Bottlenecks of Competitive Data Scraping and Delivery

How Retailers Are Losing Their Revenue

Introduction

Regardless of whether the retailer uses artificial intelligence
to change prices on a storefront or works with a team of analysts,
it needs competitive data to set the optimal prices on products.

In this white paper we uncover data quality requirements for different
retailers, which criteria to use when checking the data quality, how
the cost of data delivery varies by quality level, and which questions
to ask the data provider when signing a contract, etc.

Before we start, write down the data quality metrics you have now
to compare them afterwards. They also are supposed to be included
in SLA with your data provider.

Different Data Quality Requirements

Small retailers, whose turnover does not exceed $2 million a year, rarely monitor competitors and use competitive data for regular repricing. Therefore, the quality of data does not play a big role for them, and 90% of the qualitative data for them is quite enough.

Mid-sized retailers ($2-40 million in turnover) already catch sight of the relationship between the optimal price and the business margin. The quality of competitive data becomes an important factor for them. For such retailers, data quality below 95% is unacceptable.

The most demanding for the data quality are large retailers, companies with more than $40 million in turnover. Even the minimum percentage of poor-quality data becomes critical, as the cost of error increases. For such retailers, the required data quality index gravitates toward 100%.
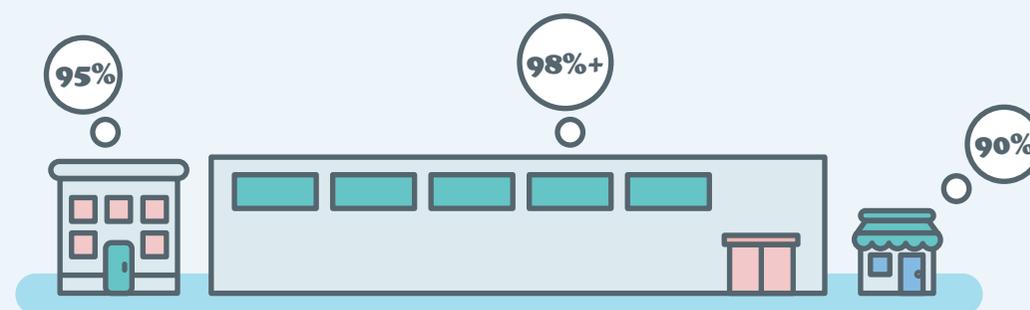
The Cost of Error For Different Retailers

Let's assume any retailer's product has only $1 margin. Here is how much profit a retailer can loose due to poor-quality data.

|	Small retailer	Middle-sized retailer	Enterprise retailer (comparison of high and low data quality)	
Products in monitoring	100	500	1000	1000
Competitors in monitoring	2	5	10	10
Monitoring frequency	daily	daily	daily	daily
Required data quality	90%	95%	90%	98%+
Products with mistakes	100*2*10%=20 products	500*5*5%=125 products	1000*10*10%=1000 products	1000*10*2%=200 products
Margin loss due to pricing errors per day	$20	$125	$1000	$200

$1 marginality is just an example: poor data quality usually affects high-margin categories or KVI-products. The cost of an error for a retailer depends on the category, the industry, the elasticity of the product, etc.

Moreover, if the retailer monitors a small number of products for many competitors (for example, 100 products with ten competitors), then even 10% of the data with errors can lead to the situation when all products are priced incorrectly (if the error data is not intersected and all data points are delivered for different products with errors).

The Foundation for Qualitative Data Collection

Correct matches are the basis to the qualitative data collection. We have already uncovered the importance of qualitative matches, and how the data provider should check these comparisons in the previous whitepaper.

If the matches are done incorrectly, the number of errors grows exponentially, and it's impossible to collect data on the desired products.

competera | Whitepaper

Manual vs. Automated Matches
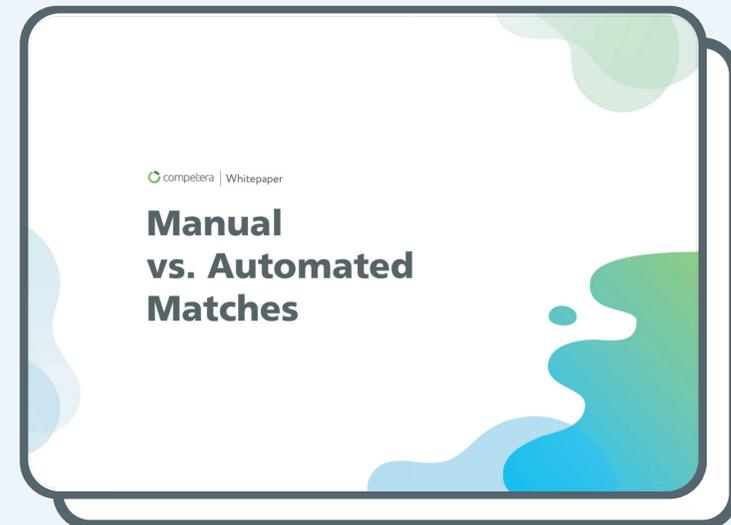
Download Free Whitepaper

How to Check Data Quality

Three main indicators help the retailer check
the quality of the collected data:

• Freshness •
• Completeness •
• Correctness •

Below we tell in detail about each of them.

Data Freshness

The freshness of data is an indicator that affects the effectiveness of pricing. For industries with high turnover and a little margin, e. g. electronics, where competitors change prices several times per day, and the customers' elasticity is very high, it's important to use the most recent market data. Otherwise, the retailer can quickly fall out of the market.

A Category Manager needs to clearly understand when the data was collected. In this case, he can take the right decision to use or ditch particular data for pricing.

Data Freshness, % =
Data collected [2 hours] before pricing /
Scheduled data volume * 100%

*Number in square brackets is a variable

0-2 hours
Healthy Data

2-4 hours
Stale Data

4-48 hours
Irrelevant Data

48+ hours
AncientData

Many data providers do not show this indicator, delivering data collected 10 minutes ago, along with collected 48 hours before repricing in the same dataset. It's impossible to build competitive and efficient pricing on such data. To allow the manager to take an intelligent pricing decision, the data should be as fresh as possible, delivered shortly before the moment when the category managers are to set new prices.
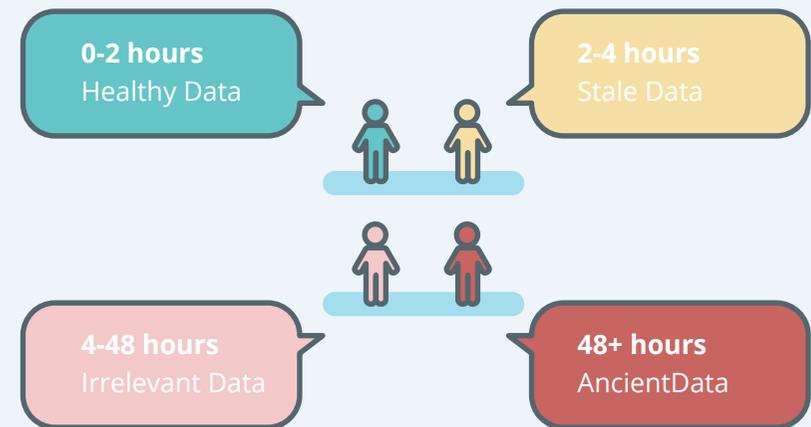
Data Completeness

The second indicator of data quality is their completeness: which percentage of the planned data volume was actually collected and delivered to the retailer.

<div align="center">

Data Completeness =
Collected volume
/ Scheduled volume * 100%

</div>

This is the indicator that we talked about at the beginning of this white paper. If the retailer collects 100,000 data points (1,000 products per 10 competitors), each uncollected percent of the data leads to thousands of poorly priced products.

If the collected data volume is less than the required one, it is important for the retailer to see how critical the missing data is:

- Is data missing from the key or secondary competitors?
- In what categories the data is missing?
- Is there any data missing on KVI-products?

Unhealthy Data Insurance

Professional data vendors use a multilevel system to deliver retailers healthy data:

If the parser does not receive a response from a web page, it tries to collect the necessary data several times. If all attempts fail, the parser automatically sends the failure report to the manual matching team.

Manual matches team, in turn, updates the information—sets the correct link to the product or confirm that the product on the selected site does not exist anymore.

Data Correctness

Another crucial indicator when collecting data. It's the correctness of the collected data. For most retailers, it's an entity that is difficult to estimate.

The data collection can be successful according to the data freshness and completeness and the retailer gets required data points yet they're not necessarily relevant. There are several cases competitor sites protect themselves from automatic parsers (bots). In this case, the parser can get distorted values: prices without discounts, no-promo prices, etc.

In order to avoid such cases, the data provider must organize selective visual controlof collected data. Especially when it comes to collecting data from sites that are constantly working on their security and come up with new ways to block parsers.

Such a check can be made by a retailer itself, e. g., if he sees a high price divergence with one of the competitors in monitoring.

$309 $305 $310 $350 $308
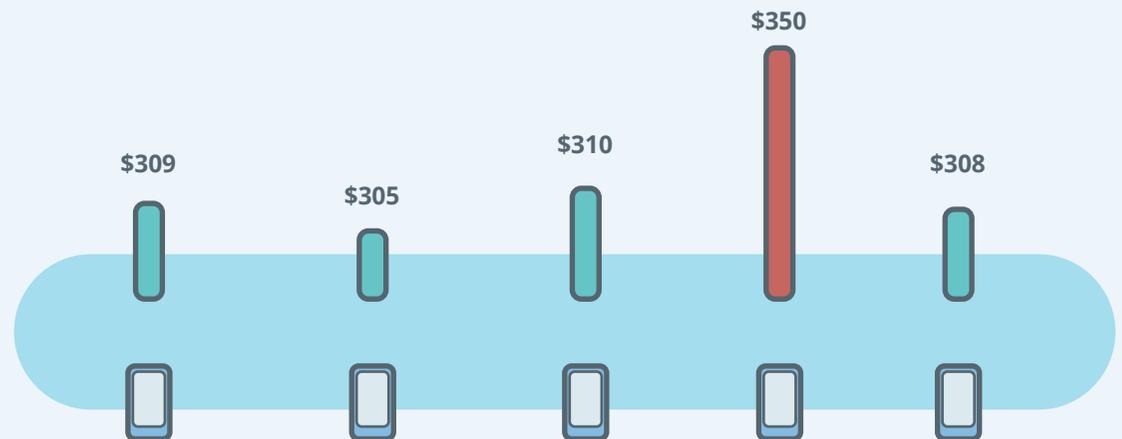
Checking the Quality of Incoming Data

To make correct pricing decisions, the retailer must accurately understand that he is getting high-quality data.

Transparency of data quality should be present at all levels: in assessing its freshness, completeness and correct collection. In this case, the Category Manager will understand that the TVs, for example, have been successfully scanned, and the small household appliances have not enough data.

Information on whether the scheduled data collection has passed successfully and how good this data needs to be shown to the retailer independently of the channel he uses (uploading .csv from the data vendor's interface, or getting them via the API).
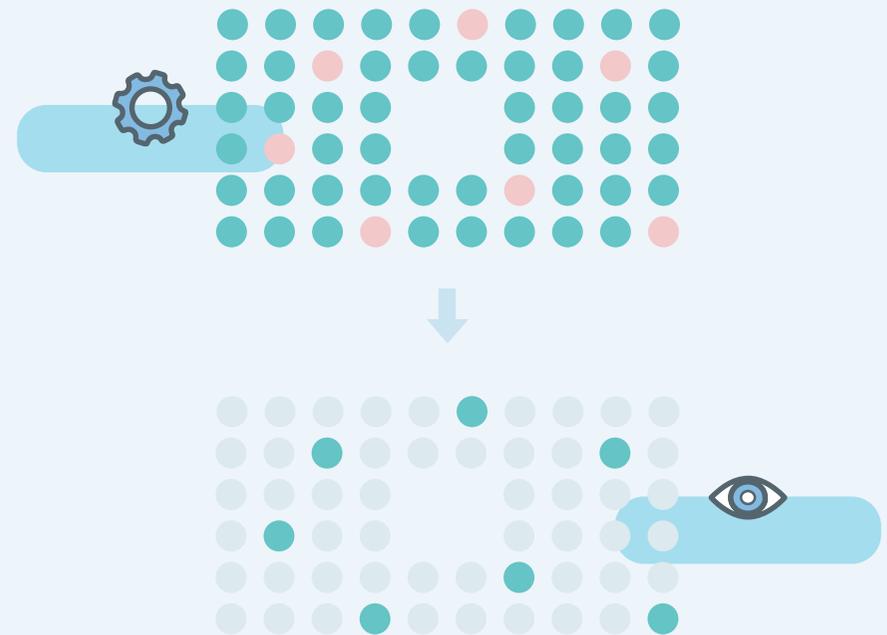
Data Transparency

If the data provider does not highlight data relevance and quality, Category Manager may not pay attention, for example, to its freshness, and change prices, based on outdated or incomplete data. Read more about competitive data quality in the Competera survey.

Cost
of Qualitative Data

Fresh, complete and the correctly collected information is expensive:
Every additional percent of data quality requires infrastructure
and team expanding from the data provider.

If the data provider signs the SLA, which guarantees not less than
98% of data quality (and, for example, that any error will be eliminated
within 24 hours), it needs to create a dedicated team for the particular
client, which will collect data and exclude errors 24/7.

In a different scenario, when the data is collected by a single team for
several clients, some of these clients will receive poor-quality data and,
therefore, will take incorrect pricing decisions.

Data collection and quality control is the resource-intensive process,
so the closer the data quality to 100%, the more expensive every data
point is.

Data vendors that provide data quality close to 100% take on additional
commitments and even "fine" themselves for the unhealthy data.

Price

Quality

90%   100%
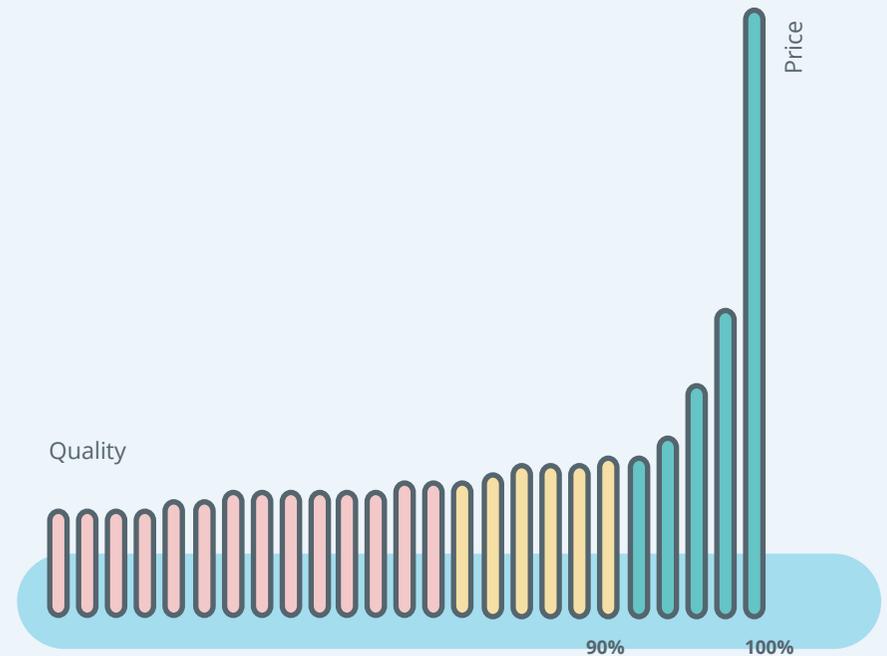
5 Questions to Ask the Data Provider

In a process of choosing a data provider the retailer should ask a series of questions to assess the potential risks:

• Is the supplier ready to sign the data quality SLA and take financial responsibility for its non-compliance?
• How does the provider checks the quality of the collected data and work with the errors that occur?
• Does the supplier provide any quality assessment tools?
• Is there an option to set a custom schedule for data collection?
• Does the supplier has a dedicated team delivering quality data?

If the supplier has answers to these questions and is ready to bear responsibility for the missing, incomplete or mistaken data; If the data quality assessment process is transparent to both parties; If the retailer can independently and quickly assess the quality of incoming data — then the retailer can work with such a supplier.
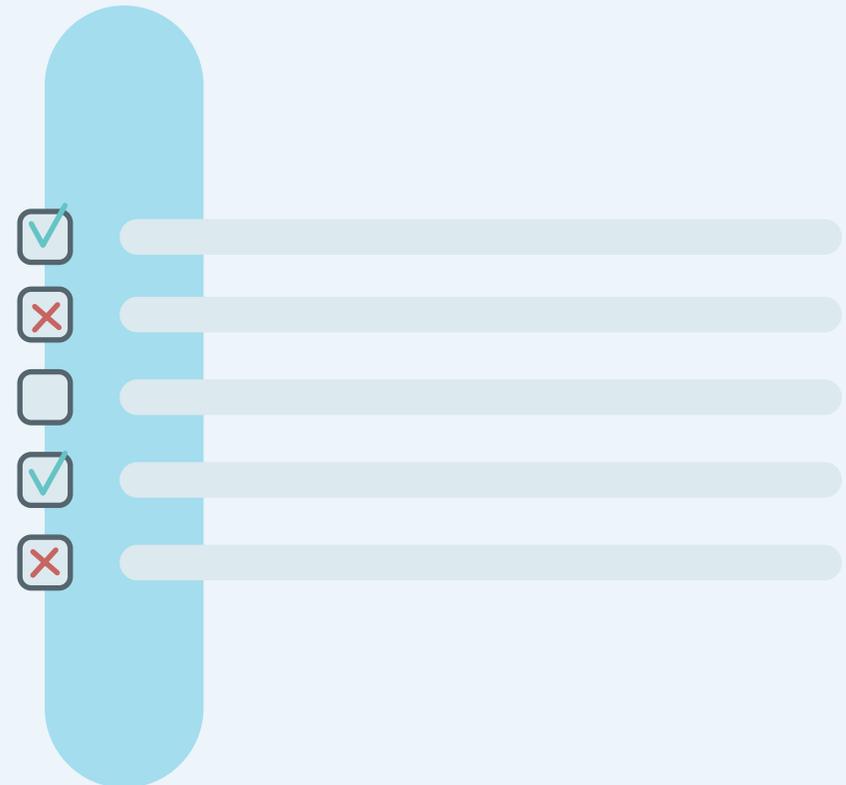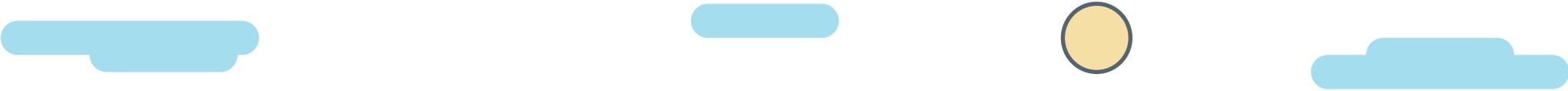
Does your SLA include
data freshness, completeness and correctness?

Does your provider transparently show
the quality of delivered data?

Request a pilot to check the data quality Competera delivers

Request Pilot