

Scanner data use in Statistics Canada's Retail Trade Program

Mathieu Thomassin and Saman Rudgar
Retail and Service Industries Division, Statistics Canada
Voorburg, Netherlands
2026-09-22



Statistics
Canada

Statistique
Canada

Canada

Outline

- Background & Motivation
- Scanner Data Use in Retail Program
- Retailer Data Feasibility
- Machine Learning Classification Approach
- Project Evolution and Current State
- Project Maintenance

Background & Motivation

History and Use of Scanner Data in the Retail Trade Program

Background & Motivation

- Historically, data collection often relied on in-person collection conducted by interviewers.
- In 2015, Statistics Canada identified an opportunity to reduce retailer response burden and improve data collection efficiency through the integration of scanner data into statistical programs, particularly those measuring price change over time, such as the Consumer Price Index (CPI), as well as retail sales and output.

Background & Motivation

- The first retailer served as a pilot "test case" for scanner data integration.
- Agency-wide discussions explored potential uses of the data across divisions
- By 2016, extensive back-and-forth between divisions created delays in defining data requirements
- To streamline the process, the focus was placed on meeting the needs of both the Consumer Price Index and the Retail Program
- Over the last 10 years, Statistics Canada has partnered with major retailers to obtain scanner data covering millions of consumer transactions

Requests for information - Retailer Point of Sale data (also known as scanner data or transactional data)

Under the authority of the [Statistics Act](#), Statistics Canada is hereby requesting the following information which will be used solely for statistical and research purposes and will be protected in accordance with the provisions of the *Statistics Act* and any other applicable law. This is a mandatory request for data.

What information is being requested?

What personal information is included in this request?

What years of data will be requested?

From whom will the information be requested?

Why were these organizations selected as data providers?

Note: All scanner data are provided voluntarily and free of charge by retailers to Statistics Canada.

Why is this information being requested?

- Scanner data represent a highly valuable source of information for statistical programs.
- Their use has enhanced the quality of estimates produced by the Retail Commodity Survey (RCS) and the Monthly Retail Trade Survey (MRTS) in several ways.
- They improve geographic coverage and representativeness by providing access to the complete universe of products sold within a store.
- They also increase the effective sample size by ensuring that all stores for which scanner data are available are included in the sample.
- In addition, scanner data improve the timeliness and accuracy of estimates while significantly reducing data collection costs.

Challenges

- While the integration of scanner data into official statistics initially presented challenges, many of these have been significantly reduced over time. Advances in machine learning and the maturation of processing methods have greatly improved product classification and the management of evolving product identification codes. Ongoing monitoring and validation remain important to ensuring data quality and consistency.
- Scanner data can significantly reduce non-sampling errors, especially coverage and non-response errors, by providing detailed transaction-level information directly from retailers. However, classification errors remain a potential source of bias when products are incorrectly assigned to commodity categories because of inaccurate descriptions, product code changes, or limitations in the classification process.
- Considerations for handling large files and data volumes, such as data transmission, storage, and processing, must be top of mind.

Scanner Data Use in Retail Program

Scanner Data Use in Retail Program

- **Retail Trade Survey** collects data on:
 - Sales and e-commerce sales
 - Number of retail locations by geography
 - Retail programs require complete transaction coverage to produce estimates of total retail sales. This level of coverage may not be necessary for other statistical programs.
- **Scanner data** provides as an alternative source:
 - Sales totals and location counts
 - Enables more timely data and reduces response burden, enabling use in the advance indicator program
 - Reduces data collection costs

North American Industry Classification System (NAICS)	Retail trade [44-45]				
Sales	Total retail sales				
Adjustments	Seasonally adjusted				
Geography	October 2025	November 2025	December 2025	January 2026	February 2026
	Dollars				
Canada (map)	70,112,774 ^A	70,985,144 ^A	70,708,740 ^A	71,590,599 ^A	72,062,984 ^A
Newfoundland and Labrador (map)	1,034,448 ^B	1,041,652 ^B	1,003,962 ^B	1,040,308 ^B	1,043,441 ^B
Prince Edward Island (map)	305,960 ^A	306,265 ^A	301,180 ^A	311,324 ^A	314,759 ^A
Nova Scotia (map)	1,882,310 ^A	1,922,846 ^A	1,899,667 ^A	1,930,292 ^A	1,974,537 ^A
New Brunswick (map)	1,602,061 ^A	1,578,334 ^A	1,593,453 ^A	1,607,706 ^A	1,622,220 ^A
Quebec (map)	15,601,219 ^A	15,648,579 ^A	15,748,859 ^A	15,824,140 ^A	16,089,157 ^A
Ontario (map)	26,570,735 ^A	26,736,104 ^A	26,739,658 ^A	26,846,318 ^A	27,127,005 ^A
Manitoba (map)	2,328,604 ^A	2,390,847 ^A	2,377,432 ^A	2,379,781 ^A	2,431,747 ^A
Saskatchewan (map)	2,180,706 ^A	2,207,896 ^A	2,186,340 ^A	2,237,588 ^A	2,270,390 ^B
Alberta (map)	8,872,432 ^A	9,247,734 ^A	8,999,012 ^A	9,372,823 ^A	9,197,631 ^A
British Columbia (map)	9,493,269 ^A	9,665,996 ^A	9,619,953 ^A	9,788,669 ^A	9,746,032 ^A
Yukon (map)	98,692 ^B	99,650 ^B	100,400 ^B	102,421 ^B	101,136 ^B
Northwest Territories (map)	79,976 ^B	79,862 ^A	82,055 ^A	83,407 ^B	78,216 ^B
Nunavut (map)	62,361 ^A	59,379 ^A	56,770 ^A	65,823 ^A	66,714 ^A

Scanner Data Use in Retail Program - ARI

Retail Trade Survey - Advance retail indicator (ARI)

Statistics Canada publishes an Advance Retail Indicator, an early estimate of monthly retail sales that provides a timely indication of retail market performance before the release of the official Monthly Retail Trade figures.

The estimate is typically based on responses received from 50% to 60% of surveyed retailers and is subject to revision as additional data become available.

Retail sales, by industry — Seasonally adjusted

	May 2025	April 2026 ^r	May 2026 ^p	April to May 2026	May 2025 to May 2026
	millions of dollars	millions of dollars	millions of dollars	% change	% change
Total retail trade (current dollars)	69,613	72,951	73,708	1	5.9
Retail e-commerce	4,874	5,102	5,025	-1.5	3.1
Total retail trade (2017 chained dollars)	59,290	59,903	60,081	0.3	1.3
Total (current dollars) excluding motor vehicle and parts dealers	50,432	53,425	54,046	1.2	7.2
Total (current dollars) excluding motor vehicle and parts dealers and gasoline stations and fuel vendors	44,414	45,938	46,330	0.9	4.3
Motor vehicle and parts dealers	19,181	19,526	19,662	0.7	2.5
New car dealers	15,133	15,383	15,463	0.5	2.2
Used car dealers	1,571	1,665	1,680	0.9	6.9
Other motor vehicle dealers	1,134	1,108	1,140	2.9	0.6
Automotive parts, accessories and tire retailers	1,343	1,370	1,379	0.6	2.7
Building material and garden equipment and supplies dealers	3,975	3,827	3,860	0.9	-2.9
Food and beverage retailers	13,047	13,267	13,332	0.5	2.2
Supermarkets and other grocery retailers (except convenience)	9,343	9,435	9,527	1	2
Convenience retailers and vending machine operators	659	678	668	-1.4	1.5
Specialty food retailers	984	1,031	1,030	-0.1	4.7
Beer, wine and liquor retailers	2,062	2,123	2,107	-0.8	2.2
Furniture, home furnishings, electronics and appliances retailers	3,591	3,351	3,373	0.6	-6.1
Furniture retailers	1,218	1,180	1,190	0.8	-2.3
Floor covering, window treatment and other home furnishing retailers	750	709	711	0.3	-5.1
Electronics and appliances retailers	1,624	1,462	1,472	0.7	-9.4
General merchandise retailers	10,113	10,739	10,847	1	7.3
Health and personal care retailers	5,881	6,562	6,612	0.8	12.4
Gasoline stations and fuel vendors	6,018	7,488	7,716	3.1	28.2
Clothing, clothing accessories, shoes, jewelry, luggage and leather goods retailers	3,777	3,951	3,989	1	5.6
Clothing and clothing accessories retailers	2,942	3,100	3,132	1	6.5
Shoe retailers	399	393	398	1.2	-0.4
Jewellery, luggage and leather goods retailers	435	457	459	0.3	5.4
Sporting goods, hobby, musical instrument, book, and miscellaneous retailers	4,030	4,241	4,317	1.8	7.1
Sporting goods, hobby, musical instrument and book retailers and news dealers	1,385	1,421	1,454	2.3	4.9
Miscellaneous store retailers	2,645	2,820	2,863	1.5	8.3
Cannabis retailers	473	488	482	-1.2	1.9

^r revised
^p preliminary

Notes(s):

Figures may not add up to totals as a result of rounding.

Source(s):

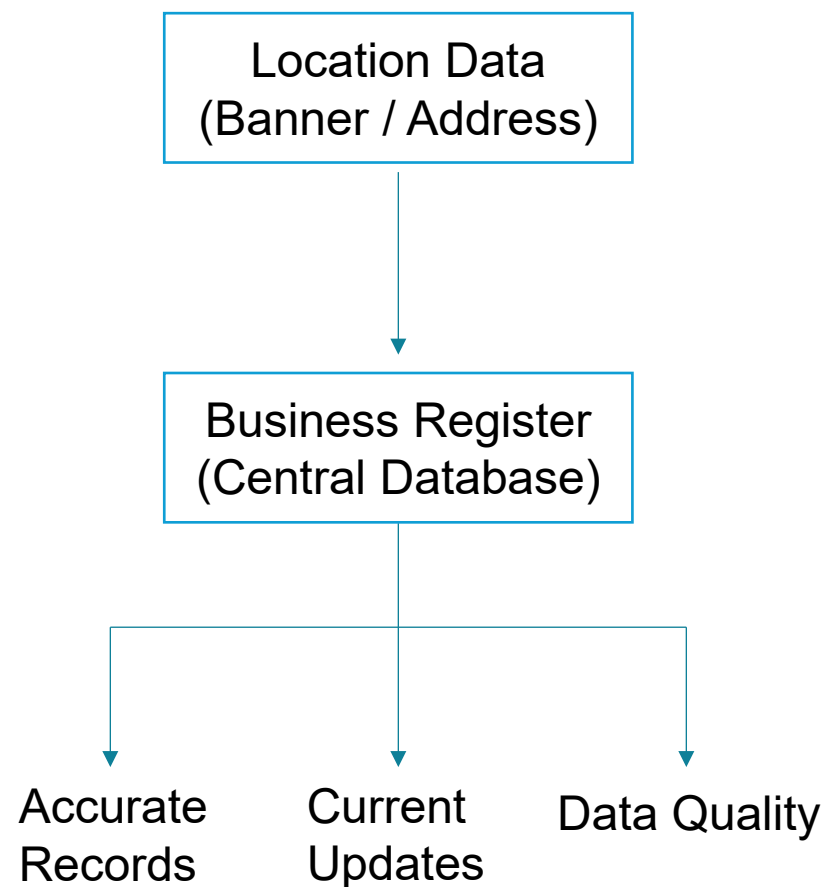
Scanner Data use in Retail Program

- **Retail Commodity Survey** collects data on:
 - Retail commodity sales at the national level
 - North American Product Classification System (NAPCS)
- **Scanner data** provides as an alternative source:
 - Sales by product allows for commodity breakdowns
 - More accurate than conventional respondent data
 - Significantly reduces response burden

Geography	Canada (map)				
North American Product Classification System (NAPCS) Canada 2022 Version	October 2025	November 2025	December 2025	January 2026	February 2026
	Dollars				
Food and beverages at retail [561111]	16,976,871 ^B	16,697,144 ^A	18,841,880 ^B	16,029,634 ^A	15,258,597 ^A
Fresh meat and poultry, at retail [5611111]	1,874,714 ^B	1,749,959 ^A	1,956,595 ^A	1,804,197 ^A	1,668,472 ^A
Fresh fish and other fresh seafood, at retail [5611112]	200,883 ^B	196,064 ^B	229,712 ^C	194,046 ^B	192,047 ^B
Fresh fruit and vegetables, at retail [5611113]	2,050,557 ^A	1,994,832 ^A	2,149,334 ^A	2,218,470 ^A	2,087,712 ^A
Eggs and dairy products (except frozen desserts), at retail [5611114]	1,852,191 ^A	1,848,178 ^A	1,961,466 ^A	1,841,815 ^A	1,733,579 ^A
Baked goods (except frozen products, cookies and crackers), at retail [5611115]	886,590 ^B	822,271 ^B	896,214 ^B	805,974 ^B	773,290 ^B
Perishable prepared foods (including fresh sliced deli meats, prepared entrees and fresh pasta), at retail [5611116]	820,163 ^A	775,310 ^A	872,953 ^A	745,171 ^A	714,486 ^A

Use of Scanner Data Beyond Retail Surveys

- The scanner data is also used to update the Business Register, information is typically required on the business's legal and operating names, address, contact details, industry classification, ownership structure, operational status, employment, revenue, and any organizational changes such as openings, closures, mergers, or acquisitions.
- For example, location data used by the Frame Team to maintain and update the Business Register.
- Helps ensure organizational records remain accurate and current
- Supports data quality across statistical programs



Retailer Data Feasibility

Assessing data availability, structure and key constraints

Retailer Data Feasibility

Coverage

Does it represent the target population?

Quality

Are key fields complete and reliable?

Timeliness

Is it timely and consistent enough for production use?

Integration

Can it link to existing systems?

Effort

How much cleaning and processing is required?

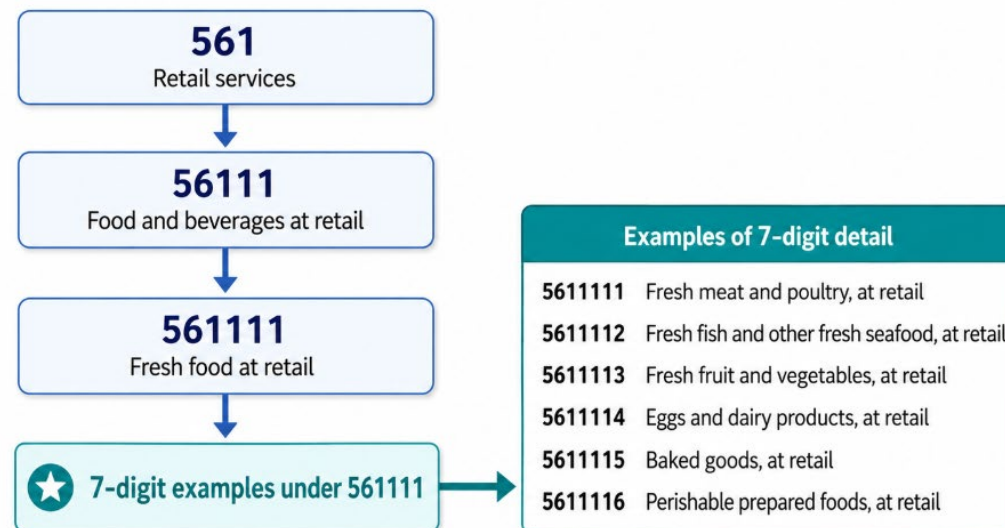
Machine Learning Classification Approach

Approach to automating product classification into NAPCS

Machine Learning Approach

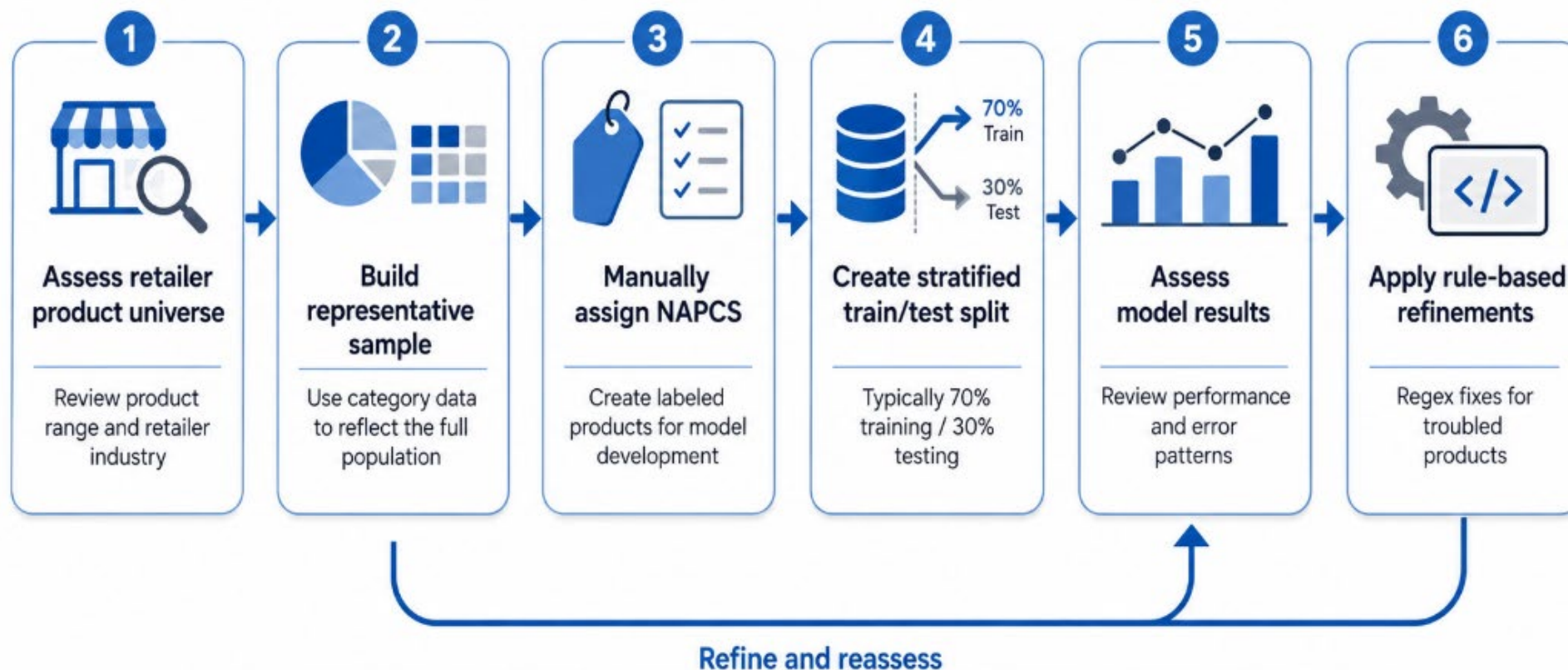
- Retail Commodity Survey relies on NAPCS (North American Product Classification System) for standardized commodity reporting. NAPCS similar to COICOP (Classification of Individual Consumption According to Purpose)
- Retailer products not aligned with NAPCS categories
- Traditional mapping approaches were quickly deemed as labor-intensive, inconsistent, and not scalable
- Machine learning for text classification offers an automated, scalable solution for mapping products to NAPCS

Example of NAPCS Hierarchy: Fresh Food at Retail



Machine Learning Approach

Model Creation and Refinement Process



Machine Learning Approach

Historical approach

- Used an XGBoost model with linear learners, a machine learning approach that identifies patterns in product information by combining multiple simple prediction models.
- Performance was weaker for retailers with limited, inconsistent, or less structured product descriptions.

Current approach

- Implemented a BERT-based model, a language model that better understands the meaning and context of product descriptions.

Future/ Experiments

- Explore the use of Large Language Models (LLMs) to further automate and enhance product classification.
- Develop a single standardized approach that can be applied consistently across all retailers.

Project Evolution

How the pipeline has evolved over time

Project Evolution

Raw .txt / SAS files

- Weekly/monthly scanner files were transformed into SAS datasets
- Processing required reading, chunking, and aggregating files into monthly totals
- Runtime often took several hours per retailer

Netezza

- Improved querying and aggregation through an on-premise data warehouse
- Still required extensive local machine processing
- Large retailer datasets remained resource-intensive

Azure Cloud

- Transactions, locations, and products are stored separately by retailer and month
- Processing is organized directly around monthly totals
- Runtime reduced to approximately 20 minutes per retailer

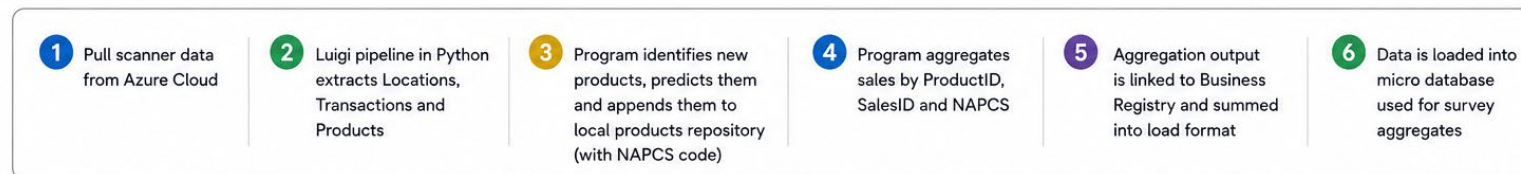
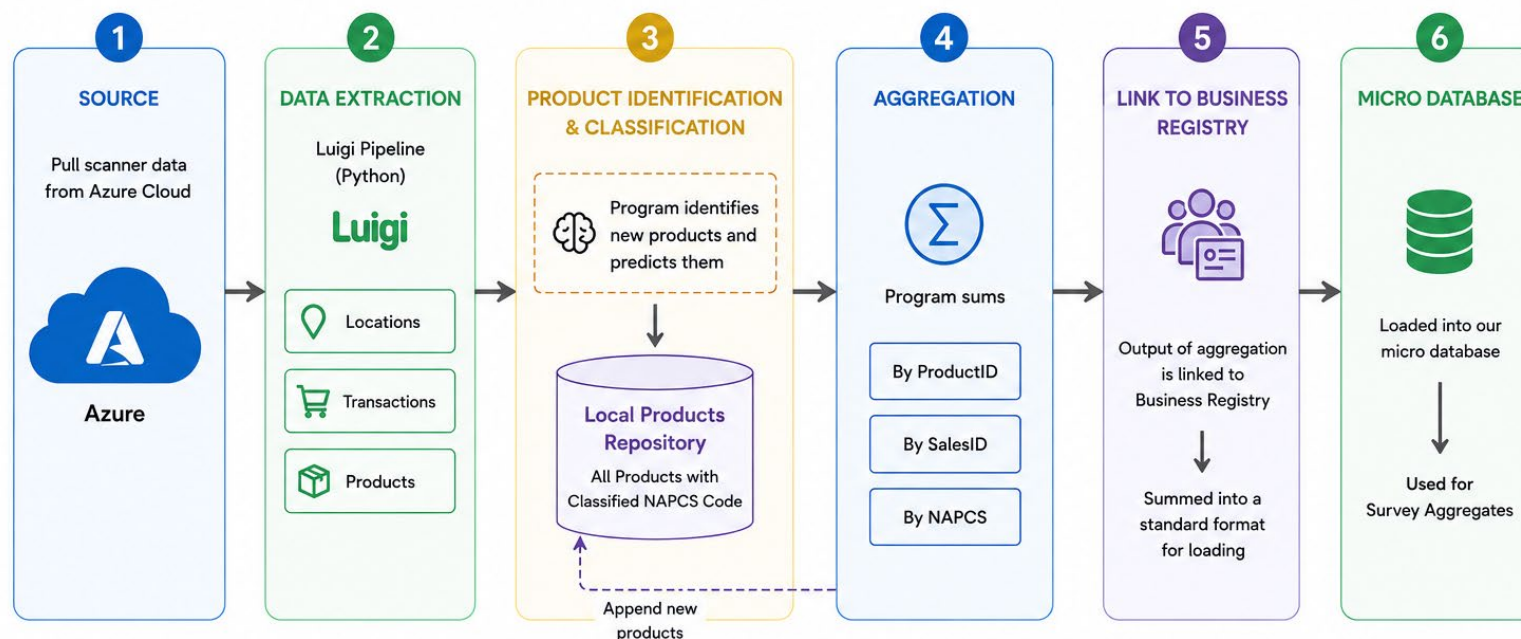
Data Source Evolution



From file-based inputs to
modern cloud infrastructure

Current Pipeline

RETAILER SCANNER PIPELINE



→ Data Flow - - -> Repository Update (New Products)

Project Maintenance

Required maintenance for the models and frame

Project Maintenance

- Monthly methodology reports detect fluctuations in products sold or changes in locations
- Models are retrained at least every 5 years and are aligned with new NAPCS releases (currently NAPCS 2022)
- Codebase is continuously modernized to remain compatible with current Python versions (currently Python 3.13)

Summary of Possible Errors			
Article Description	Site	Sales by Region	Sales by Store_ID
Good	Good	Warning	Warning
Previous previous week		2025-11-21	
Previous week		2025-11-28	
Current week		2025-12-05	

ReadMe

Article Description

Site

Address List

Sales by Region

Sales by Store_ID

+

Conclusion / Questions?

- Scanner data have transformed the production of official statistics, such as the Consumer Price Index (CPI) and retail trade estimates, by providing richer, more timely, and more representative information while reducing response burden on retailers
- The integration of scanner data into Statistics Canada's programs has required strong partnerships, robust governance, and a commitment to transparency. It has also necessitated investments in data processing and IT capabilities

STAY CONNECTED

Eh Sayers podcast

StatsCAN Plus

The Daily

My StatCan

Website

Surveys and statistical programs

Data service centres



Questions? Contact us: infostats@statcan.gc.ca



Statistics
Canada

Statistique
Canada

statcan.gc.ca

Canada 

STATISTICS CANADA

Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

statcan.gc.ca

Canada 