

Article

# Changes to the provision of microdata outputs for consumer price inflation statistics: January 2026

Development of new aggregate output statistics designed to meet the existing needs of microdata users, including prototype datasets of synthetic data for each output.

Contact:  
Consumer Price Inflation team  
cpi@ons.gov.uk  
+44 1633 456900

Release date:  
28 January 2026

Next release:  
To be announced

## Table of contents

1. [Overview](#)
2. [Introduction to new microdata outputs](#)
3. [Example datasets on new aggregate output statistics](#)
4. [Regional consumption segment indices and weights](#)
5. [Counts and proportions of indicator marker codes](#)
6. [Improvements to the average prices on the shopping prices comparison tool](#)
7. [Distributions of national and regional price levels](#)
8. [National and regional distributions of price relatives](#)
9. [Data on Changes to the provision of microdata outputs for consumer price inflation statistics](#)
10. [Related links](#)
11. [Cite this article](#)

# 1 . Overview

- We are introducing grocery scanner data into the production of UK consumer price statistics from 2026, greatly improving the quality of our accredited official inflation statistics.
- This will lead to the discontinuation of price quote microdata for the Consumer Prices Index including owner occupiers' housing costs (CPIH) and Consumer Prices Index (CPI) for COICOP Divisions 1 and 2 only.
- This article details the development of new aggregate output statistics that have been designed to meet the existing needs of users of the microdata; we have also published prototype datasets of synthetic data for each output.

## 2 . Introduction to new microdata outputs

We will introduce grocery scanner data into our consumer price inflation statistics in March 2026.

Scanner data refers to data that are collected by retailers at the point of sale (whether in-store or online). There are many benefits to using scanner data, compared with our current sources. These benefits include:

- collecting prices for all products sold by a retailer, rather than relying on a sample
- tracking prices throughout the month, instead of on a single day
- collecting more detailed and timely information on the quantity of each product purchased

Scanner data supports a more comprehensive understanding of how prices and consumer spending patterns are evolving.

We will initially introduce scanner data for approximately 50% of the grocery market. For the remaining percentage, we will continue to manually collect prices through existing (local collection) methods, which include sampling prices in the field, in-store and online.

In this article, we update users on our plans for our [Consumer price inflation consumption segment indices data and price quotes microdatasets](#). These datasets are published on the same day as our inflation figures and supplement our [Consumer price inflation, UK bulletins and datasets](#). They are not part of the accredited official statistics suite of consumer price inflation statistics. Instead, we publish these datasets to aid user research and analysis.

In March 2025, we replaced item indices with consumption segment indices, as described in our [Introducing alternative data into consumer price statistics: aggregation and weights article](#). Consumption segment indices have broader definitions for grocery indices than items.

With the planned introduction of grocery scanner data, we will no longer be able to publish the price quote microdataset in its current form. Our data-sharing agreements with retailers that provide scanner data mean price quote microdata in the price quotes microdataset will be discontinued for:

- Consumer Prices Index including owner occupiers' housing costs (CPIH) and Consumer Prices Index (CPI) Division 1 (food and non-alcoholic beverages)
- CPIH and CPI Division 2 (alcohol and tobacco)
- equivalent RPI categories

This means the coverage of our analytical microdata will be reduced. However, the introduction of large scanner datasets will greatly improve the quality of our accredited official inflation statistics.

We collected user feedback through a questionnaire and targeted focus groups as part of our [User needs from consumer price inflation item indices and price quote microdata consultation](#). In response, we have developed new aggregate output statistics to meet the existing needs of users of the microdata. The new outputs that we have developed are:

- regional consumption segment indices and weights
- counts and proportions of indicator marker codes (for example, sales and recoveries, replacements) that are manually collected in the field
- improvements to the average prices used in our [Shopping prices comparison tool](#)
- national and regional distributions of price levels
- national and regional distributions of price relatives

We present information on each of these outputs and how they are calculated in this article. We have also published two prototype datasets of synthetic data to help users explore and understand how the new outputs are intended to be presented. Our Monthly consumer prices aggregate microdata example dataset is an example microdataset for regional consumption segment indices and weights, indicator marker codes, and national and regional distributions of price levels and price relatives. Our January average shopping prices example dataset includes average prices improvements. These datasets are populated with synthetic data and do not reflect actual inflation data.

We intend to begin publishing regional consumption segment indices and indicator marker counts from March 2026, with the remaining outputs following in the summer. We will continue to provide analytical microdata for Divisions 3 to 12 in our price quotes microdataset. There will be no changes to the availability of data in the monthly consumption segment indices dataset.

These outputs are provided to support user research and understanding of consumer price inflation data and are not accredited official statistics. Smaller sample sizes at consumption-segment and regional levels may introduce volatility in the indices, compared with higher level aggregates. These indices should be interpreted in the context of longer-term trends and supplementary information.

We welcome any feedback on the outputs presented in this article to [cpi@ons.gov.uk](mailto:cpi@ons.gov.uk).

### **3 . Example datasets on new aggregate output statistics**

We have published two example datasets containing synthetic data to support understanding of the microdatasets described in this article.

Our Monthly consumer prices aggregate microdata example dataset includes worksheet tables for:

- regional consumption segment indices and weights
- counts and proportions of indicator marker codes
- national and regional distributions of price levels

Our January average consumer prices example dataset includes a worksheet table for:

- improvements to the average prices used in our [Shopping prices comparison tool](#)

Both datasets include example "Cover Sheet", "Contents" and "Notes" worksheets. The "Cover Sheet" worksheet summarises the file and includes publication dates, information on notes, acronyms, sources, contact details, and relevant links. The "Contents" worksheet provides a table of contents with links to each worksheet. The "Notes" worksheet sets out cell-specific notes and clarifications applicable to the workbook.

Some worksheets are labelled with numbers and letters (for example, "2a" and "2b"). Worksheets sharing the same numerical identifier relate to the same output, with the accompanying letter indicating different versions of that output.

## 4 . Regional consumption segment indices and weights

This worksheet table in our Monthly consumer prices aggregate microdata example dataset provides regional breakdowns of consumer price indices for consumption segments that have regional stratification. We currently publish consumption segment indices. However, these regional components are used in the live calculation of consumer price statistics and therefore offer users additional insight into regional inflation trends.

The purpose of this table is to:

- support regional inflation analysis
- enable construction of regional aggregates
- enhance transparency, particularly following the removal of grocery price quote data

Indices are produced by aggregating local collection and scanner data (where available) into regional consumption segment indices, which are then aggregated to form consumption segment indices. This process is described in our [Introducing alternative data into consumer price statistics: aggregation and weights article](#).

The table will include the following variables:

- INDEX\_DATE - the date for which the index is calculated (year and month)
- CS\_ID - consumption segment identifier
- CS\_DESC - consumption segment description (for example, "pub-hot meal")
- REGION\_ID - region identifier (Classification of Territorial Units for Statistics (NUTS) 1 code)
- REGION\_DESC - region description (for example, "South East (England)")
- CPI\_REGION\_INDEX - Consumer Prices Index (CPI) regional index for the given region and consumption segment
- RPI\_REGION\_INDEX - Retail Prices Index (RPI) regional index for the given region and consumption segment
- REGION\_WEIGHT - regional weight for the given region code and consumption segment (weights within a consumption segment sum to 1)

This table uses within-year (or unchained) indices. Consumption segments without regional stratification are excluded from the table. We have not included a CPIH\_REGION\_INDEX column in this table as the values of this column would be the same as those in the CPI\_REGION\_INDEX column. Price quote data for non-groceries items will still be available to users. Existing national level consumption segment indices will also still be available on our microdata page.

From March 2026, Classification of Individual Consumption According to Purpose (COICOP) Divisions 1 and 2 (groceries) are calculated from a combination of existing collection methods and scanner data; all other divisions in this dataset use solely existing collection methods.

As sample sizes at the consumption segment and regional level are smaller, these indices may be more volatile than higher-level aggregates. As such, these indices should be interpreted in context of longer-term trends and supplementary information.

We have published an example of this table using synthetic data in Tab 1 of our Monthly consumer prices aggregate microdata example dataset to support the understanding of this approach.

## 5 . Counts and proportions of indicator marker codes

Counts and proportions of indicator marker codes in our Monthly consumer prices aggregate microdata example dataset summarise counts and proportions of local collection indicator markers for all items across the basket. These markers are assigned by price collectors. They flag special features of price quotes, such as sales, temporary stock issues or product changes (see Section 6.2.6: Indicator codes of our [Consumer Prices Indices Technical Manual, 2019](#)).

This table helps provide context for interpreting higher level index movements, such as sudden price shifts for specific basket items or changes in product characteristics that affect comparability and may require weight or size adjustments. It also supports research into product availability (for example, environmental or conflict impacts), seasonal sale activity, product shortages, and substitution patterns. Price quotes from scanner data retailers are excluded. Scanner data are excluded from this dataset as it relies on a three-week average price, which may hide changes in promotional events or availability. Therefore, this analysis does not correspond reliably to scanner data.

Each monthly analysis includes two tabs, labelled "a" and "b" (for example, 2a and 2b), within the wider dataset. These are:

- All Quotes - counts and proportions for indicator markers for all sample data price quotes per item and COICOP5, plus an overall total
- Valid quotes - the same metrics, but where all quotes have been subset to exclude quotes flagged during quality assurance; only includes quotes that are considered "valid" and used in the given month's index calculations

The dataset has three hierarchy levels (ID\_TYPES):

- Item - the most granular level, providing metadata for individual basket items (for example, "Basmati Rice")
- COICOP5 - groups items into their corresponding COICOP5 categories (for example, "CP01111 - Rice")
- Grand Total - sums all price quotes across the full basket

The table will include the following variables:

- INDEX\_DATE - year and month the dataset corresponds to
- ID\_TYPE - row type ("Item" for item level, "COICOP5" for COICOP5 level, and "Grand Total" for total row)
- ID - row identifier
- ID\_DESC - row description
- CS\_ID - consumption segment identifier
- CS\_DESC - consumption segment description; this will be blank for COICOP5 and Grand Total
- NO\_INDICATOR to Z - counts for each indicator, with one column per indicator code (defined in Section 6.2.6: Indicator codes of our [Consumer Prices Indices Technical Manual, 2019](#))
- SUM\_OF\_QUOTES - total quotes for the row
- NO\_INDICATOR\_PROPORTION to Z\_PROPORTION - proportions for each indicator, relative to the row total

We have published an example of this table using synthetic data in Tab 2 of our Monthly consumer prices aggregate microdata example dataset to support the understanding of this approach.

## 6 . Improvements to the average prices on the shopping prices comparison tool

This output aims to improve the granularity and accuracy of average price data presented in our [Shopping prices comparison tool](#), following the introduction of grocery scanner data into Classification of Individual Consumption According to Purpose (COICOP) Divisions 1 and 2. The tool allows users to track how average prices change over time across more than 450 consumption segments in the consumer prices basket.

Average prices are currently calculated at consumption-segment level across the entire basket using local collection data, which are samples of prices collected in the field, in-store or online. With the inclusion of grocery scanner data, which are data collected by retailers at the point of sale, we will lower the level of aggregation presented for Divisions 1 and 2 to a level that aligns with basket items, but with more broad definitions ("product types"). Averages will be based on combined weighted aggregates of local collection and scanner data, where applicable. As a result, users will see price points for more specific product types, such as "Microwave rice" and "Basmati rice", rather than the broader consumption segment, such as "Rice, in all forms". Units of measurement will also be included, making comparisons more meaningful and easier to interpret.

For COICOP Divisions 3 to 12 (non-groceries), we will also introduce improvements to the calculation process by using a more robust system for processing calculations. However, these changes will not result in notable differences to published outputs in the Shopping prices comparison tool for these divisions.

### Current data collection and the introduction of scanner data into average price calculations

Average prices in our Shopping prices comparison tool are currently calculated using weighted arithmetic mean price quotes collected each January. For subsequent months, the January average price is adjusted using the Consumer Prices Index including owner occupiers' housing costs (CPIH) consumption segment index growth rate. This approach avoids misleading movements that would arise from changes to the sample of collected items.

However, users should be aware that the January sample of collected prices quotes is designed to measure price change, rather than price levels. Over longer periods, sample and quality changes over link periods can also cause average prices to behave unexpectedly. For more information, see the Average prices subsection of our [Shopping prices comparison tool page](#).

Currently, these calculations are only drawn from local collection data. These include:

- prices for basket items gathered directly in shop locations across the UK
- prices collected from our head office, where national pricing is applied,
- prices for items where most expenditure occurs online, through brochures, or similar formats

For more information, see Section 6: Collection of prices of our [Consumer Prices Technical Manual, 2019](#).

This approach ensures broad coverage across a range of retailers, including independent retailers, nationwide supermarkets chains and online sellers, which are aggregated using appropriate retailer and regional weights. However, as a sample, local collection data are less comprehensive than scanner data and do not include quantity information.

In contrast, scanner data consist of full point-of-sale records that include total expenditure and quantity sold for each product. Average prices can therefore be calculated as quantity-weighted prices by dividing total expenditure by the total quantity, giving an average price paid per unit for each product (sometimes referred to as a unit value). The scanner data being introduced will cover around half of the grocery market and provide a near-complete record of sales for participating retailers, though publication of shop-level averages is not possible for disclosure reasons.

A notable advantage of scanner data is their scale. Local collection provides between 100 to 300 price quotes per month across the UK per basket item, whereas scanner data capture a full census of transactions. This richer information will be used to calculate more representative average prices for Divisions 1 and 2, by calculating averages separately and then aggregating them using retailer weights and the methods described in the following subsection.

## Methods and limitations of combining data sources

Average prices for grocery categories will combine both scanner and locally collected data. This means that the products included in these calculations must share compatible units of measurement. Local collection data are based on basket items with well-defined weight or volume ranges, such as "Basmati rice, 500 g to 1 kg", so that comparable products can be collected consistently. In contrast, scanner data includes the full range of pack sizes and units sold in stores, ranging from single products to large multipacks, and may be priced individually, per pack, by weight, or by volume.

To ensure the two data sources are comparable, scanner products are filtered to only include those matching the units defined in the corresponding basket item description, with weights or volumes adjusted where necessary. For example, for the basket item "Basmati rice, 500 g to 1 kg", basmati rice products in the scanner data priced as "per pack" or other incompatible units are excluded. Prices expressed as "per gram" or "per kilogram" are then multiplied using an appropriate factor derived from the basket item description. These adjusted scanner prices are then expressed in the same unit as the basket item so the two data sources can be combined. For more information on size-adjusted prices, see the Calculating unit values for products subsection of Section 4 in our [Introducing grocery scanner data into consumer price statistics methodology](#).

As scanner prices are filtered to match the units specified in each basket item, the resulting figures cover comparable forms of a product type, but not all forms of that product. The CPIH is a stronger measure of price change because it uses price relatives, which are the percentage change (or ratio) of a product's price between two periods, that are then combined into indices. This approach compares rates of change across items, regardless of pack size or units.

After filtering, we apply outlier detection to remove scanner retailer averages that are more than three times higher than (or one-third of) the equivalent local collection average price. This avoids including anomalies that are common in large scanner datasets, such as miscoding or extreme promotions, that would distort averages. Therefore, some records may be excluded in a non-random manner. For example, if unusual pricing mainly occurs within specific retailers, the remaining data may be less representative and comparable.

We then calculate weighted average prices for each grocery product, using combined weighted aggregates of local collection and scanner data, where applicable. To preserve confidentiality, an average price will only be published when it has contributions from all scanner retailers and local collection data. If a scanner retailer is missing, the average price will be solely based on local collection data.

As with previous years, some average prices are manually omitted from our Shopping prices comparison tool, where the consumption segment definition is too broad. For example, "Small pet mammal" covers several pets with differing characteristics and prices.

Our Shopping prices comparison tool is designed for comparing price levels across different product types or consumption segments within the same series. It should be used alongside other price statistics to provide a more complete view of the changing cost of goods and services.

Price indices, particularly the CPIH, are the most representative indicator of overall price movements. This is because they use price ratios (relatives) that are comparable across units of measurement, allowing us to use all data. Data collected for indices is purposefully sampled to support index construction. This means that using it to produce average prices involves compromises that may reduce their representativeness. For this reason, average prices are not measures of overall inflation and price indices should always be preferred over the rate of change in average prices.

## January average consumer prices example dataset

We have published an example of this table using synthetic data in Tab 1 of our January average consumer prices example dataset to support the understanding of this approach. For subsequent months, these average prices would be adjusted using the CPIH consumption segment index growth rate. The dataset includes the following variables:

- ID\_START - year and month the ID was added to the basket; for Division 1 and 2, this is set to "202502" to align with the [introduction of consumption segments into the aggregation structure](#)
- ID\_NAME - name of product type (for example, "Bread, sliced, white" for Divisions 1 and 2) or consumption segment (for example, "Electricity" for Divisions 3 to 12)
- CONSUMPTION\_SEGMENT\_CODE - consumption segment code
- CONSUMPTION\_SEGMENT\_DESC - consumption segment description
- Category 1/Category 2 - broad categories for classification in our Shopping prices comparison tool
- COICOP levels - COICOP5, COICOP4, COICOP3, or COICOP2
- WEIGHT\_SIZE - weight and units for the average price (for example, "1000 g")
- AVERAGE\_PRICE - the calculated average price for the ID; blank entries indicate cases where the average prices has been excluded

## 7 . Distributions of national and regional price levels

These tables in our Monthly consumer prices aggregate microdata example dataset present statistical bins for constructing histogram distributions and summary statistics for box plot analysis of monthly prices for each basket item within Classification of Individual Consumption According to Purpose (COICOP) Divisions 1 and 2 (groceries).

These monthly tables are constructed across two hierarchy levels: product types and regional. Product types are aggregate measures of all regions for each product. Users can visualise price dispersion, volatility, and other characteristics relevant to price behaviour analysis by product type. At regional level, price levels are split into 12 UK Nomenclature of Territorial Units for Statistics major socio-economic regions (NUTS1) codes and similar trends can be visualised.

Each table will include the following variables:



- INDEX\_DATE - year and month of the dataset
- ID\_DESC - description of product (for example, "Bread, sliced, white")
- WEIGHT\_SIZE - the weight and size of the product (for example, "1000 g")
- CS\_ID - consumption segment code
- CS\_DESC - consumption segment name
- COICOP5\_ID - COICOP5-level identifier
- COICOP5\_DESC - COICOP5-level description
- COICOP4\_ID - COICOP4-level identifier
- COICOP4\_DESC - COICOP4-level description
- REGION\_ID - region identifier
- REGION\_DESC - description (which will be blank for product-type level)
- MEAN - weighted arithmetic mean
- MEDIAN - weighted median (50th percentile)
- MODE - weighted mode
- VARIANCE - weighted variance
- STD\_DEV - weighted standard deviation
- Q1 - weighted first quartile (25th percentile)
- Q3 - weighted third quartile (75th percentile)
- MIN - minimum value
- MAX - maximum value
- Lowest bin (for example, "less than 0.10") - the lowest bin representing the proportion of products with price levels less than the value stated (for example, less than £0.10)
- Middle bins (for example, "[0.10, 0.20)" to "[10.00, 10.10)") - intermediate bins rising in equal increments, each showing the percentage of products within their left-inclusive range (for example, 100 bins increasing in £0.10 increments)
- Highest bin (for example, "greater than or equal to 10.10") - the highest bin representing the proportion of products with price levels greater than or equal to the value stated (for example, greater than or equal to £10.10)

Price-level distributions represent the range of actual prices observed, rather than price relatives that measure the change of prices between two periods as a ratio or percentage (calculated by dividing the current month's price by the base or previous month's price). As such, there may be substantial variation in price levels across different products.

To present these distributions effectively, products are grouped by COICOP sections (COICOP3) within Divisions 1 and 2. We have created three additional subgroups in each section using 5 pence, 10 pence, and 20 pence intervals, based on price level groupings. This results in three worksheets per COICOP3 section. Two additional worksheets, using 40 pence and 80 pence intervals, are included for product types with higher price-level distributions. While these groupings lead to different bin ranges across tables, each table has the same number of bins for consistency, with bin widths adjusted to reflect the price variation within each group.

For example, COICOP3 section "Food" includes products such as "Beef, roasting joint, per kg", "Olive Oil, 500ml - 1 Litre", and "Lemon, each". There are substantial price differences between products in this section, so the data are split across three tables, each with its own bin categories. Products like "Fresh Fish, salmon fillets, per kg" may require wider bins to accommodate greater price dispersion. Products like "Lemon, each" use narrower bin widths to capture smaller price variations.

These tables are produced using both local collection and scanner data, where available. Some rows may include only local collection data, where scanner data are limited or unavailable. In cases where a row has fewer than 10 observations, the sample size is considered too small for reliable results. This limitation typically occurs only at the regional level. In such cases, a row will still be provided, but all summary statistics and bins will be marked with "[x]" for not available.

To ensure comparability between scanner data and local collection items, we use the methods described in [Section 6: Improvements to the average prices on the shopping prices comparison tool](#). Scanner products are aligned with the units defined in the corresponding basket item description, with weights or volumes adjusted where necessary. They are then filtered to those item specifications, before combined average prices are calculated.

Outliers are then removed from scanner data by excluding price quotes that are more than three times higher than (or one-third of) the calculated average price. These outlier detection thresholds and reasonings are consistent with the rules described in Section 5. After filtering, summary statistics and bin calculations are derived.

The mean prices in these tables represent actual observed monthly averages. This is unlike the methods described in Section 5, where average prices are calculated in January and then uprated monthly using Consumer Prices Index including owner occupiers' housing costs (CPIH) consumption segment indices. These figures are not comparable across monthly files because of changes in the underlying sample of items, stores and transaction volumes. For this reason, they also cannot be directly compared with the averages shown in our Shopping prices comparison tool.

Each bin represents a monthly price level grouped in variable intervals. For example, the bin "[0.40, 0.50)" corresponds to the proportion of products that are greater than or equal to £0.40 and less than £0.50. The bins are left inclusive, meaning the lower value in the bin range is included and the upper value is excluded. For each ID, entries within each bin reflect the percentage of products falling into that bin, with the sum across bins equal to 100. These bins and summary statistics are weighted using corresponding regional and retailer-type weights, where applicable. The spread of price levels may be wide in some instances because scanner data have numerous data points across a wide range of products per item, and local collection covers both independent and chain stores.

We have published an example of this table using synthetic data in Tabs 3a, 3b, 3c, 3d, and 3e of our Monthly consumer prices aggregate microdata example dataset to support the understanding of this output.

## 8 . National and regional distributions of price relatives

This table in our Monthly consumer prices aggregate microdata example dataset provides statistical bins for constructing histograms and summary statistics for box plot analysis of month-on-month price relatives within Classification of Individual Consumption According to Purpose (COICOP) Divisions 1 and 2 (groceries).

Like price-level distributions, this monthly table is structured across two hierarchy levels: product types and regional. At both levels, this table supports the identification of trends, and other characteristics relevant to price behaviour analysis. The regional level offers further insight into geographic price dispersion and price-change behaviours across regions.

The table will include the following variables:



- INDEX\_DATE - year and month of the dataset
- ID\_DESC - description of product (for example, "Bread, sliced, white")
- CS\_ID - consumption segment code
- CS\_DESC - consumption segment name
- REGION\_ID - region identifier
- REGION\_DESC - region description, which will be blank for product-type level
- MEAN - weighted arithmetic mean
- MEDIAN - weighted median (50th percentile)
- MODE - weighted mode
- VARIANCE - weighted variance
- STD\_DEV - weighted standard deviation
- Q1 - weighted first quartile (25th percentile)
- Q3 - weighted third quartile (75th percentile)
- MIN - minimum value
- MAX - maximum value
- Less than 0.33 - bin representing the proportion of products with a price relative less than 0.33
- "[0.33, 0.35) to [2.31, 2.33)" - bins in 0.02 increments, each showing the percentage of products within its left-inclusive range
- Greater than or equal to 2.33 - bin showing the percentage of products with a price relative of greater than or equal to 2.33

Month-on-month price relatives are calculated by dividing the current month's price by the previous month's price. This table is produced using both local collection and scanner data, where available, similar to [Section 7: Distributions of national and regional price levels](#). However, unlike price level distributions, price relatives do not require products from local collection and scanner data to be aligned to identical units of measurement based on the corresponding basket item description. Instead, price relatives are calculated separately for individual products within local collection and scanner data and then combined using appropriate weighting. This approach means all available data can be used from both data sources.

Each bin represents a month-on-month price relative grouped in intervals of 0.02. For example, the bin "[1.01, 1.03)" corresponds to the proportion of products that have a month-on-month price change greater than or equal to a 1% increase and less than a 3% price increase. The bins are left inclusive, meaning the lower value in the bin range is included and the upper value is excluded. For each row, entries within each bin reflect the percentage of products falling into that bin, with the sum across bins equal to 100. These bins and summary statistics are weighted using corresponding regional and retailer-type weights for both breakdown levels, where applicable.

To make interpretation easier for important reference points, the bins are shifted so that noteworthy values of interest fall at the centre of the bin. For example, most data points centre around a price relative of 1 (indicating no price change), so the bin "[0.99, 1.01)" captures this minimal price movement. A bin width of 0.02 was chosen to provide sufficient detail, while balancing a manageable number of bins. The full range of the bins spans from less than 0.33 to greater than or equal to 2.33.

Outlier detection removes extreme price changes from both data sources where the calculated price relative is greater than 4 or less than 0.25, which is equivalent to a 300% price increase and a 75% price decrease, respectively. Like the other outlier detection methods in this article, this approach allows for promotions while filtering out scanner data anomalies and implausible price changes that distort measures. For more information, see the Data cleaning subsection of [Section 4: Grocery scanner data: methods](#) in our Introducing grocery scanner data into consumer price statistics methodology.

Some rows may include only local collection data, where scanner data are limited or unavailable. Rows with fewer than 10 observations are considered to have too small a sample size for reliable results. This limitation typically occurs only at the regional level. In such cases, a row for these IDs will still be provided, but all summary statistics and bins will be marked with "[x]" for not available.

It should be noted that the month-on-month price relatives used in these distributions do not necessarily match the relatives used in consumer price inflation calculations. Local collection data uses January-to-month price relatives, rather than consecutive month-on-month changes. For scanner data, multiple base periods are used across the 25-month measurement window to account for product turnover and variable availability. Therefore, these distributions should be interpreted as indicative movements within each data source, rather than the exact relatives feeding into headline inflation.

We have published an example of this table using synthetic data in Tab 4 of our Monthly consumer prices aggregate microdata example dataset to support the understanding of this output.

## 9 . Data on Changes to the provision of microdata outputs for consumer price inflation statistics

### [Monthly consumer prices aggregate microdata example](#)

Dataset | Released 28 January 2026

Example dataset using synthetic data to introduce users to our new monthly aggregate consumer prices microdata outputs.

### [January average consumer prices example](#)

Dataset | Released 28 January 2026

Example dataset using synthetic data to introduce users to our new January average consumer prices microdata outputs.

## 10 . Related links

### [Consumer price inflation, UK](#)

Bulletin | Released monthly

Price indices, percentage changes, and weights for the different measures of consumer price inflation.

### [Consumer prices development plan: updated September 2025](#)

Article | Released 18 September 2025

An update on the proposed development plan for consumer price statistics. Includes the rationale for our priorities and their potential impact.

### [Transformation of consumer price statistics](#)

Article series | Last updated 6 August 2024

Our plans to transform UK consumer price statistics by including new improved data sources and developing our methods and systems.

### [Impact analysis on transformation of UK consumer price statistics](#)

Article series | Last updated 28 January 2026

This series demonstrates the indicative impacts on our headline consumer price statistics of introducing new data and methods into our consumer price inflation statistics.

## 11 . Cite this article

Office for National Statistics (ONS), released 28 January 2026, ONS website, article, [Changes to the provision of microdata outputs for consumer price inflation statistics](#)