



Statistics Canada

Modernisation of Retail Services Price Index



Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

Canada



Retail Services Price Index: Modernization

- Starting two years ago, the Retail Services Price Index commenced a modernization project to make it more relevant and efficient
- Modernization was tackled on two fronts:
 - Data
 - Production Systems



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada

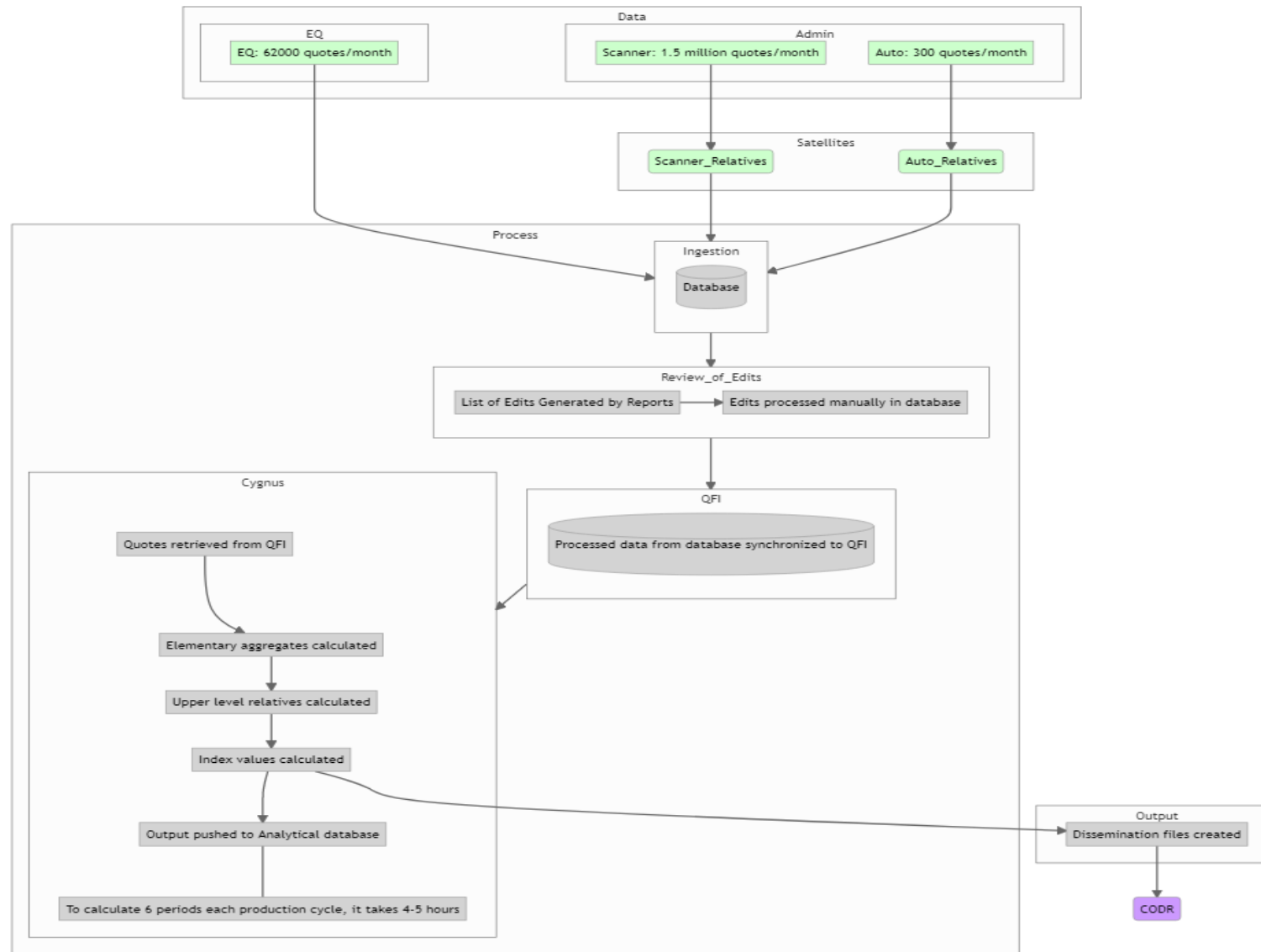


Retail Services Price Index: Data Modernization

- Effort was put into improving the quality of data
- Two major outcomes:
 - Incorporation of Scanner data to improve the quality of our index in NAICS 44511: Supermarkets and other grocery (except convenience) stores
 - Around 1.5 million quotes per month
 - Incorporation of an admin source to measure NAICS 44111: New car dealers in our next basket which was previously not measured in the Retail Services Prices Index
 - About 300 quotes per month
- In addition, products have been NAPCS coded to allow for more detailed analysis at product level indices when needed



Retail Services Price Index: System Modernisation - Flow Chart of Current Process



Retail Services Price Index: Current Process vs New Process

Current Process

Maintenance

- Complex ecosystem with multiple systems
- Changes and improvements require help of IT

Not Flexible

- Process is rigid
- Need to follow pre-defined steps
- Very cumbersome to adapt and implement new baskets

Inefficient

- Processing of data is manual
- Calculation of index takes a long time i.e. 4 to 5 hours for each run

Cost

- Costly to maintain due to multiple systems

New Process

Maintenance

- Simpler ecosystem
- Changes can be done without IT support

Flexible

- Easy to adapt
- New processes can be added easily

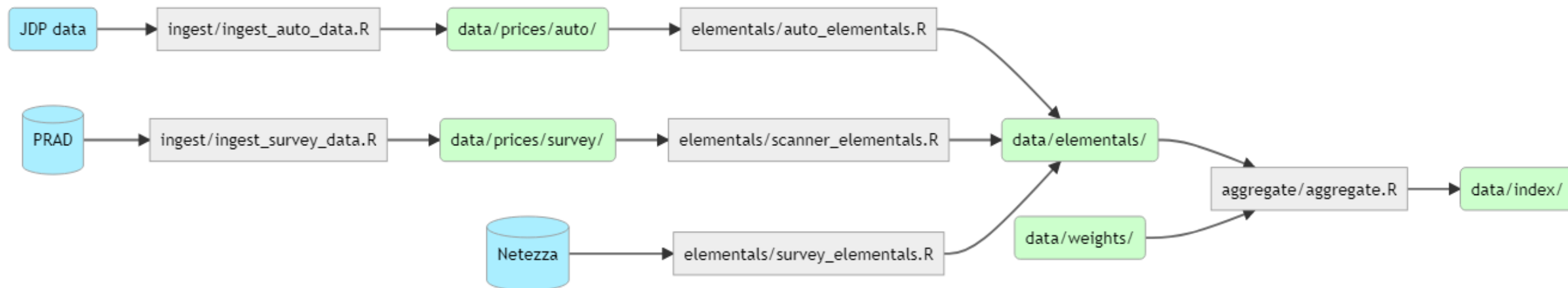
Efficient

- Much faster to calculate: minutes instead of hours
- Processing of data can be automated

Cost





- Uses open-source software
- Less IT resources needed

Retail Services Price Index: New Process Flow Chart



Retail Services Price Index: Demo




automated price indexes · rspi







 **rspi**  Project ID: 5743  Star 0  Fork 2

96 Commits 15 Branches 0 Tags 2 MB Project Storage

An automated process to make the Retail Services Price Index

master rspi / Find file Web IDE Clone

A Merge branch 'contributing-guidelines' into 'master'  Ambuj Dewan authored 1 day ago  bf75d401 

 README  CHANGELOG  CONTRIBUTING  CI/CD configuration  Add LICENSE  Configure Integrations

Name	Last commit	Last update
aggregate	Added mention of contributions to relevant ...	1 week ago
analysis	Corrected export location.	1 day ago
data	Added contributions subfolder to the data f...	3 weeks ago
disseminate	Forced readline and removed include.only fo...	1 week ago
elementals	Added mention of contributions to relevant ...	1 week ago
ingest	Survey query now uses FinalValue to reflect ...	6 days ago
sample_management	substitute markdown for Rmarkdown	6 months ago
tests	Added data dissemination directory	2 months ago
.gitignore	Initial commit	8 months ago
.gitlab-ci.yml	Updated E2E test to use latest version of pric...	2 months ago
CHANGELOG	Add CHANGELOG	8 months ago
CONTRIBUTING.md	Revised contributing guidelines	1 day ago
DATA.md	Updated readme.	5 months ago
README.md	Updated dependencies and package installa...	2 months ago

master rspi / +





Switch branch/tag

Search branches and tags



Branches



- quarterly-contributions
- ✓ master
- parallel-run
- parallel-run-chain
- parallel-run-documentation
- dev



automated price indexes · rspi · Issues



Open 14 Closed 45 All 59     Edit issues New issue



Search or filter results... Created date



(Potentially) replace substring method for stripping trailing slashes with a regex-based method (i.e. gsub)   updated 2 weeks ago



#54 · created 2 weeks ago by Ambuj Dewan  Dec 1, 2022 



Allow both empty strings and unsupplied arguments from the command line   updated 2 weeks ago

#53 · created 2 weeks ago by Ambuj Dewan  Sep 30, 2022 

Base period flexibility   updated 3 weeks ago

#52 · created 3 weeks ago by Ambuj Dewan  Dec 23, 2022 

Make auto data toggle Boolean   updated 3 weeks ago

#49 · created 3 weeks ago by Ambuj Dewan  Sep 30, 2022 



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada



Retail Services Price Index: Demo

automated price indexes > rspi > Commits

master rspi Author Search by message

07 Sep, 2022 3 commits

- A** Merge branch 'contributing-guidelines' into 'master' **bf75d401**
Ambuj Dewan authored 1 day ago
- A** Revised contributing guidelines **67845f02**
Ambuj Dewan authored 1 day ago
- A** Corrected export location. **94506ef2**
Ambuj Dewan authored 1 day ago

06 Sep, 2022 1 commit

- A** New analysis script for obtaining full series. **f55a9b62**
Ambuj Dewan authored 2 days ago

02 Sep, 2022 3 commits

- A** Merge branch 'parallel-run' into 'master' **49bb5d15**
Ambuj Dewan authored 6 days ago
- A** Merge branch 'parallel-run-final-price' into 'parallel-run' **59813de5**
Ambuj Dewan authored 6 days ago
- A** Survey query now uses FinalValue to reflect currency conversion. **2235ce0e**
Ambuj Dewan authored 6 days ago

31 Aug, 2022 2 commits

automated price indexes > rspi > Commits > bf75d401

Commit bf75d401 authored 1 day ago by **A** Ambuj Dewan

Browse files Options

Merge branch 'contributing-guidelines' into 'master'

Revised contributing guidelines

See merge request [114](#)

parents [94506ef2](#) [67845f02](#) master

1 merge request [114](#) Revised contributing guidelines

Changes 1

Showing 1 changed file with 14 additions and 8 deletions

Hide whitespace changes Inline Side-by-side

CONTRIBUTING.md +14 -8 View file @bf75d401

```
@@ -2,20 +2,26 @@ All contributions to this project are welcome, be it code, documentation, or ide
2 2
3 3 To help ensure smooth collaboration, please follow these guidelines as best you can.
4 4
5 5 - - Contributing to this project means that you'll sometimes have issues assigned to you. Please resolve issues "before"
6 6 their due date, either by making a merge request with the resolution to the issue, or assigning the issue to whoever
7 7 needs to act on it.
8 8 + - Contributing to this project means that you'll sometimes have issues assigned to you. Please resolve issues "before"
9 9 their due date, either by making a merge request with the resolution to the issue, or by assigning the issue to
10 10 whoever needs to act on it.
11 11 - - Contributing code is done through merge requests; contributing ideas and things to do is done through issues and
12 12 milestones. Please don't attach code or other important items with issues.
13 13 + - The master branch is the production branch. This branch includes all code used to produce disseminated statistics
14 14 and quarterly outputs required by analysts. Code used to produce other outputs (e.g. monthly outputs, one-off
15 15 analysis scripts) does not belong in the master branch. It belongs either in its own branch or in a separate project.
16 16 - - All contributed code for this project will be reviewed before being merged into the relevant branch.
17 17 + - Contributing code to the master branch is done through merge requests; contributing ideas and things to do is done
18 18 through issues. Only attach code snippets to an issue if you're discussing with other team members about how to code
19 19 something. Any code that you intend to be part of the project should be part of a commit.
20 20 - - Please use a clean and consistent style, and follow best practices. Code that is difficult to read is not
21 21 functional, and will not be merged.
22 22 + - Do not attach files to comments unless absolutely necessary. If you need special text formatting in your comments,
23 23 use markdown.
24 24 - - Make sure all tests pass. If a test should be modified, that needs to be documented and part of the merge request.
```



Statistics Canada

Statistique Canada

Delivering insight through data for a better Canada



Retail Services Price Index: Demo

> RSPI > 2022Q2

ne	Date modified	Type	Size
----	---------------	------	------

Name	Date modified	Type
Automation_RSPI_Demo	9/8/2022 8:45 PM	File folder

Name	Date modified	Type	Size
aggregate	9/8/2022 8:43 PM	File folder	
analysis	9/8/2022 8:43 PM	File folder	
data	9/8/2022 8:44 PM	File folder	
disseminate	9/8/2022 8:44 PM	File folder	
elementals	9/8/2022 8:44 PM	File folder	
ingest	9/8/2022 8:44 PM	File folder	
sample_management	9/8/2022 8:44 PM	File folder	
tests	9/8/2022 8:45 PM	File folder	
.gitignore	9/8/2022 8:43 PM	Text Document	1 KB
.gitlab-ci.yml	9/8/2022 8:43 PM	YML File	1 KB
Automation_RSPI_Demo	9/8/2022 8:45 PM	R Project	1 KB
CHANGELOG	9/8/2022 8:43 PM	File	1 KB
CONTRIBUTING.md	9/8/2022 8:43 PM	MD File	3 KB
DATA.md	9/8/2022 8:43 PM	MD File	8 KB
README.md	9/8/2022 8:43 PM	MD File	5 KB



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada



Retail Services Price Index: Demo

```
1 #####
2 # Ingest survey data from GPS
3 #
4 # Survey data reside in the GPS database. This script reads these data for reference period
5 # YYYYMM, and writes the resulting csv to stdout. It is a simple utility script that wraps a SQL
6 # query to PRAD, and that is doing all the work.
7 #
8 # Dependencies: RODBC (>= 1.3-16)
9 #####
10
11 # Enclose entire script in curly braces to force readline
12 {
13 {
14 #---- 0 Setup ----
15 library(RODBC, include.only = c("odbcDriverConnect", "sqlQuery"))
16
17 if(!interactive()) {
18 # set reference period
19 if(is.na(commandArgs(TRUE)[1])) {
20 stop("Error: You must provide a reference period.")
21 } else if(!grepl("^[1-9]{1}[0-9]{3}([01-9]{1}){1}|1[0-2]{1}){1}$", commandArgs(TRUE)[1])) {
22 stop("Error: The reference period must be in YYYYMM format.")
23 } else {
24 ref_pd_yyyyymm <- commandArgs(TRUE)[1]
25 # tsqL doesn't have a way to represent year months, so set the reference period YYYYMM
26 # as the first day of the year month YYYY-MM-01
27 ref_period <- as.Date(paste0(ref_pd_yyyyymm, "01"), format = "%Y%m%d")
28 }
29 }
30 # set working directory (if provided)
31 if (!is.na(commandArgs(TRUE)[2])) setwd(commandArgs(TRUE)[2])
32 } else {
33 # Set reference period
34 repeat {
35 ref_pd_yyyyymm <- readline("Reference period (YYYYMM): ")
36
37 if(ref_pd_yyyyymm == "") {
38 stop("You must provide a reference period.")
39 } else if(!grepl("^[1-9]{1}[0-9]{3}([01-9]{1}){1}|1[0-2]{1}){1}$", ref_pd_yyyyymm)) {
40 stop("The reference period must be in YYYYMM format.")
41 } else {
42 # tsqL doesn't have a way to represent year months, so set the reference period YYYYMM
43 # as the first day of the year month YYYY-MM-01
44 ref_period <- as.Date(paste0(ref_pd_yyyyymm, "01"), format = "%Y%m%d")
45 break
46 }
47 }
48 }
49 }
50 }
51 }
52 #---- 1 Query ----
53 # This query can stand on its own; it doesn't need R
54 query <-
55 "-- Set reference period (first day of the year month, e.g., '2021-01-01')
56 DECLARE @ref_period date = %s;
57
```

The screenshot shows the RStudio interface with the following components:

- Environment:** Shows "Global Environment" and "Environment is empty".
- Files:** Shows a directory tree for "Automation_RSPI_Demo" with files like .gitignore, aggregate, analysis, Automation_RSPI_Demo.Rproj, CHANGELOG, CONTRIBUTING.md, data, DATA.md, disseminate, elementals, ingest, README.md, sample_management, and tests.

