



upply

OUR BENCHMARK
METHODOLOGY



AIR FREIGHT

CONTENTS

1. NATURE AND SOURCES OF COLLECTED DATA

2. GEOGRAPHICAL COVERAGE

3. ESTIMATION OF THE MARKET PRICE DISTRIBUTION* BY OUR MACHINE LEARNING ALGORITHMS

a. Price definition

b. Median price modelling: Data Science algorithms for transport price transparency

c. Estimated market price distribution*

d. History of price estimates

4. CONFIDENCE INDEX

a. A to E letter grade

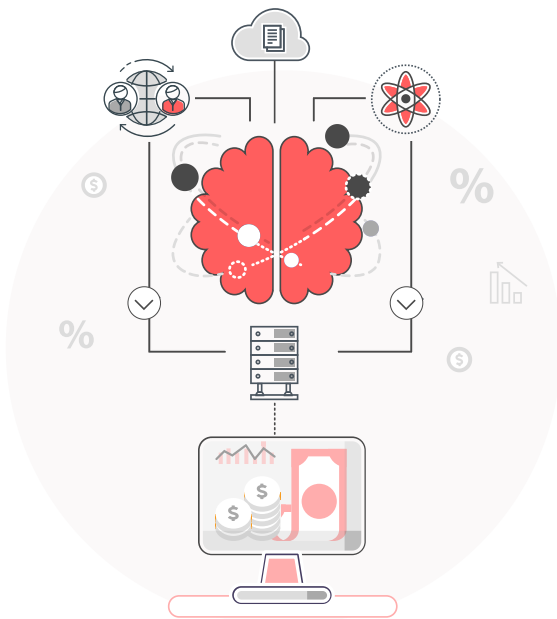
b. Score construction

Our price benchmark service provides you with a wealth of millions of prices for transport operations provided by our partners and users.

1. NATURE AND SOURCES OF COLLECTED DATA

Our Upply database now contains over **750 million prices from invoiced transactions.**

It is populated by over **4,000 companies** and updated weekly.



This data is collected automatically on a daily basis from specialised partners and our users. In order to guarantee a secure analysis, the data is **encrypted and stored anonymously.**

This pricing information is derived from billing data. Each price in our database corresponds to a shipment made on a given date.

2 main data sources

1

Our partners

These are shippers, carriers, freight forwarders or design offices with whom we have signed a partnership agreement. We are connected to their information systems and we retrieve their invoicing data. In return, these partners benefit from access to our SMART solution at a preferential rate.

2

Our SMART users

Several thousand users use our solution every day to analyse their transport prices. These prices are collected, filtered to verify their veracity and fed into our database.



Good to know

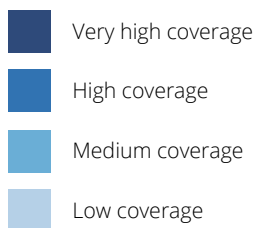
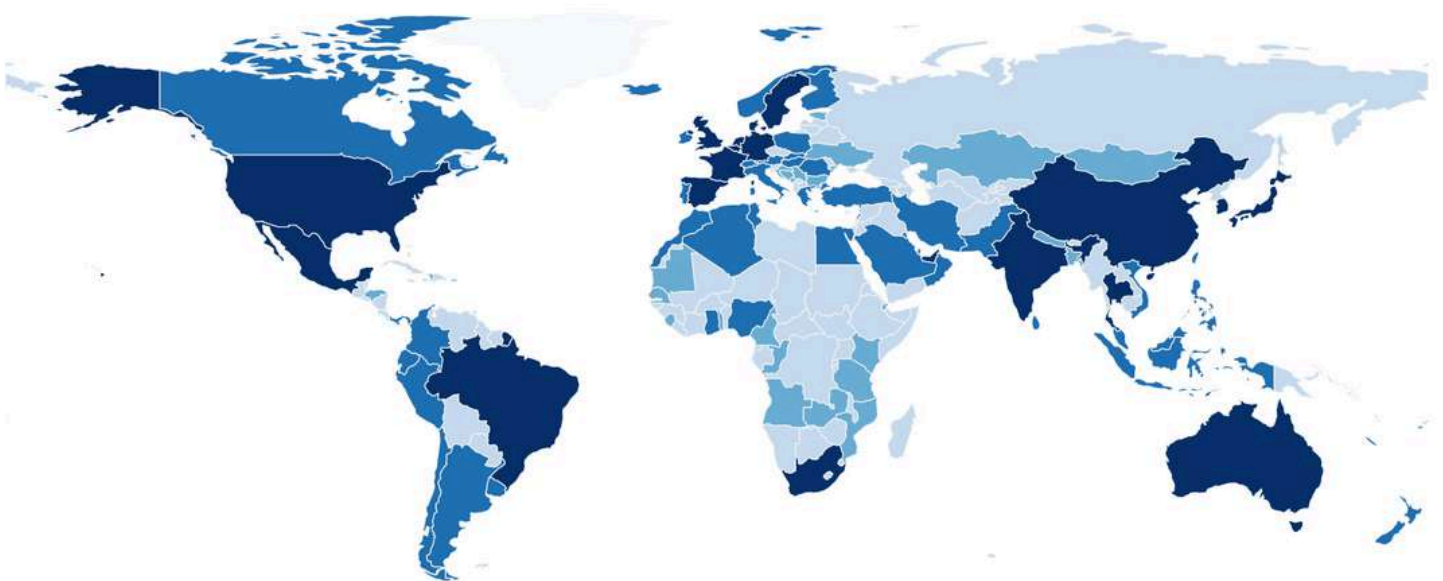
To ensure unbiased price analysis, we process raw data to harmonise formats and remove outliers.

This processing ensures the quality of the data used by our proprietary algorithmic models, and therefore the reliability of the price estimates provided by Upply.

2. GEOGRAPHICAL COVERAGE

The data collected daily from our various sources allows us to have a representative market in terms of price but also geographical coverage on **a global scale**: we collect data from more than **340 airports**.

Worldwide coverage



Source: Upfly

This geographical coverage allows users to compare transport prices on any desired origin-destination, anywhere in the world.

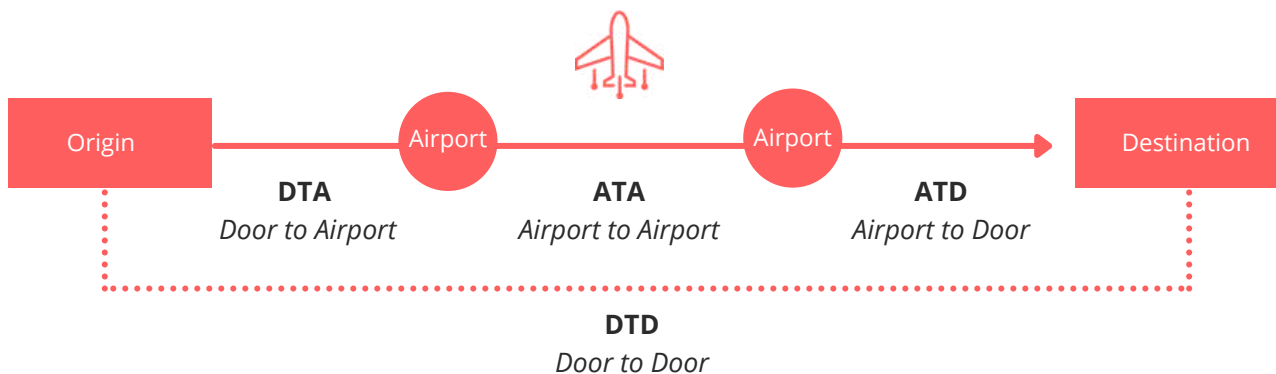
3. ESTIMATION OF THE MARKET PRICE DISTRIBUTION* BY OUR MACHINE LEARNING ALGORITHMS

a. Price definition

The price is broken down **into 3 parts**:

- Door to Airport - **DTA** (pre-carriage)
- Airport to Airport - **ATA** (main leg)
- Airport to Door - **ATD** (on-carriage)

The Door to Door price is the sum of these three prices.



The ATA price includes:

- Air freight rate (which may include any transshipments)
- Airport handling charges at departure and destination terminals
- The main security and fuel surcharges
- Possible transshipments due to the chosen air route

b. Median price modelling: Data Science algorithms for transport price transparency

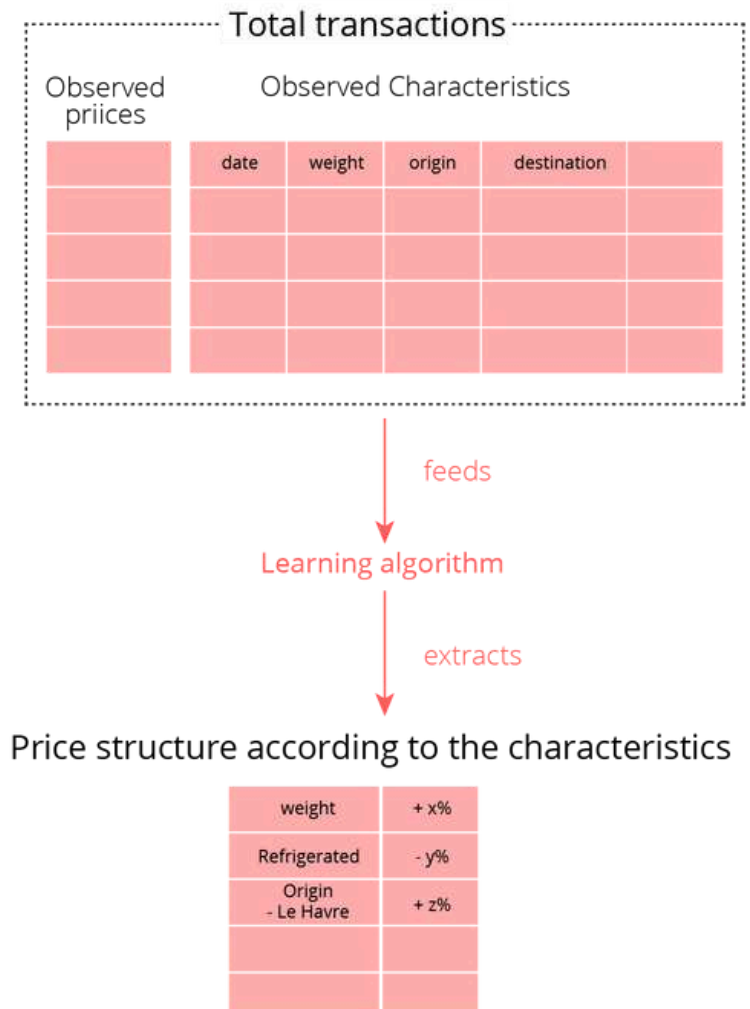
An algorithm to extract the logic of price construction

By definition, the data collected by Upply does not cover all possible queries with the same level of quality. Thanks to proprietary algorithms developed by our team of data scientists, Upply is able to calculate transport prices for any query in the world. Without these algorithms, the results would be biased or incomplete.

Upply uses a **so-called "learning" algorithm** to extract all the value from the data collected. From this data and thanks to the statistical modelling work carried out upstream, Upply's algorithm is able to model the price construction logic by **taking into account thousands of parameters simultaneously**: origin, destination, weight, type of goods, date, etc.

Once the price construction logic is extracted, our algorithm estimates the price distribution for any query, even those that have not been directly observed.

This is the strength of Upply's proprietary learning algorithms.



How can we be sure of the quality of an algorithm?

a. The amount of data

Since the algorithm relies on data to "learn" the logic of prices, the quality of its estimates varies according to the amount of underlying data.

b. Market reality check*

To choose the best possible algorithm, we use a classical method of evaluating our models: we choose a set of transactions for which we know the characteristics, as well as their actual observed prices. From the characteristics of the transactions, the algorithm calculates a price distribution. By comparing these estimates provided by our models with the actual observed prices, we can measure the accuracy of our algorithm against the reality of the market*.



*Upply data

Characteristics taken into account by the algorithm

Geographical data

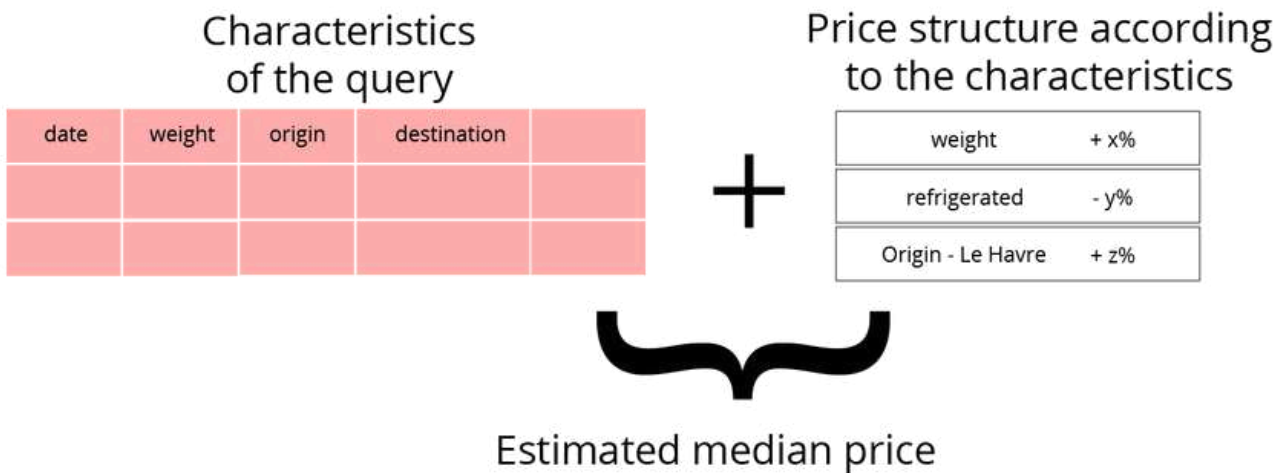
- Airports of origin and destination.
- Geographical regions, built around the main trade corridors (import/export).
- Each airport included in our geographical coverage is associated with a region.
- Our statistical modelling adapts its geographical accuracy to the amount of data available on the selected airport. If the airport does not have enough data available, the modelling will extend its scope to the regional level.

Temporal data

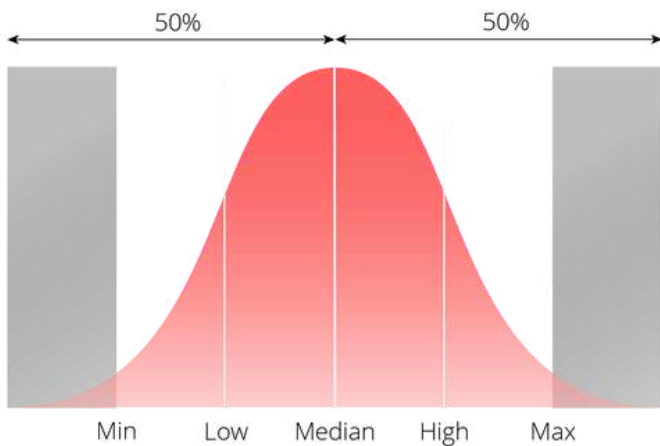
The median price is estimated per week or per month depending on the frequency and quantity of data we receive.

c. Estimated market price distribution*

Once the median price is estimated, our algorithms reconstruct the distribution of prices according to the distribution observed for the origin/destination pair.

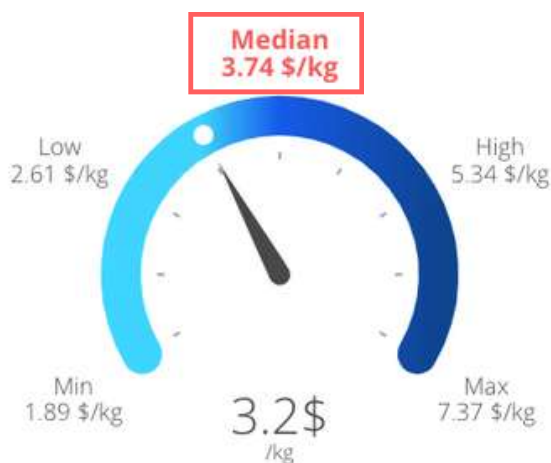


Median price



Median price

Half of the observed prices are below this value, the other half above.

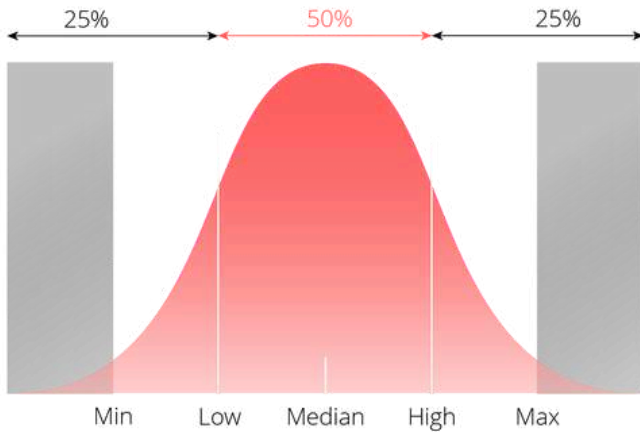


Example of reading

Half of the observed prices are below 3.74\$, and half above.

Low & high prices

Half of the observed prices are between these values.

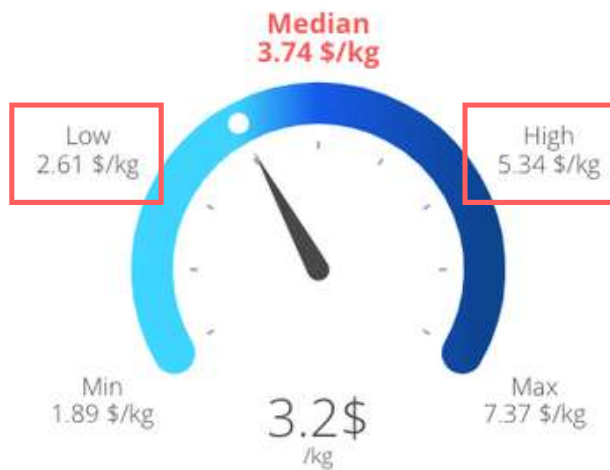


Low price

25% of observed prices are below this value and 75% are above

High price

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



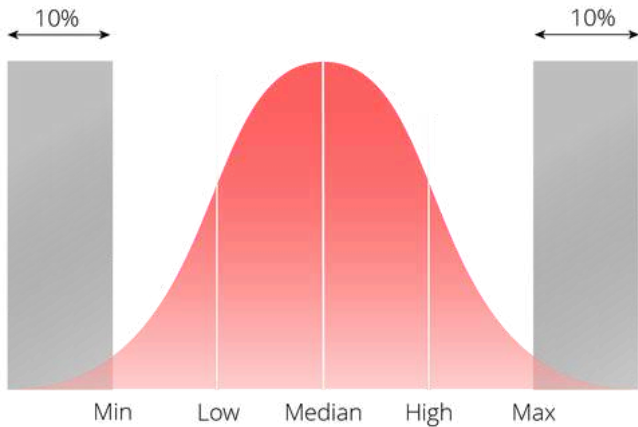
Example of reading

25% of observed prices are below 2.61\$. 75% of the prices are above this price.

75% of the prices are below 5.34\$, 25% are above this price.

50% of the prices are between 2.61\$ and 5.34\$.

Minimum & maximum prices



Minimum & maximum prices

Minimum and maximum values of observed prices after exclusion of outliers (10% of lowest and 10% of highest values)

d. History of price estimates

We receive new price data daily from our partners and users. This allows us to have the most up-to-date and accurate estimates possible. By integrating this data, we re-estimate the price distribution for earlier dates (e.g. a price estimate may differ depending on the date of the comparison).

The Upply price history thus represents the **distribution of the 5 prices (Min, Low, Median, High and Max) for a date over the last two years.**



4. CONFIDENCE INDEX

Upply measures the quality of the estimated prices thanks to a confidence index. This was developed by our team of data scientists and is a guarantee of transparency: it informs our users of the reliability we have in the price displayed. This confidence index reflects **the amount of available data** as well as **the level of accuracy of the algorithm**.

This confidence index is based on **3 complementary criteria**:

- 1 - **The number of data items available** on the line being searched.
- 2 - **The diversity of data sources** on the line searched. In other words, the number of different companies that provided us with transport prices.
- 3 - **The frequency of receipt of data sources**: the more regularly the data arrive, the higher the confidence index.

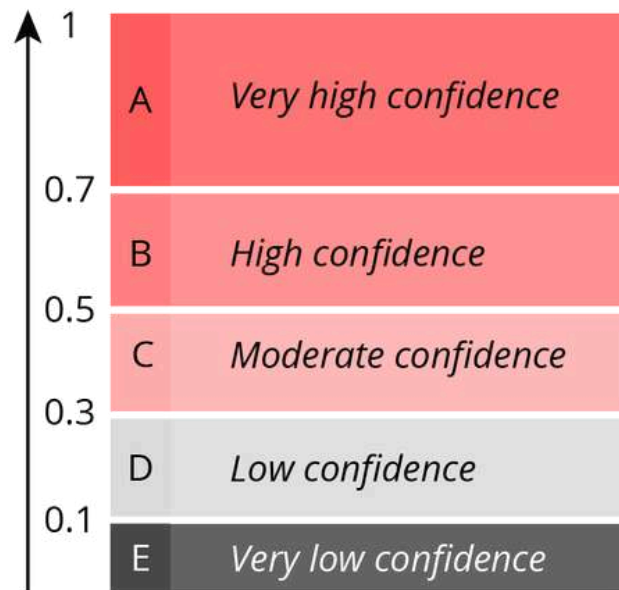
a. A letter grade from A to E

The confidence index is characterised by **a letter from A, the highest confidence, to E, the lowest confidence**.

Upply's rating is based on a score that varies **from 0 to 1**.

For each transaction, a score between 0 and 1 is obtained and it is categorized into one of 5 letters:

- A – Very high confidence
- B – High confidence
- C – Moderate confidence
- D – Low confidence
- E – Very low confidence



b. Score construction

A score depending on the parameters of the Comparator

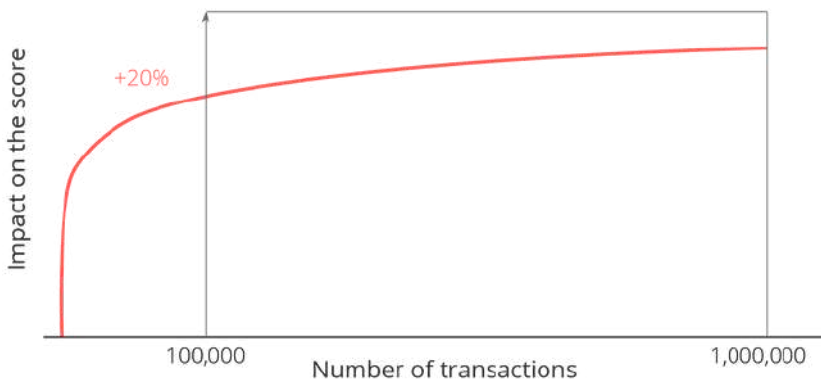
The confidence we have in the prices estimated by our algorithms for a query depends on the set of parameters chosen by the user.

In the air freight context:

- The line: airport of origin, airport of destination, corridor involved
- The weight of the load, as well as the quantity
- Transport of dangerous goods
- Type of service: ATA, DTD

Taking into account the number of transactions

The more data we have, the better our algorithm is. However, after a certain threshold, adding data is no longer as impactful. We have therefore created a scaling system that allows us not to overvalue large amounts of data.



For example, between a country with 100,000 transactions and a country with 1,000,000 transactions, the score will be 20% higher for the latter.

Conversely, the score will be 5% higher for a country with 2 million transactions, compared to a country with 1 million transactions, because 1 million transactions allows for sufficient quality and representativeness for price estimation.

*Upply data

Level of accuracy of the algorithm

The level of accuracy of the algorithm allows us to know if our estimate is close to the reality of the market*. We compare the prices estimated by our algorithm to the prices observed in the transactions. By definition, the accuracy level corresponds to the number of transactions for which the observed price is between the Low and High Price Bounds out of the total number of transactions.

In our example above, the accuracy is 4 out of 5 transactions or 80%.

