

What Behind the Property Prices in Calgary: A Community-level Analysis

Alan Li, Li Chen, Xinzhen Tang, Ze Yu

1 Introduction

Canada, known for being an immigration-friendly country, has experienced a significant population increase over the years, primarily due to the growth of international immigration^[1]. This influx of people has contributed to a housing crisis across the nation, with the city of Calgary witnessing a surge in property prices in recent decades^[2]. By the end of 2023, the price of a townhouse is expected to increase by 17.2% year-over-year, reaching \$449,000^[3]. It is widely acknowledged that supply and demand dynamics play a crucial role in determining long-term property prices^[4]. The supply side is influenced by factors such as the number of homes for sale and new buildings available, while the demand side is affected by the number of households, economic growth, mortgage availability, interest rates, and more. Additionally, community-specific factors, including crime rates, public services, and demographic features, can also impact property prices at specific time points^[5].

Whether we are international immigrants or local residents, the need to rent or purchase a home is inevitable at some point in our lives. Making this decision requires considering various factors at a specific time. This realization inspired us to conduct a community-level analysis of property prices in Calgary, focusing on data from 2019. In this study, we examine the factors affecting property prices from three perspectives: safety factors, public services, and demographic factors.

Through this investigation, we aim to provide valuable insights into the factors behind property prices in Calgary, thus helping both local residents and international immigrants make informed decisions when it comes to renting or purchasing a home. By shedding light on these community-level factors, we hope to contribute to a better understanding of the housing market in Calgary and address the challenges faced by individuals and families in finding suitable and affordable housing options.

2 Guiding Questions

After the data screening and further exploration, we refine our guiding questions based on those proposed in our proposal report. Guiding questions have been structured into two parts with factors such as household income and transportation removed due to the data availability and consistency.

The first two questions are about an overall view of Calgary's property price across different communities:

Data 604 Working with Data at Scale Group Project Report

1. What is the variation in unit property prices across communities in Calgary?
2. How do property prices differ among residential, non-residential, and farmland properties in Calgary?

The following three questions are regarding affecting factors:

3. How does safety, in terms of crime rate and disorder, impact property prices in the different communities?
4. What role do public services, such as education, community center and medical facilities, play in determining property prices in Calgary communities?
5. How do demographic factors, including gender, population, and language, influence property prices across communities in Calgary?

3 Individual Datasets

3.1 Data Sources and Licensing

All datasets utilized in this study are publicly accessible from the Open Calgary Dataset website and the use of these datasets is permitted. All of these datasets are licensed under the following URL: <https://data.calgary.ca/d/Open-Data-Terms/u45n-7awa>.

The primary datasets used and assigned among four group members are listed below:
Historical Property Assessments (2019)^[6] (by team member: Xinzhen Tang).
Census by Community (2019)^[7] (by team member: Alan Li).
Schools in Communities^[8], Health Clinics and Hospitals Community Services^[9] (by Li Chen).
Community Disorder Statistics^[10], Community Crime Statistics^[11] (by team member: Ze Yu).

3.2 Property Assessment Dataset

The dataset utilized for this project consists of historical assessed values of residential, non-residential, and farm land properties in Calgary. These properties are registered parcels at Alberta's Land Titles Office^[6]. We selected the 2019 dataset to analyze the features of property prices across various Calgary communities. The dataset contains over 500,000 rows, with key fields including assessed value, community code, community name, land size, property type, and multipolygon.

We employed the Python Pandas library to extract and clean the key columns from the original dataset (see Supplementary file 01_calgary_property_price_overview_update.ipynb, Part I: Data Initial Cleaning and Transformation). We then removed rows with unavailable (None) values in the assessed

Data 604 Working with Data at Scale Group Project Report

value, land size, and land use designation fields. A crucial aspect of the data cleaning process involved identifying residential properties. Although the assessment class provided information on property types, there was an overlap between residential and non-residential land types, which could lead to inaccuracies in residential property prices due to larger land sizes in non-residential properties. To address this issue, we excluded assessed values that did not strictly correspond to the residential land type as defined by the City of Calgary^[12]. This resulted in a cleaned property assessment dataset ready for SQL queries.

For the SQL query process, we first identified distinct community code fields, which served as the keys to merge three datasets processed by other team members (see Supplementary file 01_calgary_property_price_overview_update.ipynb, Part II: Queries Using SQL). We then calculated average property total values, land sizes, and unit prices for the three property categories across different communities. This provided a set of property assessed values that were ready for visualization and integration with factor datasets.

3.3 Public Service Dataset

The Open Calgary website holds two base data sources regarding our project. One is “Community Services”, and the other is “Schools in Communities”. The basic information of these two data sets is listed below:

- 1, Community Services,
https://data.calgary.ca/Services-and-Amenities/Community-Services/x34e-bcjz/about_data. It contains hospitals, libraries, community centers, and other public services in a community, locations are stored in a spatial format.
- 2, Schools in Communities,
https://data.calgary.ca/Services-and-Amenities/Schools-in-Communities/xmep-aa_sr. It contains school information in a community, and locations are stored in a spatial format.

The data sets we are going to explore contain geographic point information instead of the community to which each entry belongs. And, since our project uses community code, a code based on community initials, as the primary key, some steps must be done in advance to map the geographic points into areas that can be calculated and compared with the multi-polygon information of each community. Some key functions that implemented the mapping task are in the section. (see Supplementary file 02_Public Service DataSet Process and Analysis.ipynb, Part I, 1, Geographical data process)

After mapping all entries to communities, we need to migrate data into the database engine. We use PostgreSQL to store and process public service data. At the final step

Data 604 Working with Data at Scale Group Project Report

of this particular part, we create a view to store the final result of public services to facilitate further analysis. (see [Supplementary file 02_Public Service DataSet Process and Analysis.ipynb](#), PART I, 2, Migrate to Database (PostgreSQL))

3.4 Community Safety Dataset

Safety data include information like Community name, Disorder count and Crime count. This data originally sourced from 2 different dataset and then combined using both geographic method and community name mapping (NAME).The attribute “NAME” is also used as key to join with other datasets.

The original disorder dataset:

(https://data.calgary.ca/Health-and-Safety/Social-Disorder-Statistics/4evm-wx9a/data_preview) includes attributes like Community name, event count, month and year (2018-2023).

The original crime dataset:

(https://data.calgary.ca/Health-and-Safety/Community-Crime-Statistics/78gh-n26t/data_preview) includes attributes like Community name, category, crime count, month and year (2018-2023).

For our project, we need the total crime count and total disorder count for each community in the year 2019. The data is cleaned and extracted by joining the two dataset using SQL Queries, documented in supplementary file 03 Calgary Property Price Safety.ipynb: "Extract Necessary Data Using SQL Queries"

After cleaning the empty crime type and combining the two dataset, 267 communities remained in this data source; others were dropped either because of the lack of key feature information or the geographic information which makes it impossible to add useful information to the community.

3.5 Demographic Dataset

The demographic dataset utilized in this analysis originates from the 2019 community census sourced from the open Calgary dataset available at https://data.calgary.ca/Demographics/Census-by-Community-2019/rkfr-buzb/about_data. This dataset encompasses essential information such as community names, gender distribution, age demographics, and languages spoken at home.

Employing the community code as the primary key, we leveraged SQL's join function to amalgamate the demographic dataset with property values. This integration

Data 604 Working with Data at Scale Group Project Report

encompassed various factors including population density, gender ratio, and languages spoken at home, all of which are pertinent to property pricing dynamics.

The process commenced with data cleaning and the merging of the census data with the property value file, documented in detail within the supplementary file 04 Calgary Property Price Demographic.ipynb "Part I: Data Cleaning in the Calgary Property Price Demographic.ipynb" file. During this phase, a unique index was generated employing the community code as the primary key to facilitate the seamless merging of the two datasets.

Subsequently, SQL queries were employed to identify key influencers within community census data, documented in supplementary file 04 Calgary Property Price Demographic.ipynb:"Part II: Queries Using SQL." These queries delve into notable findings such as demographic distribution, gender ratio distribution, and the impact of various language groups on property values.

4 Data exploration

4.1 Multiple Datasets Merge

Each of our datasets contains an attribute of Community Name or Community Code which is used as a key for joining. Every team member has already cleaned their own dataset, so just need to extract necessary data and no extra data cleaning needed for the merged dataset. Steps of merging data are documented in supplementary file 05 Calgary Property Price Merged.ipynb: Part1.

To see the impact on property price of each factor, we need to create a table of property prices and investigate terms for each factor. Steps of extracting data of factor vs property price are documented in supplementary file 05 Calgary Property Price Merged.ipynb: Part2. The data can also be extracted from the merged table directly.

4.2 Overview of Calgary's Property Price

This part provides an overview of Calgary's property price spatial distributions and histograms for three categories, residential, non-residential and farmland, based on the data generated by cleaning and SQL queries. Several plots including geometry plots, histogram, and boxplot are utilized to display the results (Codes in Supplementary file 01_calgary_property_price_overview_update.ipynb, Part III: Result visualization).

Our data analysis reveals some interesting patterns in property prices. For residential unit prices, we observe that they are higher in the downtown area and its surrounding regions. As we move away from the downtown, the unit prices decrease, but they

Data 604 Working with Data at Scale Group Project Report

increase again in areas near the city's edge. In contrast, non-residential unit prices tend to decrease from downtown towards the city's edge. As for farmland, the unit prices are slightly higher in the north-east and south-east compared to the east, but the differences are relatively small.

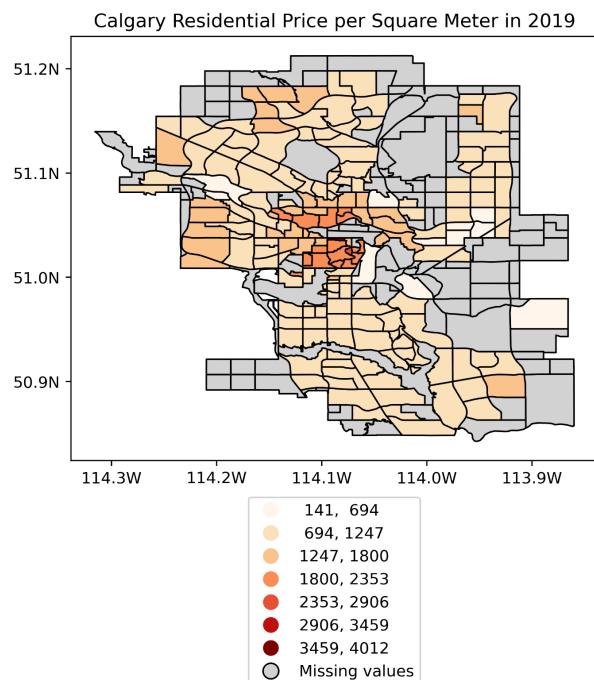


Fig. 1 Calgary Residential Unit Price

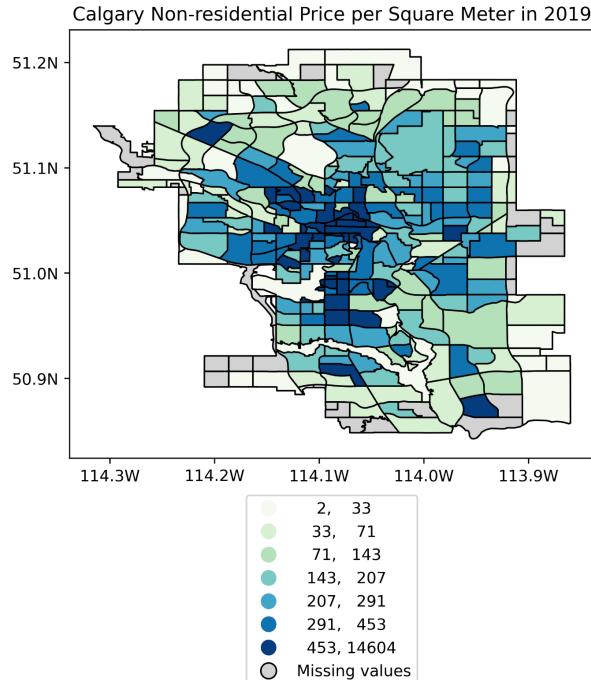


Fig. 2 Calgary Non-residential Unit Price

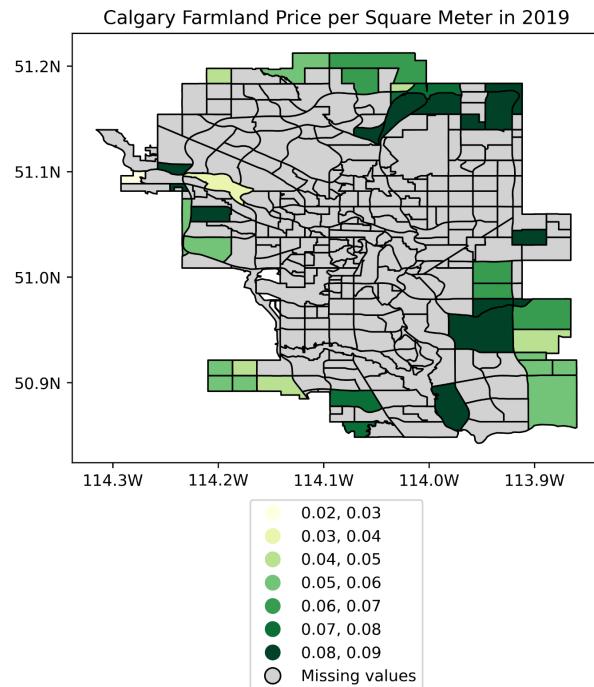


Fig. 3 Calgary Farmland Unit Price

Data 604 Working with Data at Scale Group Project Report

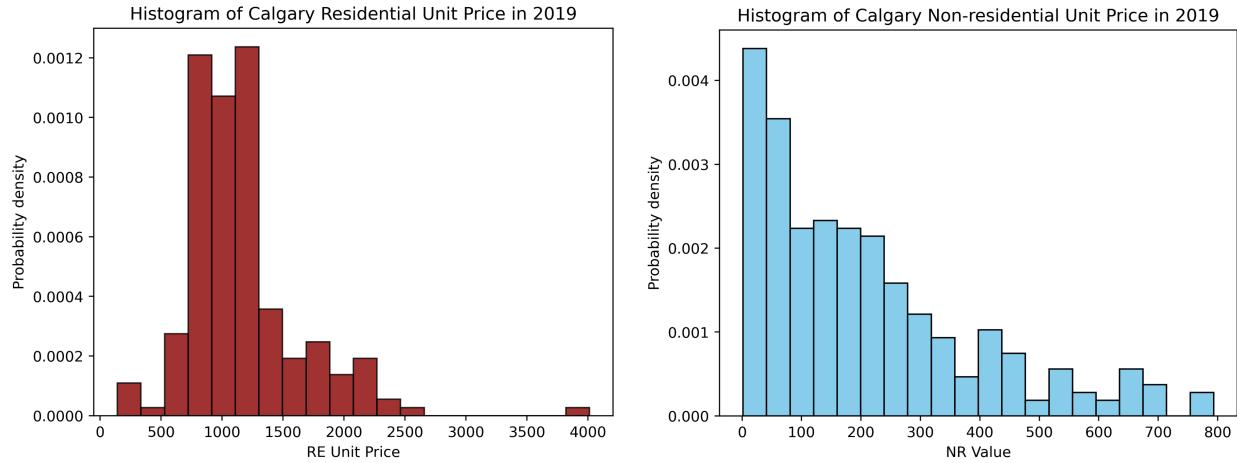


Fig. 4 Histogram of Calgary Residential Unit Price Fig. 5 Histogram of Calgary Non-residential Unit Price

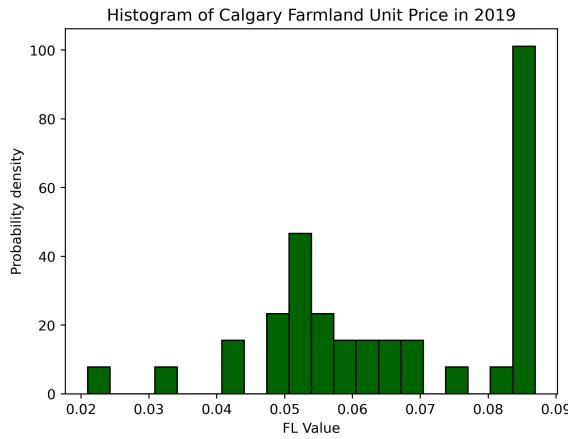


Fig. 6 Histogram of Calgary Non-residential Unit Price

The histograms for property prices illustrate different distributions. Residential unit prices are centered around 1,000 dollars and follow a normal distribution, while farmland prices are left-skewed, and non-residential prices are right-skewed. Residential unit prices exhibit a larger magnitude and quantile range compared to non-residential prices, whereas the price range for farmland is quite narrow.

We also analyzed the distribution of residential total prices and sizes. Both variables are right-skewed, with the highest frequency observed for sizes around 500 square meters and total residential prices of approximately 500,000 dollars. We identified the top 10 communities with the most expensive and the cheapest residential properties. The Red Carpet community has the highest average residential price, with an assessed value

Data 604 Working with Data at Scale Group Project Report

exceeding 9 million dollars. On the other hand, the Belmont community has the lowest average residential price, with an assessed value of under 300,000 dollars.

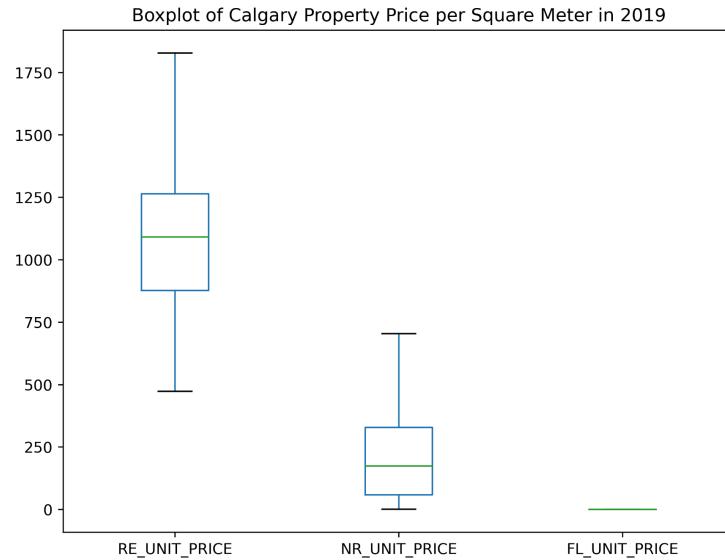


Fig. 7 Boxplot of Calgary Property Unit Price

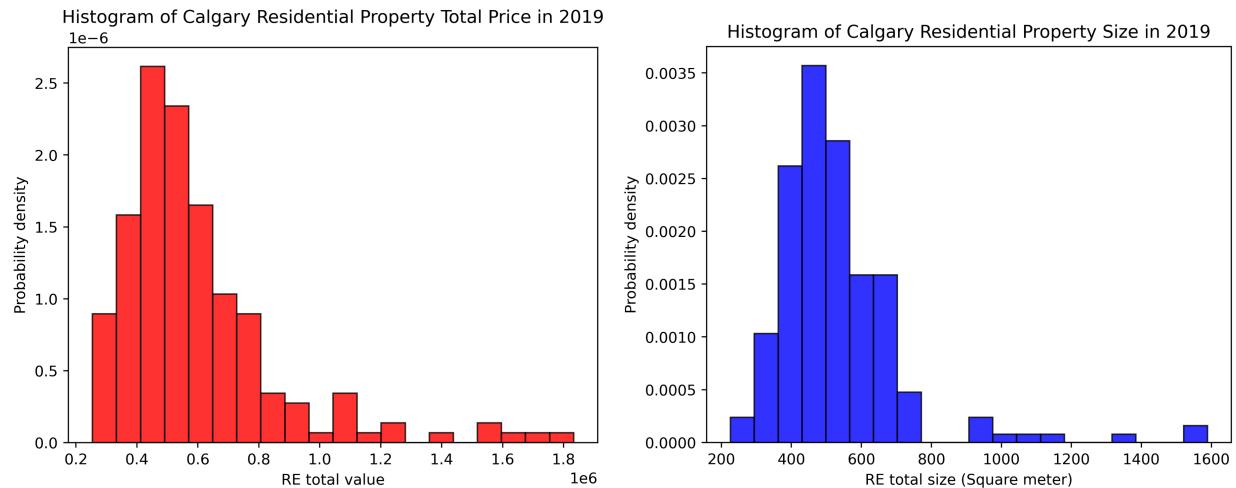


Fig. 8 Histogram of Calgary Residential Unit Price Fig. 9 Histogram of Calgary Residential Property Size

Data 604 Working with Data at Scale Group Project Report

Table 1 top 10 communities with the most expensive residential properties

COMM_CODE	CLASS	RE_VALUE	RE_AREA	RE_UNIT_PRICE
RED	Residential	9.316000e+06	55974.720000	166.432275
MDH	Residential	4.390000e+06	25790.000000	170.221016
BEL	Residential	2.140297e+06	1354.494937	1580.144311
BRT	Residential	2.011378e+06	958.601938	2098.241019
EAG	Residential	1.832763e+06	1588.687629	1153.633259
UMR	Residential	1.739527e+06	765.942771	2271.092795
RID	Residential	1.625930e+06	630.398529	2579.209930
CLI	Residential	1.591650e+06	944.960000	1684.357010
EPK	Residential	1.527422e+06	681.170565	2242.348329
ROX	Residential	1.418591e+06	658.609756	2153.918083

Table 2 top 10 communities with the cheapest residential properties

COMM_CODE	CLASS	RE_VALUE	RE_AREA	RE_UNIT_PRICE
BLM	Residential	254035.555556	312.479111	812.968120
LIV	Residential	284626.395534	307.576077	925.385351
SET	Residential	290325.459318	280.896588	1033.567056
12A	Residual Sub Area	295333.333333	1585.433333	186.279251
FHT	Residential	298137.671681	472.139961	631.460364
ABB	Residential	302865.032487	399.142056	758.790081
PEN	Residential	304062.148338	470.215754	646.643898
DOV	Residential	304544.989099	430.138517	708.016085
FAL	Residential	305400.462963	383.856061	795.611935
ERI	Residential	306714.321608	351.817437	871.799659

4.3 Residential Property Price and Public Services

4.3.1 Geographical distribution of public services

First, we glance at the data we are working on.

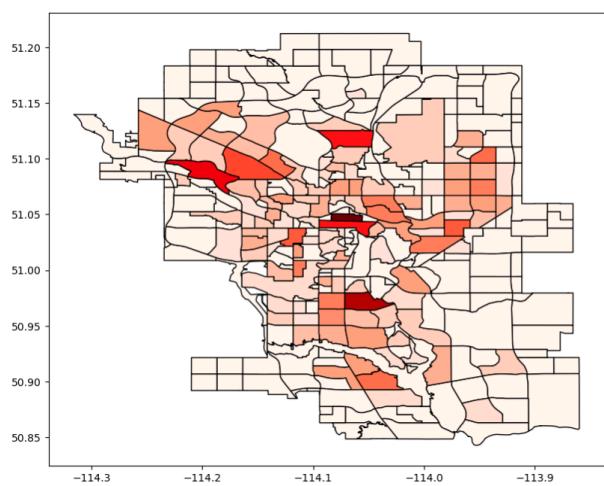


Fig.10 Public services distribution map

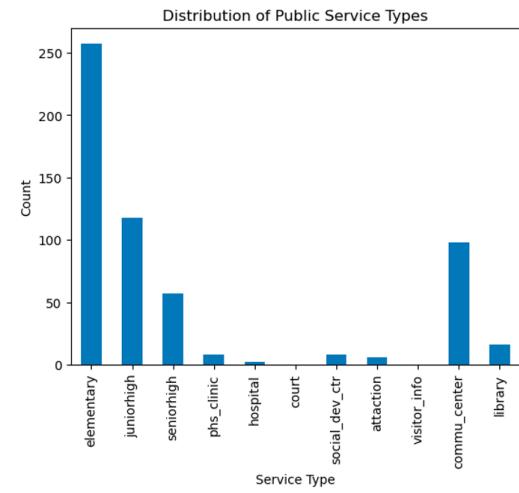


Fig. 11 Different public service distribution

Data 604 Working with Data at Scale Group Project Report

As seen from the chat, courts, and visitor info have small counts, therefore we can remove them. And schools consist of most of the public services, so we can dive into schools. (see Supplementary file 02_Public Service DataSet Process and Analysis.ipynb, Part II, chapter 2- chapter 3)

4.3.2 Dive into Schools in the data set

On average, there are 2 elementary schools, and 1 junior high school in a community block, while high schools are scattered in the city.

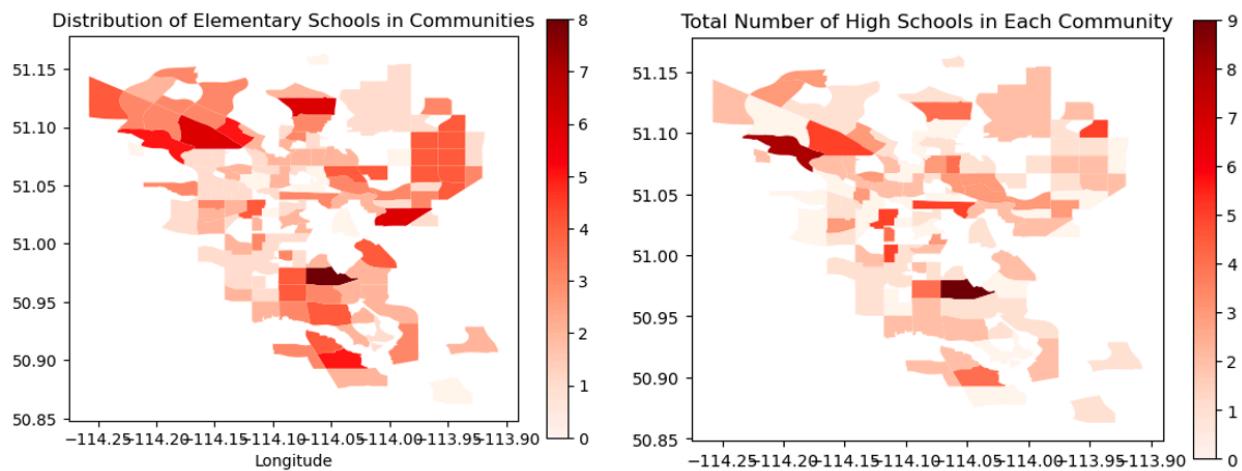


Fig. 12 Elementary Schools and High Schools Distribution

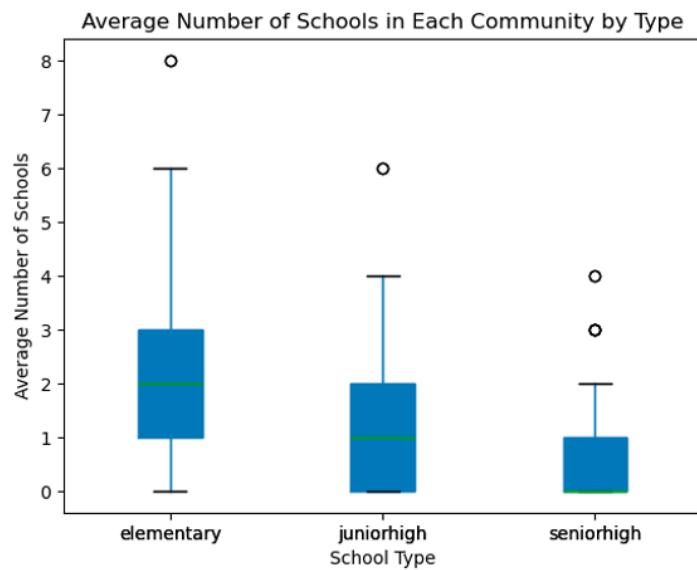


Fig. 13 Average Number of Schools in Communities

Data 604 Working with Data at Scale Group Project Report

(see Supplementary file 02_Public Service DataSet Process and Analysis.ipynb, Part II, chapter 2- chapter 3)

4.3.3 Investigate relations between each factor

Schools, libraries, clinics, and social development centers are highly correlated, however, property value and property value/square do not have strong relations with these factors. (see Supplementary file 02_Public Service DataSet Process and Analysis.ipynb, Part II, chapter 4- chapter 5)

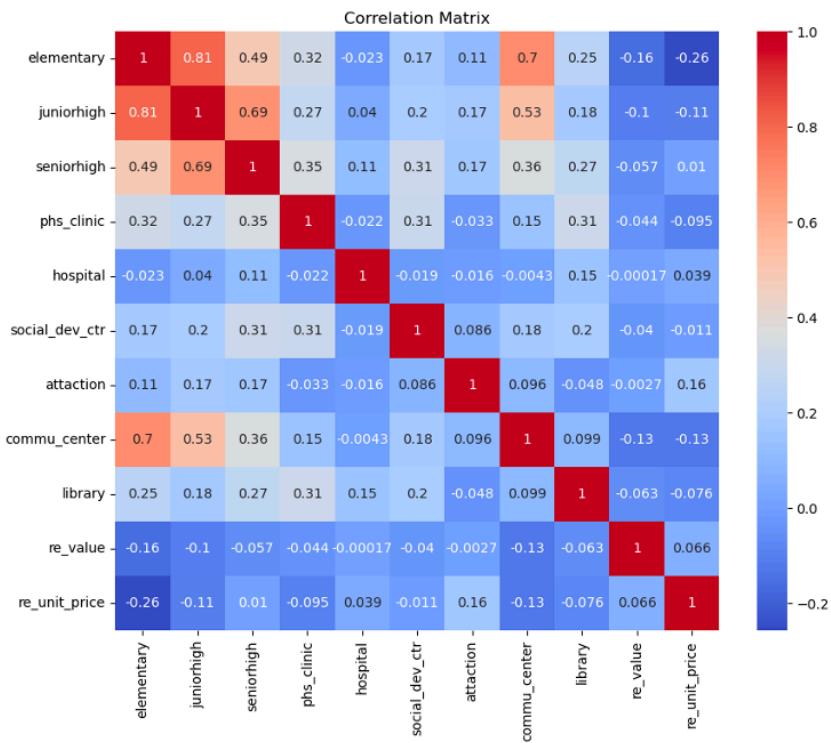


Fig. 14 Correlation Matrix of All Factors

4.3.4 Statistical analysis of the data set

To gain more statistical meaning of the exploration, we can use statistical methods to analyze the data, hoping to find some connections behind these numbers. (see Supplementary file 02_Public Service DataSet Process and Analysis.ipynb, Part II, chapter 5).

Data 604 Working with Data at Scale Group Project Report

As shown in the table, attraction and schools seem to have relations with the price at 0.05 level of significance. So, we can investigate the prices related to school and attraction information.

OLS Regression Results						
Dep. Variable:	re_unit_price	R-squared:	0.133			
Model:	OLS	Adj. R-squared:	0.090			
Method:	Least Squares	F-statistic:	3.044			
Date:	Thu, 28 Mar 2024	Prob (F-statistic):	0.000206			
Time:	10:53:50	Log-Likelihood:	142.95			
No. Observations:	188	AIC:	-265.9			
Df Residuals:	178	BIC:	-233.5			
Df Model:	9					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.2832	0.012	22.966	0.000	0.259	0.308
elementary	-0.0393	0.011	-3.507	0.001	-0.061	-0.017
juniorhigh	0.0241	0.019	1.289	0.199	-0.013	0.061
seniorhigh	0.0157	0.020	0.789	0.431	-0.024	0.055
phs_clinic	-0.0064	0.049	-0.131	0.896	-0.103	0.090
hospital	0.0142	0.085	0.167	0.867	-0.153	0.182
court	-1.566e-17	1.38e-16	-0.114	0.910	-2.88e-16	2.56e-16
social_dev_ctr	-0.0053	0.041	-0.128	0.898	-0.086	0.076
attaction	0.0996	0.043	2.327	0.021	0.015	0.184
visitor_info	0	0	nan	nan	0	0
commu_center	0.0222	0.024	0.930	0.353	-0.025	0.069
...						
Notes:						
[1]	Standard Errors assume that the covariance matrix of the errors is correctly specified.					
[2]	The smallest eigenvalue is 5.7e-35. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.					
Output is truncated. View as a scrollable element or open in a text editor . Adjust cell output settings ...						

Fig. 15 Regression Results Analysis

4.3.4 Final Conclusion of Public Service vs. Property Prices

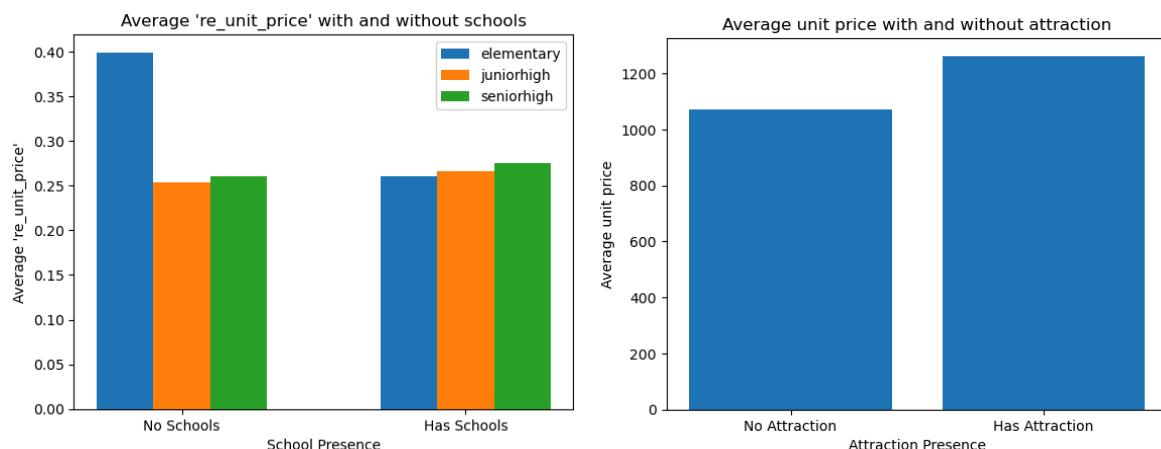


Fig. 16 Factors That Have Significant Impact on Property Prices

Data 604 Working with Data at Scale Group Project Report

It is noticeable that residential unit prices tend to be higher in areas with attractions or without elementary schools, whereas prices exhibit less variance between communities with and without junior and senior high schools.

Moreover, communities with high schools appear to have slightly higher unit prices compared to those without. One possible explanation for this trend is that communities without elementary schools may have lower population density, influencing property values. As for high schools, their distribution tends to be more dispersed, resulting in average prices reflecting broader market trends rather than localized factors.

4.4 Residential Property Price and Community Safety

In Part III, visualization (Codes in Supplementary file 03_Calgary Property Price Safety.ipynb):

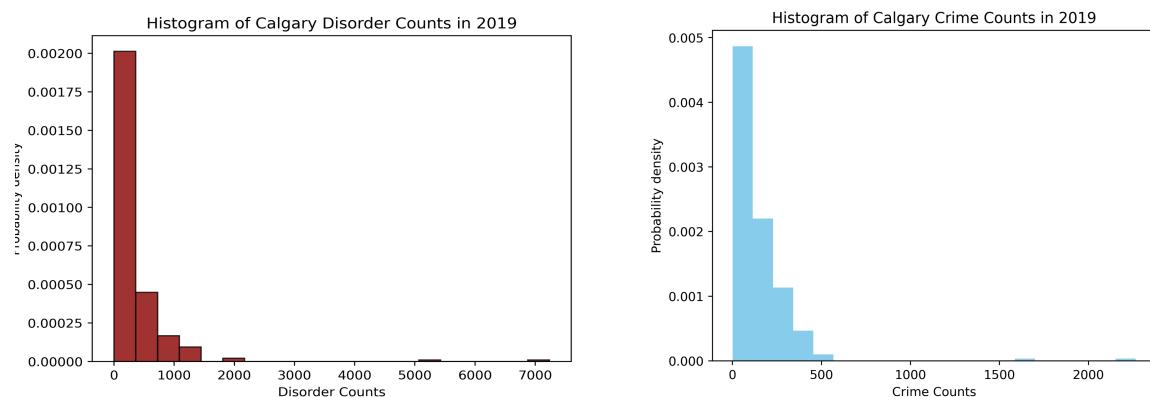


Fig. 17 Histogram of Calgary Disorder/Crime counts in 2019

The horizontal axis represents disorder/crime counts. The vertical axis represents the probability density.

Both of the histograms display a right-skewed distribution, indicating that in most communities disorder/crime counts are low. The highest probability density is in the first bar, which includes the lowest range of disorder/crime counts. As the disorder/crime counts increase, the probability density decreases sharply. Most of the communities have a disorder count <2000 and a crime count <500 . Only two communities get extremely high values in both graphs: Beltline with a disorder count over 7000 (crime count over 2000) and Downtown Commercial Core with a disorder count over 5000 (crime count over 1500).

Data 604 Working with Data at Scale Group Project Report

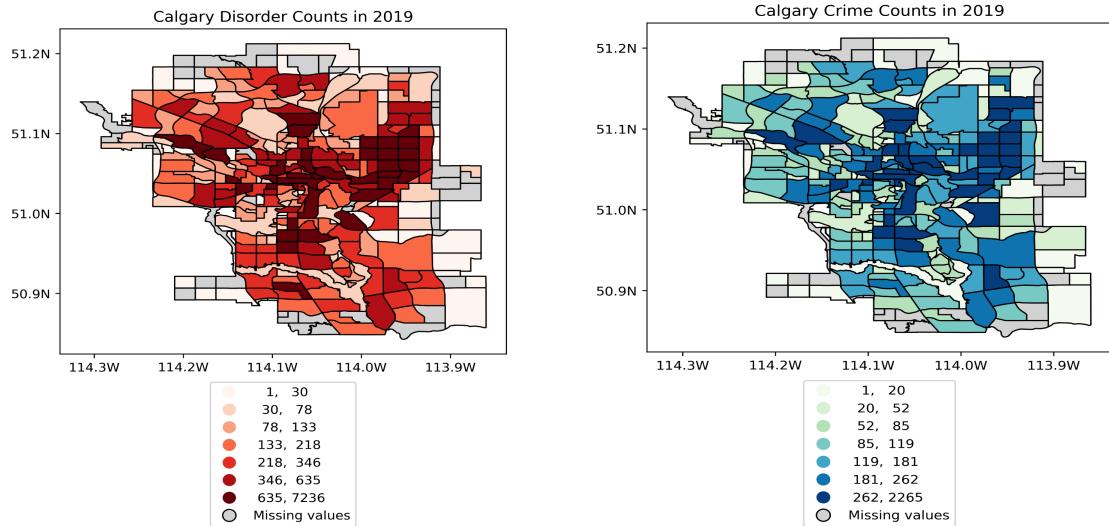


Fig. 18 Spatial Distributions of Calgary Disorder/Crime counts in 2019

The map shows that disorder/crime counts vary across Calgary. Comparing the two maps, we can see that the spatial distributions of community disorder counts and crime counts are quite similar in Calgary 2019. Darker shades are concentrated in central areas and northeast areas, indicating higher counts of disorder and crime in these communities. The distribution is not uniform, with some regions having significantly higher counts than others. The Downtown of the city, which is likely to be more densely populated and urbanized, shows the highest disorder/crime counts. The presence of lighter shades on the outskirts of the city implies lower disorder/crime counts in these areas. Additionally, there are some regions marked with the "Missing values" symbol, which means that data for these areas are not included in the analysis.

Next, we analyze if there is a relationship between Property Price and Disorder/Crime Counts

Data 604 Working with Data at Scale Group Project Report

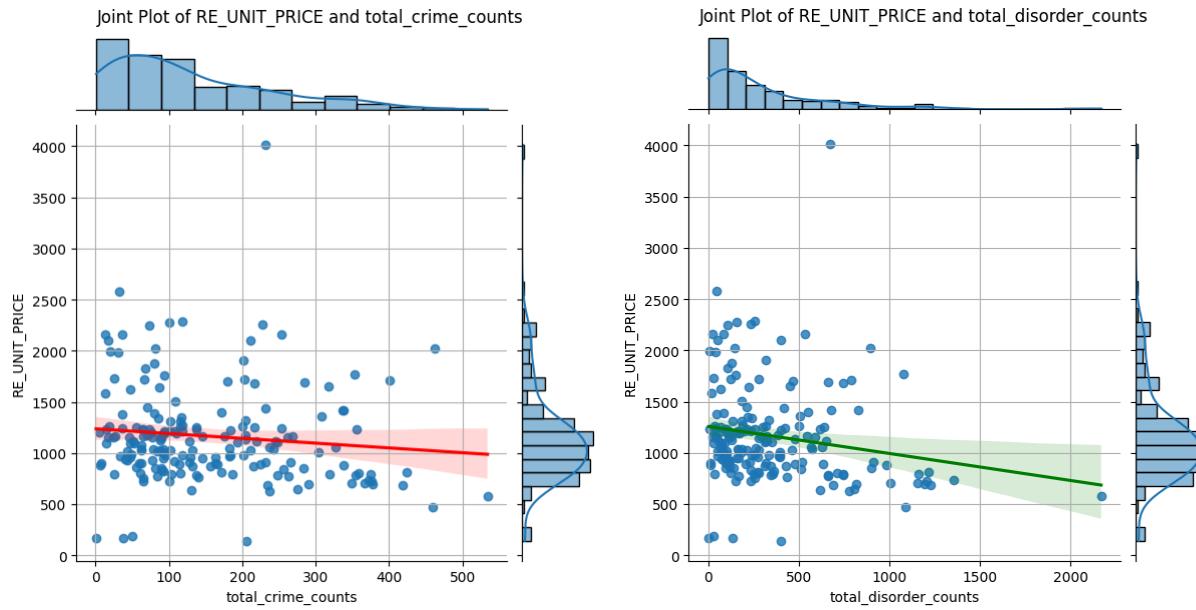


Fig. 19 Join Plots of RE_UNIT_PRICE VS Disorder/Crime counts

The images are joint plots showing the relationship between RE_UNIT_PRICE and disorder/crime counts. From the scatter plots and fitted lines, we can conclude that there is a weak negative correlation between RE_UNIT_PRICE and disorder/crime counts. Comparing the two fitted lines, the slope's value of disorder line seems smaller, meaning disorder counts might have a stronger negative correlation with RE_UNIT_PRICE than crime counts. The density histogram of RE_UNIT_PRICE shows that most communities have a residential unit price between 500 to 1500. The results of density histograms of disorder/crime counts are identical to our previous graphs.

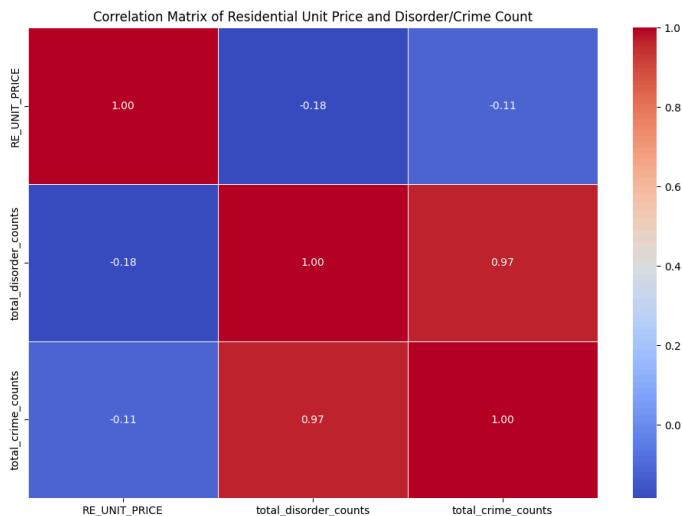


Fig. 20 Correlation Matrix of RE_UNIT_PRICE, Disorder Counts and Crime counts

The image is the correlation matrix of Residential Unit Price, Total Disorder Counts, and Total Crime Counts. The values in the matrix represent the correlation coefficients between the variables. We can see that disorder counts and crime counts are strongly positively

Data 604 Working with Data at Scale Group Project Report

correlated. However, they both have a very weak negative correlation with Residential Unit Price (-0.18 and -0.11) which are identical to our interpretation of scatter plot.

4.5 Residential Property Price and Demographic Factors

In Part III, visualization(Codes in Supplementary file 04_Calgary_Property_Price_Demographic.ipynb:

- Demographic density and gender ratio:

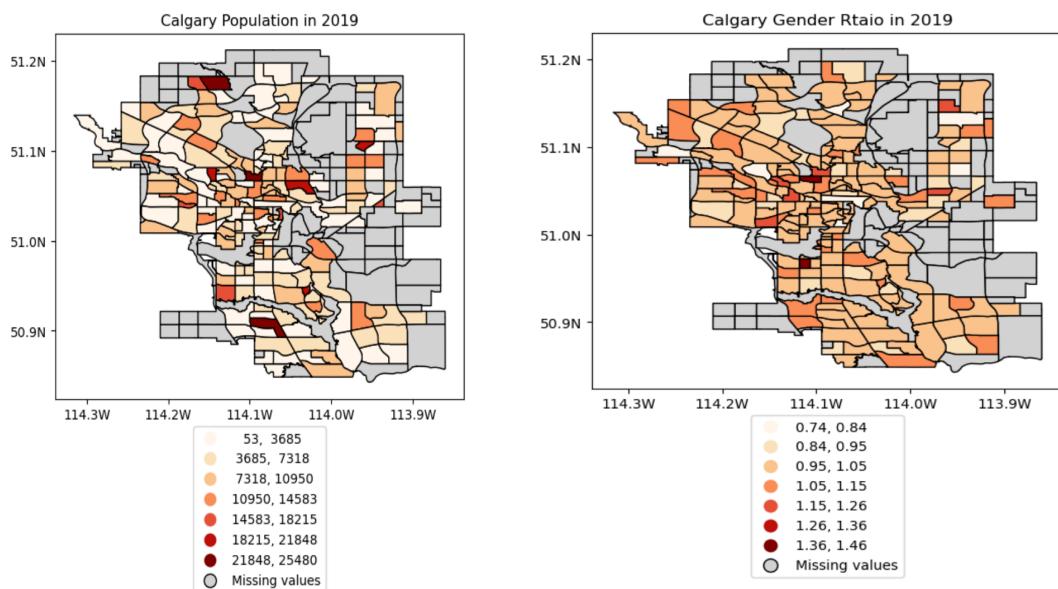


Fig. 21 Calgary Population in 2019 vs Calgary Gender Ratio in 2019

The table shows the total population across various communities as recorded in the 2019 census. The downtown area stands as the focal point of dense habitation, radiating outward gradually towards the city's periphery. Notably, the northwest and southwest sectors exhibit higher population concentrations compared to their counterparts in the northeast and southeast regions.

In terms of gender distribution, Calgary maintains a general equal distribution, with a balanced ratio between males and females. While certain pockets around the downtown core may display slight variations favoring one gender over the other, the majority of communities hold around a 1:1 ratio.

Data 604 Working with Data at Scale Group Project Report

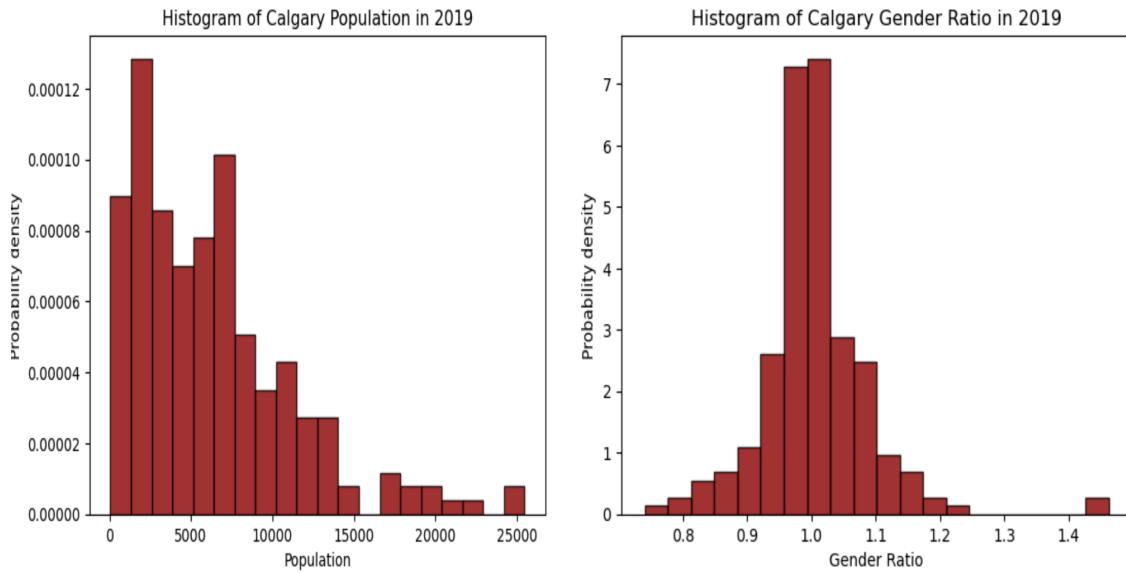


Fig. 22 Histogram of Calgary Population in 2019 vs Histogram of Calgary Gender Ratio in 2019

The histograms showing population distribution in Calgary reveal that the majority of areas exhibit populations under 10,000, a trend commensurate with the size of these communities. The second plot highlights a near-even distribution between genders, indicating a balanced representation across the city's demographic landscape.

- Scatter Plots

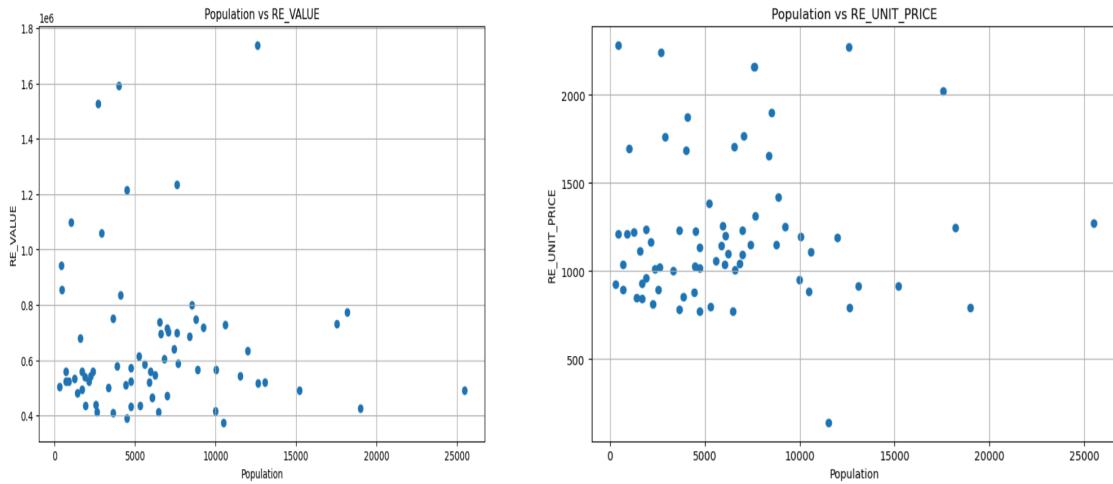


Fig. 23 Population vs RE_VALUE vs Population vs RE_UNIT_PRICE

These two plots explain the relationship between population and residual property value. The first plot illustrates this relationship, while the second plot examines the correlation between population and residual unit price. Upon the scatterplots, it

becomes evident that no strong linear relationship exists between the two variables.

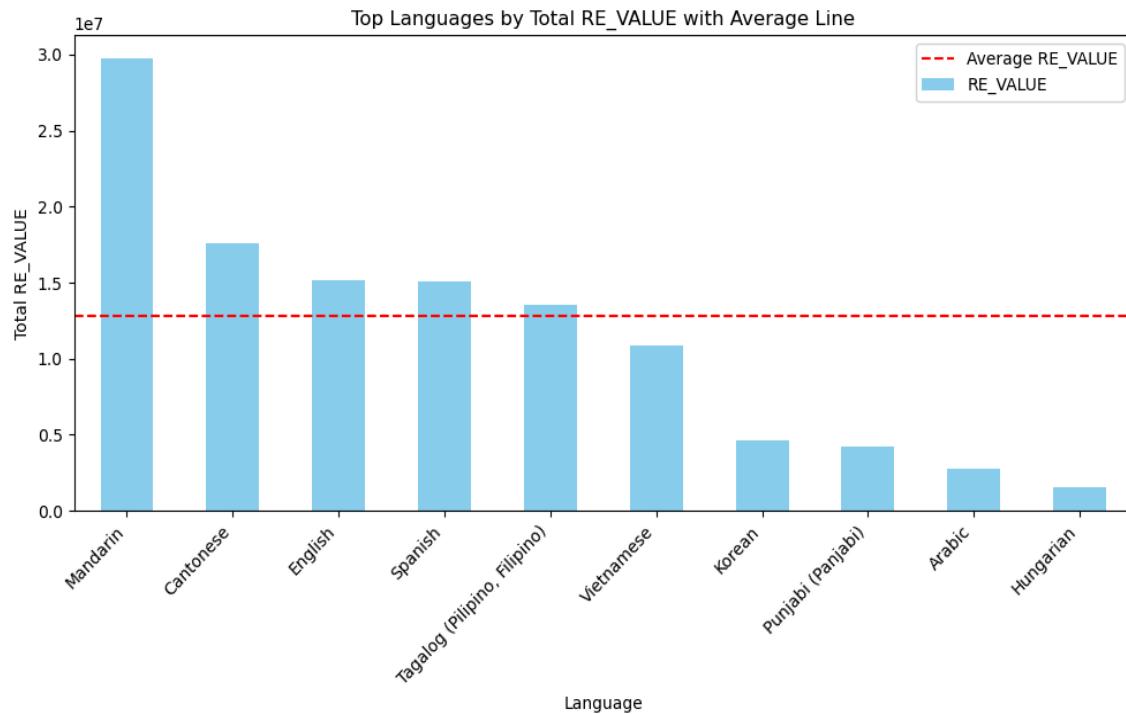


Fig. 24 Top Languages by Total RE_VALUE with Average Line

Upon analyzing the language preferences associated with residual property values, it becomes apparent that Chinese language speakers exhibit a tendency to purchase properties with higher values. Conversely, English, Spanish, and Tagalog speakers hold around the average price range. Languages such as Korean, Punjabi, Arabic, and Hungarian are correlated with properties priced below the average market value.

5 Discussion

5.1 Learned Techniques and reflection

Throughout this group project, I have gained valuable experience in various techniques and technologies, such as data cleaning and preprocessing using Python Pandas library, SQL queries for data extraction and manipulation, and data visualization for pattern recognition. Reflecting on our work, we realize that employing advanced machine learning algorithms and geospatial analysis tools could potentially enhance our understanding of property price trends and underlying factors. If given the opportunity to approach this project again, we would explore these advanced techniques and invest more time in feature engineering to uncover deeper insights and improve our analysis. (By Xinzheng Tang)

Data 604 Working with Data at Scale Group Project Report

Throughout the project, I gained valuable expertise in merging tables using primary keys and visualizing data through geospatial maps. This process involved integrating disparate datasets utilizing a common primary key, such as community codes, to create comprehensive analytical models. (By Alan Li)

I enhanced my programming skills in Python and SQL which are commonly used in Data Science. I also practiced writing clean, efficient, and readable code during this project and throughout this course. (By Li Chen)

The course and project provide me with a strong foundation in working with relational databases. I become familiar with the syntax and structure of SQL queries, filtering and sorting data, joining tables and generating spatial distribution. I enhance my knowledge and skills necessary to work with relational databases effectively throughout the project. (By Ze Yu)

5.2 Future Work

- The demographic analysis undertaken in this project was constrained by the absence of household income data, a crucial factor known to significantly influence property values. Acknowledging this limitation, further investigation and analysis related to household income would be imperative to gain a more comprehensive understanding of the dynamics impacting property values.
- Consider examining additional demographic or socioeconomic factors that may contribute to variations in residential property prices.
- Investigate potential correlations between school quality, academic performance, and residential property prices to gain deeper insights into the observed trends.
- The residential property price does not consider the difference between house and apartment, and there are also some non-residential properties that may have residential units. Such overlap would have an impact on the accuracy of the average residential property prices. Residential properties can be further categorized into 15 sub-types based on the City of Calgary's official definition to study the residential property price in depth.
- To get more details of how safety levels impact property price, we may also want to include factors like number of homeless and traffic conditions.

6 Conclusion

In conclusion, our study employed multiple datasets to investigate property prices in Calgary in 2019 and their potential influencing factors, such as public services, safety, and demographic factors. We utilized SQL queries for data aggregation and joins, and Python programming for data cleaning and visualization.

Data 604 Working with Data at Scale Group Project Report

Our findings reveal distinct patterns in property prices. Residential unit prices are higher downtown, decrease as we move outwards, and increase again towards the city's edge. Non-residential prices decrease from downtown to the edge, while farmland prices exhibit minimal variation. Furthermore, we observed that residential prices follow a normal distribution, farmland prices are left-skewed, and non-residential prices are right-skewed. The highest frequency of residential total prices and sizes is around 500 sqm and \$500,000, with the Red Carpet community having the highest residential prices and Belmont having the lowest.

We discovered that residential unit prices tend to be higher in areas with attractions or lacking elementary schools, while the presence of junior and senior high schools has less impact on prices. Communities with high schools show slightly higher unit prices than those without. This could be attributed to lower population density in communities without elementary schools, which could influence property values. The distribution of high schools appears more widespread, resulting in average prices reflecting broader market trends rather than localized factors.

The 2019 Calgary census data indicates that downtown, northwest, and southwest areas are densely populated compared to the northeast and southeast. Gender distribution is relatively balanced across the city, with minor variations in specific areas. Most communities have populations under 10,000. Our analysis of the relationship between population and property value, using histograms and scatter plots, reveals no strong linear correlation.

Finally, an examination of Calgary's disorder and crime data shows that most communities have less than 2,000 disorder counts and less than 500 crime counts, with higher-count communities concentrated in central and northeast areas. We observed a weak negative correlation between disorder/crime counts and property prices.

Overall, our study provides valuable insights into the factors affecting property prices in Calgary and their varying impacts across different property types and communities.

Supplementary files

Codes of data cleaning, queries, and visualization:

01_calgary_property_price_overview_update.ipynb
02_Public Service DataSet Process and Analysis.ipynb
03_Calgary Property Price Safety.ipynb
04_Calgary Property Price Demographic.ipynb
05_Calgary Property Price Merged.ipynb

Data 604 Working with Data at Scale Group Project Report

References:

- [1] <https://www150.statcan.gc.ca/n1/daily-quotidien/221026/dq221026a-eng.htm>
- [2] Hardin, H. (2023). Breaking the Immigration Taboo: Canada needs to reduce immigration dramatically until the housing crisis is resolved, especially in Vancouver. *Inroads: A Journal of Opinion*, (53), 159-178.
- [3] <https://www.nesto.ca/mortgage-basics/calgary-housing-market-outlook/>
- [4] Leishman, C. (2024). Understanding the Role of New Housing Supply Through Macro, Micro and Behavioural Perspectives. In *The Routledge Handbook of Housing Economics* (pp. 121-132). Routledge.
- [5] Li, N., Li, R. Y. M., & Nuttapong, J. (2022). Factors affect the housing prices in China: a systematic review of papers indexed in Chinese Science Citation Database. *Property Management*, 40(5), 780-796.
- [6] https://data.calgary.ca/Government/Historical-Property-Assessments-Parcel-/4ur7-wsgc/about_data
- [7] https://data.calgary.ca/Demographics/Census-by-Community-2019/rkfr-buzb/about_data
- [8] <https://data.calgary.ca/Services-and-Amenities/Schools-in-Communities/xmep-aasr>
- [9] <https://data.calgary.ca/Services-and-Amenities/Calgary-Health-Clinics-and-Hospitals/tsqf-wjr5>
- [10] https://data.calgary.ca/Health-and-Safety/Community-Disorder-Statistics/h3h6-kgme/data_preview
- [11] https://data.calgary.ca/Health-and-Safety/Community-Crime-Statistics/78gh-n26t/data_preview
- [12] <https://www.calgary.ca/planning/land-use/districts.html>