CPI Alternative Data Sources (ADS) Initiative

www.statcan.gc.ca

Telling Canada's story in numbers

Jonathan Wylie,
Consumer Prices Division

Nov 9, 2017

Presentation to 4th annual Conference on Big Data

Statistics Canada / Statistique Canada

Canada

Context

Statistics Canada has undertaken a modernization initiative that aims to:

- Ensure more timely and responsive statistics.
- Develop and release more granular statistics.
- Ensure cost effective products and service delivery.

Project Plan

Phase 1 — Simple Implementation

- Replace existing sample
- Infrastructure/tools to house & extract prices

Phase 2 — Full Implementation

- Leverage quantities
- Use more prices
- Use new methods

Use it ALL

All outlets, all prices

Data Sources

Four main data sources

1. Scanner data

2. Web scraping

3. Application Programming Interfaces (API)

4. Administrative data

Scanner Data Challenges

- ## Acquisition
  - Data is sensitive and security is important

- ## "Big" nature of data
  - More significant IT storage and processing requirements

- ## Timeliness
  - How well does the availability of the data source correspond with the monthly production calendar?

- ## Methodology
  - Integration of scanner data with data collected instore

Web Scraping Challenges

- Methodological challenges (availability of weights)

- IT infrastructure and process

  - Set-up costs may be high

  - Responsiveness to web site changes

Next steps

- Moving ahead with ~20% replacement of field collected prices with alternate data in 2018.

- Balance publication timeliness with availability of data sources.

- Planning IT infrastructure that will meet the project needs.

- Acquiring talent with data science skillset.