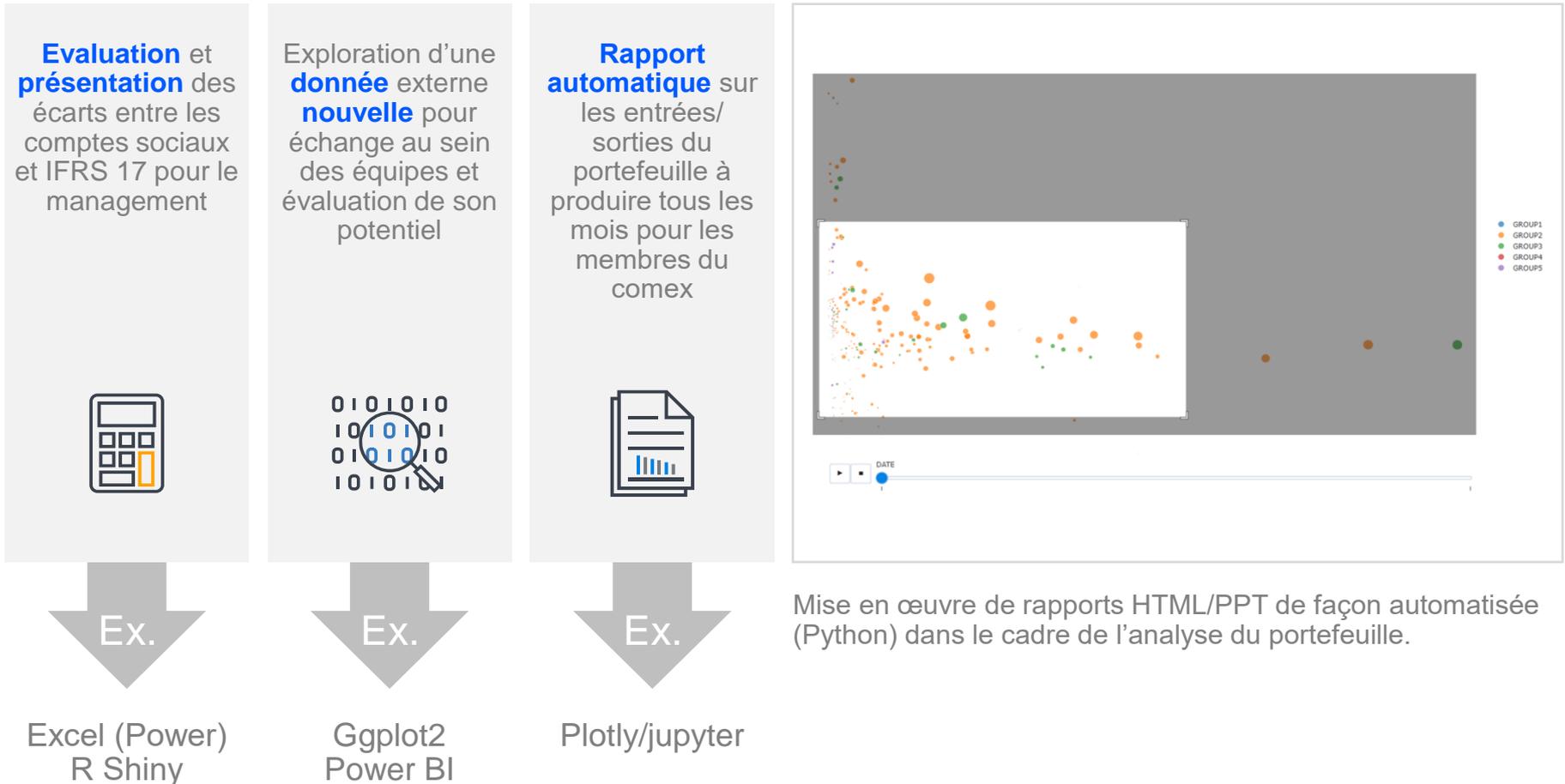


Dashboard Power BI pour l'analyse de la BAAC (data.gouv.fr), représentation des accidents de la route pour les véhicules légers.

Quelques cas d'étude autour de la data viz'

Illustrations

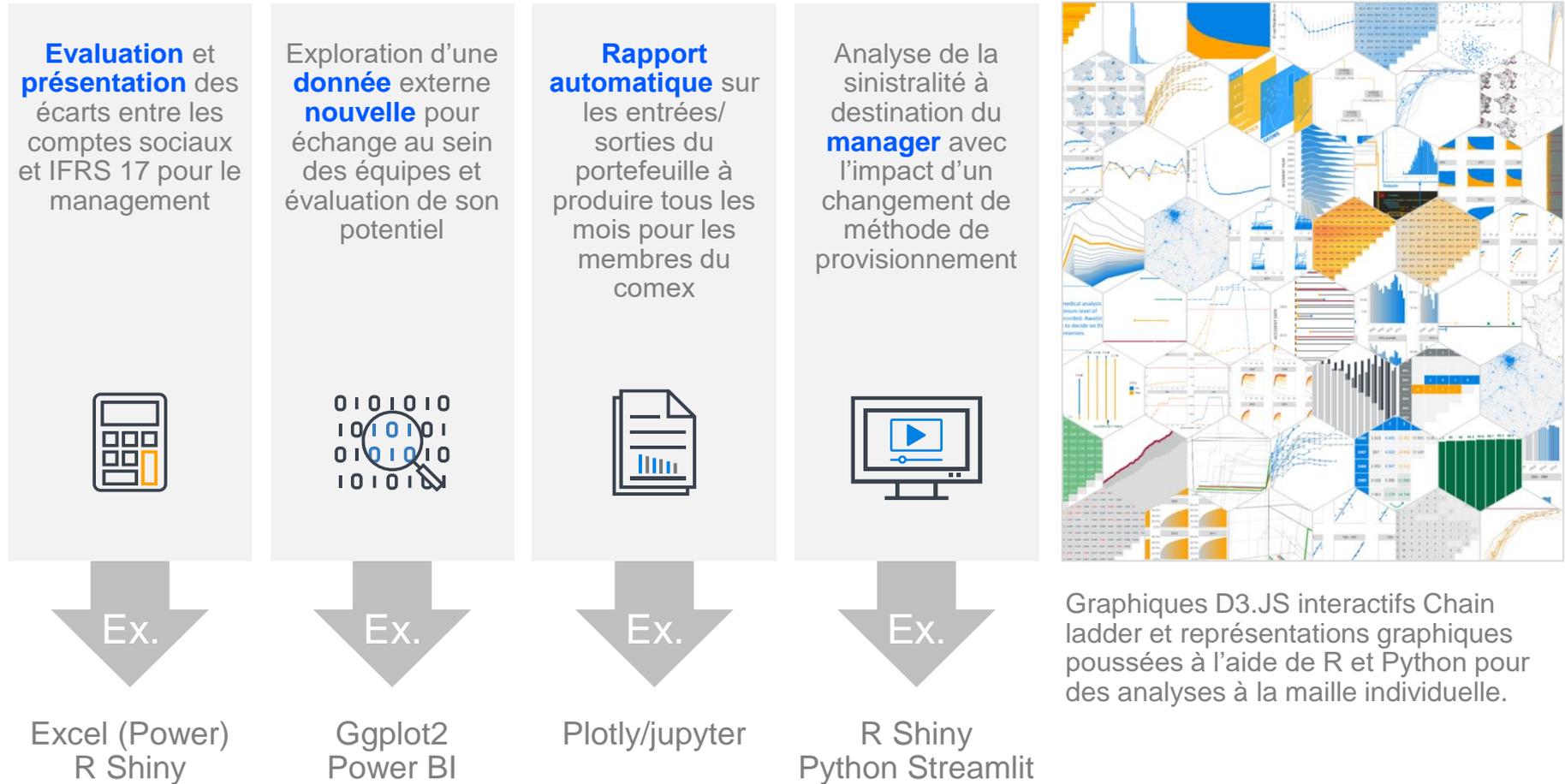
- Le **choix de l'outil** dépend de la cible et des utilisateurs.



Quelques cas d'étude autour de la data viz'

Illustrations

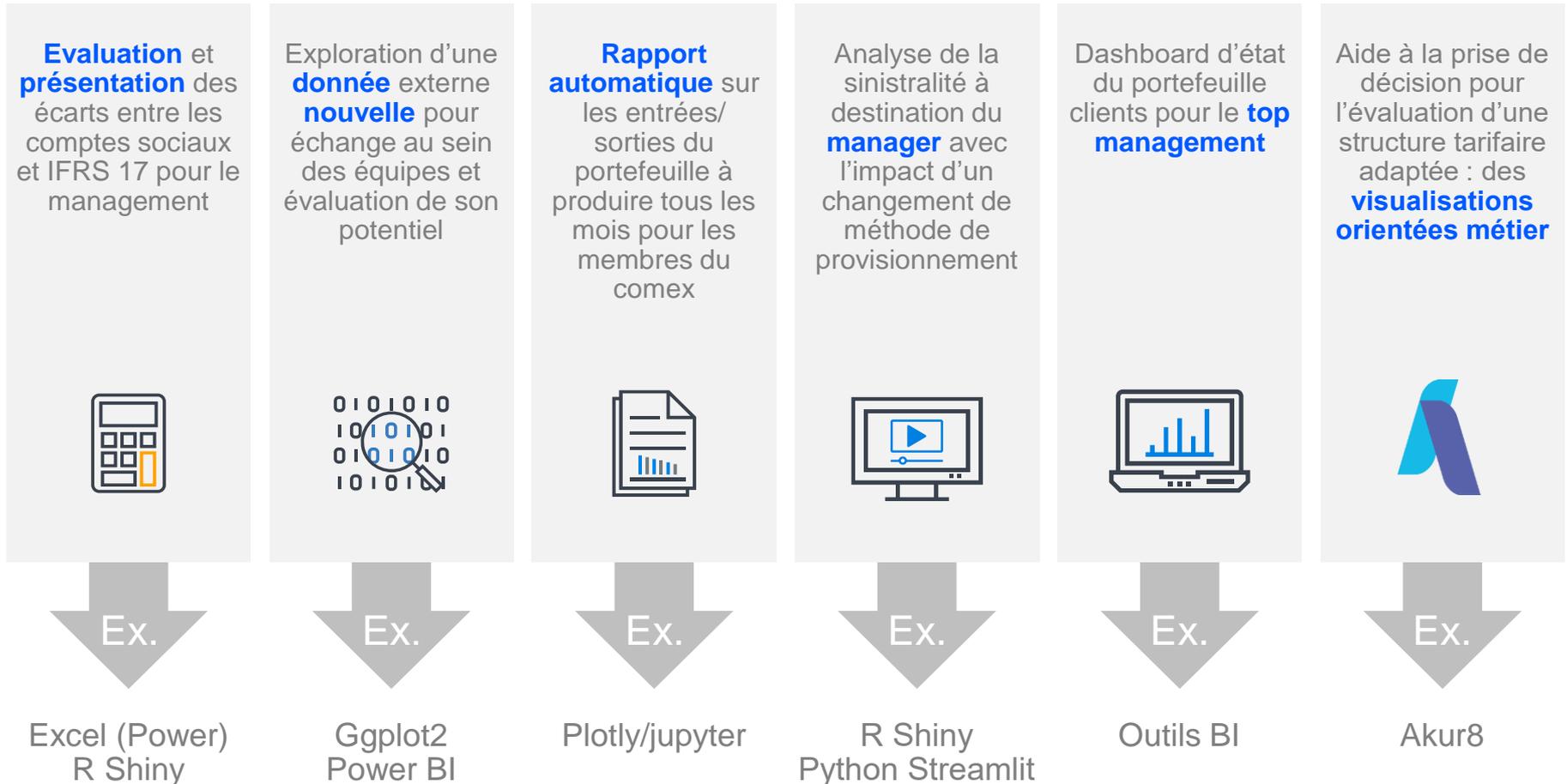
- Le **choix de l'outil** dépend de la cible et des utilisateurs.



Quelques cas d'étude autour de la data viz'

Illustrations

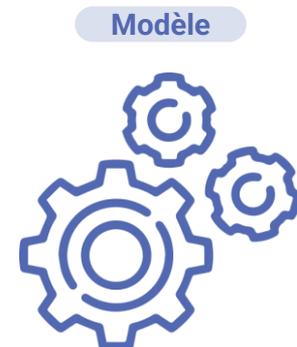
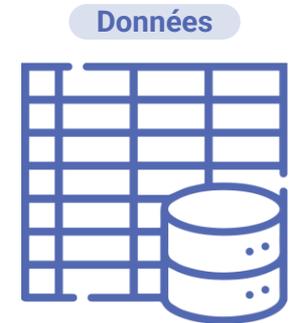
- Le **choix de l'outil** dépend de la cible et des utilisateurs.



Visualisation de données : l'exemple de la **sensibilité aux prix**

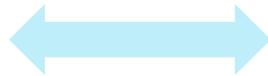
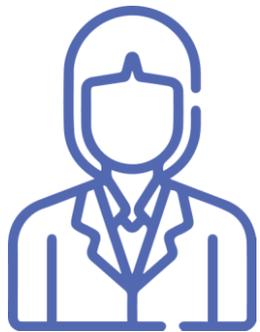
Visualisation et Tarification

Lire directement les tables ?

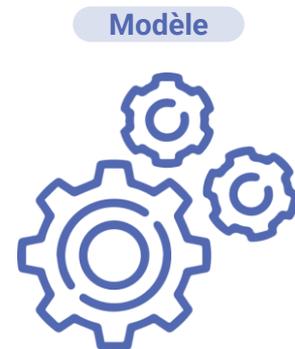
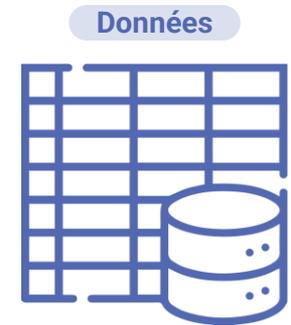
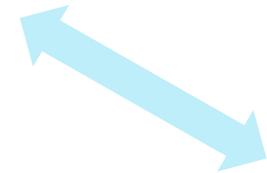


Visualisation et Tarification

Utiliser un score ?



GINI = 23%



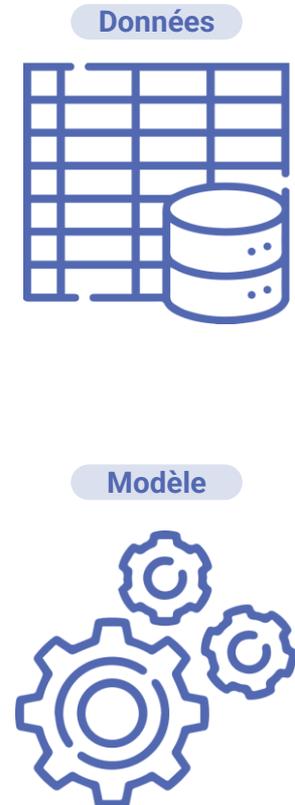
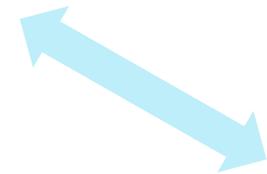
Visualisation et Tarification

Utiliser beaucoup de scores ?



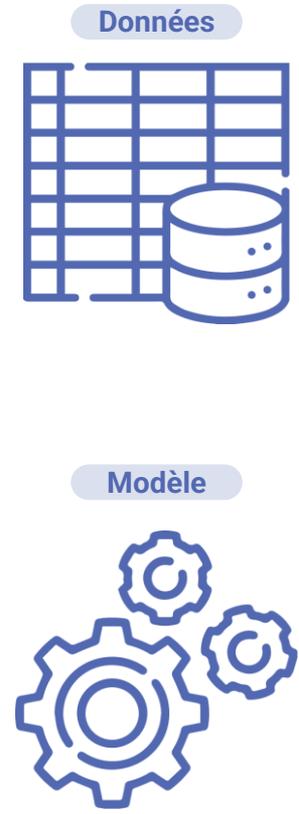
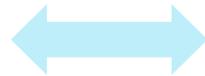
GINI = 23%
EDR = 2.1%
RMSE = 0.13
MAE = 0.22
MRE = 0.23

...



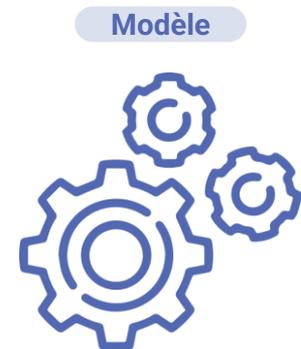
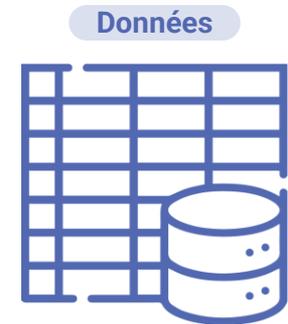
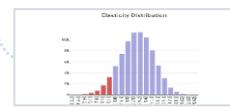
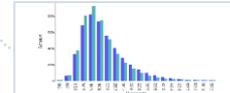
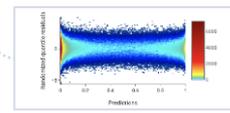
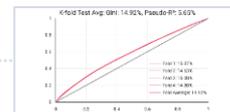
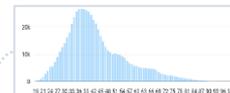
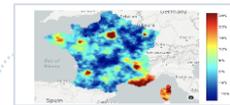
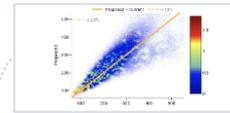
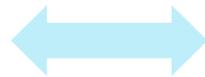
Visualisation et Tarification

La solution : le « Graphique » ?



Visualisation et Tarification

La solution : le « Graphique » ?



Synthèse

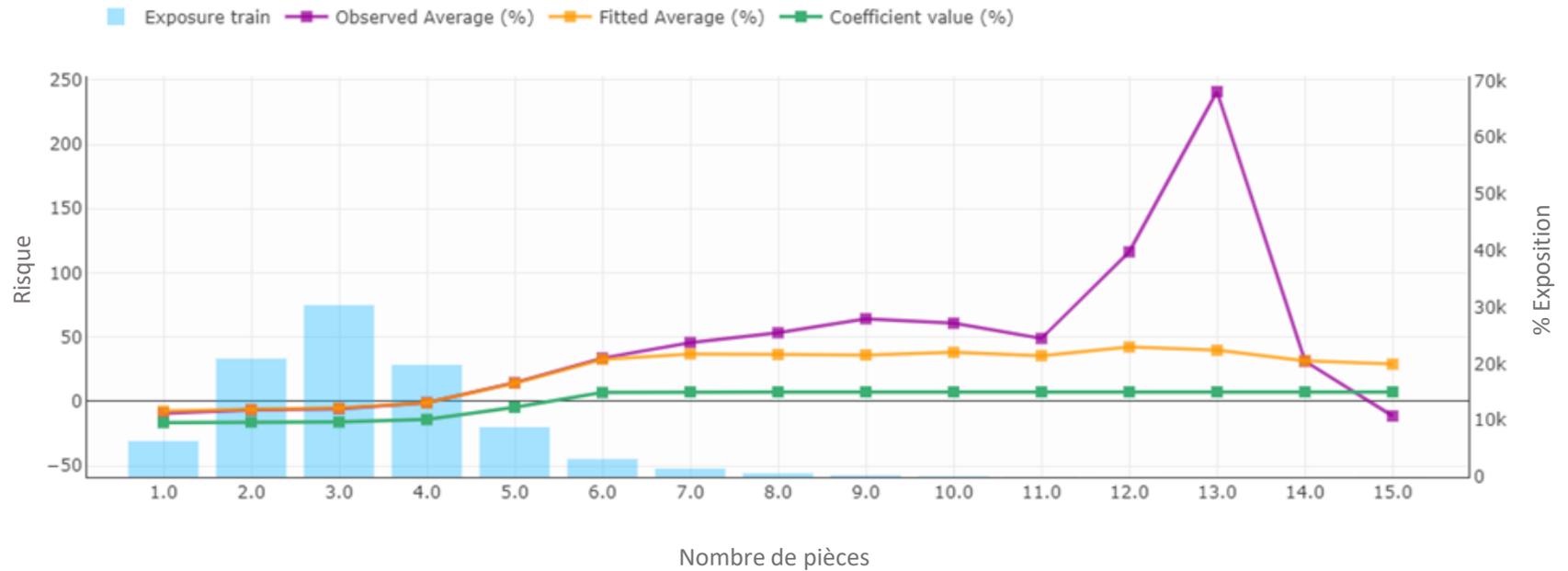
- Les visualisations doivent être présentes à chaque étape de la tarification.
- Le modélisateur et le décideur ne doivent **jamais perdre contact avec les données**.
- Les graphiques doivent être **adaptés au public** : modélisateur expert, décideur, audience large... « recycler » les graphiques d'un public à un autre est rarement efficace.
- De même, le graphique doit être **adapté au type de données, de modèles ou de KPI** représenté.
- Quelle que soit la personne qui visualise le graphique, celui-ci doit avoir un but : détecter des anomalies dans les données, une erreur de modèle, passer un message, proposer une action tarifaire...
- Construire un bon graphique est un **investissement**. En particulier, les graphiques de visualisation, intermédiaires dans la modélisation, sont souvent sous-estimés. Lors de la modélisation, **chaque graphique visualisé évite une erreur potentielle**.

Exemple : Visualiser le risque...



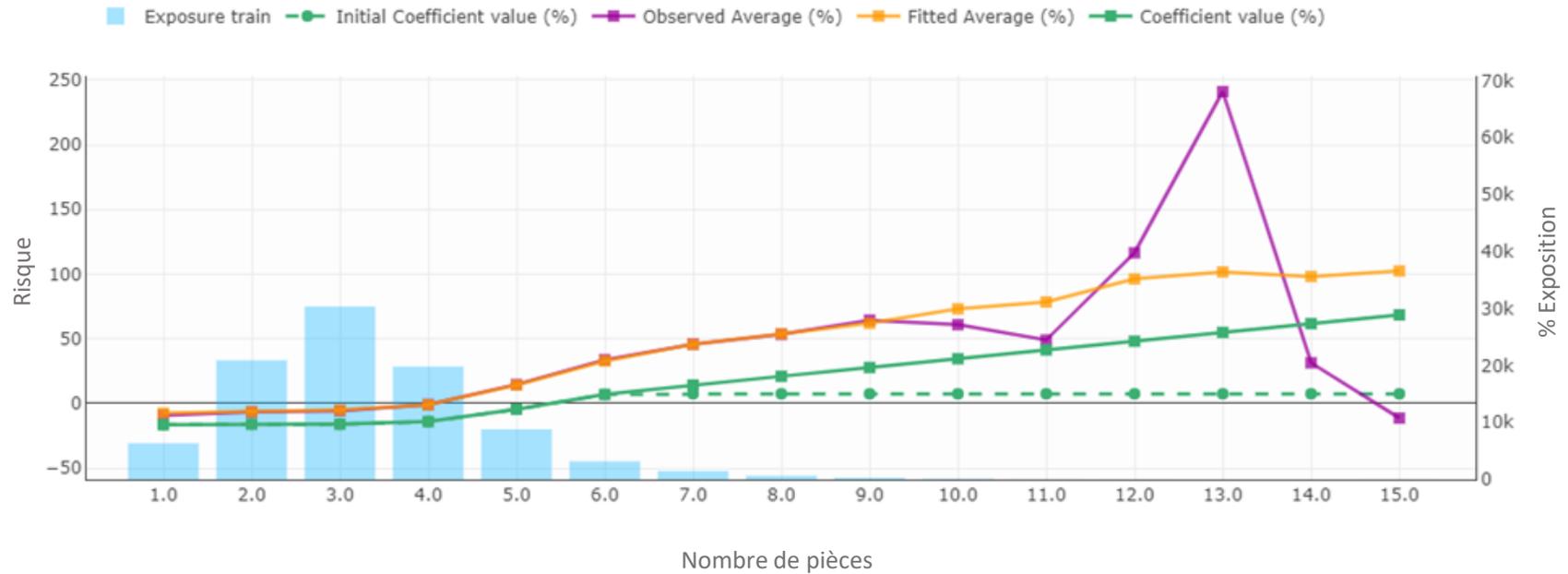
Visualisation des risques

L'importance d'une vue globale



Visualisation des risques

L'importance d'une vue globale

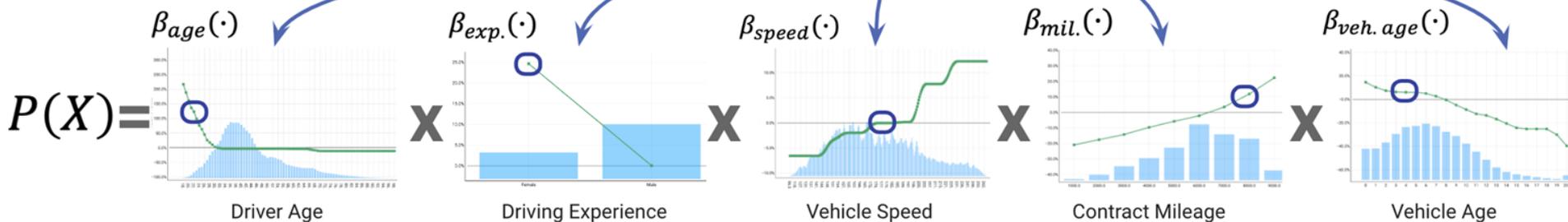


Visualisation des risques

Cette visualisation reflète le modèle utilisé !

- Le modèle est multiplicatif et peut être visualisé variable par variable.

$$Predictions = \prod_i \beta_i(X_i)$$

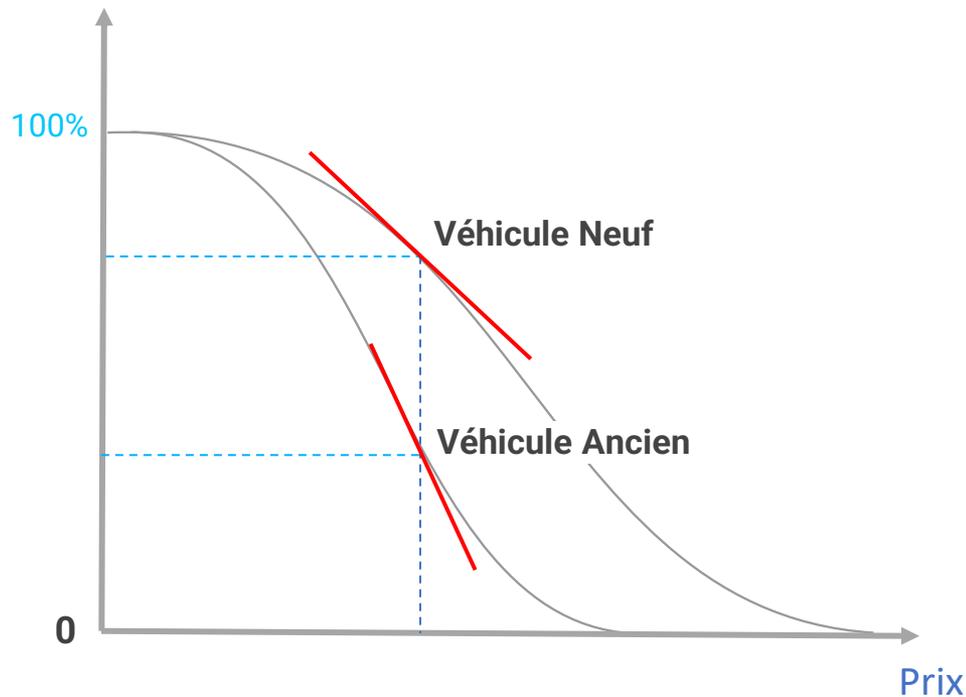


... et la **demande** !

Visualisation de la demande

Qu'est ce qu'un modèle de Demande ?

Probabilité d'Achat



Véhicule Neuf

- Taux de Conversion : Élevé
- Sensibilité aux Prix : Faible

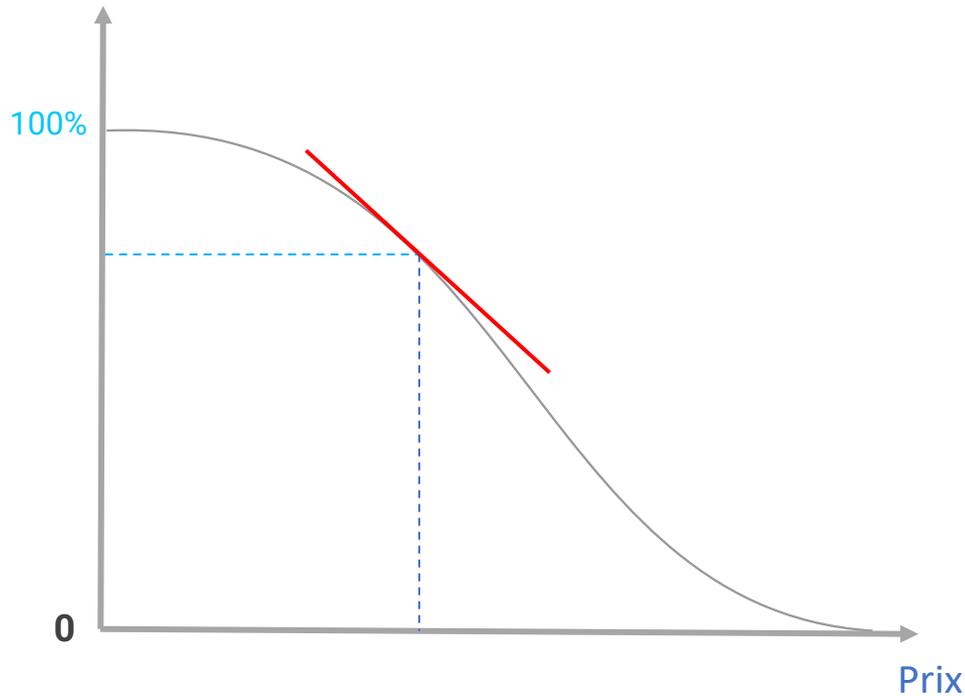
Véhicule Ancien

- Taux de Conversion : Faible
- Sensibilité aux Prix : Élevée

Visualisation de la demande

Comment le valider ?

Probabilité d'Achat



Ce modèle est-il bon ?

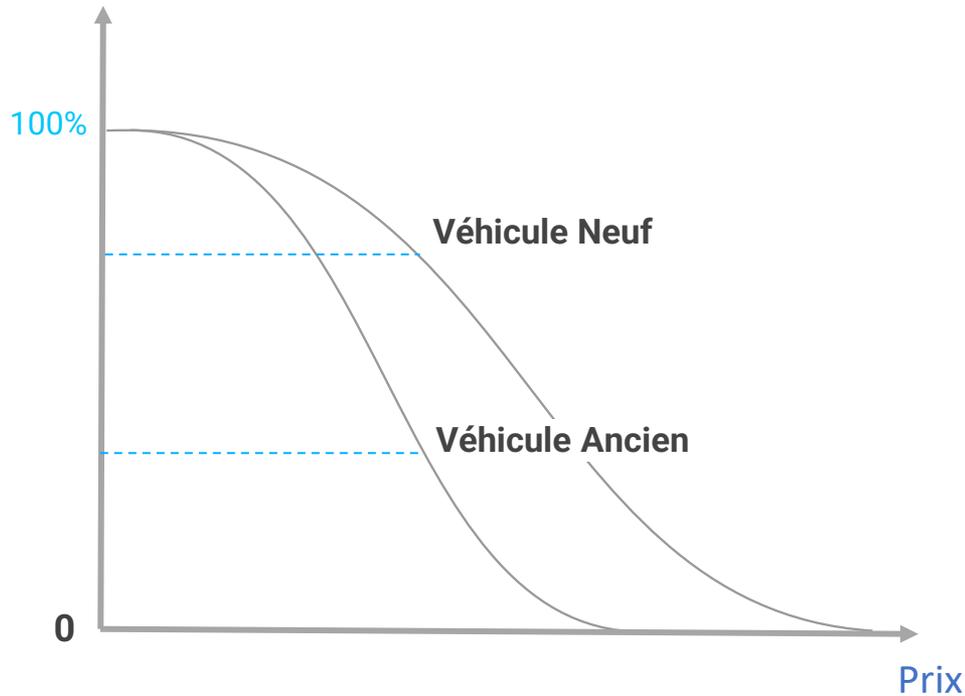
Le taux de conversion est-il fiable ?

La pente est-elle fiable ?

Visualisation de la demande

Visualiser le modèle de conversion

Probabilité d'Achat

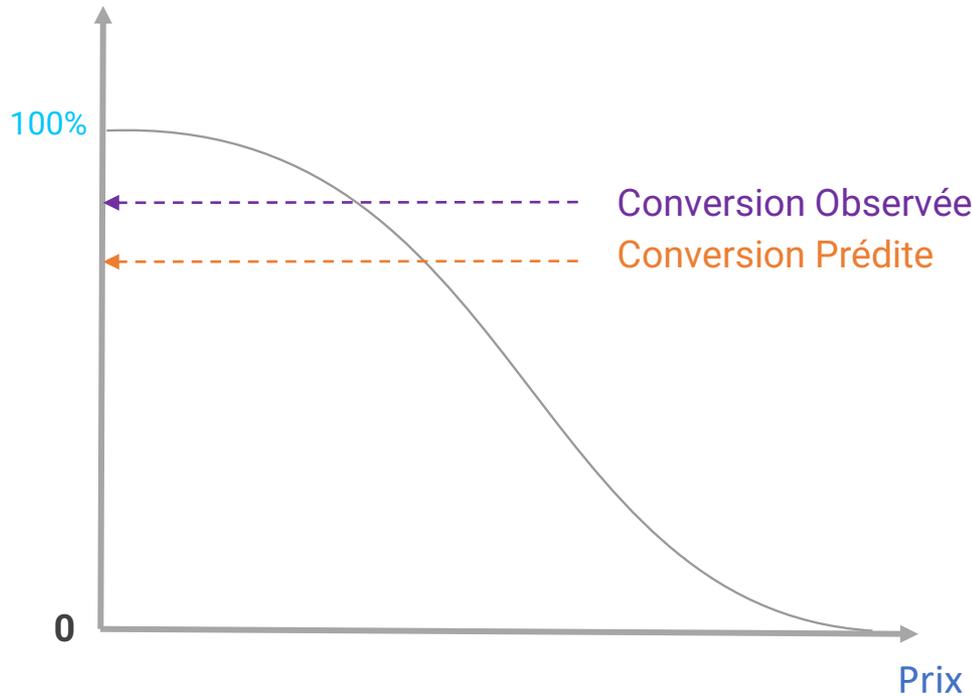


- ✓ Pouvoir prédictif
- ✓ Visualisation directe

Visualisation de la demande

Valider le modèle de conversion

Probabilité d'Achat

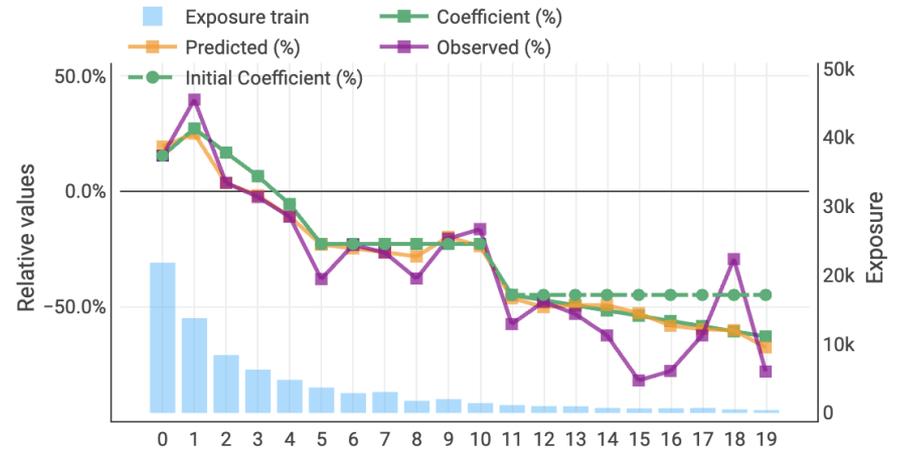
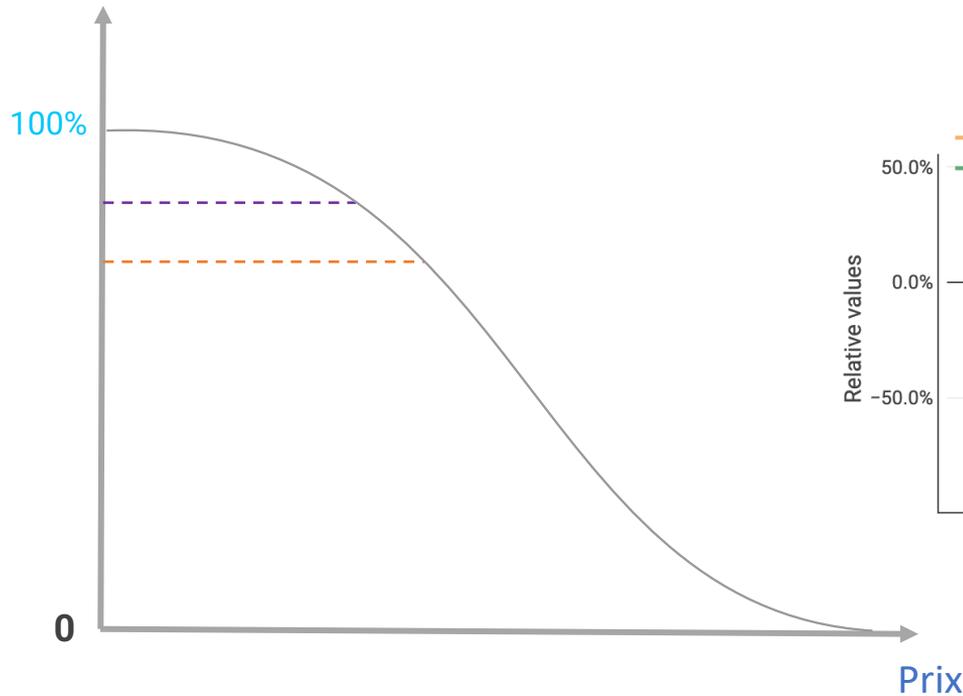


➔ Pouvoir prédictif

Visualisation de la demande

Valider le modèle de conversion

Probabilité d'Achat

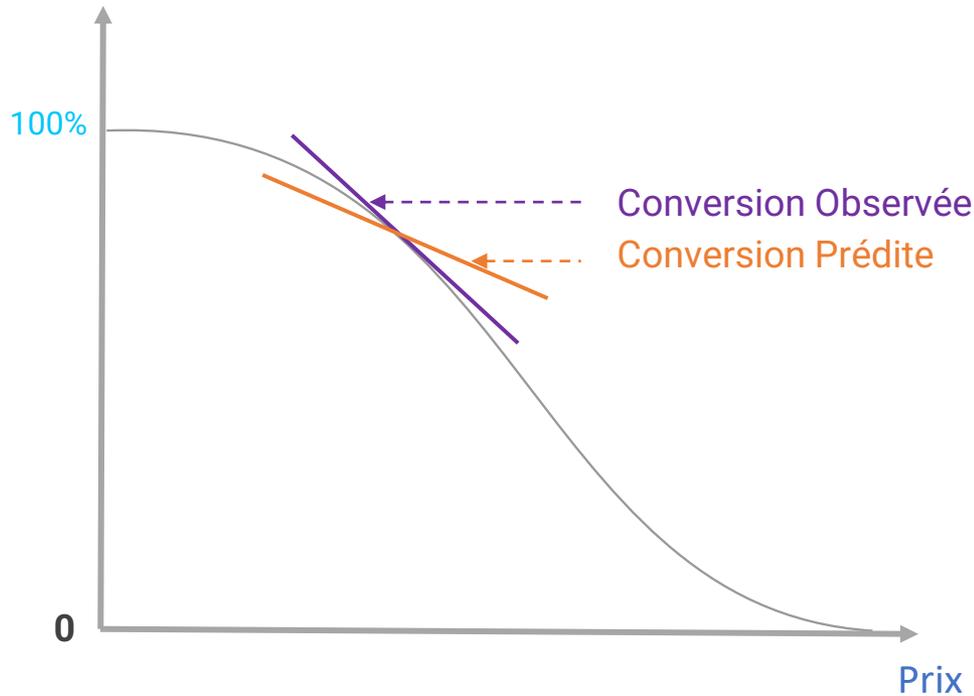


Visualisation de la demande

Pourquoi visualiser un modèle de demande ?

- On ne visualise pas un point mais une pente

Probabilité d'Achat



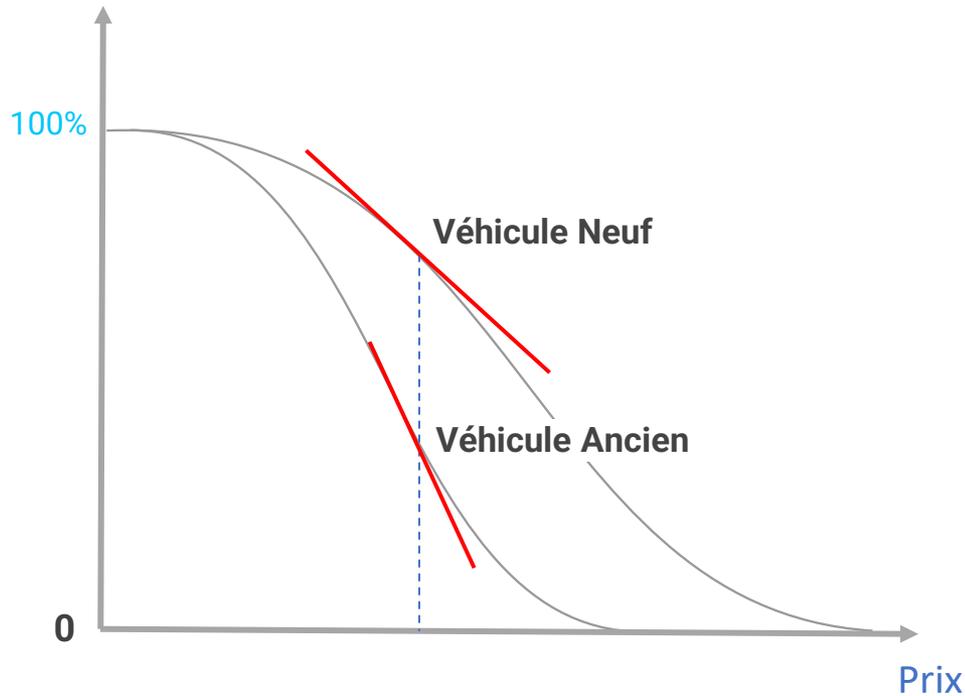
➔ Pouvoir prédictif ???

Visualisation de la demande

Pourquoi visualiser un modèle de demande ?

- On ne visualise pas un point mais une pente

Probabilité d'Achat



✗ Pouvoir prédictif

? Visualisation directe

Visualisation de la demande

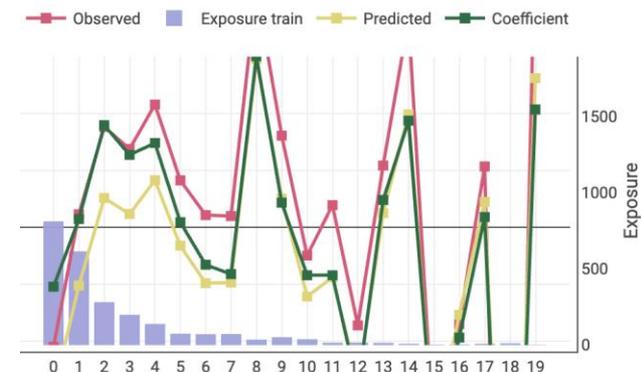
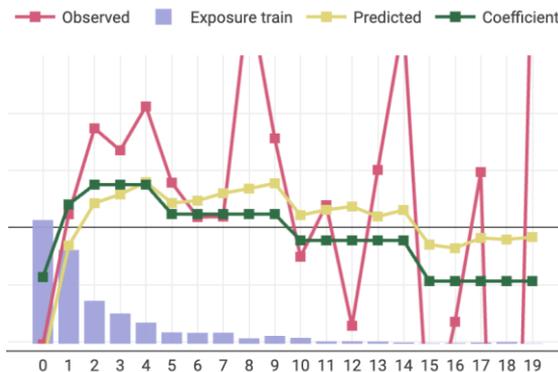
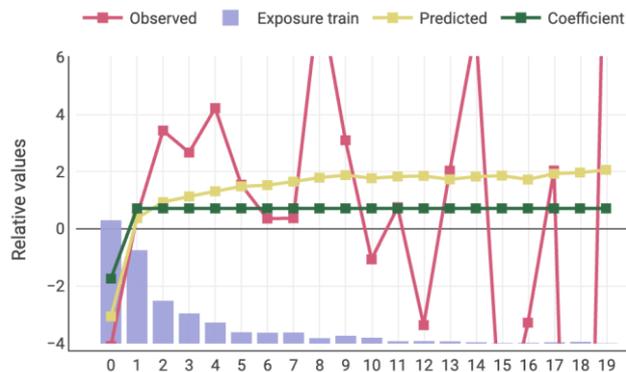
Le score ne permet pas de distinguer ces modèles

- Les différences de score ne sont pas significatives entre ces trois modèles, malgré un comportement très différent.
- **Sans visualisation, le choix ne serait pas évident.**
- Les décisions prises en utilisant ces différents modèles seraient, en revanche, extrêmement différentes.

AUC Score : 81.11%

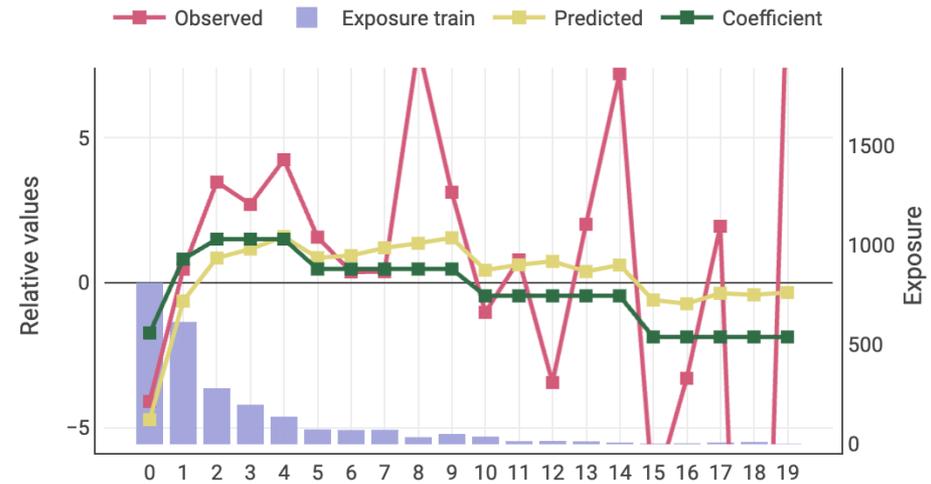
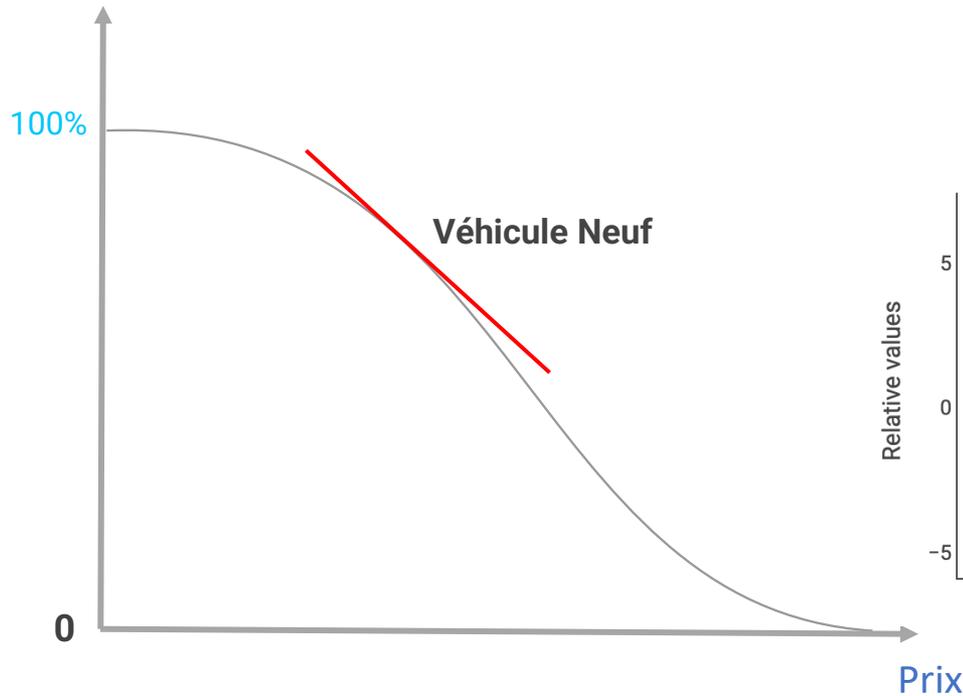
AUC Score : 81.13%

AUC Score : 81.12%



Visualiser le modèle d'élasticité ?

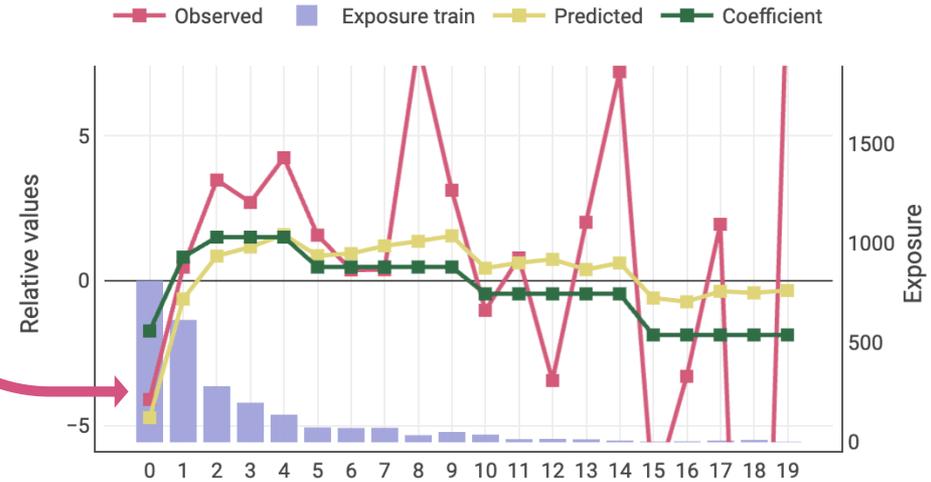
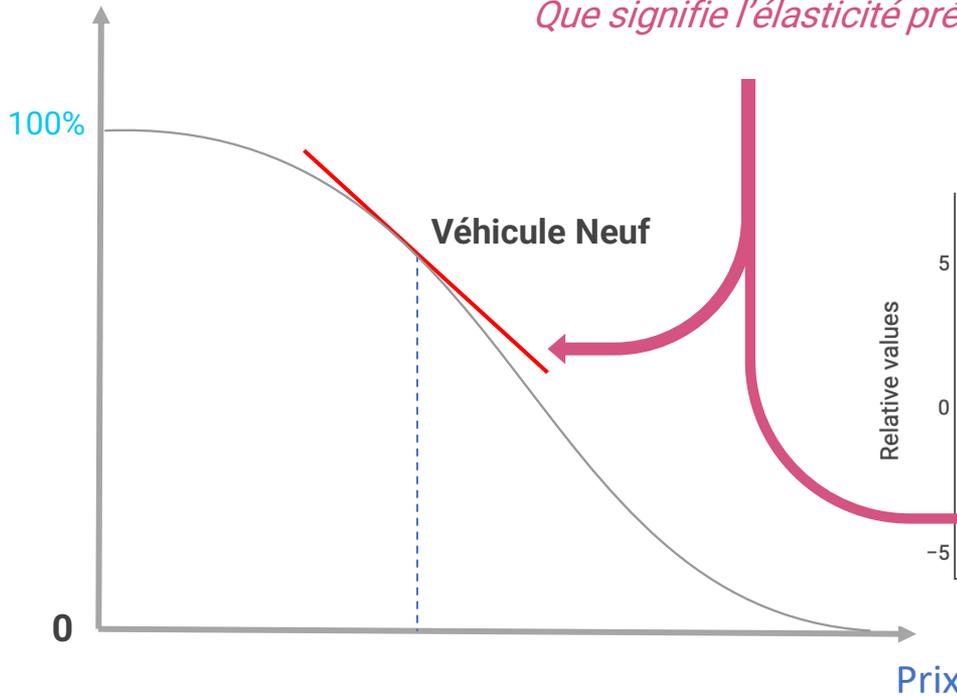
Probabilité d'Achat



Visualiser le modèle d'élasticité ?

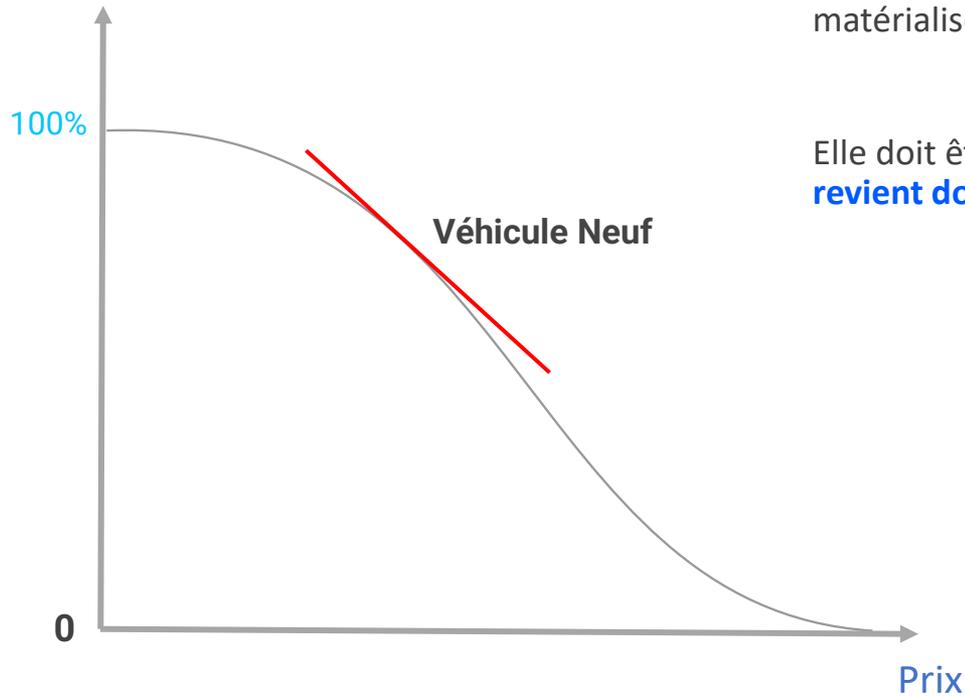
Probabilité d'Achat

*Que signifie l'élasticité observée ?
Que signifie l'élasticité prédite ?*



Visualiser le modèle d'élasticité ?

Probabilité d'Achat

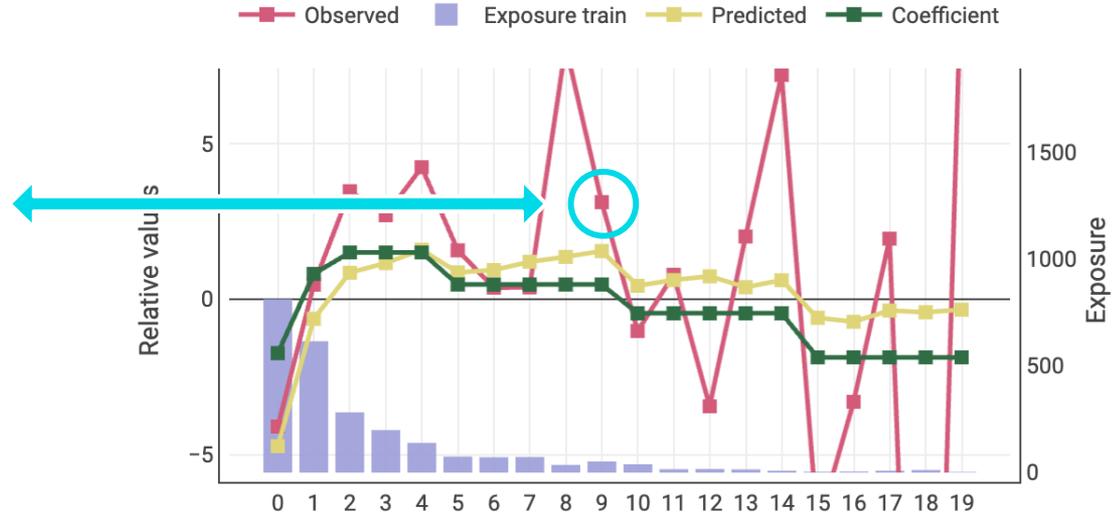
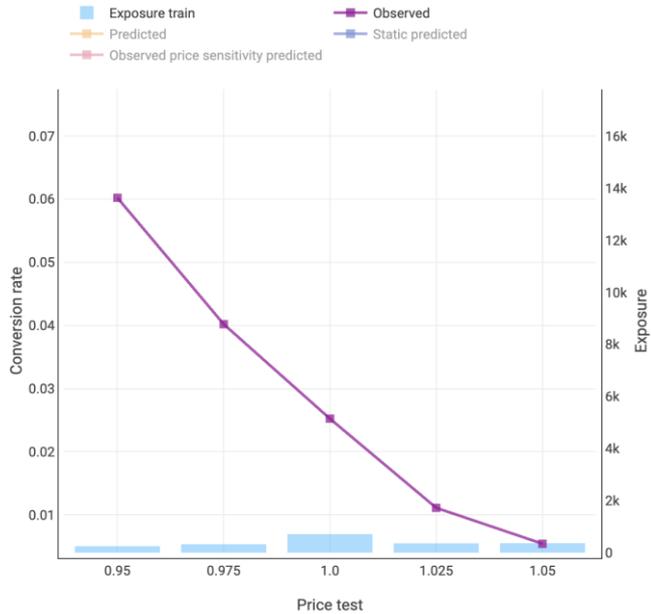


L'élasticité est une mesure théorique : elle n'est pas matérialisée pour un client donné.

Elle doit être estimée par un modèle : la **représenter revient donc à représenter un modèle**.

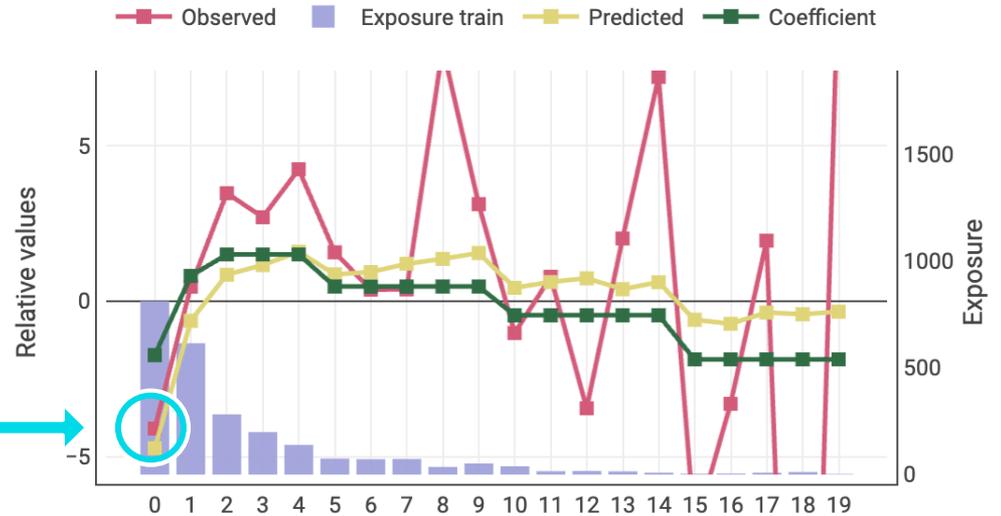
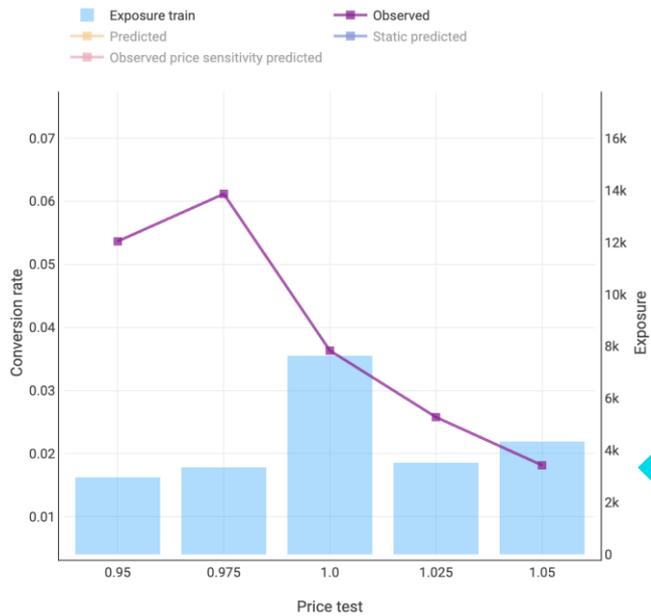
Visualiser le modèle d'élasticité ?

Qu'observe-t-on ?



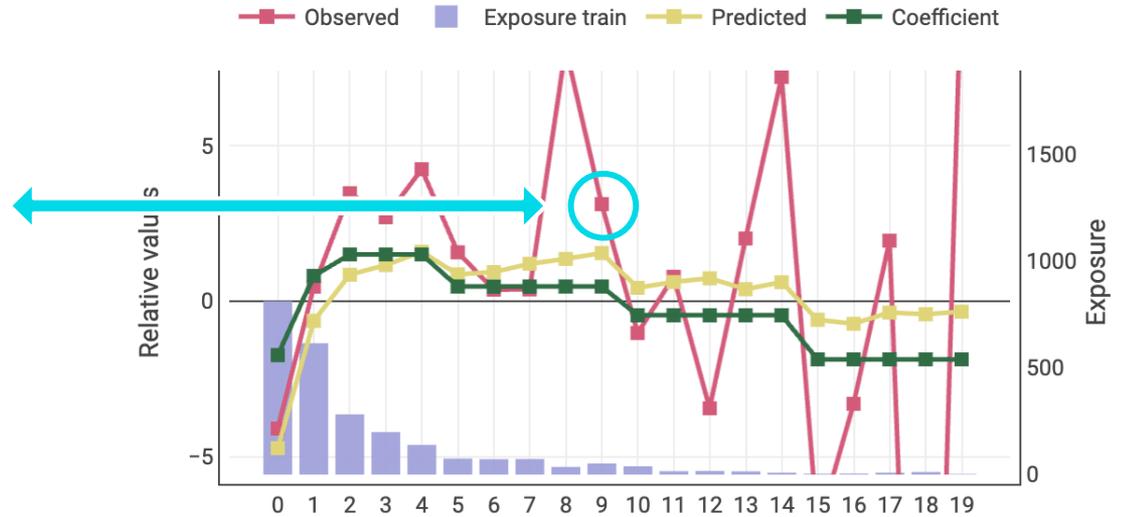
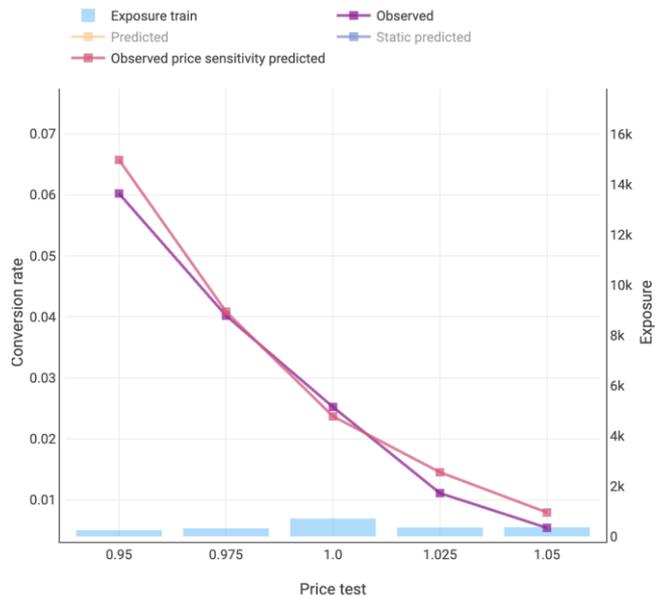
Visualiser le modèle d'élasticité ?

Qu'observe-t-on ?



Visualiser le modèle d'élasticité ?

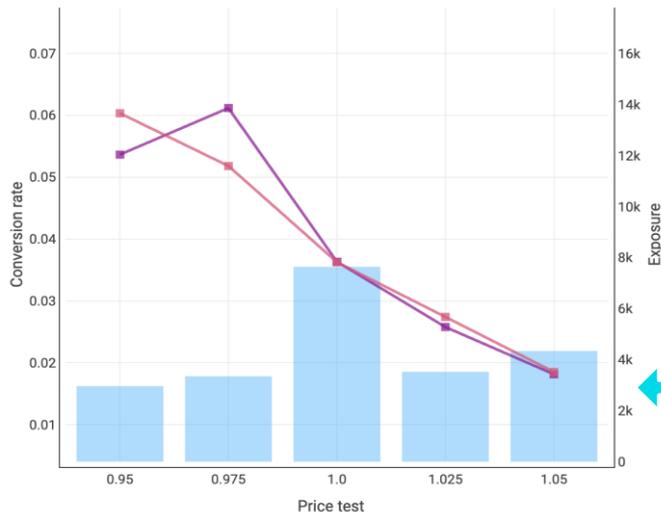
Qu'observe-t-on ?



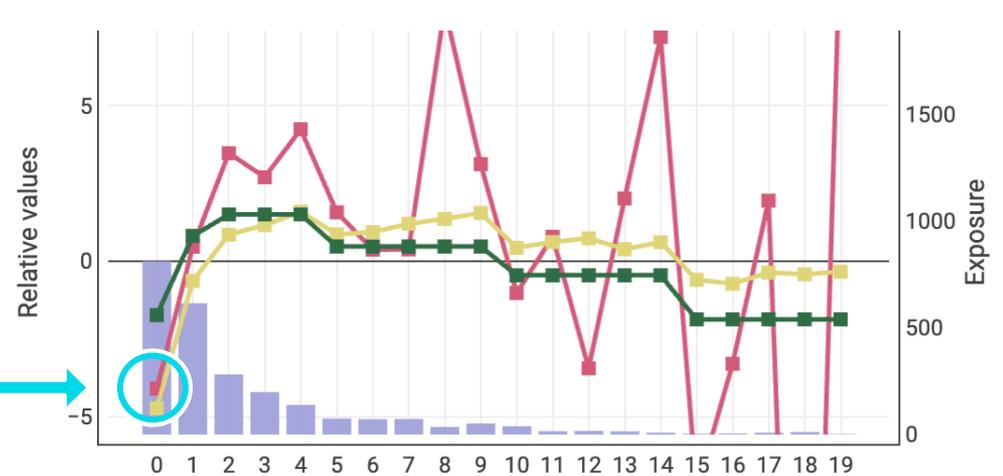
Visualiser le modèle d'élasticité ?

Qu'observe-t-on ?

Exposure train
Predicted
Observed price sensitivity predicted
Observed
Static predicted



Observed Exposure train Predicted Coefficient

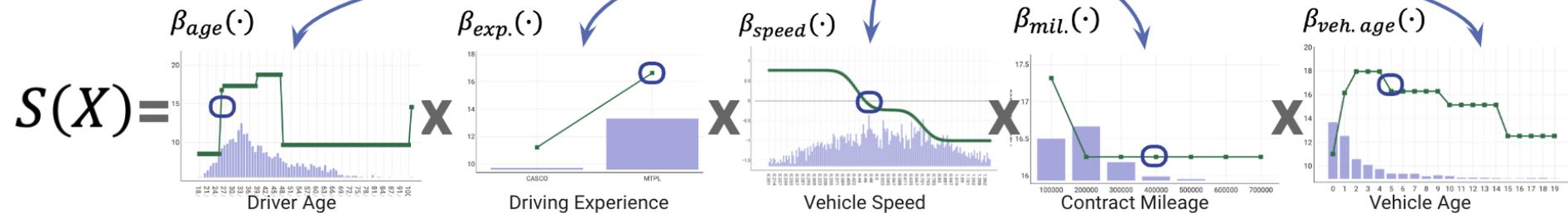


Visualiser le modèle d'élasticité ?

Cette visualisation reflète le modèle utilisé !

- Le modèle est multiplicatif et peut être visualisé variable par variable.

$$Sensibilité\ Prix(X) = \prod_i \gamma_i(X_i)$$



Synthèse

- Les visualisations doivent être présentes à chaque étape de la tarification.
- Le modélisateur et le décideur ne doivent **jamais perdre contact avec les données**.
- Les graphiques doivent être **adaptés au public** : modélisateur expert, décideur, audience large... « recycler » les graphiques d'un public à un autre est rarement efficace.
- De même, le graphique doit être **adapté au type de données, de modèles ou de KPI** représenté.
- Quelle que soit la personne qui visualise le graphique, celui-ci doit avoir un but : détecter des anomalies dans les données, une erreur de modèle, passer un message, proposer une action tarifaire...
- Construire un bon graphique est un **investissement**. En particulier, les graphiques de visualisation, intermédiaires dans la modélisation, sont souvent sous-estimés. Lors de la modélisation, **chaque graphique visualisé évite une erreur potentielle**.

Rémi BELLINA
Chief Data Scientist & Manager Actuariat
Milliman

remi.bellina@milliman.com



Guillaume BERAUD-SUDREAU
Responsable R&D & Actuariat
AKUR8

guillaume.beraud@akur8-tech.com

