

## **Project Proposal**

Harjot Dhaliwal (30051859)

Lukas Escoda (30211208)

Shabbir Khandwala (30219011)

Gurdeep Panag (30101520)

University of Calgary

DATA 604: Working with Data at Scale

Instructor: Dr. Ovens

November 6th, 2023

## Introduction

Our subject domain encompasses the province of Alberta and its municipalities. We're focusing on analyzing various CSV datasets organized by municipality and year. These datasets cover diverse metrics, which we've categorized into three overarching domains: housing, health, and business. The goal is to create four merged datasets that put useful and relevant data side-by-side to address each of our guiding questions rather than making a singular master dataset. These potential combinations of data would enable us the flexibility to explore various aspects. Whereas the rigidity of a singular master dataset would cause information loss at scale and prevent us from investigating smaller components of our inquiry.

Our project serves as an open investigation into understanding the performance of different Alberta municipalities across a range of metrics. By doing so, we aim to combat citizen apathy and foster greater citizen engagement. This valuable information benefits established residents in Alberta, providing insights into their local communities and the overall state of affairs. It empowers them to identify areas for improvement and where provincial intervention may be needed.

On the other hand, potential newcomers to Alberta seek comprehensive data to inform their decisions about relocating. We aspire that our investigated data aligns with their interests and assists in making informed choices based on their priorities.

Ultimately, our project offers a comprehensive view of Alberta's recent years, guiding our focus towards municipalities and industries in need of attention and improvements.

## Dataset

Our datasets come from National portals for open data, specifically the Government of Alberta. The license for our datasets are under the Open Government License from the Government of Alberta. This allows us to copy, modify, publish, translate, adapt, distribute or otherwise use the Information in any medium, mode or format for any lawful purpose (identify the Government of Alberta as the source of the materials).

The dataset is a repository of information containing various subjects, meticulously categorized by format, publisher, topic, language, and licensee. To illustrate, we have attached our table headers and a few rows for each sector in the **Appendix**. Our primary focus revolves around three specific sectors: housing, health, and business/economy at the municipal level. We have split up work on these datasets as follows; Harjot will be working on the housing datasets, Gurdeep will be working on the health datasets and Shabbir and Lukas will be working on the business/economy datasets.

As we downloaded the data in CSV format, it is conveniently presented in a tabular form. Each industry contains multiple csv files that will be joined together in SQL, primarily on the "municipality" and "year" columns, creating one master table for each guiding question. For each guiding question, we employ approximately five to six datasets (**Appendix**) and will create merged tables in order to address each guiding question. Regarding the average size of our data, each dataset typically comprises around 6,000 rows on average with 20 files in total.

## Methodology

The core industries that we deem fit are the prime factors for the success of a province as mentioned which were Health, Business/Economy and Housing. For all of these sectors, we have identified four key guiding questions, each investigated individually by one of us:

1. Shabbir Khandwala - The Government of Alberta allocates resources to enhance conditions and support its residents across various municipalities. What is the influence of government funding on business growth, assessing how these investments may contribute to increased productivity within the municipalities of Alberta?
2. Harjot Dhaliwal - Housing conditions in Alberta are a critical consideration for individuals deciding whether to reside in or move to the province. What are the factors driving the housing challenges, including average housing costs and vacancy rates in Alberta? We aim to identify municipalities that excel in providing attractive housing options, while conducting year-to-year comparisons.
3. Gurdeep Panag - To assess the quality of life in Alberta's municipalities we would like to look into the various health factors. How do the annual trends in life expectancy, greenhouse gas emissions, air quality, and population growth per capita change over time and by municipality in Alberta and how do these influence the overall success of Alberta? By examining these indicators, we aim to identify municipalities that stand out in offering superior health and well-being to their residents.
4. Lukas Escoda - To understand business incorporations in Alberta, How do the impact of tax rates, median income, and average rent affect the economic success of Alberta's municipalities? This analysis will help us identify which municipalities provide more favorable economic conditions for businesses year by year.

The datasets that we have identified to help us with our pursuit of answering these questions are (but not limited to) the following. Each dataset provides us with information per municipality year over year and might have been reused in more than one appendix:

1. Average Rent by Municipality: Displays rental rates for different unit types in non-subsidized rental buildings with three or more units, categorized by municipality, year, and unit size (**Appendix A**)
2. Building Permits by Municipality: Lists the number and value of all municipal building permits granted in the calendar year, by building type (commercial, industrial, institutional and residential) and municipality. For this indicator seasonally adjusted estimates are shown. (**Appendix A**)
3. Dwelling Units by Municipality: Lists the number and value of all municipal building permits granted in the calendar year, by building type (commercial, industrial, institutional and residential) and municipality. For this indicator seasonally adjusted estimates are shown. (**Appendix A**)
4. Housing Starts by Municipality: Lists the number of housing starts by year, municipality, and housing type. A Housing Start is defined as the beginning of construction work on a building, usually when the concrete has been poured for the whole of the footing around the structure, or an equivalent stage where a basement will not be part of the structure. (**Appendix A**)
5. Percent Single Family Houses by Municipality: Percentage of dwelling units that are single-family houses, by census year and municipality. (**Appendix A**)
6. Property Assessments by Municipality: The total value of a municipality's equalized assessment, by year, property type, and municipality and municipal district. The total value of a municipality's equalized assessment is equal to the total taxable assessment divided by the assessment level of that municipality. (**Appendix A**)

7. Residential Share of Property Assessments by Municipality: The percentage of total municipal property assessments represented by residential property, by year and municipality or municipal district. This measurement can be used as an approximate guide to the proportion of residential vs non-residential property assets in a jurisdiction. (**Appendix A**)
8. Vacancy Rates by Municipality: Lists the residential vacancy rate for units in non-subsidized rental buildings containing three or more rental units, by year and municipality. (**Appendix A**)
9. Air Quality by Municipality: Provides air quality indexes and health risk ratings over several years for several municipalities. (**Appendix B**)
10. Births, Marriages & Deaths by Municipality: Provides the most comprehensive births and deaths history for each municipal district over several years. (**Appendix B**)
11. GHG Emissions by Municipality: Provides greenhouse gas emissions for various emission types for Alberta's municipalities over several years. (**Appendix B**)
12. Life Expectancy by Municipality: Provides life expectancy estimates in years for two genders for Alberta's municipalities over several years. (**Appendix B**)
13. Population by Municipality: Provides population estimates for Alberta's municipalities separated by the age group that they belong to, over several years. (**Appendix B**)
14. Average Rent by Municipality: Rental rates for different unit types in non-subsidized rental buildings with three or more units, categorized by unit size (**Appendix C**)
15. Median Income: Lists the median income of families, including couple families, lone-parent families, and people without children, by municipality. (**Appendix C**)
16. Tax Rate: Provides mill rates for Alberta municipalities, representing property tax amounts per assessed dollar value. (**Appendix C**)
17. Business Incorporations: Shows yearly counts of incorporations pertaining to businesses formed under the Business Incorporations Act. (**Appendix C**)
18. Number of Businesses: The number of businesses with employees on record, by size (number of employees) and industry. (**Appendix D**)
19. Major Projects by Municipality: An approximate annual value of capital projects above \$5 million in December of each year, by sector. (**Appendix D**)
20. Median Income: Lists the median income of families, including couple families, lone-parent families, and people without children, by municipality. (**Appendix D**)
21. Employment Census: Lists the employment rate, unemployment rate, and participation rate by census year, gender, and municipality and municipal district. The employment rate refers to the number of persons employed expressed as a percentage of the total population 15 years of age and over and participation rate measures the total labour force relative to the size of the working-age population. (**Appendix D**)
22. Population by Municipality: Estimates of the number of people including Canadian citizens and non-permanent residents, by year, age, age group and gender. (**Appendix D**)

In our project, we use Python, SQL, and libraries like SQLAlchemy, Pandas, and Plotly to analyze data from a MariaDB database hosted on the University of Calgary Cloud. We focus on data preparation, cleaning, and transformation to create merged datasets that contain relevant information for each of the questions. To achieve this, we aim to combine datasets, connect them by location and year, format the relevant columns in an appropriate manner and drop the unnecessary ones and conclude our investigation by producing meaningful visualizations using SQL queries.

## Appendix

### Appendix A: Housing

#### Average Rent

CSDUID	CSD	Period	IndicatorSummaryDescription	Rental Unit Type	UnitOfMeasure	OriginalValue
4808011	Red Deer	1987	Average Residential Rent (2-bedroom units)	1 - bedroom	\$	377
4808011	Red Deer	1988	Average Residential Rent (2-bedroom units)	1 - bedroom	\$	379
4808011	Red Deer	1989	Average Residential Rent (2-bedroom units)	1 - bedroom	\$	388

#### Building Permits

CSDUID	CSD	Period	IndicatorSummaryDescription	Building Type	UnitOfMeasure	OriginalValue
4807044	Sedgewick	1992	Building Permits (value of permits)	Commercial	\$	2500
4807044	Sedgewick	1993	Building Permits (value of permits)	Commercial	\$	174
4807044	Sedgewick	1994	Building Permits (value of permits)	Commercial	\$	0

#### Dwelling Units

CSDUID	CSD	Period	IndicatorSummaryDescription	UnitOfMeasure	OriginalValue
4807044	Sedgewick	1994	Dwelling Units		364
4807044	Sedgewick	1995	Dwelling Units		378
4807044	Sedgewick	1996	Dwelling Units		380

#### Housing Starts

CSDUID	CSD	Period	IndicatorSummaryDescription	Housing Type	UnitOfMeasure	OriginalValue
4808011	Red Deer	1987	Housing Starts	Apartment		58
4808011	Red Deer	1988	Housing Starts	Apartment		24
4808011	Red Deer	1989	Housing Starts	Apartment		35

#### Major Projects

CSDUID	CSD	Period	IndicatorSummaryDescription	Sector	UnitOfMeasure	OriginalValue
4808011	Red Deer	2008	Major Projects	Pipelines	\$	150
4808011	Red Deer	2013	Major Projects	Pipelines	\$	10
4808011	Red Deer	2014	Major Projects	Pipelines	\$	10

#### Percent Single Family Homes

CSDUID	CSD	Period	IndicatorSummaryDescription	UnitOfMeasure	OriginalValue
4807044	Sedgewick	2001	Single Family Houses (% of all housing units)		0.87671
4807044	Sedgewick	2006	Single Family Houses (% of all housing units)		0.80822
4807044	Sedgewick	2011	Single Family Houses (% of all housing units)		0.81818

#### Property Assessments

CSDUID	CSD	Period	IndicatorSummaryDescription	Property Type	UnitOfMeasure	OriginalValue
4807044	Sedgewick	1997	Property Assessments	Farmland	\$	15100
4807044	Sedgewick	1998	Property Assessments	Farmland	\$	14360
4807044	Sedgewick	1999	Property Assessments	Farmland	\$	9040

#### Residential Share of Property Assessments

CSDUID	CSD	Period	IndicatorSummaryDescription	UnitOfMeasure	OriginalValue
4807044	Sedgewick	1997	Residential Share (of property assessments)		0.78702
4807044	Sedgewick	1998	Residential Share (of property assessments)		0.74046
4807044	Sedgewick	1999	Residential Share (of property assessments)		0.73709

#### Vacancy Rates

CSDUID	CSD	Period	IndicatorSummaryDescription	UnitOfMeasure	OriginalValue
4808011	Red Deer	1988	Residential Vacancy		0.023
4808011	Red Deer	1989	Residential Vacancy		0.007
4808011	Red Deer	1990	Residential Vacancy		0.007

## Appendix B: Health

### Air Quality by Municipality

CSDUID	CSD	Period	IndicatorSummaryDescription	Air Quality Health Index	Health Risk	UnitOfMeasure	OriginalValue
4808011	Red Deer	2011	Air Quality (percent high-quality hours)	1	High Quality		0.05019
4808011	Red Deer	2012	Air Quality (percent high-quality hours)	1	High Quality		0.08876
4808011	Red Deer	2013	Air Quality (percent high-quality hours)	1	High Quality		0.09738
4808011	Red Deer	2014	Air Quality (percent high-quality hours)	1	High Quality		0.12925
4808011	Red Deer	2015	Air Quality (percent high-quality hours)	1	High Quality		0.09625
4808011	Red Deer	2016	Air Quality (percent high-quality hours)	1	High Quality		0.27628

### Births, Marriages, and Deaths by Municipality

Year	Municipality	Vital Event	Value
2001	Airdrie	Birth by Residence of Mother	243
2001	Airdrie	Marriage by Place of Event	76
2001	Airdrie	Death by Residence of Deceased	90
2001	Athabasca	Birth by Residence of Mother	53
2001	Athabasca	Marriage by Place of Event	27
2001	Athabasca	Death by Residence of Deceased	54

### Greenhouse Gas Emissions by Municipality

CSDUID	CSD Name	Year	Emission Type	Indicator	Value
4801003	Cypress County	2014	CO2	Greenhouse Gas Emissions	74,600
4801003	Cypress County	2014	CO2 Equivalent	CO2 Eq. Emissions	74,600
4801003	Cypress County	2014	CH4	Greenhouse Gas Emissions	38.5778
4801003	Cypress County	2014	CH4 Equivalent	CO2 Eq. Emissions	964.445
4801003	Cypress County	2014	N2O	Greenhouse Gas Emissions	1.343
4801003	Cypress County	2014	N2O Equivalent	CO2 Eq. Emissions	400.214

### Life Expectancy by Municipality

CSDUID	CSD	Period	IndicatorSummaryDescription	Gender	UnitOfMeasure	OriginalValue
4807044	Sedgewick	2011	Life Expectancy	Female	Years	83.27000
4807044	Sedgewick	2012	Life Expectancy	Female	Years	82.89000
4807044	Sedgewick	2013	Life Expectancy	Female	Years	82.51000
4807044	Sedgewick	2014	Life Expectancy	Female	Years	83.43000
4807044	Sedgewick	2015	Life Expectancy	Female	Years	81.82000

### Population by Municipality

CSDUID	CSD	Period	IndicatorSummaryDescription	Age	Age Group	Gender	UnitOfMeasure	OriginalValue
4807044	Sedgewick	2001	Population	0	0-4	Female		6.00000
4807044	Sedgewick	2002	Population	0	0-4	Female		6.00000
4807044	Sedgewick	2003	Population	0	0-4	Female		8.00000
4807044	Sedgewick	2004	Population	0	0-4	Female		5.00000
4807044	Sedgewick	2005	Population	0	0-4	Female		6.00000

## Appendix C: Business/Economy

### Average Rent

CSDUID	CSD	Period	Type_of_rent	Rental Unit Type	UnitOfMeasure	Average Rent value
4814024	Edson	2015	Average Residential Rent (2-bedroom units)	Bachelor	\$	823.0
4814024	Edson	2016	Average Residential Rent (2-bedroom units)	Bachelor	\$	777.0
4814024	Edson	2017	Average Residential Rent (2-bedroom units)	Bachelor	\$	730.0

### Median Income

CSDUID	CSD	Period	Type_of_family	Median income value
4800000	Not Stated, Alberta	2018	Couple Families Income (Median)	93980.0
4800000	Not Stated, Alberta	2018	Lone-Parent Income (Median)	52490.0
4800000	Not Stated, Alberta	2019	All Families Income (Median)	90720.0

### Tax Rate

CSDUID	CSD	Period	Tax Rate Type	Tax value
4807044	Sedgewick	1998	Non-Residential	13.434
4807044	Sedgewick	1999	Non-Residential	12.494
4807044	Sedgewick	2000	Non-Residential	12.112

### Business Incorporations

CSDUID	CSD	Period	Incorporations
4811065	Redwater	2019	14.0
4811065	Redwater	2020	12.0
4811065	Redwater	2021	15.0



## Appendix D: Business/Economy

### Number of Businesses

CSD UID	CSD	Period	IndicatorSummaryDescription	Size	Industry	NAICS	UnitOfMeasure	Original Value
4806016	Calgary	1998	Number of Businesses	Large (200+)	Accommodation and food services	72		15
4806016	Calgary	1999	Number of Businesses	Large (200+)	Accommodation and food services	72		14
4806016	Calgary	2000	Number of Businesses	Large (200+)	Accommodation and food services	72		19

### Major Projects

CSDUID	CSD	Period	IndicatorSummaryDescription	Sector	UnitOfMeasure	Original Value
4808011	Red Deer	2008	Major Projects	Pipelines	\$	150
4805041	Kneehill County	2011	Major Projects	Pipelines	\$	8.7
4802001	Warner County No. 5	2009	Major Projects	Pipelines	\$	60

### Median Income

CSDUID	CSD	Period	IndicatorSummaryDescription	UnitOfMeasure	Original Value
4807044	Sedgewick	2000	All Families Income (Median)	\$	54000
4807044	Sedgewick	2000	Couple Families Income (Median)	\$	56800
4807044	Sedgewick	2000	Lone-Parent Income (Median)	\$	33300

### Employment Census

CSDUID	CSD	Period	IndicatorSummaryDescription	Gender	UnitOfMeasure	OriginalValue
4807044	Sedgewick	1981	Participation Rate	Female		0.389
4807044	Sedgewick	1986	Participation Rate	Female		0.397
4807044	Sedgewick	1991	Participation Rate	Female		0.576

## Population

CSD UID	CSD	Period	IndicatorSummaryDescription	Age	Age Group	Gender	UnitOfMeasure	Original Value
4807044	Sedgewick	2022	Population	0	0-4	Female		8
4807044	Sedgewick	2022	Population	0	0-4	Male		9
4807046	Lougheed	2022	Population	0	0-4	Female		2

## References

- Air Quality Index by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program>
- Average Rent by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/average-rents-by-municipality>
- Births and Deaths by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program>
- Building permits by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/building-permits-by-municipality>
- Dwelling units by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/dwelling-units-by-municipality>
- Greenhouse Gas Emissions by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program>
- Housing starts by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/housing-starts-by-municipality>
- Incorporations by Municipality. (2023). *Government of Alberta*.  
<https://open.alberta.ca/opendata/incorporations-by-municipality>
- Life Expectancy by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program>
- Major projects by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/major-projects-by-municipality>
- Median income by Municipality. (2023). *Government of Alberta*.  
<https://open.alberta.ca/opendata/median-income-by-municipality>
- Municipal Tax Rates by Municipality. (2023). *Government of Alberta*.  
<https://open.alberta.ca/opendata/municipal-tax-rate-by-municipality>
- Open Government Program. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program>
- Percent Single Family Houses by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/percent-single-family-houses-by-municipality>

Property Assessments by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/total-equalized-assessment-by-municipality>

Population by Municipality. (2023). *Government of Alberta*. <https://www.alberta.ca/open-government-program>

Residential Share of Property Assessments by Municipality. (2023). *Government of Alberta*. <https://www.alberta.ca/open-government-program/residential-share-of-property-assessments-by-municipality>

Vacancy Rates by Municipality. (2023). *Government of Alberta*.  
<https://www.alberta.ca/open-government-program/vacancy-rates-by-municipality>