StatCan CEWS Midterm Presentation

Bohan Gao
Eric Baxter
Vicens Paneque

Team

Bohan Gao
Data Scientist

- MDS | UBCO (2021)
- B.Sc. Computer Science | UBCO (2020)

Eric Baxter
Data Scientist

- MDS | UBCO (2021)
- B.Sc. Mathematics | University of Ottawa (2020)
- B.Sc. Psychology | University of Ottawa (2018)

Vicens Paneque
Data Scientist

- MDS | UBCO (2021)
- MBA | EGADE Business School (2017)
- B.A. Financial Management | Tec de Monterrey (2010)

Client

Statistics Canada

COVID-19
Canada Emergency Wage Subsidy Analysis
&
API Feasibility Study

Project

Part 1: Canada Emergency Wage Subsidy (CEWS)

- Interactive dashboard with power BI
- Exploratory report on the urban/rural divide

Part 2: Feasibility Study of APIs Use for Economic Recovery Tracking

Presentation Outline

Part 1: CEWS

- CEWS overview
- Data
- Research questions and goals
- Our approach
- Roadblocks
- Sample results

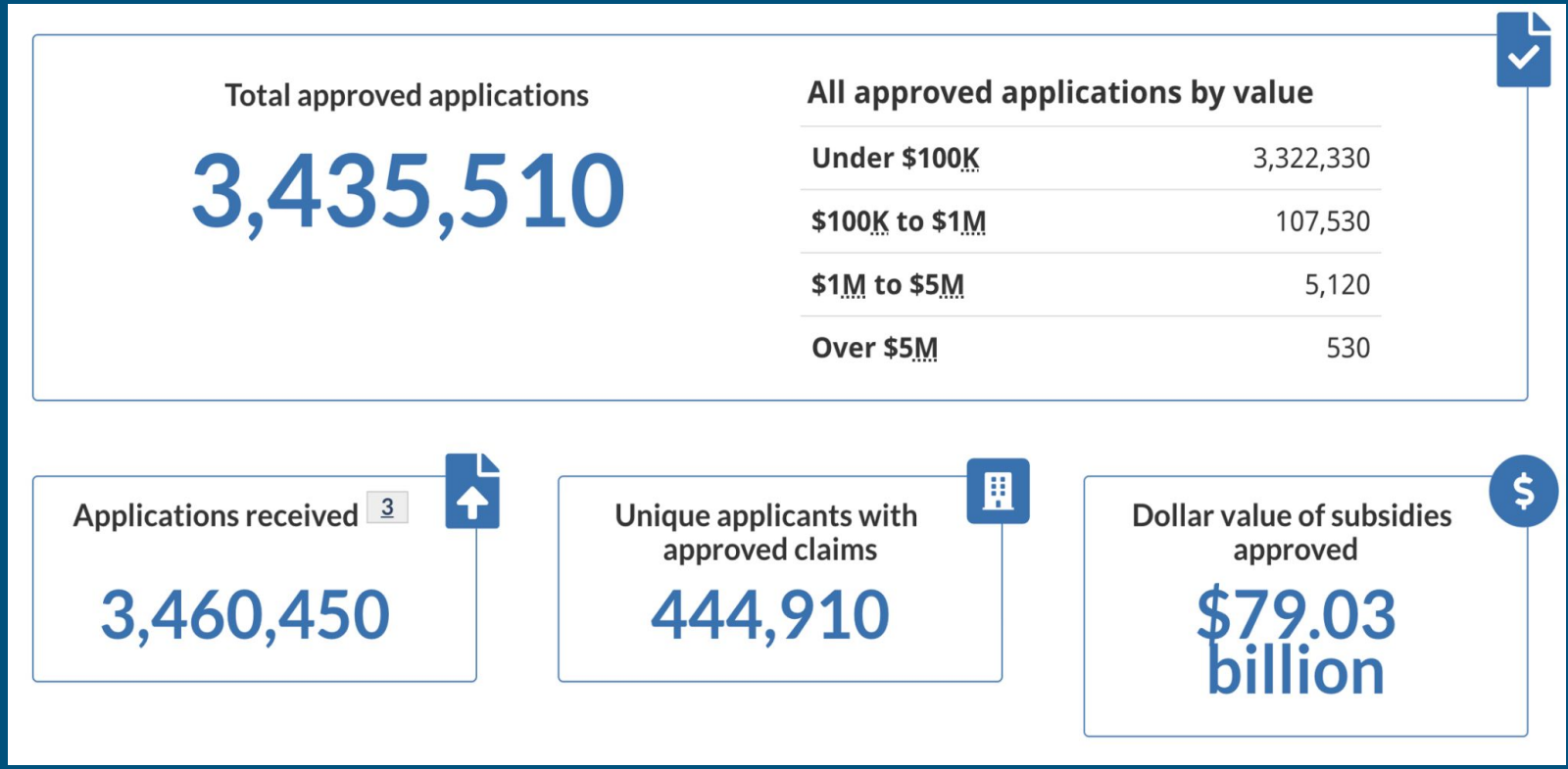
Presentation Outline

Part 2: API Feasibility Study

- Overview
- Research questions and goals
- Our approach
- Roadblocks
- Sample results

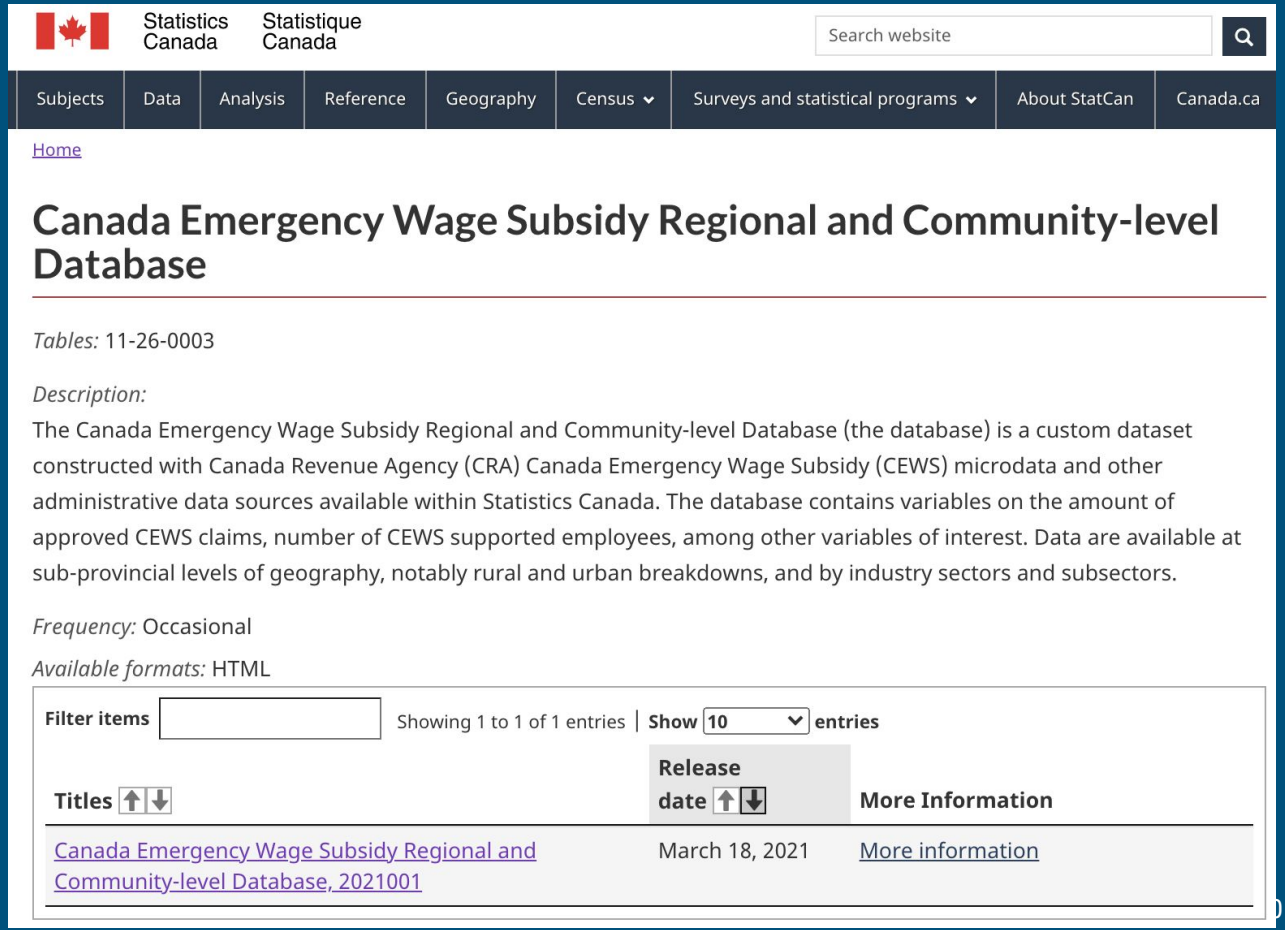
Part 1: CEWS

Canada Emergency Wage Subsidy Overview



Data

Canada Wage Subsidy Regional and Community-Level Database



Statistics Canada Statistique Canada

Search website

Subjects | Data | Analysis | Reference | Geography | Census | Surveys and statistical programs | About StatCan | Canada.ca

Home

Canada Emergency Wage Subsidy Regional and Community-level Database

Tables: 11-26-0003

Description:

The Canada Emergency Wage Subsidy Regional and Community-level Database (the database) is a custom dataset constructed with Canada Revenue Agency (CRA) Canada Emergency Wage Subsidy (CEWS) microdata and other administrative data sources available within Statistics Canada. The database contains variables on the amount of approved CEWS claims, number of CEWS supported employees, among other variables of interest. Data are available at sub-provincial levels of geography, notably rural and urban breakdowns, and by industry sectors and subsectors.

Frequency: Occasional

Available formats: HTML

Filter items | Showing 1 to 1 of 1 entries | Show 10 entries

Titles	Release date	More Information
Canada Emergency Wage Subsidy Regional and Community-level Database, 2021001	March 18, 2021	More information

```
cews.head(3)
```

	Start_date_of_CEWS_period	RegionCode	RegionName	RuralUrbanFlag	CMACAFlag	IndustryCode	IndustryName	Number_business_locations	Subsidy_amount	Supported_employees	CEWS_rehire_count
0	2020-03-15	10	Newfoundland and Labrador	Not applicable	Not applicable	11	Agriculture, forestry, fishing and hunting	30	823,000	362	0
1	2020-03-15	10	Newfoundland and Labrador	Not applicable	Not applicable	111	Crop production	10	X	90	0
2	2020-03-15	10	Newfoundland and Labrador	Not applicable	Not applicable	112	Animal production and aquaculture	10	X	X	0

```
cews.sample(3)
```

	Start_date_of_CEWS_period	RegionCode	RegionName	RuralUrbanFlag	CMACAFlag	IndustryCode	IndustryName	Number_business_locations	Subsidy_amount	Supported_employees	CEWS_rehire_count
231185	2020-06-07	1211006	Cumberland, Subd. B	RURAL	Not applicable	99	Other and Missing NAICS	5	X	X	0
190058	2020-05-10	3534005	Bayham	RURAL	Not applicable	23	Construction	10	84,000	30	0
508051	2020-09-27	1208008	East Hants	RURAL	Not applicable	238	Specialty trade contractors	5	43,000	39	0

Standard Geographic Classification 2016

- 1-digit: Canada
- 2-digit: Province
- 5- digit: Census agglomeration or census metropolitan area
- 7-digit: Census subdivision

North American Industry Classification System

- “All industries”
- 2-digit: Industry sector
- 3-digit: industry subsector

Data

```
cews.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 562491 entries, 0 to 562490
Data columns (total 11 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   Start_date_of_CEWS_period   562491 non-null  object
 1   RegionCode                  562491 non-null  object
 2   RegionName                  562491 non-null  object
 3   RuralUrbanFlag              562491 non-null  object
 4   CMACAFlag                   562491 non-null  object
 5   IndustryCode                562491 non-null  object
 6   IndustryName                562491 non-null  object
 7   Number_business_locations   562491 non-null  object
 8   Subsidy_amount              562491 non-null  object
 9   Supported_employees         562491 non-null  object
 10  CEWS_rehire_count           562491 non-null  object
dtypes: object(11)
memory usage: 47.2+ MB
```

```
cews['Subsidy_amount'].value_counts(normalize=True)
```

```
X              0.675029
0              0.010516
17,000         0.001916
20,000         0.001888
22,000         0.001865
                 ...
6,824,000      0.000002
17,151,000     0.000002
89,020,000     0.000002
12,221,000     0.000002
149,094,000    0.000002
Name: Subsidy_amount, Length: 16468, dtype: float64
```

```
cews['Supported_employees'].value_counts(normalize=True)
```

```
X          0.665020
0          0.010779
12         0.003127
13         0.003095
14         0.003052
             ...
6,385      0.000002
11,659     0.000002
235,392    0.000002
8,469      0.000002
7,270      0.000002
Name: Supported_employees, Length: 12991, dtype: float64
```

Research Questions and Goals

- Design an interactive CEWS dashboard in PowerBI

- Write an exploratory research paper

Our Approach

Dashboard:

- Develop a number of pages allowing users to explore the data on any dimensions, at any level of granularity

Paper:

- Focus on rural Canada
  - Which industries were disproportionately affected in rural areas?
  - Which rural areas received the most subsidies, and what do they have in common?

Roadblocks

- Structure of data was not interpretable
  - Due to the hierarchical nature, individual subsidies are counted in multiple rows
    - I.e., once in the “Canada” row, once in the “B.C.” row, once in the “Kelowna” row
  - Could not remove aggregated rows due to information loss from suppression at more granular levels
  - **Solution:** Develop a set of “indicator” columns
    - Querying a specific level for each indicator ensures no data duplication

Roadblocks

- Initially unclear expectations from the client
  - Initial instruction for the report was “explore the data”
  - Much work thrown away before finding a thread the client was interested in

- Power BI licensing / Mac version
  - Many useful features are only available for paid versions of Power BI
    - GIS software if very limited in the free version
  - There is no Mac version of the software

Current Progress

- Report is finished, aside from formatting and proofreading
- Dashboard is mostly finished
  - Small tweaks after each client meeting

GIS Analysis: Level 1 Industry(Industry Code: XX) Subsidy Distribution

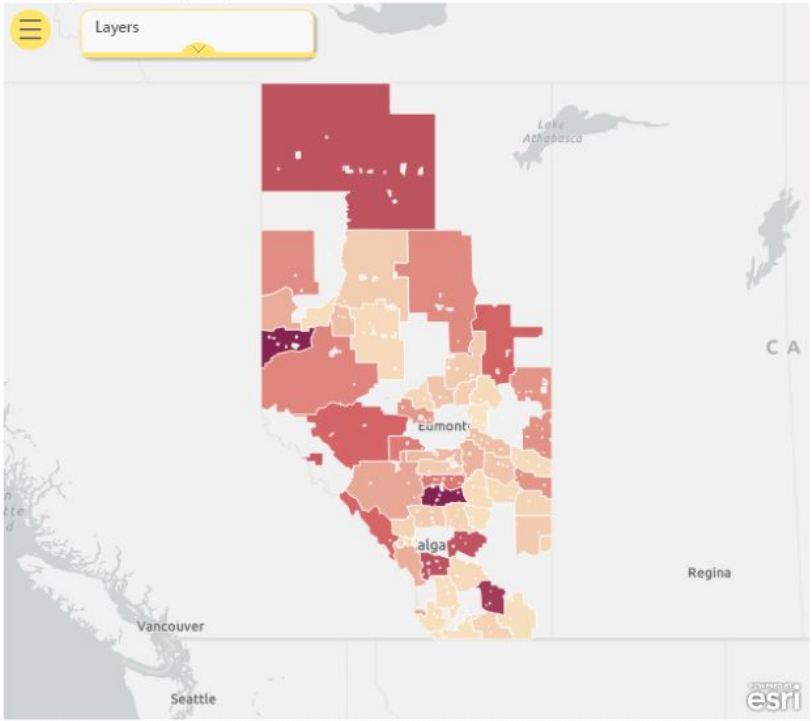
Province/CSD region
province_regionName, RegionName
Alberta

RuralUrbanFlag
RURAL

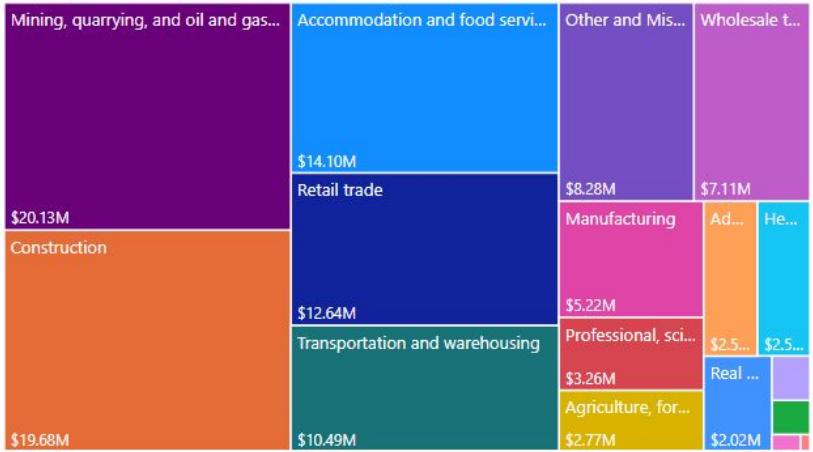
Start_date_of_CEWS_period
March 15, 2020

IndustryName
All

Subsidy_amount by RegionName



Subsidy_amount by IndustryName



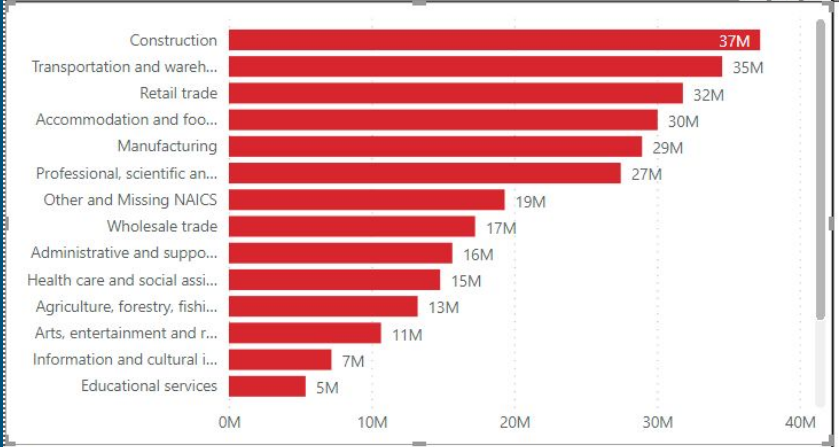
IndustryName	Subsidy_amount
Health care and social assistance	$2,525,000
Real estate and rental and leasing	$2,023,000
Information and cultural industries	$520,000
Arts, entertainment and recreation	$404,000
Educational services	$155,000
Finance and insurance	$50,000

Statistics Canada Statistique Canada

CEWS Comparison Tool

*Note: Subsidy values may not be entirely accurate due to suppressed data.

Kelowna British Columbia Subsidy

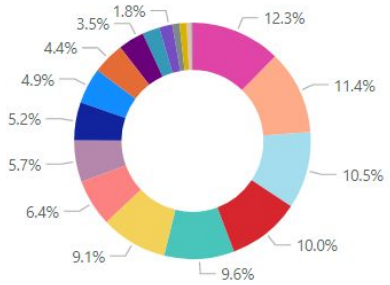


Period: All

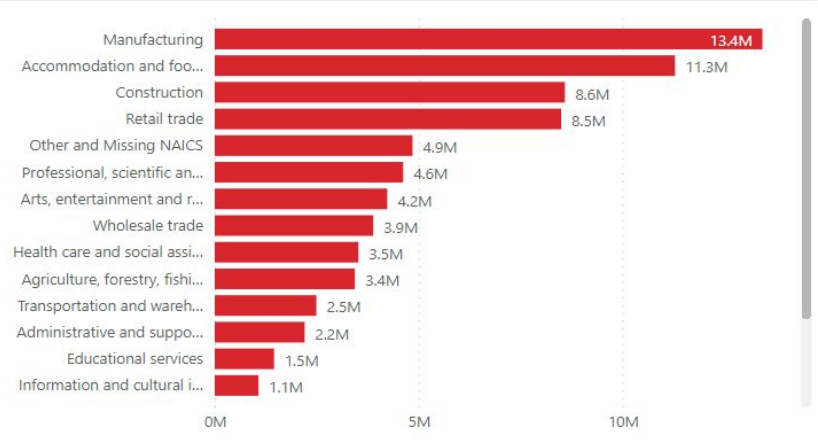
Metric: Subsidy

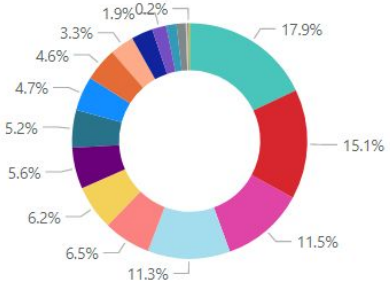
Geography: CMA/CA (Geographic Aggregati...

Industry Aggregation: Industry Sector (2 Digit NIACS)



Chilliwack British Columbia Subsidy



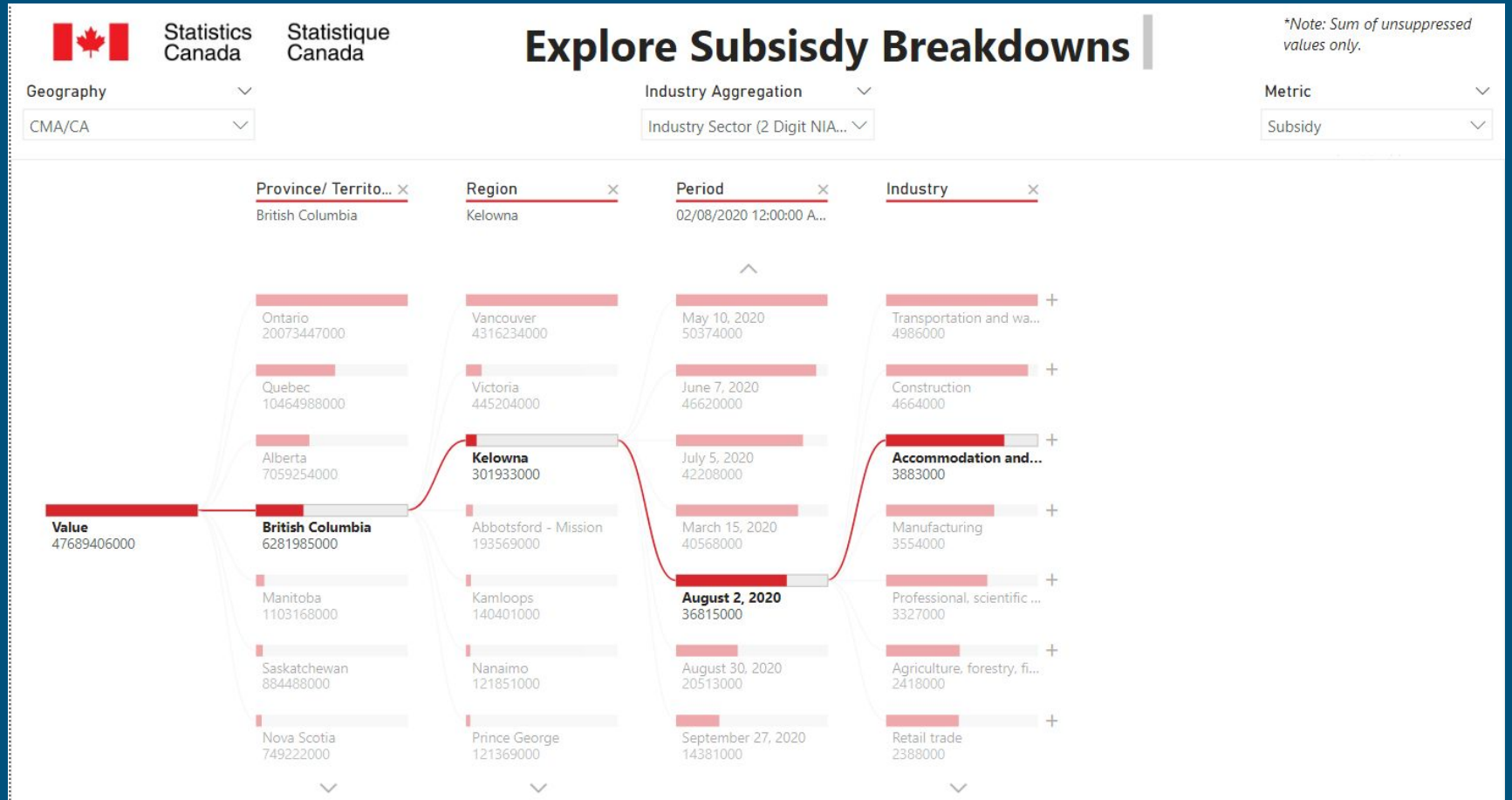


Period: All

Metric: Subsidy

Geography: CMA/CA (Geographic Aggregati...

Industry Aggregation: Industry Sector (2 Digit NIACS)

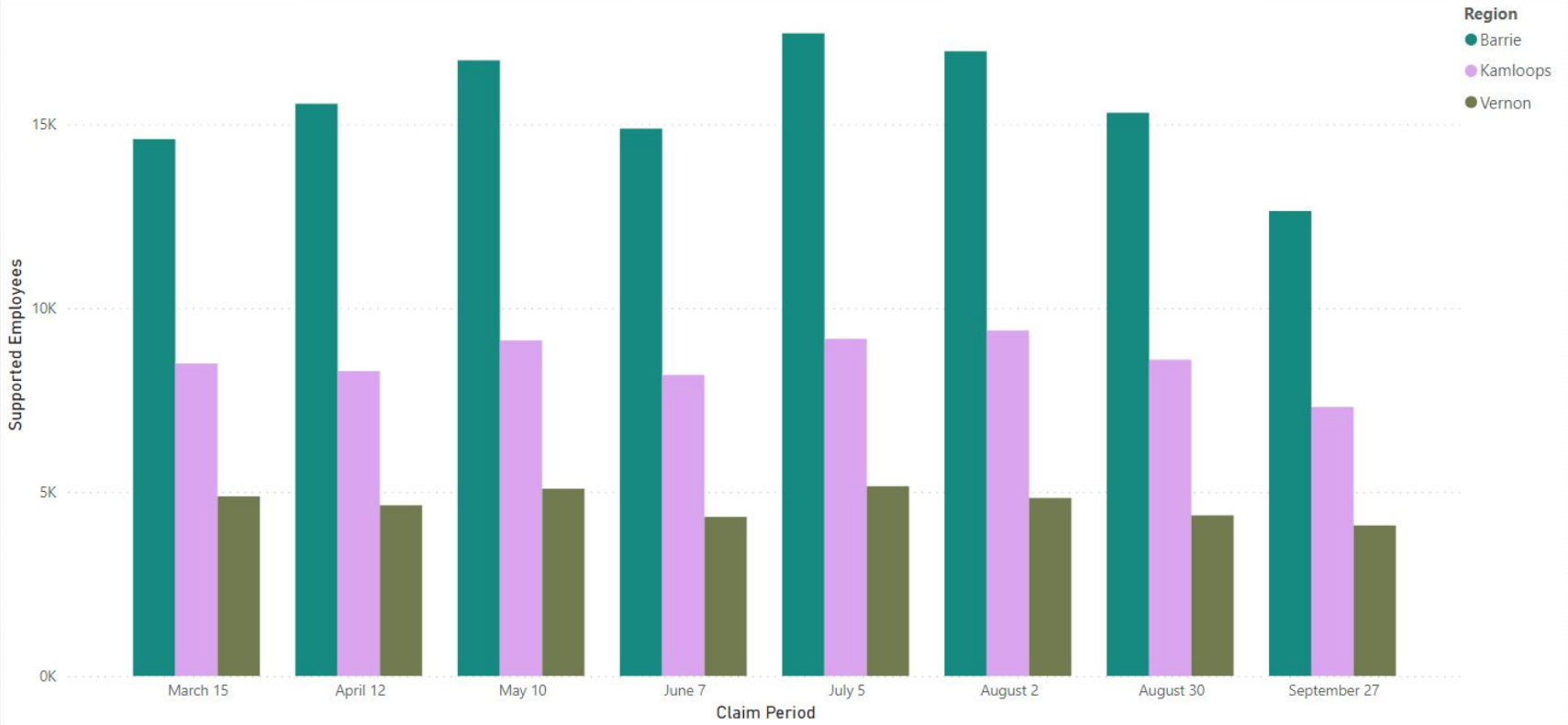

Statistics Canada
Statistique Canada
Explore Subsisdy Breakdowns
*Note: Sum of unsuppressed values only.
Geography
CMA/CA
Industry Aggregation
Industry Sector (2 Digit NIA...
Metric
Subsidy
Province/ Territo...
British Columbia
Region
Kelowna
Period
02/08/2020 12:00:00 A...
Industry
Value
47689406000
British Columbia
6281985000
Ontario
20073447000
Quebec
10464988000
Alberta
7059254000
Manitoba
1103168000
Saskatchewan
884488000
Nova Scotia
749222000
Kelowna
301933000
Vancouver
4316234000
Victoria
445204000
Abbotsford - Mission
193569000
Kamloops
140401000
Nanaimo
121851000
Prince George
121369000
August 2, 2020
36815000
May 10, 2020
50374000
June 7, 2020
46620000
July 5, 2020
42208000
March 15, 2020
40568000
August 30, 2020
20513000
September 27, 2020
14381000
Accommodation and...
3883000
Transportation and wa...
4986000
Construction
4664000
Manufacturing
3554000
Professional, scientific ...
3327000
Agriculture, forestry, fi...
2418000
Retail trade
2388000

Statistics Canada Statistique Canada

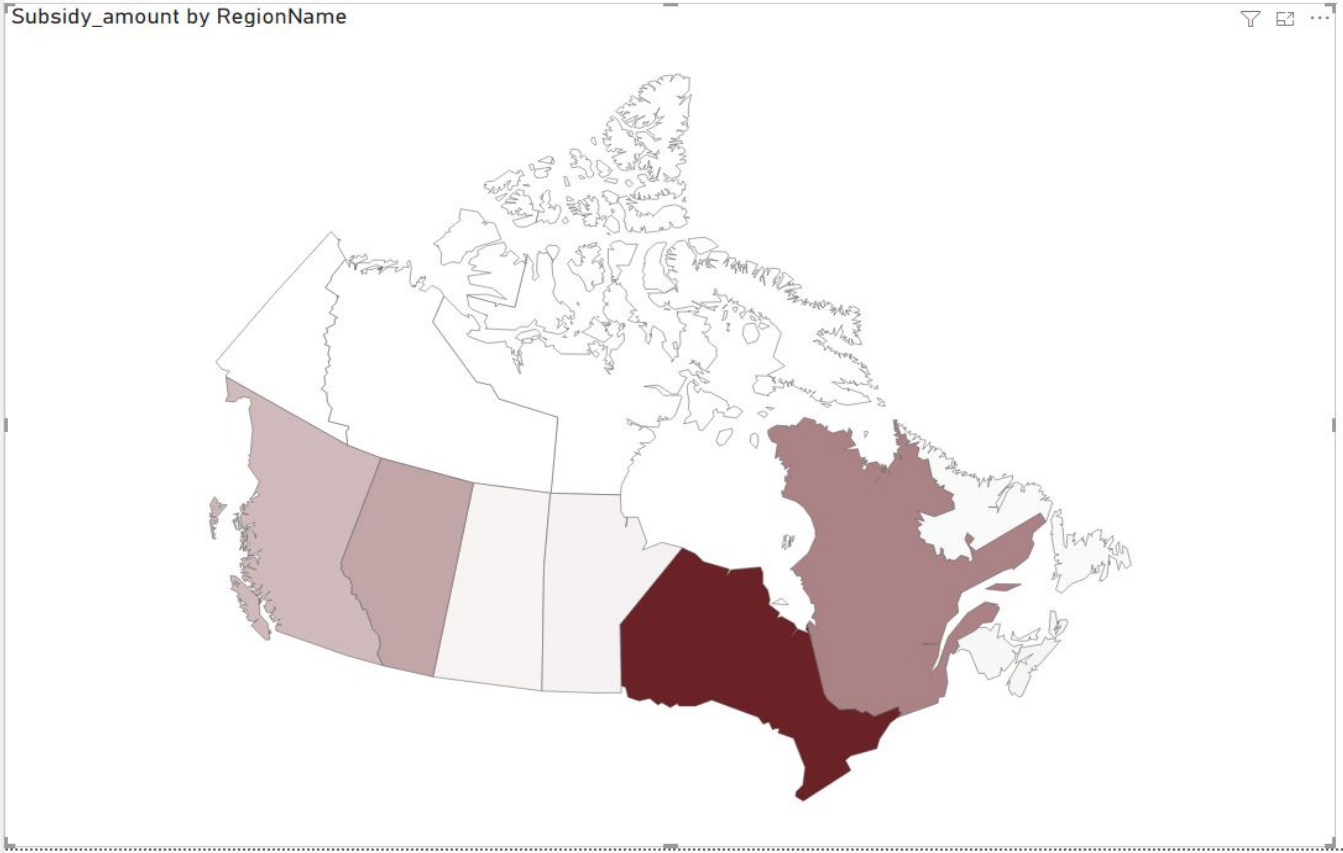
CEWS Time Series by Region

*Note: Subsidy values may not be entirely accurate due to suppressed data.

Geographic Aggregation: CSD | Region: Multiple selections | Metric: Supported Employees | Industry Aggregation: All Industries | Industry: All



Dashboard Sample 1: GIS Analysis on Each Province's Subsidy Amount & It's Average Effects Over Periods.



56bn
Subsidy_amount

21.72K
Average of Number_business_locations

2.16K
Average of CEWS_rehire_count

293.17K
Average of Supported_employees

GIS Analysis: Level 1 Industry(Industry Code: XX) Subsidy Distribution

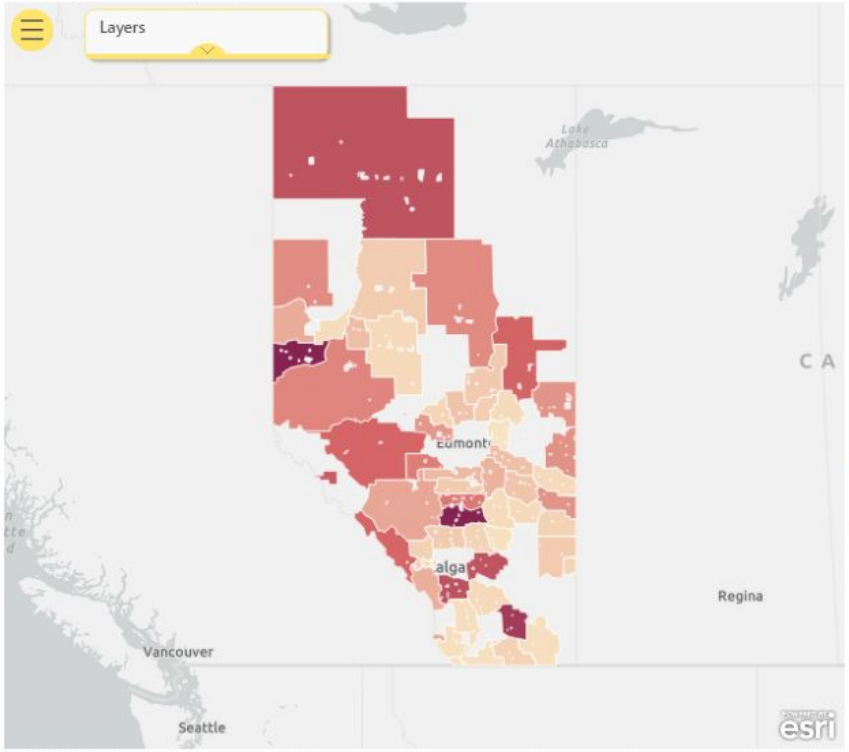
Province/CSD region
province_regionName, RegionName
Alberta

RuralUrbanFlag
RURAL

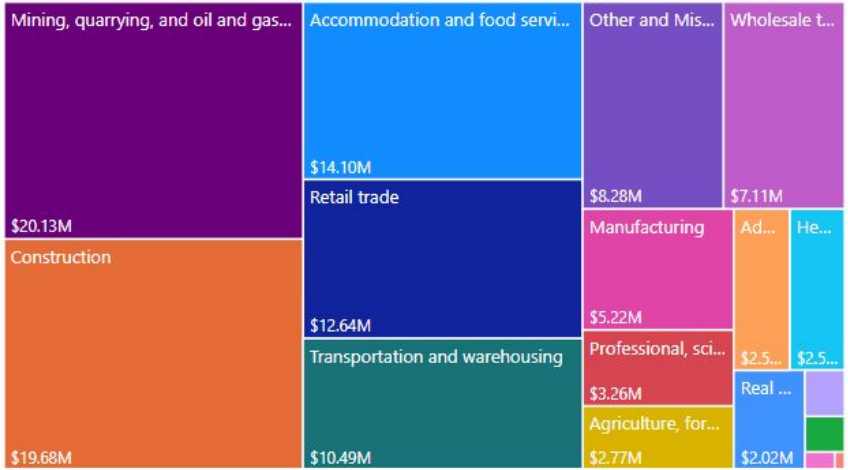
Start_date_of_CEWS_period
March 15, 2020

IndustryName
All

Subsidy_amount by RegionName



Subsidy_amount by IndustryName



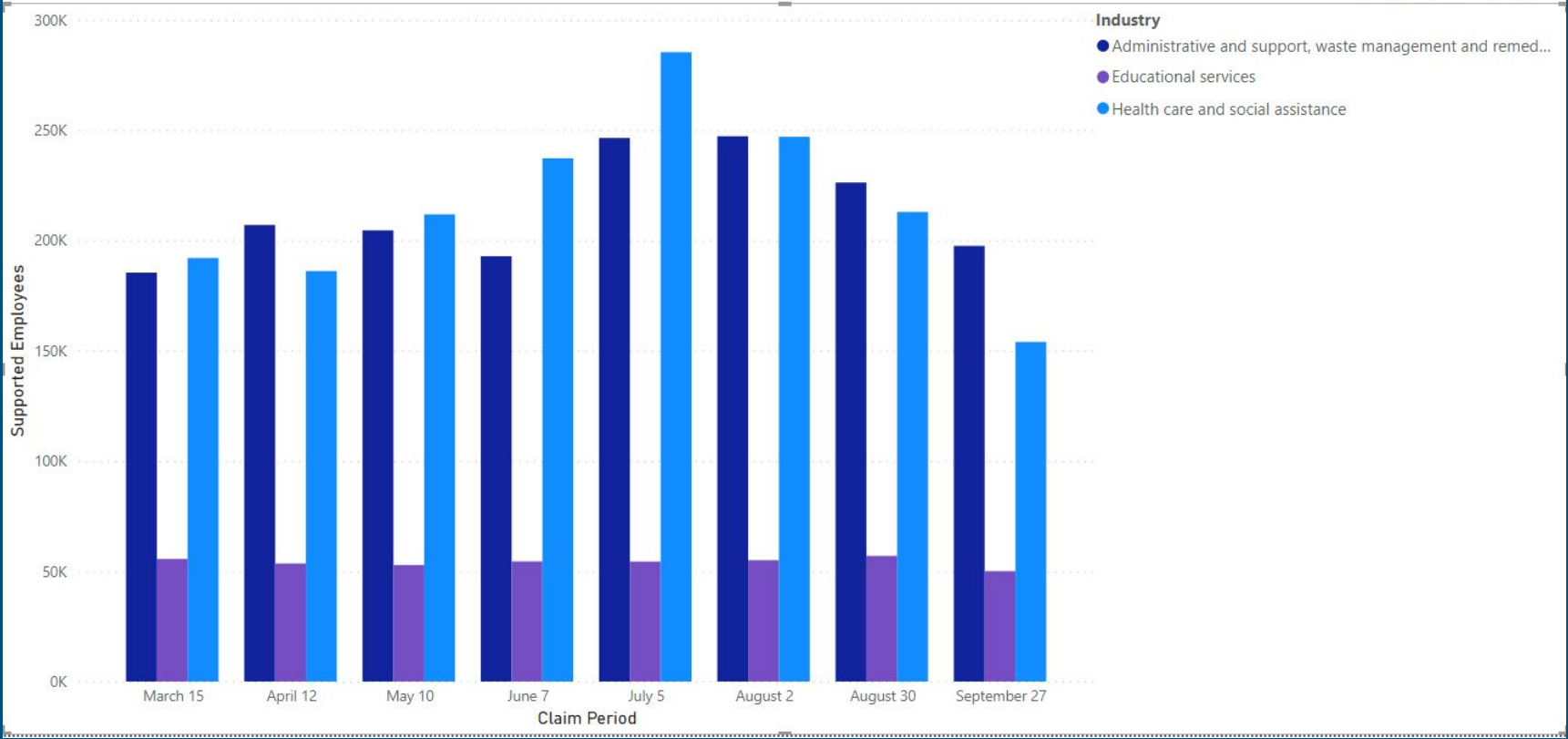
IndustryName	Subsidy_amount
Health care and social assistance	$2,525,000
Real estate and rental and leasing	$2,023,000
Information and cultural industries	$520,000
Arts, entertainment and recreation	$404,000
Educational services	$155,000
Finance and insurance	$50,000

Statistics Canada Statistique Canada

CEWS Time Series by Industry

*Note: Subsidy values may not be entirely accurate due to suppressed data.

Geographic Aggregation: Canada | Region: Canada | Metric: Supported Employees | Industry Aggregation: Industry Sector (2 Digit NIACS) | Industry: Multiple selections



Part 2: API Feasibility

Overview


STATISTICS CANADA
DATA STRATEGY
Delivering insight through
data for a better Canada
2019 to 2022
Statistics Canada
Statistique Canada
Canada

Research Questions and Goals

- Study and report on the potential use of data collected from various APIs to create a real-time business indicator for economic activity and recovery in the context of COVID-19.

- Choose one or two APIs and develop prototype software

Our Approach

- Twitter
- Google
- Airbnb
- Uber
- Doordash/ Skipthedishes
- Reddit
- Yelp
- Facebook
- Foursquare
- Eventbrite

Our Approach

- Depending on the type of data extracted, numerical analysis, or text-based sentiment analysis and content extraction could be used.
  - Tracking of economic activity within regions
  - Examine post-COVID success of industries and compare to pre-COVID

Roadblocks

- Many APIs are expensive to use for these general purposes

- Personal data not available without user authorization (i.e., Facebook likes, Twitter networks)

Current Progress

- Exploring the potential of some APIs, with much focus on Twitter

Thank You!

Questions and Demonstrations