



Implementation of Application Programming Interface and Administrative Data in the Canadian CPI

UN GWG on Big Data for Official Statistics

Workshop on Scanner Data and Official Statistics

Kigali, Rwanda, 29 April – 1 May 2019



Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

Canada



Outline

- Overview
- Use of Application Programming Interface data in the Air Transportation price index
- Use of administrative data in the gasoline price index



Overview

- **Application Programming Interface**
 - Standardized data calls to companies that form the backbone of the travel booking industry
- **Administrative data:**
 - Data gathered by a company on prices and product characteristics
 - Time and geography dimensions
 - Shared with potential users (as National Statistical Agencies) for the production of (official) statistics
- **Highly technology-based**



Use of Application Programming Interface data in the Air Transportation price index



Context

- Internet manual collection
 - Packaged holidays and cruises (1.21%): 2013; 2 third party websites
 - Traveller accommodation (0.89%): 2015; individual hotels websites
 - Air transportation (1.49%): 2018; 4 websites
- Statistics Canada was looking at an improved collection for Airfares, traveller accommodation and packaged holidays
 - Cost effective production of CPI

What's an API?

- Application Programming Interfaces (API) facilitate the exchange of information between data suppliers, programmers and app developers
- A Global Distribution System (GDS) is a network that enables transactions between travel industry service providers and travel agencies



API data acquisition

- Initially, contacted 4 GDSs for API access, but two of them not feasible:
 - Google Flights – Shutting down their API
 - Amadeus – Not interested because we are not a Travel Agency
 - Amadeus now operates a self-service API: <https://developers.amadeus.com/self-service>
 - Option to develop and test for free
 - You can try using this Travelport API demo for free: <https://demo.travelportuniversalapi.com/>
- Two GDS agreed to do business



HQ: Southlake, Texas, US



HQ: Langley, UK



Access to their APIs for:

- Air fares
- Traveller accommodation
- Car rentals

- Annual contracts, auto-renew
- 30 days notice required if either side wants to terminate



Internet manual collection and API collection: Cost comparison

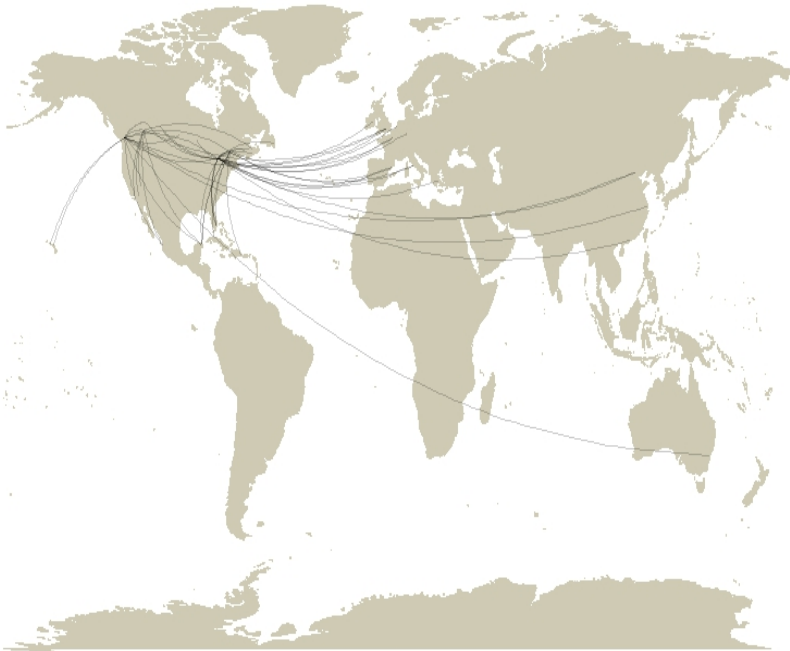
	Internet manual collection		
Traveler accommodation	Internet collection team 1	<ul style="list-style-type: none"> • API support and access fees • Usage fees • One-time certification fees • Initial API set up fees 	<ul style="list-style-type: none"> • API support and access fees • Usage fees • Initial API set up fees
Car rentals	Internet collection team 2		
Airfares	Internet collection team 3		

- Internet manual collection is very labour intensive and costly
- API collection requires advanced programming skills and is very cost-efficient
 - Collection costs reduced by 1/8
 - Number of individual price observations multiplied by 100 or 1000!

APIs mean more city-pairs, carriers and flight characteristics...collected quickly!

Current Airfares Collection

53 manually collected city-pairs



Sabre Airfares Potential

3852 round-trip destinations originating from a Canadian city



APIs vastly expand our access to airfare data

- More availability for city pairs when designing the product
 - Originally had 53 city pairs
 - Now essentially unlimited (180 city pairs selected)
- More availability of carriers
 - Originally had 4 carriers
 - Now essentially unlimited (more than 20 carriers included)
- More fare details available
 - Better knowledge of what we price
- More volume of data collected
 - Prices collected every day of the month
 - More booking lags priced
- Lower production cost
 - Low collection cost
 - Important investment to build a processing system
 - Low maintenance cost



Use of administrative data in the gasoline price index



Context

- Gasoline is a unique product in the CPI
 - Prices are volatile and can change notably day to day
 - Other products' price movements can be captured reasonably accurately with one observation per month
 - Gasoline is priced multiple times per month
- Need for a better way to collect gasoline prices
 - Cost effective production of CPI

Previous gasoline price collection

- Based on a collection month composed of 4 pricing weeks
 - No prices collected for some weeks of the month
 - Sometimes, prices on previous month days are included in the current month
- If there are different price fluctuations over the month, then all of them are not well reflected in the price index



Previous gasoline price collection

- Large fluctuations in daily gasoline prices





Data confrontation

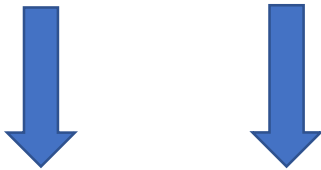
- Our CPI gasoline data confronted with data from a Company that monitors and collects gasoline prices on a daily basis in many Canadian cities
 - Has always been accurate, differences noted due to methodology differences
 - Daily data: Monday to Friday (excluding holidays)
 - Available for 70 cities plus select aggregations



Cost comparison

Data collection using company's administrative data

- Annual collection cost reduced by 1/5
- More cities, different outlets per day per city
- Collection 20 days / month
- Number of price quotes multiplied by more than 8 (regular unleaded)



- Information on:
 - Price with and without taxes
 - Fuel grade (Regular, Mid-Grade, Premium, Diesel)
- Data received in .xml format at 4 PM every week day except holidays

StatCan field collection

- A number of representative cities, number of outlets varies by city
- Collection 7 days / month
- An acceptable number of price quotes (regular unleaded)



Benefits of using administrative data

- The index will:
 - reflect prices for the entire calendar month
 - be more accurate
 - be more in line with what the “outside world” expects



Price aggregation

- Arithmetic average monthly price per city
 - Days in a month are not weighted
- Weighted geometric mean in a geo strata
 - Population count will be used to weight cities within a geo strata
- Population count by city was chosen over volume.



Questions?

Thank you!