upply

# NOTRE MÉTHODOLOGIE
## DE COMPARAISON DES PRIX

**ROUTIER AMÉRIQUE DU NORD**

# Sommaire

1. NATURE ET SOURCES DES DONNÉES COLLECTÉES

2. COUVERTURE GÉOGRAPHIQUE

3. ESTIMATION DE LA RÉPARTITION DES PRIX DE MARCHÉ* PAR NOS ALGORITHMES DE MACHINE LEARNING

a. Définition du prix

b. Modélisation du prix médian : les algorithmes de Data Science au service de la transparence des prix de transport

c. Estimation de la répartition des prix sur le marché*

d. Historique des estimations de prix

4. INDICE DE CONFIANCE

a. Une note en lettre allant de A à E

b. Construction du score

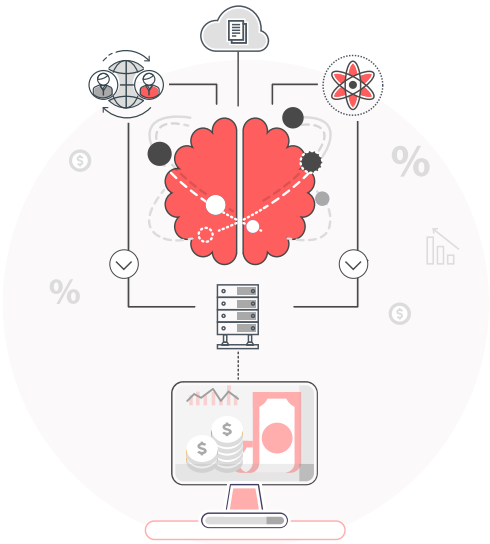
*Date de publication : février 2022*

*données Upply

Notre Comparateur de prix met à votre disposition toute la richesse des millions de prix issues d'opérations de transport communiqués par nos partenaires et nos utilisateurs.

## 1. NATURE ET SOURCES DES DONNÉES COLLECTÉES

Notre base de données Upply comporte aujourd'hui plus de **350 millions de prix issus de transactions facturées.** Elle est alimentée par plus de **4 000 entreprises** et **mise à jour quotidiennement.**



Ces données sont collectées automatiquement et quotidiennement auprès de partenaires spécialisés ainsi qu'auprès de nos utilisateurs. Afin de garantir leur analyse en toute sécurité, **les données sont cryptées et stockées de manière anonyme.**

Ces informations de prix proviennent de données de facturation. Chaque prix de notre base de données correspond à un transport effectué à une date donnée.

## 2 sources de données

1 **Nos partenaires**
Il s'agit de chargeurs, transporteurs, commissionnaires de transport ou bureaux d'études avec lesquels nous avons passé un accord de partenariat. Nous sommes connectés à leur système d'information et nous récupérons leurs données de facturation. En contrepartie, ces partenaires bénéficient d'un accès à notre solution Comparer & Analyser à un tarif préférentiel.

2 **Nos utilisateurs Comparer & Analyser**
Plusieurs milliers d'abonnés utilisent quotidiennement notre solution pour analyser leurs prix de transport. Ces prix sont collectés, harmonisés et filtrés pour en vérifier la véracité et viennent alimenter notre base de données.

### À savoir

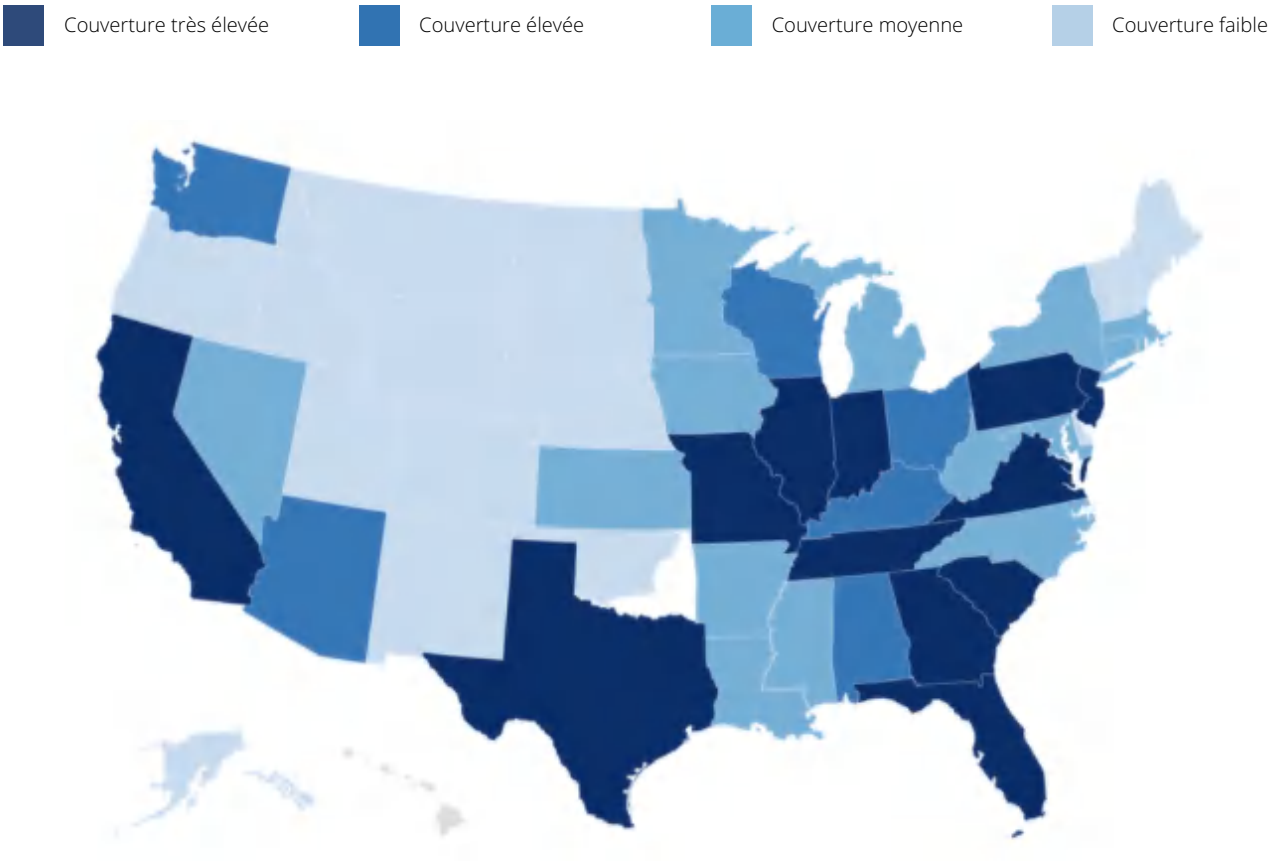
Plus de 80% de nos prix proviennent de l'utilisation de nos clients et des données fournies par nos partenaires.

Enfin, pour assurer une analyse non biaisée des prix, nous traitons des données brutes afin d'harmoniser les formats et supprimer les observations aberrantes. Ces traitements garantissent la qualité des données utilisées par nos modèles algorithmiques propriétaires, et donc la fiabilité des estimations de prix apportées par Upply.

## 2. COUVERTURE GÉOGRAPHIQUE

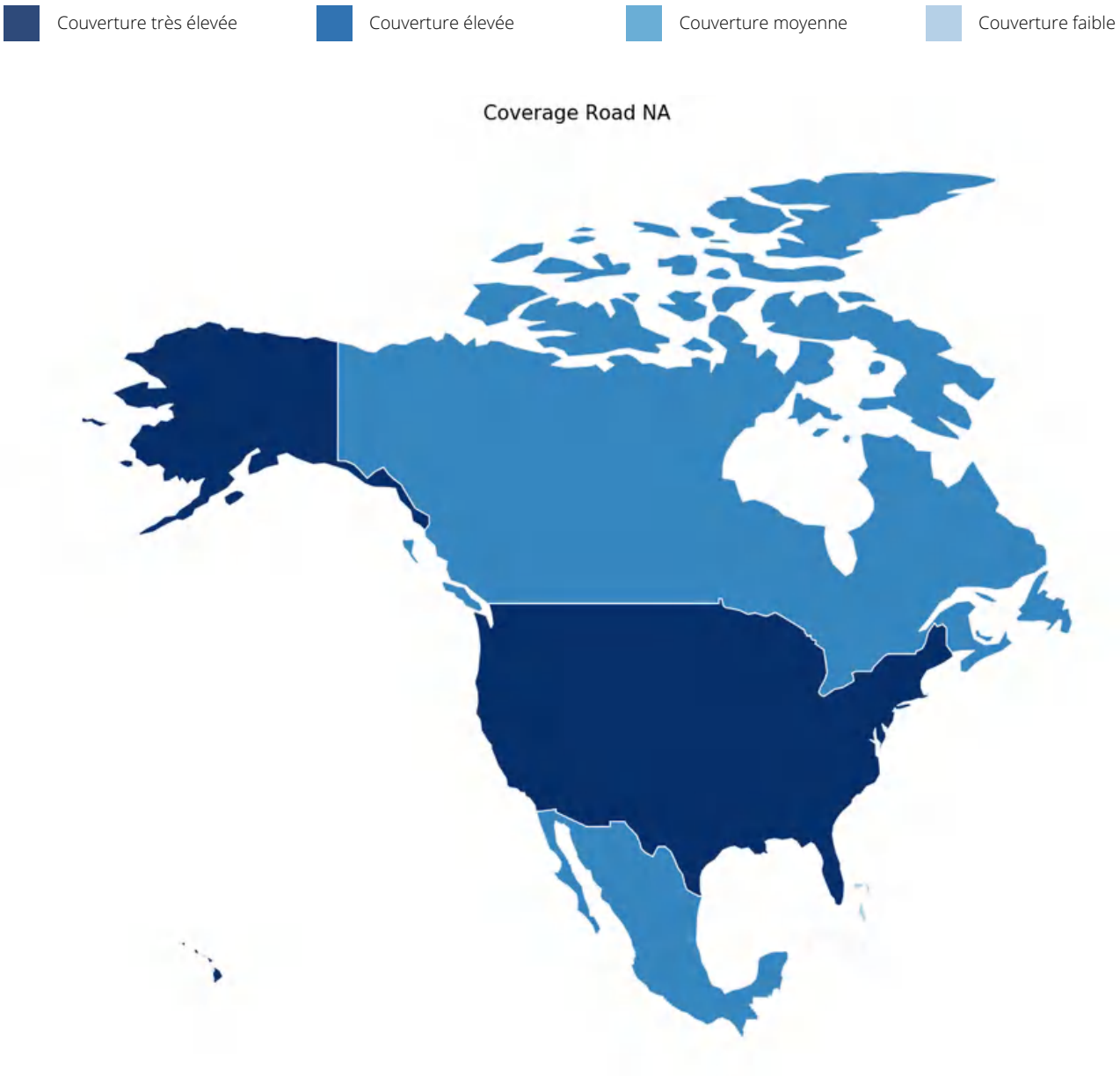
Les données collectées quotidiennement auprès de nos différentes sources nous permettent d'avoir une représentativité du marché en termes de prix mais aussi de **couverture géographique à l'échelle des États-Unis, Canada et Mexique.**

Couverture Upply aux États-Unis pour le transport de fret routier




Source : Upply

## Couverture Upply au Canada et au Mexique pour le transport de fret routier




Source : Upply

Nous offrons une couverture complète en Amérique du Nord (États-Unis Mexique, Canada), pour le transport routier de marchandises. Cette couverture géographique permet aux utilisateurs de comparer des prix de transport sur n'importe quelle origine-destination souhaitée.

# 3. ESTIMATION DE LA RÉPARTITION DES PRIX DE MARCHÉ* PAR NOS ALGORITHMES DE MACHINE LEARNING

## a. Définition du prix

- **Chargement FTL**
  Vous pouvez comparer rapidement le coût d'un transport FTL (*Full Truck Load – camion complet).*

- **Deux marchés**
  - Le marché des prix **réguliers et spot** qui vont représenter le marché complet nord-américain.
  - Le marché des prix **spot uniquement.**

- **Carburant**
  Les prix Upply comprennent le coût du carburant par défaut.

- **Type de camion**
  Vous avez le choix entre 3 types de camions : sec, réfrigéré, plateau.

## b. Modélisation du prix médian : les algorithmes de Data Science au service de la transparence des prix de transport

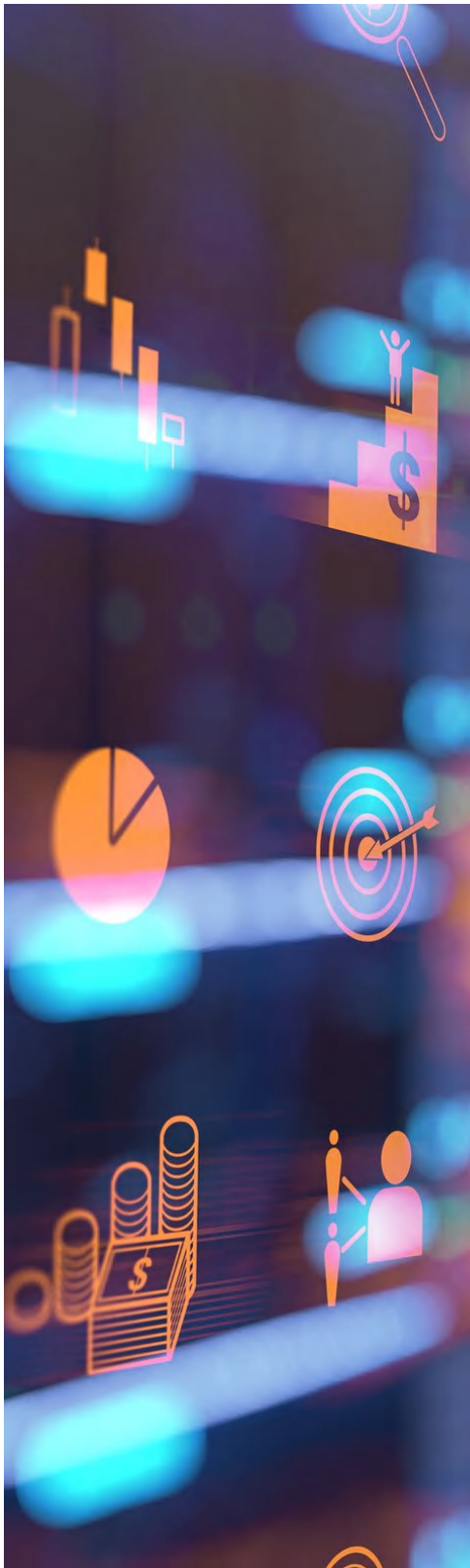
**Un algorithme pour extraire la logique de construction des prix**

Par définition, les données collectées par Upply ne couvrent pas avec le même niveau de qualité l'ensemble des requêtes possibles. Grâce aux algorithmes propriétaires développés par notre équipe de Data Scientists, Upply est capable de calculer les prix de transport pour n'importe quelle requête en Amérique du Nord. Sans ces algorithmes, les résultats seraient biaisés ou incomplets.

Upply utilise **un algorithme dit « d'apprentissage »** afin d'extraire toute la valeur à partir des données collectées. À partir de ces données, l'algorithme d'Upply est capable de modéliser la logique de construction des prix en prenant en compte **des milliers de paramètres simultanément**: date, origine, destination, poids, type de conteneur, etc.

Une fois la logique de construction des prix extraite, notre algorithme estime la répartition des prix pour n'importe quelle requête, même celles qui n'ont pas été directement observées.

C'est la force des algorithmes d'apprentissage propriétaires d'Upply.



Ensemble des transactions

| Prix observés | Caractéristiques observées | | | | |
|---|---|---|---|---|---|
| | date | poids | origine | destination | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

nourrit

Algorithme d'apprentissage

extrait

Structure du prix en fonction des caractéristiques

| | |
|---|---|
| Poids | + x% |
| Sec | - y% |
| Origine - Atlanta | + z% |
| | |
| | |

## *Comment peut-on s'assurer de la qualité d'un algorithme ?*

### *a. La qualité des données*

Puisque l'algorithme s'appuie sur les données pour « apprendre » la logique des prix, la qualité de ses estimations varie en fonction de la quantité de données sous-jacentes.

### *b. La confrontation à la réalité du marché**

Pour choisir le meilleur algorithme possible, nous utilisons une méthode classique d'évaluation de nos modèles : nous choisissons un ensemble de transactions pour lesquelles nous connaissons les caractéristiques, ainsi que leurs prix réellement observés. À partir des caractéristiques des transactions, l'algorithme calcule une répartition du prix. En comparant ces estimations fournies par nos modèles avec les prix réellement observés, nous pouvons mesurer la précision de notre algorithme par rapport à la réalité du marché*.

*données Upply

## *Les caractéristiques prises en compte par l'algorithme*

### Données géographiques

- Villes d’origine et de destination.
- Zone d'origine parmi les grandes régions des États-Unis, du Mexique, du Canada, comme par exemple "Midwest", "South West", etc.

### Données temporelles

Le prix médian est estimé par semaine ou par mois en fonction de la fréquence et de la quantité des données que nous recevons.

### Données de transport

- Type de camion
- Poids du chargement

## c. Estimation de la répartition des prix sur le marché*

Une fois le prix médian estimé, nos algorithmes reconstituent la répartition des prix en fonction de la distribution observée pour le couple origine/destination.

Caractéristiques de la requête

| date | poids | origine | destination | |
|---|---|---|---|---|
| | | | | |
| | | | | |

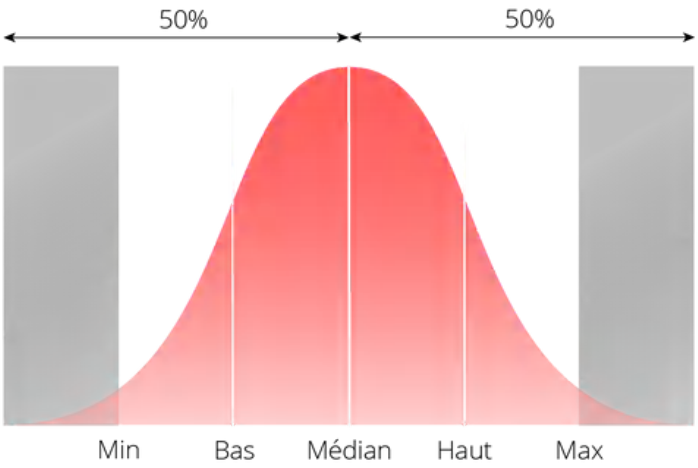
+

Structure du prix en fonction des caractéristiques

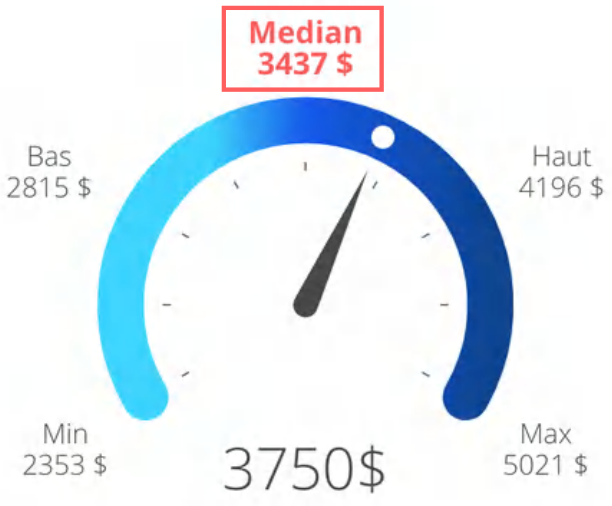
| | |
|---|---|
| Poids | + x% |
| Réfrigéré | - y% |
| Origine - Le Havre | + z% |

Prix médian estimé

## Prix médian



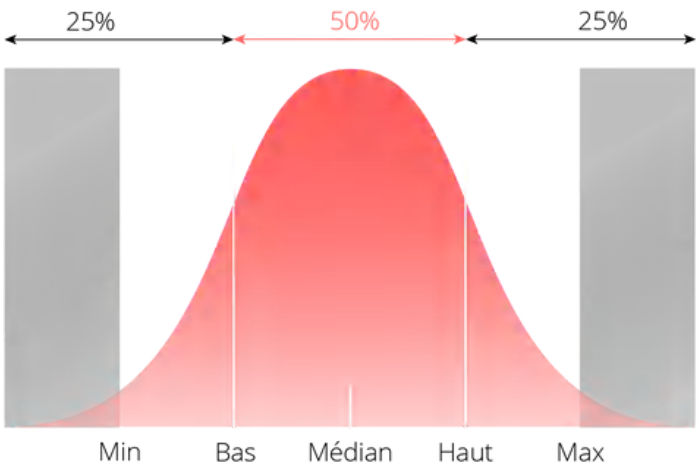

**Prix médian**
La moitié des prix observés est située en-dessous de cette valeur, l'autre moitié au-dessus.




**Exemple de lecture**
La moitié des prix observés est située en-dessous de 3 437$, l'autre moitié au-dessus.

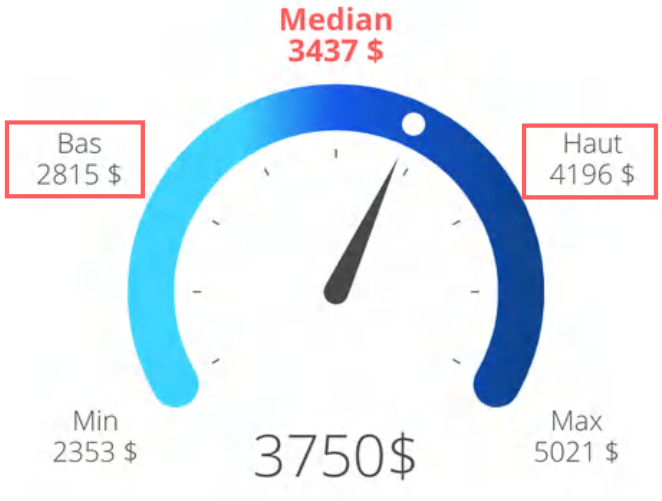
## Prix bas et haut

La moitié des prix observés est située entre ces valeurs.




**Prix bas**
25% des prix observés sont situés en-dessous de cette valeur et 75% sont situés au-dessus

**Prix haut**
75% des prix observés sont situés en-dessous de cette valeur et 25% sont situés au-dessus

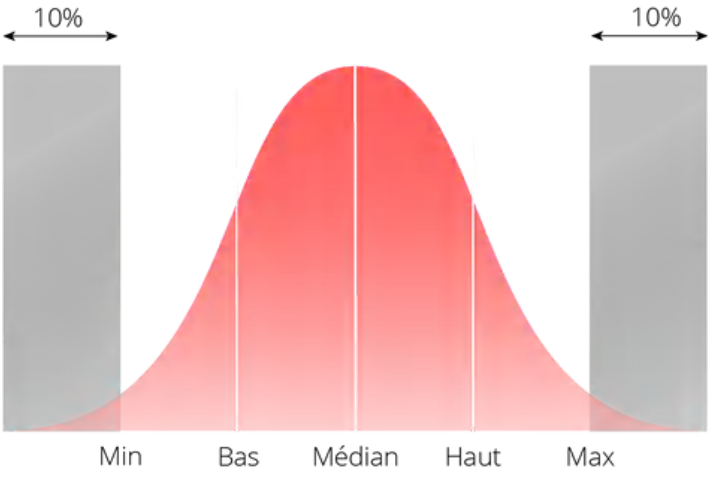



**Exemple de lecture**
25% des prix observés sont inférieurs à 2 815$, 75% des prix sont supérieurs.

75% des prix sont inférieurs à 4 196$, 25% sont supérieurs.

50% des prix se situent entre 2 815$ et 4 196$.

## Prix minimum et maximum




**Prix minimum & maximum**
Valeurs minimales et maximales des prix observés après exclusion des valeurs aberrantes (10% des valeurs les plus basses et 10% des valeurs les plus élevées)

## d. Historique des estimations de prix

Nous recevons quotidiennement de nouvelles données de prix transmises par nos partenaires et utilisateurs. Cela nous permet d'avoir les estimations les plus à jour et les plus précises possibles. En intégrant ces données, nous ré-estimons la distribution des prix pour des dates antérieures (ex : une estimation de prix peut différer en fonction de la date de réalisation de la comparaison).

L'historique des prix Upply représente ainsi **la répartition des 5 prix (Min, Bas, Médian, Haut et Max)** pour une date au cours des **deux dernières années.**

## 4. INDICE DE CONFIANCE

Upply mesure la qualité des prix estimés grâce à un indice de confiance. Celui-ci a été développé par notre équipe de Data Scientists et est un gage de transparence : il renseigne nos utilisateurs sur la fiabilité du prix affiché. Cet indice de confiance reflète la **quantité de données disponibles** ainsi que le **niveau de précision de l'algorithme.**

Cet indice de confiance est basé sur **3 critères complémentaires :**

1 – **Le nombre de données** disponibles sur la ligne recherchée.

2 – **La diversité des sources de données** sur la ligne recherchée. En d'autres termes, le nombre d'entreprises différentes nous ayant fourni des prix de transport.

3 – **La fréquence de réception des sources de données** : l'indice de confiance est d'autant plus élevé que les données arrivent régulièrement.

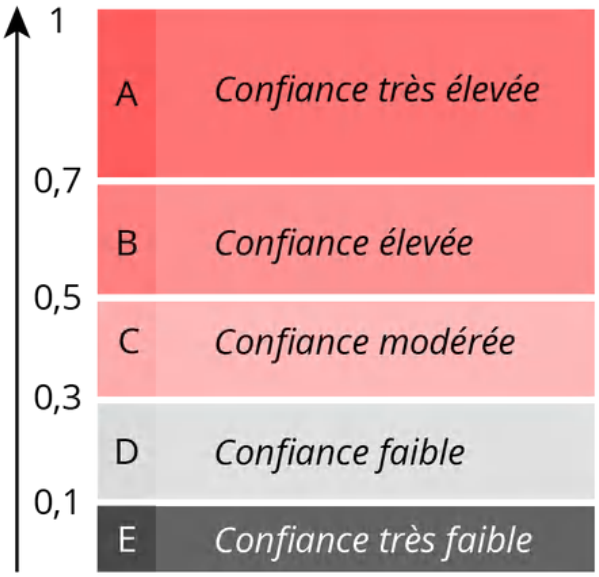
### a. Une note en lettre allant de A à E

L'indice de confiance se caractérise par **une lettre allant de A, pour la confiance la plus élevée, à E, la confiance la plus faible.**

L'attribution de la note par Upply dépend d'**un score qui varie de 0 à 1.**

Pour chaque transaction, un score entre 0 et 1 est obtenu et il est catégorisé en une des 5 lettres :

A – Confiance très élevée
B – Confiance élevée
C – Confiance modérée
D – Confiance faible
E – Confiance très faible




## b. Construction du score

### *Un score dépendant des paramètres du Comparateur*

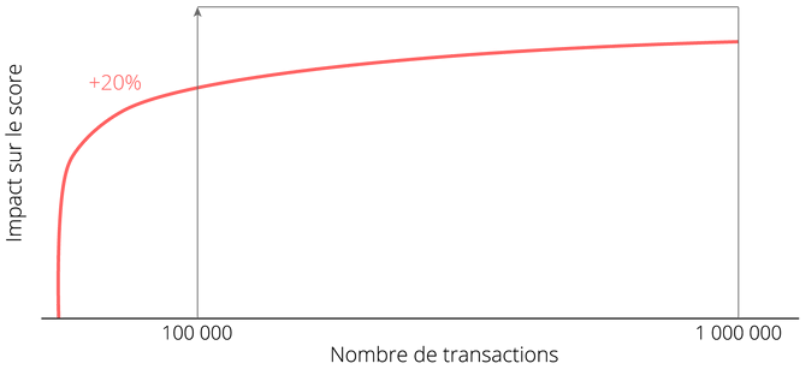
La confiance que nous avons dans les prix estimés par nos algorithmes pour une requête dépend de l'ensemble des paramètres choisis par l'utilisateur.

Dans le cadre du routier :

- La ligne : lieu d'origine, lieu de destination, corridor concerné
- Le poids du chargement ainsi que la quantité

## *Prise en compte du nombre de transactions*

**Plus nous avons de données, meilleur est notre algorithme.** Cependant à partir d'un certain seuil, l'ajout de données n'est plus aussi impactant. Nous avons donc créé un système d'échelle qui permet de ne pas survaloriser les trop grandes quantités de données.




Par exemple, entre un pays d'origine avec 100 000 transactions et un pays d'origine avec 1 million de transactions, le score sera 20% plus élevé pour ce dernier.

À l'inverse, le score sera 5% plus élevé pour un port avec 2 millions de transactions, comparé à un port avec 1 million de transactions car 1 million de transactions permettent d'avoir une qualité et une représentativité suffisante pour l'estimation des prix.

*données Upply

## *Niveau de précision de l'algorithme*

Le niveau de précision de l'algorithme nous permet de savoir si notre estimation est proche de la réalité du marché*. Nous comparons les prix estimés par notre algorithme aux prix observés dans les transactions. Par définition, le niveau de précision correspond au nombre de transactions pour lesquelles le prix observé se situe entre les Bornes Prix Bas et Haut sur le nombre total de transactions.

Dans notre exemple ci-dessous, la précision est de 4 transactions sur 5, soit 80%.