A Grassroots Approach to Calculating Real-time Local Costs of Living: Proof of Concept

April 2023

#

Щ Щ

#

Щ. Щ.

Charles Plante, Stan Yu, Scott Moeller, Michael Jean, Ripa Danielyan, Jackie Rorquist, and Scott Wells



21S

Acknowledgements

The team would like to thank the Saskatchewan Urban Municipalities Association for motivating this project and the other volunteers at Data for Good Regina and the Data for Good Canada network for helping and making this project possible.

About Data for Good Regina

Data For Good is a collective of do-gooders who want to use their powers for good to help improve our communities through data. We are a national not-forprofit organization with chapters nationwide that helps other not-for-profit and non-governmental organizations harness the power of their data to make more informed and better decisions in their quest to make their communities flourish.

Copyright

Attribution-NonCommercial CC BY-NC



Suggested Citation

Charles Plante, Stan Yu, Scott Moeller, Michael Jean, Ripa Danielyan, Jackie Rorquist, and Scott Wells. 2023. A Grassroots Approach to Calculating Realtime Local Costs of Living in Saskatchewan: Proof of Concept. Data for Good Regina.

Contact

Charles Plante, Data for Good Regina, Regina, Saskatchewan, E-mail: charles.plante@usask.ca.

Table of **Contents**

Introduction Section 1. The Opportunity Existing Measures of Cost of Living and Their Limits Section 2. The Problem and the Data 1. Identifying the Goods and Services Required to Meet Basic Standards 2. Collecting data about these Costs at the Lowest Level of Geography Possible 3. Summarizing These Costs for Different Regions and Times 4. Relating the Needs of Different People and Households to the Costs of **Different Goods and Services** 5. Interactively Visualizing this Information to Support Accessibility and Insight 6. Sharing Overall Indicators of the Cost of Living for Different Regions, Times, and Needs Section 3. A Novel Solution Demonstration, Community Housing Rental Costs in Saskatchewan in 2022 Mapping Available Housing Costs to Needs What is "Suitable" Housing? Section 4. Evaluation

Section 5. Scaling Up	13
Data Collection	13
Data Management and Analysis	14
Data Visualization	14
Online Hosting	14
Outreach, Communications, and Promotions	14

Appendix A: Detailed Summary of the Data Collection, Analysis, and

Visualization

15

1

1

2

3

3

3

4

4

5

5

5

6 10

12

13

Introduction

Communities in Saskatchewan need real-time local information about overall costs of living to plan and make evidence-informed decisions. Currently, the only routinely updated metrics communities can access are provided by Statistics Canada. They are not as timely nor sensitive as they need to be to capture regional realities, especially rural ones. At the same time, Big Data is changing everything and is making evidence-informed decision-making the most accessible it has ever been. Over the course of 2022, our Data for Good working group, consisting of community-minded and volunteer analysts and data scientists, set out to explore whether we could work with modern data tools to calculate and share real-time local cost of living estimates for Saskatchewan.

Specifically, Data for Good established a working group and set out to see if we could:

- 1.Collect cost of living data at lower levels of geography.
- 2. Do it faster than annually—maybe even in real-time.

We elected to focus our efforts on housing rental costs for as many communities in Saskatchewan as we could in 2022. This proof of concept provides an overview of the key aspects of what our team did, what we achieved, and a roadmap for how this work could be modulated and scaled going forward. Some investment by the Saskatchewan community is needed, but only a modest amount.

Section 1. The Opportunity

The digitalization of commerce has made price information more centralized and accessible than ever before. In the past, if we wanted to collect information on the costs of used vehicles or rent. we would have had to scan and record the classified ads in multiple newspapers and publications. Today, most advertisements are online, mainly on large-scale online marketplaces such as Auto Trader and Facebook. Similarly, most department and grocery stores host online versions of their products and prices. Although some of this data is scraped privately to support various business analytics services, there are not currently any efforts underway to collect and organize this information to support local communities and decision-makers.

At the same time, the proliferation of affordable processing power and the wide availability of open-source data analysis software and tools have radically democratized the use of data to support the invention of innovative data collection and analytics. These technological developments have been key precursors to establishing and growing Data for Good, a nonprofit that brings together people with data expertise and/or a desire to learn to work on data projects that benefit their local and regional communities.

In 2022, sparked by a request from the Saskatchewan Urban Municipalities Association (SUMA), Data for Good Regina established a working group to explore possibilities for improving existing local measures of the cost of living in Saskatchewan. The Data for Good volunteer team consisted of people with various skills, including data scraping, data processing, data analysis (i.e. analytics), and data visualization.

Existing measures of cost of living and their limits

Right now, the only routinely updated metrics communities have to understand the cost of living in their communities are Statistics Canada's Consumer Price Index, which updates monthly at the provincial level, and the Market Basket Measure (MBM) (which has also been recognized as Canada's official measure of poverty), which is more sensitive to local variation but is only updated annually and two years after the fact. These measures are not as sensitive as they need to be for regional differences in costs and not sufficiently timely. These limitations became especially apparent during the COVID-19 pandemic when realities on the ground shifted dramatically, almost overnight.

Existing measures are also largely insensitive to the fact that different kinds of people and households have different needs. For example, the MBM is calculated based on the consumption needs of a generic healthy family of four consisting of two adults and two school-aged children. The measure is applied to households of different sizes using a mathematical formula that is insensitive to changes in their composition.

Indicator	Market Basket Measure	Consumer Price Index	An improved measure
Level of resolution	Five regions: rural, < 30,000, 30,000 to 99,999, Saskatoon, and Regina	Saskatchewan, Saskatoon, Regina	All towns and rural municipalities
Update frequency	Annually, two years after the fact	Monthly, two months after the fact	Monthly or more, within a month
Components	Food, shelter, transportation, clothing, and "other"	Food, shelter, water, fuel and electricity, household operations, furnishing and equipment, clothing and footwear, transportation, health and personal care, recreation, education and reading, alcohol and tobacco	Food, shelter, water, fuel and electricity, household operations, furnishing and equipment, clothing and footwear, transportation, health and personal care, recreation, education and reading, alcohol and tobacco
Reference family	Family of four, two adults and two school-aged children	N/A	Flexible

Table 1. Existing routine indicators of cost of living, their strengths and their limitations, compared to a better measure.

For instance, the MBM does not account for the fact that according to the National Occupancy Standard created by the Canada Mortgage and Housing Corporation (CMHC), a family of three consisting of two married adults and one child requires a two-bedroom home to meet their housing need, while a family of three consisting of one single adult and two school-aged children of opposite sex require a threebedroom home.

The MBM also only collects data on the costs of a handful of goods and services compared to the Consumer Prices Index. Specifically, the MBM considers six categories: food, shelter, transportation, clothing, and an "other" category that includes things like household furnishing, equipment and personal hygiene products. The MBM has been criticized for being insufficiently sensitive to costs associated with child care and costs associated with more specialized needs, including different health needs, education and training.

A more accurate indicator of the cost of living would collect data at the level of towns and rural municipalities and draw on updated monthly or weekly information. It would also be flexible enough to examine the costs of living for people and households with different needs. Table 1 summarizes and compares the current and proposed indicators.

Section 2. The Problem and Profile the Data

By trial and error, the Data for Good team of volunteers determined that creating an improved measure of the cost of living for communities in Saskatchewan would necessitate six steps:

- 1. Identifying the goods and services required to meet basic living standards.
- 2. Collecting data about these costs at the lowest level of geography possible.
- 3. Summarizing these costs for different regions and times.
- 4. Relating the needs of different people and households to the costs of different goods and services.
- 5. Interactively visualizing this information to support accessibility and insight.
- 6. Sharing overall indicators of the cost of living for different regions, times, and needs.

1. Identifying the goods and services required to meet basic living standards.

As illustrated in Table 1, many kinds of goods and services contribute to living costs. An effective cost-of-living measure should consider as many of these as possible. In recent years, groups doing living wage calculations in Saskatchewan have identified additional categories of costs not covered by the MBM, including childcare, health, education, and savings for emergencies. A modular indicator, one that a user could easily add or remove the costs of different goods or services from, depending on their needs, could accommodate the incremental inclusion of new goods and services as data on them became available.

2. Collecting data about these costs at the lowest level of geography possible.

Different kinds of goods and services are purchased in multiple ways and varying amounts, which impacts the kinds of cost information that can be collected to support the cost of living calculations. For example, public transportation costs, such as the cost of a monthly or annual bus pass, tend to be set by and provided relatively freely on transit-provider websites, and the amount does not change very often. It is thus relatively easy to determine and update the cost of monthly or annual bus passes in a particular community. In contrast, the cost of used vehicles is set by a marketplace, and this number is not provided in one place. On websites like AutoTrader, which provides an online marketplace, many prices are listed but not in a way that makes them easy to work with to determine overall costs in communities. Listings are posted by sellers, and final selling prices may be lower.

An online marketplace like AutoTrader could provide data on local prices directly to not-forprofit teams like ours that are working on calculating local prices. However, web scraping is another option without this kind of partnership. Algorithms can be written to gather highly granular online price data at regular intervals.

3. Summarizing these costs for different regions and times.

The costs of individual goods and services are not all we need when calculating local living costs-that is, we are not interested in the cost of only one apple at one grocery store. Rather, we are interested in the average cost of an apple across all grocery stores in an area. We are also primarily interested in the cost today as opposed to some time in the recent past. Once we have gathered information on the local costs of goods and services, an additional step is needed to integrate this information and determine what it can tell us. Advanced statistical techniques can be used at this stage, ranging from simply calculating averages or medians over recent prices to implementing nowcasting tools that use past data to forecast prices in the present and near future. Techniques can also be developed to account for possible bias in cost data such as correcting for the aforementioned issue of inflated listing on AutoTrader. More complex statistical techniques require more historical data to be effective.

Another crucial technical question that arises is how to define communities geographically. Statistics Canada provides several geographical options for statistical analysis that may or may not overlap with how price information is collected and reported, or how communities understand their boundaries. Geographic definitions also need not be the same for all cost categories. For example, people may be willing to travel several hundred kilometers for some goods and services.

4. Relating the needs of different people and households to the costs of different goods and services.

Costs of living are not actually the same for every household. To calculate costs for any given household, we have to begin by considering what kind of household it is and its needs. For example, the National Occupancy Standard identifies what gualifies as suitable housing for households of different sizes, depending on the number of adults and children. Similarly, Health Canada identifies what gualifies as a nutritious food basket for people of different ages and activity levels. However, national standards of need are not available for every cost category and in their absence, we should consider these needs as they are understood locally. For example, no national mobility standard identifies the minimum transportation requirements of different households. Nonetheless, people in Lanigan, Saskatchewan, and many small towns like it, are well aware that they do not need to consider municipal public transportation costs when calculating living costs locally. However, they may suggest considering the cost of inter-city transportation options.

5. Interactively visualizing this information to support accessibility and insight

One way to ensure a cost-of-living measure can support local needs to the greatest extent possible is to share it with communities so they can easily read and interpret it and, if necessary, modify it. Business intelligence software has become an effective tool for visualizing information in a way that is accessible and can be manipulated by users. Local communities most likely want to understand how their costs compare to others and how their prices change over time. This can be effectively illustrated using heat maps, which use colour to contrast cost levels between regions, and trend figures that show changes over time. An interface could also be added to allow users to add and subtract different cost categories and to zoom in or out of communities. The underlying data could also be made freely available for download. Ideally, the front end could be hosted online and viewable without installing software or browser extensions.

6. Sharing overall indicators of the cost of living for different regions, times, and needs.

Once these pieces are in place, the calculation and reporting of overall costs for different households within different communities should be relatively straightforward and largely automatable. This could also introduce opportunities for collating and sharing this information with communities and stakeholders in other formats, such as short reports or creating additional visualizations. As long as the data is entered routinely, cost estimates and reports can be updated and disseminated in real-time. These results could be used to support additional analysis, even by communities themselves.

Section 3. A Novel Solution

Over the course of 2022, the Data for Good working group developed and trialled solutions for each of the steps identified in Section 2. We began by reviewing the Market Basket Measure and each of its six cost categories and considered where and how we could improve upon it. It was decided that housing costs, and especially rental costs, were an area where the team saw substantial room for improvement. Rental prices tend to fluctuate greatly and may not always be available for some housing types, especially those in smaller communities with few listings. However, much of the rental market today is listed on Facebook. The team contacted Facebook to share local rental listing price information but did not receive a response. Therefore the team turned to web scraping tools to extract the information we needed from the platform.



Figure 1. Overview of the steps that go into collecting and reporting on local costs in Saskatchewan.

The team scraped Facebook for rental price information approximately weekly between May 2022 and November 2022. This information was then uploaded to Google Drive and processed using a combination of R and Excel, then visualized using a GIS map in PowerBI.

The team also explored the feasibility of different steps in other cost categories, including home purchases, used vehicles, gas prices, public transportation and clothing. We contacted several websites to share price information with our group but found it difficult to make connections. Web scraping was viable, however, and we routinely scraped home and used vehicle prices throughout the fall of 2022. Public transportation prices were manually recorded from municipal transit web pages.

It became evident to the team that rural shoppers would likely have to travel into the provincial city centres to complete certain kinds of shopping, including clothing purchasing. Clothing prices were not scrapped but the team identified a list of leading sellers and their locations throughout the province.

In what follows, we summarize the proof of concept we settled on with respect to housing rental costs. Appendix A provides a detailed summary of all the avenues explored and the work completed by the team.

Demonstration: Community housing rental costs in Saskatchewan in 2022

Currently, the most reliable source of information Saskatchewanians have for local rental costs is the CMHC. The data they collect on local rental are primarily shared in an online Statistics Canada data table and the CMHC's annual rental market reports (CMHC also shares their data in their <u>online dashboard</u>). The data provided in these sources come from CMHC's <u>Rental Market Survey</u>, carried out each year in October and only for communities with populations greater than 10,000. For Saskatchewan, this includes Estevan, Lloydminster, Moose Jaw, North Battleford, Prince Albert, Regina, Saskatoon, Swift Current, Weyburn, and Yorkton.

While CMHC's rental costs offer a gold standard for data collection on rental cost information and a helpful benchmark for understanding costs in Saskatchewan, their methods are costly and do not show variation over time. CMHC usually reports the results of their October survey three months later, in January of the following year. Then communities have to wait another 12 months for the next update. CMHC also only provides data for the largest communities in the province. In a disproportionately rural province like Saskatchewan, costs for many, if not most of our communities, are never available from CMHC.

By scraping Facebook, the Data for Good working group was able to calculate point-intime estimates of rental costs for 100 communities throughout the province and tracked developments in prices weekly for each of four different types of rental units: 1bedroom, 2-bedroom, 3-bedroom, and 4 or more bedrooms. We scraped Facebook Marketplace for prices and types of rental units nearly weekly between May and December of 2022. We then cleaned and organized this data, calculated local cost estimates and uploaded them to an online dashboard created using PowerBI.



Figure 2. Screenshot of Data for Good working group Saskatchewan housing rental cost demonstration dashboard.

Notes: Current selection displayed in columns on the right is the average of all rental listings in Saskatchewan. Source: Data for Good Saskatchewan Cost of Living working group.

Access the Data for Good working group Saskatchewan housing rental cost demonstration dashboard via the following link: https://regina.dataforgood.ca/cost-of-living/

One of the challenges we encountered was that not every community always has rental units listed on Facebook Marketplace in every category of bedroom type. This is especially true for smaller and more remote communities. We calculated local costs by taking the median of the last five weekly median values for each type of rental unit and carried past estimates forward in weeks when no observations were recorded. Figure 3 illustrates observed weekly medians and trends in our cost estimates for 3-bedroom rental costs in the 35 communities for which we collected at least ten rental listings for that type of rental unit. To calculate rental housing cost estimates for all regions of the province, we aggregated community numbers for each of eighteen contiguous regions spanning the whole province (precisely, <u>Statistics Canada's eighteen</u> <u>census divisions</u>). The same method of calculating the median of weekly medians was used to calculate costs for these regions. The results for all rental unit types are presented in Figure 4.



Figure 3. Weekly median 3-bedroom rental and trends in reported estimated 3-bedroom cost estimates for communities with at least 10 3-bedroom listings.

Notes: Points indicate weekly medians and lines indicate reported estimated costs; weekly medians may be based on as few as one listing. Source: Data for Good Saskatchewan Cost of Living working group.

Figure 4. Cost of rental housing for all four types of rental units in eighteen regions throughout Saskatchewan between May and December of 2022.

Notes: Points indicate weekly medians and lines indicate reported estimated costs; weekly medians may be based on as few as one listing. Source: Data for Good Saskatchewan Cost of Living working group.

We used <u>PowerBI</u> to visualize our data and host it online so that anyone in Saskatchewan can access, manipulate, and glean insight from it. PowerBI, and interactive data visualization software like it, provides users with powerful tools for mixing, matching, and contrasting our cost estimates to understand costs across different rental unit types and for different combinations of regions and communities. We illustrate some of this functionality in Figure 5 and the tool can also be accessed at this <u>link</u>. To check the validity of our scraping method, we compared the results of our initial Facebook Marketplace data scrapes with the latest rental cost estimates provided by CMHC (which were for 2021). Table 2 presents differences for June 27th, one of the first days that we were able to collect reliable data. Impressively, differences were relatively minor for most communities (except Prince Albert and North Battleford) and, on average, less than 10%.

Figure 5

Notes: The complete dashboard is available at this link. Source: Data for Good Saskatchewan Cost of Living working group.

We also explored the feasibility of generating estimates of costs in other categories using a similar process and costs, including food, transportation, clothing, and housing purchase. For a detailed summary of the data collection tools, methods, and process, please see Appendix A.

Mapping available housing costs to need

Our demonstration dashboard does not map costs to the needs of different households, although future dashboards easily could by integrating the National Housing Standard into their interface. For example, a user could input the number of children and adults in their family, and then the dashboard could return the cost of the rental housing they need in different communities. The National Housing Standard's definition of suitable housing is provided below. The standard defines housing as suitable by relating different household compositions, in terms of numbers of adults and children of different sexes, to the number of bedrooms in the rental unit. Our process and dashboard already collect and list costs by the number of bedrooms. Table 3 provides some examples of how this mapping could work.

	СМНС	Facebook	Difference	% Difference
Estevan	\$1,100.67	\$1,175.00	-\$74.33	7%
Lloydminster, Alberta part	\$1,287.50	\$1,150.00	\$137.50	-11%
Lloydminster, Saskatchewan part	\$1,276.75	\$1,150.00	\$126.75	-10%
Moose Jaw	\$1,145.50	\$1,242.00	-\$96.50	8%
North Battleford	\$958.00	\$1,250.00	-\$292.00	30%
Prince Albert	\$1,011.25	\$1,325.00	-\$313.75	31%
Regina	\$1,361.00	\$1,400.00	-\$39.00	3%
Saskatoon	\$1,351.75	\$1,550.00	-\$198.25	15%
Swift Current	\$1,030.75	\$1,000.00	\$30.75	-3%
Yorkton	\$1,119.00	\$1,100.00	\$19.00	-2%

Table 2. Comparison of Canadian Mortgage and Housing Corporation rental cost estimates for a three-bedroom home in 2021with one day of Facebook Marketplace in 2022 (June 27th).

Notes: All amounts are presented in nominal Canadian Dollars and the time of data collection. The percent difference divides the difference by the CMHC value. Source: Statistics Canada, Table: 34-10-0133-01, Canadian Housing and Mortgage Corporation, 2021 Rental Market Survey.

What is "suitable" housing?

According to the the 2021 National Occupancy Standard, suitable housing refers to:

- A maximum of 2 persons per bedroom.
- Household members, of any age, living as part of a married or common-law couple share a bedroom with their spouse or common-law partner.
- Lone parents, of any age, have a separate bedroom from their children.
- Household members aged 18 or over have a separate bedroom, except those living as part of a married or common-law couple.
- Household members under 18 years of age of the same sex may share a bedroom, except lone parents and those living as part of a married or common-law couple.
- Household members under 5 years of age of the opposite sex may share a bedroom if doing so would reduce the number of required bedrooms. This situation would arise only in households with an odd number of males under 18, and odd number of females under 18 and at least one female and one male under the age of 5.

Example Family Type	Total occupants	Number of bedrooms
Two coupled adults	2	1
Single adult	1	1
Two unattached adults	2	2
Two parents with one or two infant-aged children	3-4	2
Two parents with one child-aged boy	3	2
Two parents with two child-aged boys	4	2
Two parents with up to two opposite-sex aged children of each sex	4-6	3
Any of the above maximums with one additional infant or child-aged boy or girl	+1	+1
Any of the above with one additional unattached adult	+1	+1

Table 3. Mapping of different household types to accommodation types available on Facebook Marketplace in Saskatchewanbased on National Occupancy Standard (NOS).

Section 4: Evaluation

Our grassroots Data for Good working group set out to see if we could improve on existing local measures of the cost of living in Saskatchewan provided by Statistics Canada. Given our limited resources, we focused on identifying a workable solution in the category of housing rental costs that could provide a template for advances in other categories of costs, while also exploring the feasibility of generating estimates of costs in other categories using a similar process and costs.

Compared to leading estimates of local rental costs provided by Statistics Canada and CMHC, our team of volunteers collectively worked just a few hours a week and were able to generate cost estimates for far more communities and weekly for four rental unit types. We were also able to display these data in a user-friendly dashboard that visualizes and makes them easy to manipulate further to gain additional insight. The dashboard provides a skeleton to which additional cost categories could be added and which could be augmented to map costs directly to different household needs. Together with the data identification, collection, and analysis pipeline we developed and implemented, our proof of concept is a great success and provides a clear template for more sustained local investment in data collection and reporting.

With less than a year of housing rental cost data, the calculation of local costs was necessarily rudimentary. Our estimates could be significantly refined if our pipeline could run for many more periods. For example, we could use small-area estimation methods that consider spatial correlations in housing prices to improve estimates in low-population areas. Or, we could use wellestablished time-series methods specifically designed to separate signal from noise, for example, by identifying and considering seasonal patterns. Additionally, although not the focus of this project, the data we have collected motivates new questions about housing in Saskatchewan, particularly in rural communities. For example, when there are no housing listings in a region for an extended period of time, what does it mean for the people living there? Are they still able to find housing through word of mouth? What about newcomers to these communities?

Section 5: Scaling Up

This project developed a pipeline that can be replicated, and, with modifications and improvements, much of it can be automated. The project also explored cost categories beyond the cost of rental housing, including housing purchases and transportation, that could be added to the data presented in Version 1 of the Cost of Living dashboard for Saskatchewan. In the future, users could ideally indicate which cost categories they want to view and explore. They could also indicate their household composition and location, and the dashboard would automatically populate information on their needs and associated costs. To scale up this project, the following elements are needed:

Data collection

• The team found that data was most obtainable through web scraping existing relevant websites. Collecting data through community partnerships, or trying to find existing datasets, proved to be more challenging and time-consuming and may require a longer time horizon.

- Automation within web scraping can be considered; however, websites and their security features, which actively detect and disrupt automated scraping, are always evolving. Thus human resourcing is required to routinely verify and update scraping tools.
- There continues to be potential for partnerships with organizations that already have data and are willing to share it for existing cost categories.
- Addition of new categories of cost that are not reflected in Statistics Canada's MBM or Consumer Price Index.

Data management and analysis

- Data analysis for this project was conducted in R. Other applications can be used, such as Python or Excel. However, embedding automation in the calculation will make it easier to update the analysis over time.
- Generate more refined calculations and report on the actual availability of goods and services in different regions; also consider augmenting with nowcasting methods and small area estimation methods as new data become available.
- Nowcasting and small area estimation methods use data on past and adjacent areas to generate meaningful estimates for smaller or less populated regions with less information.
- Additional mapping of cost categories to different household needs.

Data visualization

- This project determined that presenting the data and analysis on an interactive GIS map was the appropriate product for general users. To generate this map, web-mapping software such as Power BI or ArcGIS can be used.
- More refined data visualization of cost of living by category and in combination; for example, the addition of the functionality to add and subtract different cost categories.
- Add interactive functionality to allow users to specify their household composition and location, then have the dashboard report their needs and costs.

Online Hosting

- A hosting and storage solution for the data and GIS map is required.
- As much as possible, pursue strategies to automate the data collection, analysis, and visualization process that can be integrated and updated on the platform selected.

Outreach, Communications, and Promotions

- Ensure strategy and capacity are in place to reach the communities that will be using the data while being mindful that different regions prefer different communication approaches.
- A knowledge translation strategy is recommended to ensure users of the data understand the technical components of the product.
- Regular product communication to communities to ensure the use of the product, including promoting the product at community events, presenting at annual meetings, hosting webinars/lunch and learn events, and meeting with community leaders.

Our aim was to develop a prototype of a process and a platform for reporting local costs of living, that was more timely and more detailed than what was currently being provided. We also wanted to show this could be done on a modest budget. We hope our proof of concept provides an inspiration and template for future initiatives and investments in community data in Saskatchewan and elsewhere.

Appendix A: Detailed Summary of the Data Collection, Analysis, and Visualization

In Version 1 of the Saskatchewan Cost of Living dashboard, the cost categories of rental housing, property sales, and vehicle sales were captured. The following section outlines each step of the overall pipeline in further detail.

1. Web Scraping

Cost Category	Source of Data	Data Collection Dates	Variables Collected	Scraping Tool
Rental Housing	Facebook Marketplace	May, 2022 to November, 2022	location (town name), # of beds, price, title (of the Facebook ad), date the ad was scraped, URL (of the ad)	Python (using selenium web scraping library)
Housing Purchase	Zillow	May, 2022 to November, 2022	location (town name), number of beds, price, title (of the Facebook ad), date the ad was scraped, and URL (of the ad)	Jupyter Notebook, Python
Vehicle Sales	AutoTrader	May, 2022 to November, 2022	Price, location, year, age of the vehicle, make and model, and class of vehicle	Python (using selenium web scraping library)

2. Data Analysis

• The analysis was conducted using R and Excel.

3. Data Storage

• All data were stored as CSVs in Google Drive.

4. Data Visualization

Add-on Applications	Modeling Process	Visualization Process
2021 Statistics Canada's Census Boundary file Mapbox Visual Mapshaper change the projection of the map to wgs84) Create a sheet with <u>every town in</u> Caskatchewan and ts division Add the <u>GPS</u> coordinate for each own in the data set using <u>Awesome</u> Table	 Upload the cleaned data set- Create two identical data sets. Name one "Division" and the other "Coordinates". Each sheet has the following columns: Place, Latitude, Longitude, and Division from the complete list of towns. Create a data set for each category for each cost category (i.e. the cost of the rent). Each category should have the following columns: Place, Latitude, Longitude, Division, Median by Division for the week and Median by Town for the week. Ensure the "Division" data set has a many-to-many relationship with the "Division" column in each of the category data sets. Ensure the "Coordinates" data set has a many-to-many relationship with the "Town" column in each of the category data sets. 	 Map the "Divisions" column from the "Divisions" data set to the location and tooltips fields and the "Median by Division for the week" column in the color and tooltips fields. Set the "Latitude" and "Longitude" columns from the "Coordinates" data set to the Latitude and Longitude fields and the "Median by Town for week" column in the color and tooltips fields. Also, set the "Town" column in the "Coordinates" data set in the tooltips field. For the Divisions table, set the "Divisions" column from the "Divisions" data set to the Values fields and the "Median by Division for the week" column from the category data set to the Values field. For the Town's table, set the "Towns" column from the "Coordinates" data set to the Values fields and the "Median by Division for the week" column from the category data set to the Values field. For the Town's table, set the "Towns" column from the "Coordinates" data set to the Values fields and the "Median by Town for the week" column from the category data set to the Values field. For the Date table, set the "Date" column from the category data set to the Values fields and the "Median cost" column from the category data set to the Values field. Add a line chart - use the "week" column for the avis field and "median
	Add-on Applications	Add-on ApplicationsModeling Process2021 Statistics anada's Census oundary file Mapshaper change the projection of the map to wgs84) Create a sheet vith every town in askatchewan and s division Add the GPS coordinate for each own in the data sets using Awesome able- Upload the cleaned data set. Create two identical data sets. Name one "Division" and the other "Coordinates". Each sheet has the following columns: Place, Latitude, Longitude, and Division from the complete list of towns Create a data set for each category for each cost category for each cost category for each cost category for each cost category for the veek and Median by Division, Median by Division for the week and Median by Division" column in each of the category data sets Ensure the "Division" data set has a many-to-many relationship with the "Division" column in each of the category data sets.

cost" for the values field