THE UNIVERSITY OF BRITISH COLUMBIA

Statistics Canada | Statistique Canada

DATA-599 Capstone Project

Analysis of Rural Business Performance

Final Report

Tingwen Hua, Yilin Sun

Bowen Yang, Song Zhang

June 21, 2022

Contents

Executive Summary

Statistics Canada released the Rural Canada Business Profiles (RCBP) database for the first time on March 11, 2022. This database contains counts and financial data on small businesses and medium businesses in Canada for 2017, 2018 and 2019, aggregated by industry, geography, location, and some other relevant variables. This capstone project is expected to explore the new database along any dimensions and variables and gain some insights on the rural Canada business performance, then to augment findings by comparing RCBP with other public datasets.

We utilized Pandas and NumPy in Python to preprocess the 36 tables and total 859 variable columns in RCBP database, extracting 37 key variables from revenue, expense, profit/loss and financial ratios aspects. Using Altair in Python and Tableau, we explored and visualized the performance of the key variables in industry, geography, location, Incorporation status, and time dimensions, and delivered an analytical paper. By Tableau, we developed the RCBP dashboard to intuitively demonstrate the exploration results. Based on the key variables, we used multilinear regression models and backward selection to detect the relationship between the net profit/loss and other key variables and do the cross-analysis between unemployment rate and RCBP database. Lastly, using Dash in Python, we implemented an overview dashboard to visualize the RCBP database and other related datasets.

Based on the analysis, some characteristic findings are concluded:

- Agriculture, forestry, fishing and hunting industry had the greatest scale while the worst profit in rural areas
- Management of companies and enterprises was the most liquid and solvent industry in rural areas
- The distribution of direct and indirect expense in small and medium businesses had the opposite structures
- Alberta, Ontario and Quebec were the top three economically active provinces for small rural businesses
- Nunavut, British Columbia and Northwest Territories are the top three provinces whose small rural businesses are most profitable
- The Cross-analysis model indicates that as the total number of profitable businesses increase, the unemployment rate will decrease and if the debt-to-equity ratio is higher, the unemployment rate will be lower

These findings can be the directions for future work. Some studies on demographic, sociological, and ethnocultural fabric might be connected with these findings and be utilized to obtain more insights.

1. Introduction

Although most people still live in urban areas of Canada, which are close to the border of the United States or along Lake Ontario, rural areas comprise the vast majority of Canada's landmass[1]. Therefore, understanding the rural business performance is very important. It not only relates to the growth of population in the rural areas but also can provide some indication regarding migration and development trends.

From 2016 to 2021, the rural population has increased by 0.4% to 6.6 million, while the share of the population living in rural areas decreased from 18.7% in 2016 to 17.8% in 2021[2]. Although more people live in rural areas nowadays, most people still prefer urban cities. Thus, it is meaningful to analyze the rural business performance to see if society structure changes would have any influence on it.

This project is based on the Rural Canada Business Profile (RCBP) database provided by Statistic Canada in March 2022. By utilizing the data science techniques, including exploratory data analysis(EDA), data visualization and modeling, it aims to resolve the following research questions:

- What characteristics can be explored from the RCBP database?
- What variables have more significant influence on the net profit of rural businesses ?
- Is there a relationship between the unemployment rate and business performance?

In order to answer the first research question, we firstly performed EDA on the RCBP dataset to get as much information as we can. Then we chose 4 interesting and outstanding aspects from our analysis and dived deeper into them. Based on our EDA, we designed a RCBP dashboard to show the visualizations and an overview dashboard as a combination of RCBP and other related datasets.

In order to answer the second research question, we set up a multiple linear regression model to better understand the performance of Canadian rural businesses. We chose the net profit as the response variable in our model and further detected which variables have significant influences on the response variable.

In order to answer the second research question, we set up a multiple linear regression model to investigate the relationship between the business performance and unemployment rate. This model can estimate the unemployment rate based on the business financial information.

[1] "Population Growth in Canadas' rural areas, 2016 to 2021",Statistic Canada,https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-x/2021002/98-200-x2021002-eng.cfm
[2] "Population Growth in Canadas' rural areas, 2016 to 2021",Statistic Canada,https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-x/2021002/98-200-x2021002-eng.cfm

This report begins by the related works completed based on the RCBP datasets, which provides some general idea of the businesses information in Canada and some cross analysis Statistic Canada have done by using this dataset so far. Then it introduces the data we have used to analyze in this project. Next it explains the tools, methodology, techniques we have used to approach our goals. After that, it demonstrates the analysis of our findings, delivered products and conclusions. At the end, we provide some suggestions on the future directions and analysis.

2. Background and Related Work

The RCBP dataset was released in March 2022 by Statistics Canada, so it is a new dataset without too much related research. Statistics Canada published an analysis paper[3] on March 11, 2022 to show some initial analysis on the dataset. In their research, they focused on the counts of businesses and revenue performance of both small and medium businesses, with a concentration on the rural businesses. In their research, they found that small and medium businesses in rural areas generated, respectively, 17.1% ($139,300 million) and 14.2% ($69,300 million) of the revenues of all businesses in Canada in their respective size class in 2019. And for the counts of businesses, they stated that between 2017 and 2019, the numbers of both small and medium businesses increased at a slower rate in the country's rural areas than in its urban areas.

For the cross-analysis research question, as the rural and urban is the most important indicator for our research, we need to find a source that has similar definitions of rural and urban in order to conduct accurate research. Statistics Canada has made a Rural Canada Statistics Portal[4] that contains several data sources that are related to rural Canada. We finally selected the unemployment rate data source[5] to do the cross-analysis with our RCBP dataset, as both of the datasets have similar definitions on rural and urban, and it is an important key indicator to analyze rural Canada businesses.

[3] "A profile of businesses in rural Canada, 2017 to 2019.",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/21-006-x/21-006-x2022003-eng.htm
[4] "Rural Canada Statistics", Statistics Canada, https://www.statcan.gc.ca/en/subjects-start/society_and_community/rural_canada
[5] "Table 14-10-0375-01 Employment and unemployment rate, annual.", Statistics Canada, https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410037501

3. Data

3.1 Data description

3.1.1 Rural Canada Business Profiles

The dataset we used in this project is the Rural Canada Business Profiles(RCBP). It is a brand new database, which was first published by Statistics Canada on March 11, 2022. Now the available time period is from 2017 to 2019. RCBP provides financial profiles for Small and Medium Businesses(SMB) in rural and urban Canada, wherein the small businesses are the businesses with total annual revenues of $30,000 to $5,000,000, and the medium businesses are the businesses with total annual revenues of $5,000,001 to $20,000,000. In addition, rural Canada are the areas outside of Canada's Census Metropolitan Areas (CMA) and Census Agglomerations (CA), and urban Canada are the areas inside of Canada's Census Metropolitan Areas (CMA) and Census Agglomerations (CA)[6].

The dataset provides the main variables in following categories as tables[7]:

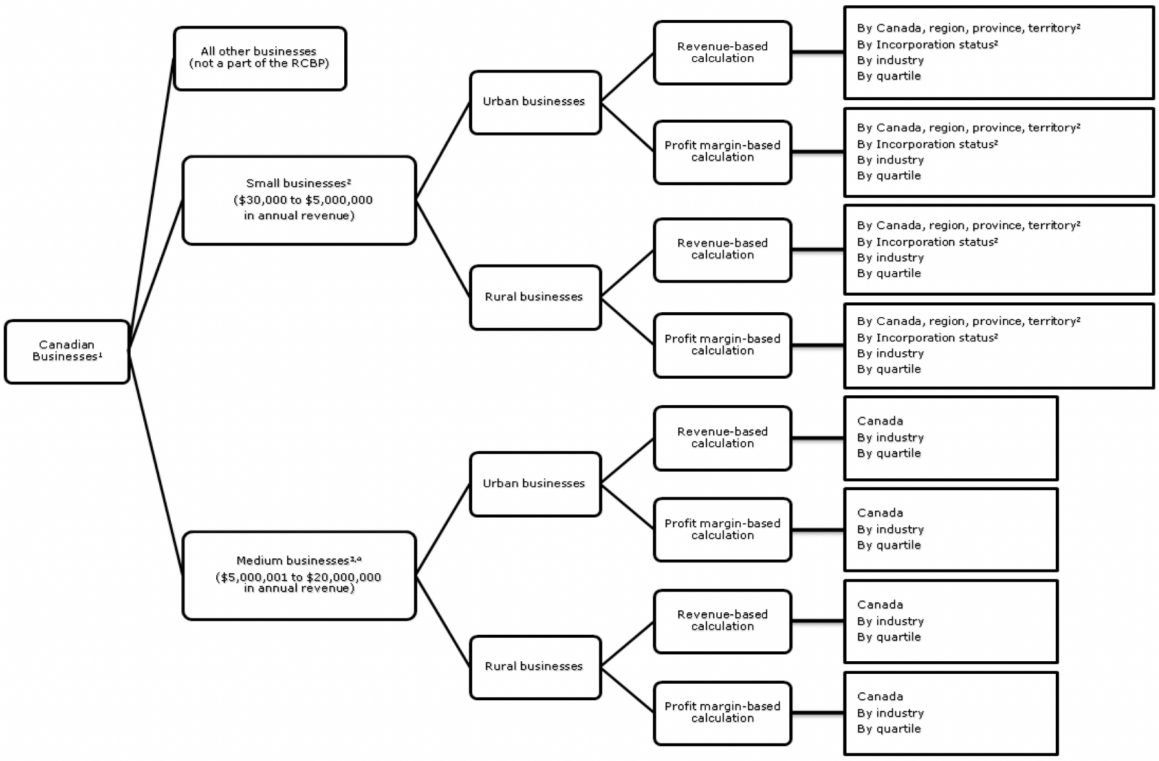
- Counts of businesses
- Revenue breakdowns
- Expense breakdowns
- Balance sheet items (assets, liabilities, equity breakdowns)
- Financial ratios

The data are managed by the following dimensions[8]:

- Small or medium size of business by annual revenues
- Various levels of geography (Canada, region, province/territory)
- Rural and urban areas
- Industry
- Incorporation status
- Profitable and non-profitable businesses

The structure of the RCBP database can be shown in Figure 1.

[6] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[7] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[8] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm



1. Includes all businesses that have submitted a Canadian tax return for the reference tax year.
2. Small businesses are classified by incorporation status into two groups: unincorporated and incorporated businesses. Data are also provided for the sum of these two categories.
3. As all medium businesses are incorporated, there are no breakdowns by incorporation status.
4. Province/territory breakdowns are not provided for medium businesses to protect confidentiality.

Figure 1: The structure of the Rural Canada Business Profiles Database[9]

3.1.2 Other data source

To do the cross-analysis of RCBP with other public Statistics Canada data, the annual employment and unemployment table was chosen. It is also available on Statistic Canada website[10]. It contains the unemployment rate data from 2011 to 2021, with dimensions of geography, sex and age group. It is used to find some relations between the unemployment rate and business performance.

[9] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[10] "Table 14-10-0375-01 Employment and unemployment rate, annual.", Statistics Canada, https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410037501

3.2 Data manipulation

3.2.1 Choose tables

The RCBP database has two sets of calculations. One is revenue-based and the other is profit margin-based. But the two calculations are only used to calculate the data quartiles. For the variable values, they are presented without regard to quartiles, so we choose revenue-based tables to do our analysis(profit margin-based tables will give the same results).

3.2.2 Data wrangling

In order to deal with the RCBP dataset, we self-defined two functions. The first function is **tab_reader** function, which is used to read all the tables. The second function is **selectData** function, which is used to select tables, combine 3 years data, and drop the irrelevant columns initially. One can easily adjust the function to select the desired data.

After that, the variables that can demonstrate the performance of the businesses were selected and one can choose to join or not join the variables according to the analytical purpose. The sample code can be seen in Appendix A.

3.2.3 Data cleaning

To protect the confidentiality of the businesses, some data are suppressed and shown as “X”. Also there are some data missed, as they are not available or not applicable. We treat the missing data as NA, and change the variable values to be numbers for analysis convenience. The code can be seen in Appendix A.

3.2.4 Final data

After data manipulation, we got two clearer datasets. One is for medium businesses, the other is for small businesses. And we kept the original dimensions of the two datasets. Figure 2 shows the structure of the clean datasets. To be noted that as the analysis methods and purposes change, the data manipulation code can be slightly adjusted.

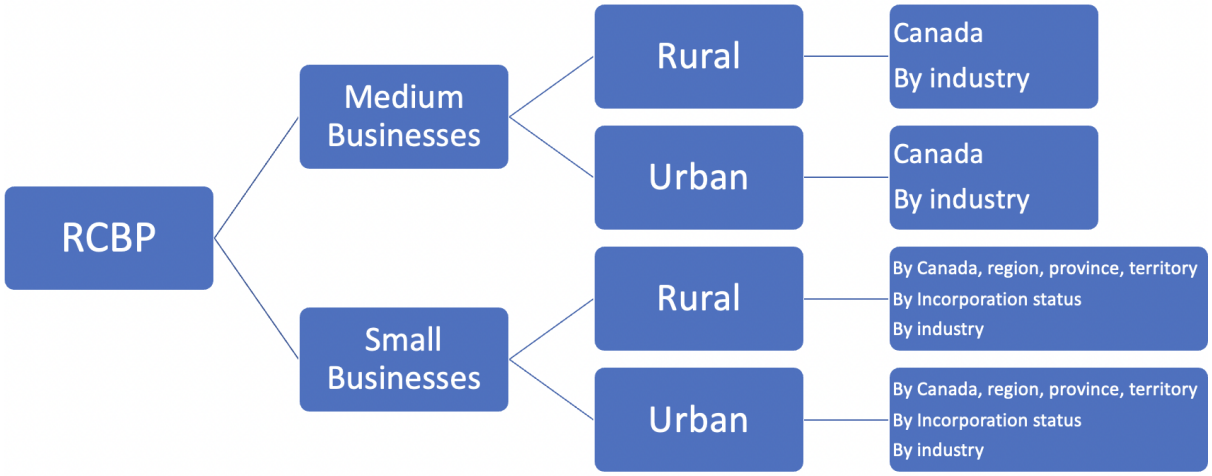


Figure 2: The structure of the clean RCBP datasets

4. Tools, Methodology, Techniques

According to our data source, this project can be divided into two phases. Phase I is the *Rural Canada Business Profiles* (RCBP) analysis phase for answering our first and second research question, while phase II is the cross-analysis phase for investigating our third research question. Each phase contains several steps to solve the corresponding questions. The table shown below will provide an overview of tools, methodology and techniques that were involved in each step.

All words with bold and italic font style in this section have corresponding links/descriptions in Appendix B for providing detailed information.

Phase I: RCBP Analysis	
Step	**Tools, methodology and techniques**
Data Processing	Python Pandas, Numpy
Data Visualization	Python Altair, Tableau
Dashboard Implementation	Python Altair, Dash, dash-bootstrap-components, Tableau
Dashboard Deployment	Heroku, GitHub, Tableau public

Modeling	Rstudio, Linear regression model, Backward selection
Phase II: Cross Analysis	
Step	**Tools, methodology and techniques**
Data Processing	Microsoft Excel, Python Pandas, Numpy
Data Visualization	Python Altair
Modeling	Linear regression model

We will provide detailed explanations of each step below.

4.1 Phase I

4.1.1 Data Processing

Generally speaking, data processing is the first step in data analysis. It is not only the exploration of the data but also the preparation for further analysis. Our data processing process includes data loading, data cleaning and data wrangling. All codes in this step were written in ***Python*** because of its powerful data processing libraries ***Pandas*** and ***Numpy***.

Since the data of the Rural Canada Business Profiles (RCBP) are stored in multiple csv (comma separated values) files, we implemented a ***Python*** function tab_reader to load data into ***Pandas*** dataframes firstly [***Appendix B[1]***]. For the data cleaning and wrangling part, we filtered out the incomplete rows and replaced the missing value with zero; we combined the data from 2017 to 2019 together to analyze the data year by year; we unpivoted expenses dataframe from wide to long format for preparing total expenses distribution exploration; and we removed aggregated geographical areas' data to avoid double counting in our later visualizations and analysis[***Appendix B[1]***]. Through these cleaning and wrangling operations, we generated a few new dataframes, and these new dataframes were the foundation of the following steps .

4.1.2 Data Visualization

Data visualization is an important part of exploratory data analysis (EDA), it enables people to understand data intuitively and quickly. For finding characteristics and gathering a deeper understanding of different

variables in the RCBP dataset, we chose data visualization as our second step.

Our group members have different backgrounds, and different preferences of visualization tools. Therefore, we selected ***Python Altair*** and ***Tableau***, which are 2 different tools, to make the visualizations. Both ***Python Altair*** and ***Tableau*** have their own advantages. For ***Python Altair***, it is also a library in python, so the dataframes generated in the previous step can be directly used in ***Altair*** functions to create corresponding charts. In terms of ***Tableau***, it allows users to create visualizations by simple "clicking" and "dragging" without coding anything.

Based on our exploratory purpose, we used bar charts to compare values across categories and industries, line charts to investigate changes in data over time, and geographic charts to analyze the distribution of data over geographic areas, and heat maps to explore the relationship between different variables. Finally, we summarized our visualizations and results of exploration in the analysis report[***Appendix B[2]***]. In the analysis report, the first 15 visualizations are made by ***Python Altair*** and the rest 11 visualizations are made by ***Tableau***.

4.1.3 Dashboard Implementation

Although we have drawn 26 visualizations and summarized them into an analysis report in the second step, there are still many aspects that were not covered. For allowing users to select their interested dimensions and providing a better overview of the RCBP database, as our client required, we implemented "RCBP Dashboard" and "Overview Dashboard" separately.

For "Overview Dashboard", it was implemented by using ***Dash***, ***Altair*** and ***dash-bootstrap-components*** libraries in ***Python***. The ***Dash*** library was used as the foundation to establish the dashboard web application. Most of the functionalities of "Overview Dashboard" were implemented with the power of the ***Dash*** library. More specifically, the interactions between filters and charts were implemented by callback functions of ***Dash***, while the clickable dashboard screenshots were implemented by placing the ***html.Img*** (image) component inside the ***html.A*** (Link) component. The ***dash-bootstrap-components*** library was employed for consistency and aesthetic purposes. For example, the layout was set up by ***Row***, ***Col*** and ***Card*** components, while buttons and dropdown lists were created by using the **Button** component. Therefore, each component has similar styles. Similar to the data visualization step, the ***Altair*** library was applied to plot charts shown in the dashboard. Line charts were used to show the changes of employment population and unemployment rate over selected time periods, while the bar chart was used to present population and dwelling count values.

For "RCBP Dashboard", it was implemented by using ***Tableau***, a third-party supplied platform to plot

data visualizations and design dashboards. We connected the Tableau to our clean dataset and utilized its convenience features to plot graphs and then combined them together as a completed Dashboard. It also can apply different filters to our desired plots and control them separately. Another important function of Tableau Dashboard is choosing plots by dropdown list, which allows us to combine related information into one space and include as much information as possible.

4.1.4 Dashboard Deployment

For convenience and communication purposes, we hope our dashboards can be visited like a public web site without any additional setting and configuration. Therefore, we need to deploy our dashboards somewhere to make it public. As the two dashboards were implemented by different techniques and tools, their deployment will be detailed below separately.

For "Overview Dashboard", it was deployed to ***Heroku***, "a platform as a service (PaaS) that enables developers to build, run, and operate applications entirely in the cloud"[11]. It supports the "one click" deployment from ***GitHub*** Repository and works well with ***Dash Python*** web application[***Appendix B[3]***]. More specifically, we placed our code of "Overview Dashboard" into an individual ***GitHub*** repository and linked the repository with ***Heroku***. Each time we want to update our dashboard, we just need to push the change to the repository and click the "Deploy Branch" button on the ***Heroku*** dashboard. By this way, the code can be easily managed, and the deployment process can be convenient. In consideration of the stability of our dashboard, we did not enable automatic deploys.

For "RCBP Dashboard", it was deployed to ***Tableau public*** which is powered by Tableau platform. It is very easy to deploy the Tableau dashboard as we just need to click on the publish button in the toolbar. The limitation of this publication is that users will see the Tableau website information along with our dashboard. The most desired solution for our clients will be a clean dashboard as a web page shown on their own Stats Canada website. A possible solution is to implement a web page by using ***HTML***, ***Javascript*** and ***CSS***, and then embed dashboards into it. The sample web page can be found in our GitHub repository, while the link is provided in appendix[***Appendix B[4]***].

4.1.5 Modeling

In order to answer the second research question about what variables have more significant influence on the net profit of rural businesses, we set up a multiple linear regression model. As the dataset has more than hundred variables, we choose 14 key variables as the explanatory variables to do the modeling.

[11] "Heroku: Cloud Application Platform", Heroku, https://www.heroku.com

These variables include the key variables about revenue, expenses, assets, liabilities and financial ratios. In order to find which variables have the direct correlation with the net profit loss, we use 2 methods in our variables selection process.

The first method is correlation matrix. It is simply a table which displays the correlation coefficients for different variables. We can find the key variables that have strong correlation with net profit loss from the correlation chart.

The second method is ***backward selection***. We add all 14 key variables as the explanatory variables and run a ***linear regression model*** first, and cut down the variables that don't have significant correlation with the net profit loss until all the explanatory variables have significant correlations with the response variable.

4.2 Phase II

4.2.1 Data Processing

Similar to Phase I, data processing is also the first step in our cross-analysis. All codes in this step were written in **Python** because of its powerful data processing libraries ***Pandas*** and ***Numpy***. Since we have finished most of Phase I works at this stage, the RCBP data has been cleaned and wrangled to a well organized format, it is easy to select necessary variables for cross-analysis. However, the unemployment rate information is included in a large csv file, which contains lots of other unnecessary information. Thus, we used ***Microsoft Excel*** to extract unemployment rate related data firstly and save them into a new csv file. Then we loaded the generated csv file into ***Pandas*** dataframe in ***Python***, and modified column names to match the corresponding columns in RCBP dataframes. Finally, we used the ***merge*** function in ***Pandas*** library to join the unemployment rate dataframe with the selected RCBP dataframe and filtered it by different conditions. The result dataframes were saved to csv files for later steps.

4.2.2 Data Visualization

After data processing, we have dataframes that contain all variables we need for analysis of the relationship between RCBP data and unemployment rate. To explore their relationships, we chose data visualization as our second step. Based on our purpose, we drew scatter plots by using ***Python Altair*** library to look for any trends and patterns in our data. Some "outliers" that have significant unemployment rates were found from visualization. After removing the outliers by applying a threshold 10.5%, the remaining data showed roughly linear correlations[[***Appendix B[5]***]].

4.2.3 Modeling

In order to answer the third research question, we set up a ***linear regression model*** that can use some business financial indicators to make some simple conclusions on the unemployment rate in Canada[***Appendix B[6]***]. Based on common sense, we believe that if the economics are blooming in that region, then the unemployment rate will be relatively low. Therefore, we choose the number of profitable businesses, the average total equity and debt to equity ratio in the Rural Canada Business Profile (RCBP) dataset which reflects the probability and liquidity of companies as our independent variables. Our response variable is the unemployment rate from the annual employment and unemployment table, which are also available on Statistic Canada website.

The data we chose from RCBP are the aggregated data including all industries within different regions and provinces. We joined the data with the unemployment rate according to the region and province. Since we do not have provincial data from medium businesses in RCBP, we won't have enough data sources to train the model for medium businesses. Therefore, our model focuses on the performance of small businesses and the unemployment rate in the same region.

5. Analysis and Interpretation

5.1 Characteristics of Rural Business Performance

In order to discover the characteristics of rural business performance, we explored the RCBP database from revenue, expense and profitability aspects, and implemented an exploratory data analysis (EDA). Based on our EDA, we composed an analytical paper which included comprehensive analysis of the RCBP datasets.

In addition to our EDA, we developed two interactive dashboards for the RCBP dataset, which are small business dashboard and medium business dashboard. Both dashboards present the outstanding financial records from 2017 to 2019. Users can use the dropdown lists to choose the filters and investigate specific information.

We also created an overview dashboard for users as an instruction guide for RCBP dashboard and an information reference. Users can see a detailed description of our key terms, data sources,filter applications and other related datasets for further interests.

After a comprehensive analysis, we found five outstanding facts which we applied to a deeper analysis.

5.1.1 Agriculture, forestry, fishing and hunting industry had the greatest amount while the worst profit in rural areas

Agriculture, forestry, fishing and hunting had the greatest total number of businesses and the greatest total revenue in rural areas from 2017 to 2019. In this section (5.1.1), we will only talk about the results for 2019, as the features in these three years are very similar[***Appendix C***]. In 2019, Agriculture, forestry, fishing and hunting had 51,950 small and medium businesses in rural areas, accounting for 16.24% of Canadian small and medium rural businesses, which was the greatest industry in rural Canada. In addition, its total revenue in 2019 was $42.99 billion for the small and medium rural businesses, accounting for 20.61% of the total revenue of all these businesses, which was also the greatest one.

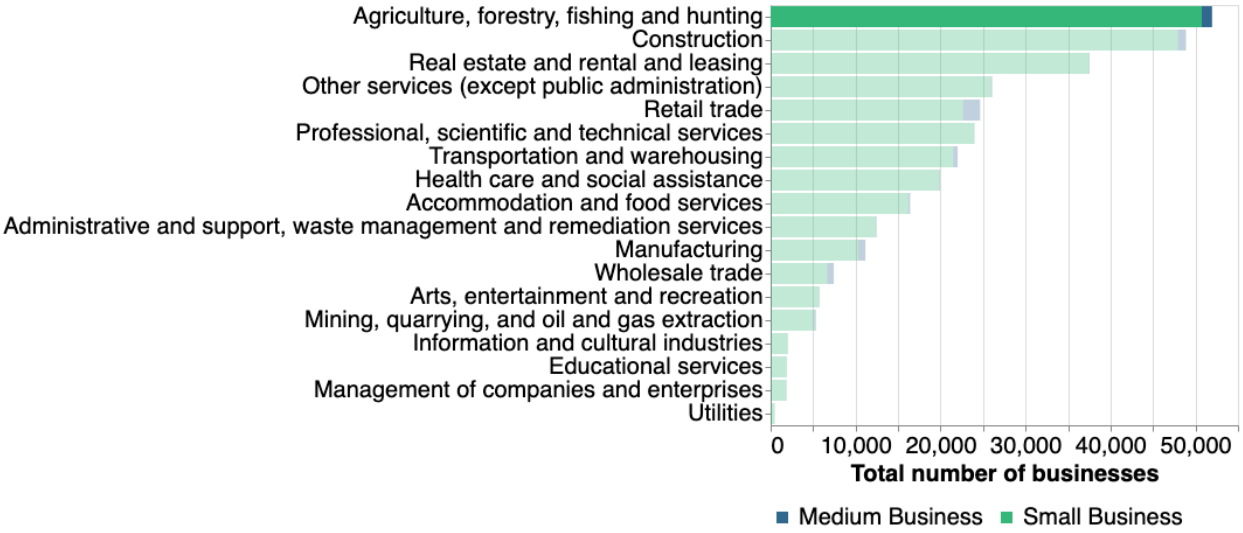


Figure 3: Total number of small and medium rural businesses of different industries in 2019

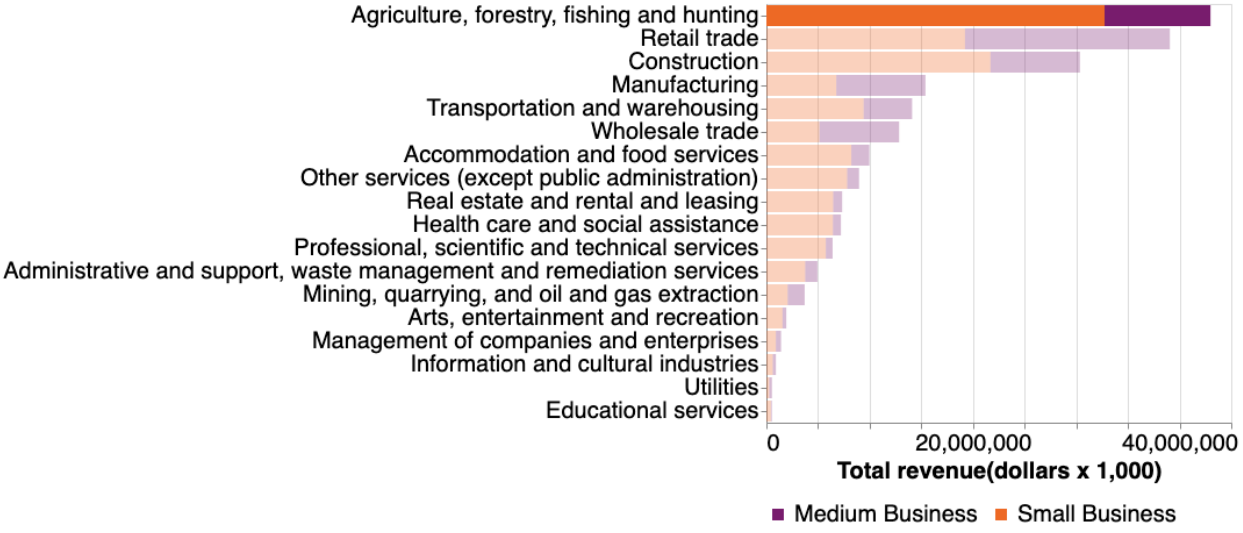


Figure 4: Total revenue of small and medium rural businesses of different industries in 2019

However, when we looked at the net profit and loss of different industries in 2019, Agriculture, forestry, fishing and hunting was the only industry that had a negative net profit in both small (-$89,800) and medium (-$1.063million) rural businesses, which means it was the only loss-making industry in rural Canada. Especially for the small businesses, we have one more variable—percent of profitable small businesses to measure the profit situation of different industries. Agriculture, forestry, fishing and hunting industry had the lowest percent of profitable businesses in both rural(51.5%) and urban(53.9%).

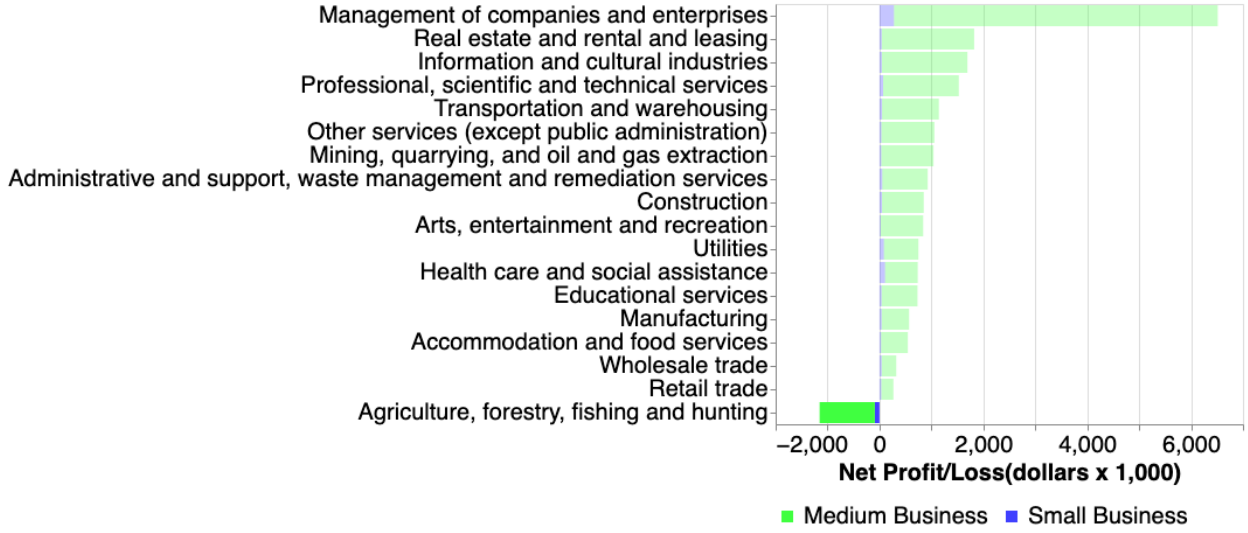


Figure 5: Net profit and loss of small and medium rural businesses of different industries in 2019

To further study the performance of Agriculture, forestry, fishing and hunting industry in rural Canada, we checked its gross margin, which is calculated as (sales of goods and services - costs of sales) * 100)/ (sales of goods and services) and provides a relative measure of the profitability or profit margin[12]. In 2019, the gross margin of small rural businesses in the Agriculture, forestry, fishing and hunting industry was 95.5%, which was the highest among the industries; and that of medium rural businesses in the Agriculture, forestry, fishing and hunting industry was 82.5%, which came third place among the industries. Therefore, the gross margin demonstrated that the profitability of the Agriculture, forestry, fishing and hunting industry was very excellent.

[12] https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm

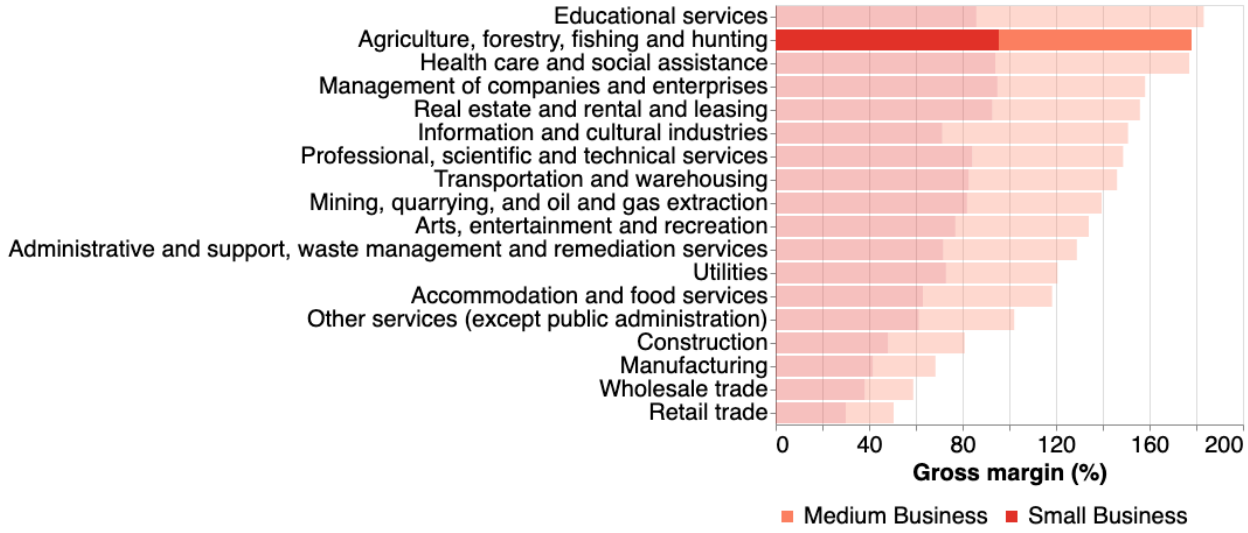


Figure 6: Gross margin of small and medium rural businesses of different industries in 2019

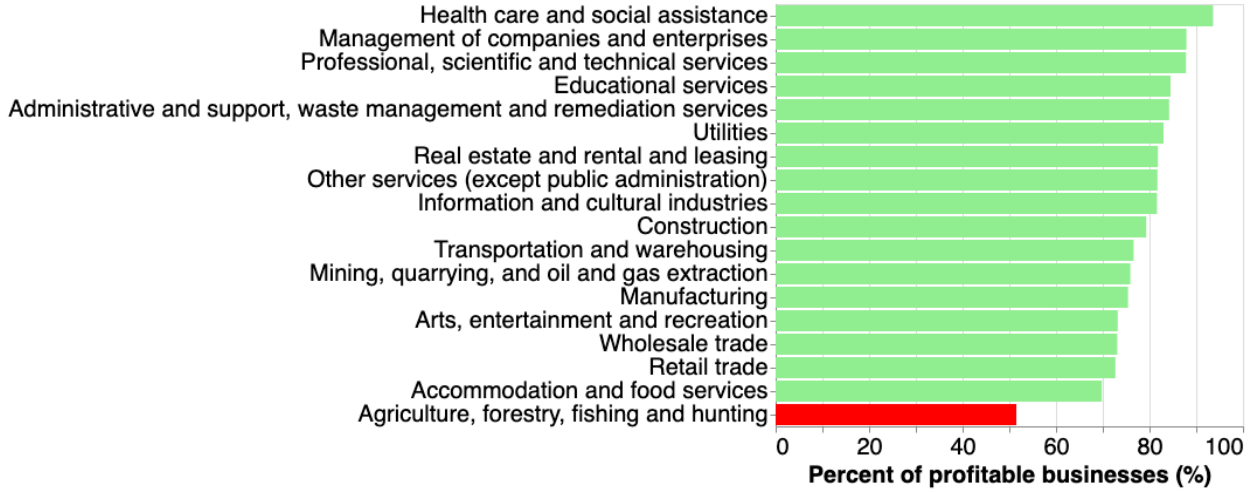


Figure 7: Percent of profitable businesses of small rural businesses of different industries in 2019

As the net profit equals the total revenue minus total expense, and the total revenue and the gross margin of Agriculture, forestry, fishing and hunting industry performed very well among all the industries, we decided to further check the expenses of the industries. The total expenses of the Agriculture, forestry, fishing and hunting industry was undoubtedly the largest. So we stepped further to analyze the cost of sales (direct expenses) and operating expenses (indirect expenses). In 2019, The cost of sales percent to revenue was very low for the Agriculture, forestry, fishing and hunting industry, which was 4.5% for small rural businesses (the smallest among the industries) and 16% for medium rural businesses (the fourth smallest among the industries). But its operating expenses percent to revenue was very high, with 109.4% for small rural businesses and 96.6% for medium rural businesses, and both the two numbers were the largest among small rural and medium rural businesses. As the operating expenses include all

expenses incurred in the course of running the business, the loss in the Agriculture, forestry, fishing and hunting industry might be the result of the "expensive" running cost.

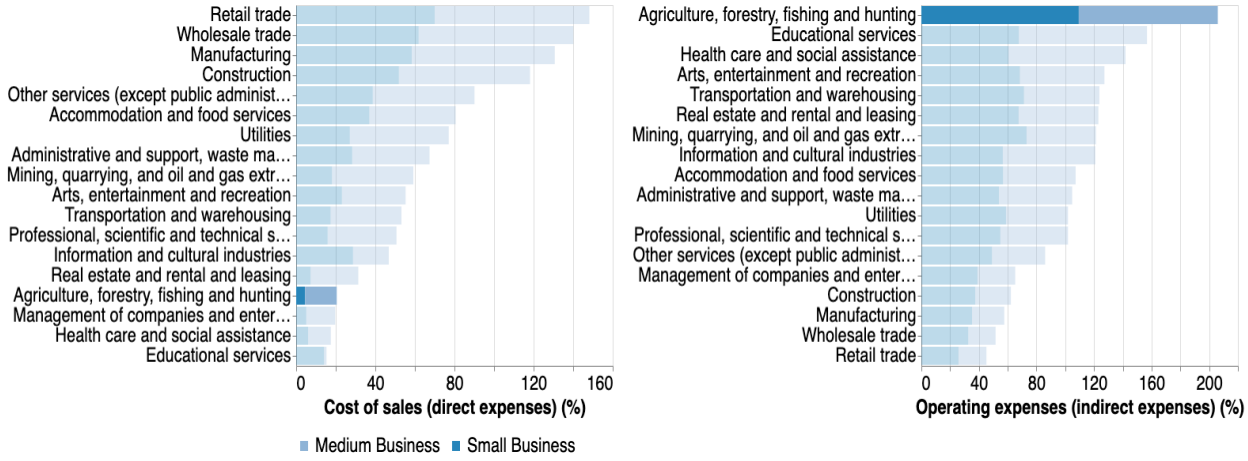


Figure 8: Expenses to revenue of small and medium rural businesses of different industries in 2019

5.1.2 Management of companies and enterprises Industry has the best debt situation and highest average revenue for rural medium businesses

In contrast to other industries, the Management of companies and enterprises industry has a better revenue and debt situation in rural areas according to our analysis.

The management industry has a higher average revenue among other industries in rural areas for medium businesses. Even though the total revenue of Management of companies and enterprises only accounted for 0.7% in the total revenue and 0.6% in the total number of businesses of the rural areas, its average revenue ranked first among the industries in rural areas in 2019 for medium businesses. The average revenue is more than 10.52 million dollars for medium rural businesses in the management industry, which is 1.13 times of the average industry revenue in rural areas. For small rural businesses, the average revenue is about 0.5 million dollars in the management industry. It is not in the top 5 industries considering the average revenue for all small rural businesses. The businesses in the management industry are not in good performance in contrast to the medium businesses due to the bad performance of unincorporated businesses in small businesses. Figure 9 shows the average revenue of rural businesses in 2019 based on different industries.

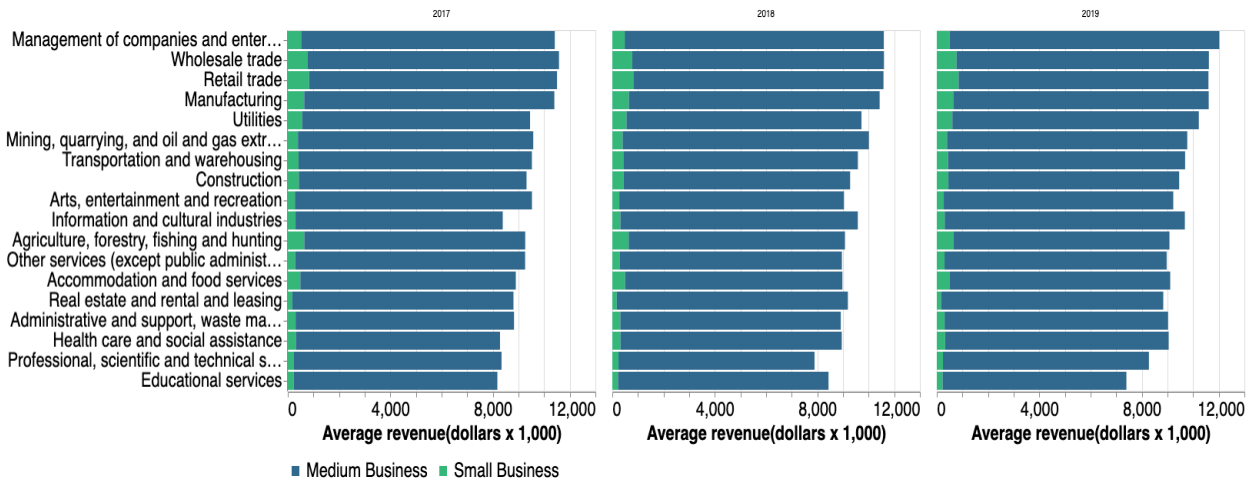


Figure 9: Average Revenue of rural businesses based on different industries in 2019

When we consider the debt situation to see the businesses liquidity and solvency, we found that both the current debt to equity[13] and debt to equity ratio[14] are low for both small and medium rural businesses of the management industry. In 2019, the current debt to equity and debt to equity ratio of the management industry are both the lowest among the other industries in rural areas. The current debt to equity is 15.1% (medium business) and 16.7% (small business) for the businesses of the rural management industry. It is significantly lower than that of all industries' average, which is 76%(medium business) and 43.6%(small business). For the debt to equity ratio, it is 0.3 times (medium business) and 0.4 times (small business) for the businesses of the rural management industry. It is significantly lower than that of all industries' average, which is 1.4 times (medium business) and 1.2 times (small business). That means the management industry is in a good condition of liquidity and solvency from the debt perspective. Figure 10 and 11 shows the current debt to equity and debt to equity ratio of rural medium and small businesses.

[13] Calculated by (current liabilities)*100%/(equity), which can show the liquidity of a business
[14] Calculated by total liabilities/total equity, which can show the solvency of a business

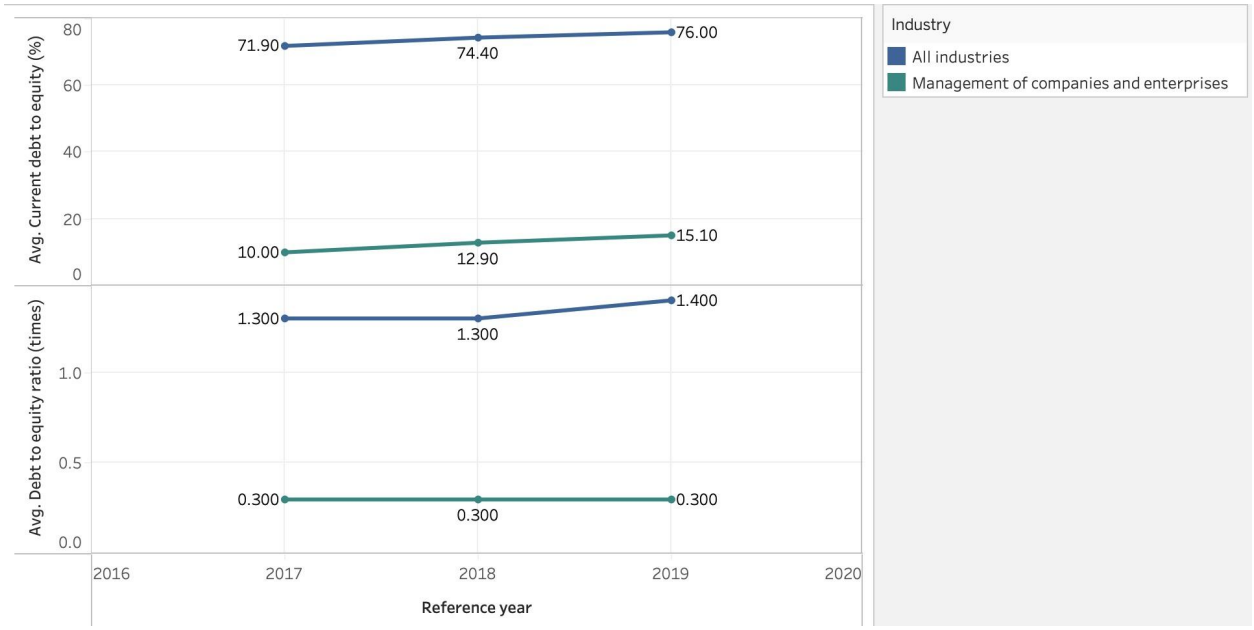


Figure 10: Current Debt to equity and Debt to equity ratio of management and all industry average of rural medium businesses

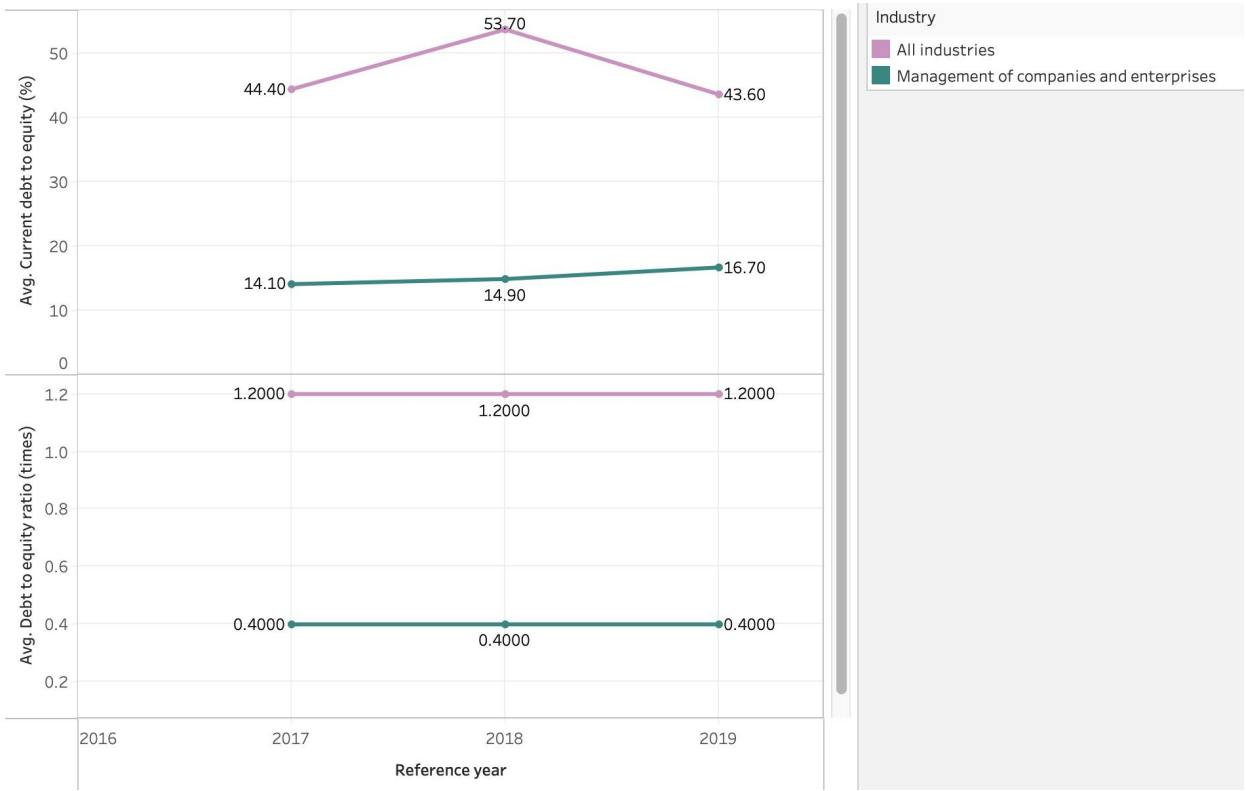


Figure 11: Current Debt to equity and Debt to equity ratio of management and all industry average of rural small businesses

However, not all businesses in the management industry are in a good financial situation. The non-profitable small businesses in rural areas of the management industry faced the greatest net loss in 2019, even though the industry had the greatest net profit for the profitable small businesses. The greatest net loss of the non-profitable small businesses are mainly caused by the high expense in contrast to other industries and profitable small businesses in the management industry. Even though the expenses of the management industry increased in both rural and urban areas of small businesses, the expenses for profitable small businesses are significantly lower than that of other industries, that is a reason that leads to greatest net profit for the profitable small businesses. Figure 12 and 13 shows

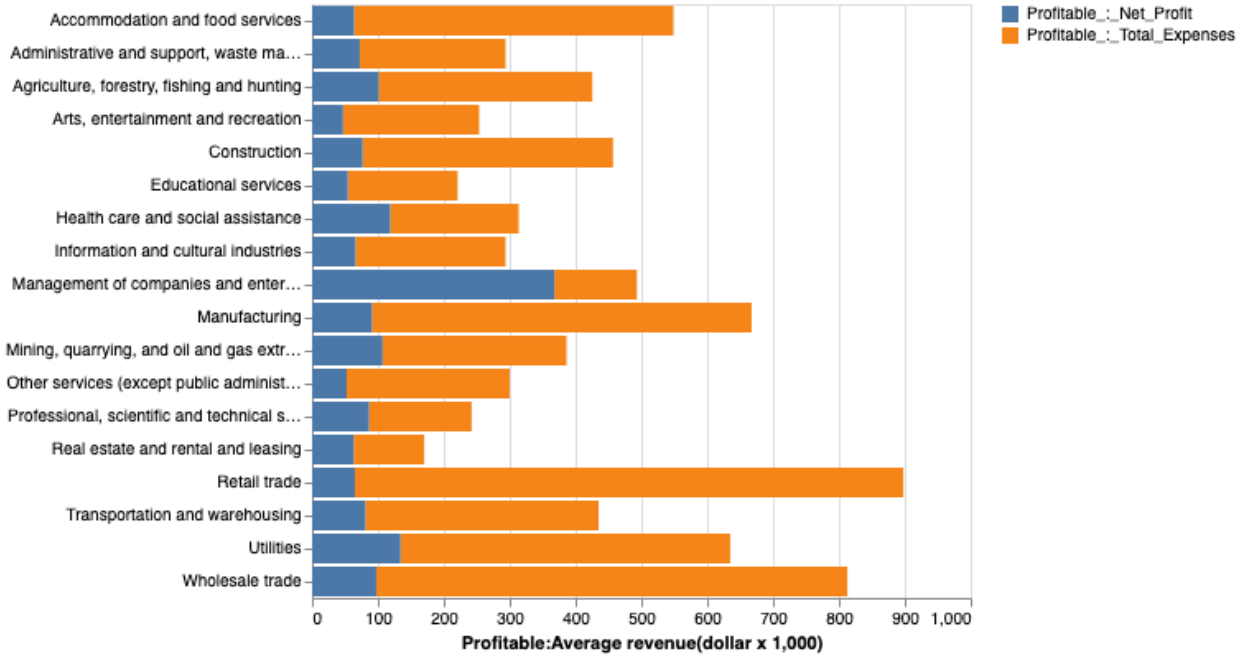


Figure 12: Profitable businesses net profit of different industry of small rural businesses in 2019

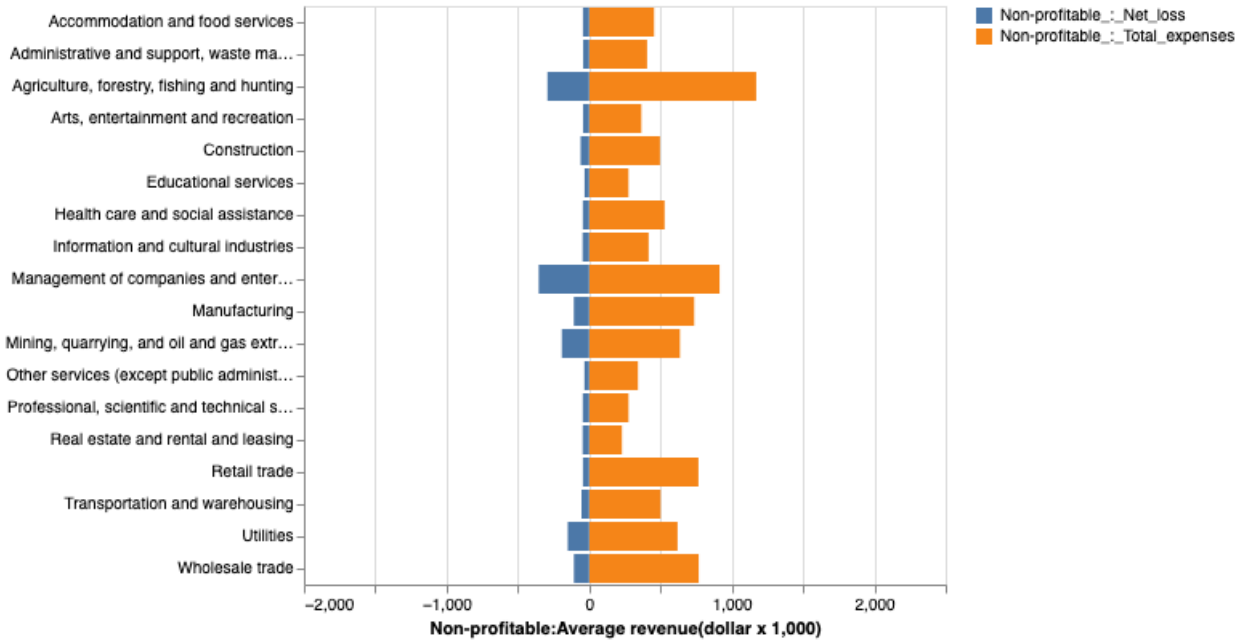


Figure 13: Non-profitable businesses net loss of different industry of small rural businesses in 2019

5.1.3 Expense consist of three levels of expense categories

Expense is an important factor affecting business performance. In the Canada Rural Business Profile (RCBP) database, there are three levels of expense categories. The highest level is the total expenses, which "includes all expenses incurred by a firm to generate revenues in the normal operation of the business."[15] For the second level of expense categories, the total expenses are divided into 2 categories, the cost of sales (direct expenses) and operating expenses (indirect expenses)[16]. The direct expenses are "calculated as wages and benefits + purchases, materials and subcontracts + opening inventory - closing inventory"[17], while the indirect expenses consist of "remuneration paid as labor and commissions, amortization and depletion, repair and maintenance costs, utility fees and telephone and telecommunication fees, rent, interest and bank fees, professional and business fees, advertising and promotional costs, delivery fees and shipping and warehouse expenses, insurance costs, and other indirect expenses"[18]. These specific categories of expenses make up the third level of expenses. According to each level, we performed a corresponding specific analysis and they are described below.

[15] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[16] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[17] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm
[18] "Rural Canada Business Profiles Database 2017-2019 Metadata and User Guide",Statistics Canada, https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm

The average total expenses of medium businesses in both rural and urban areas did not have significant differences. The average total expenses for rural medium businesses over three years is 8925767, while for urban medium businesses, this value is 8870333. The difference was less than 1% (about 0.62%) of their total expenses values, which means they had similar performance on reducing expenses. However the situation changed for small businesses, the average total expenses of small businesses in rural areas was higher than small businesses in urban areas. Over three years, the average total expenses for all rural small businesses was 412800, while for urban medium businesses, this value was about 339600. This obvious difference indicated urban small businesses did better on saving expenses than rural small businesses.

In terms of direct and indirect expenses, the distribution was different for medium and small businesses. For medium businesses, the average direct expenses were 5411.06K for rural areas and 5583.8K for urban areas, which was larger than the average indirect expenses (3514.7K for rural areas and 3286.5K for urban areas) (Figure 14). For small businesses, the average direct expenses were 144.1K for rural areas and 137.2K for urban areas, which was lower than the average indirect expenses (268.7K for rural areas and 202.3K for urban areas). (Figure 15). Comparing Figure 14 with Figure 15, it is easy to find that medium businesses and small businesses had opposite distribution for their expenses. Direct expenses was the major cost of medium businesses, while small businesses had their majority of expenses on indirect aspects. Due to the limitation of data and time available for this project, the reason for this phenomenon was not found. But it is a good direction for future research that can help businesses to develop targeted strategies to reduce expenses.

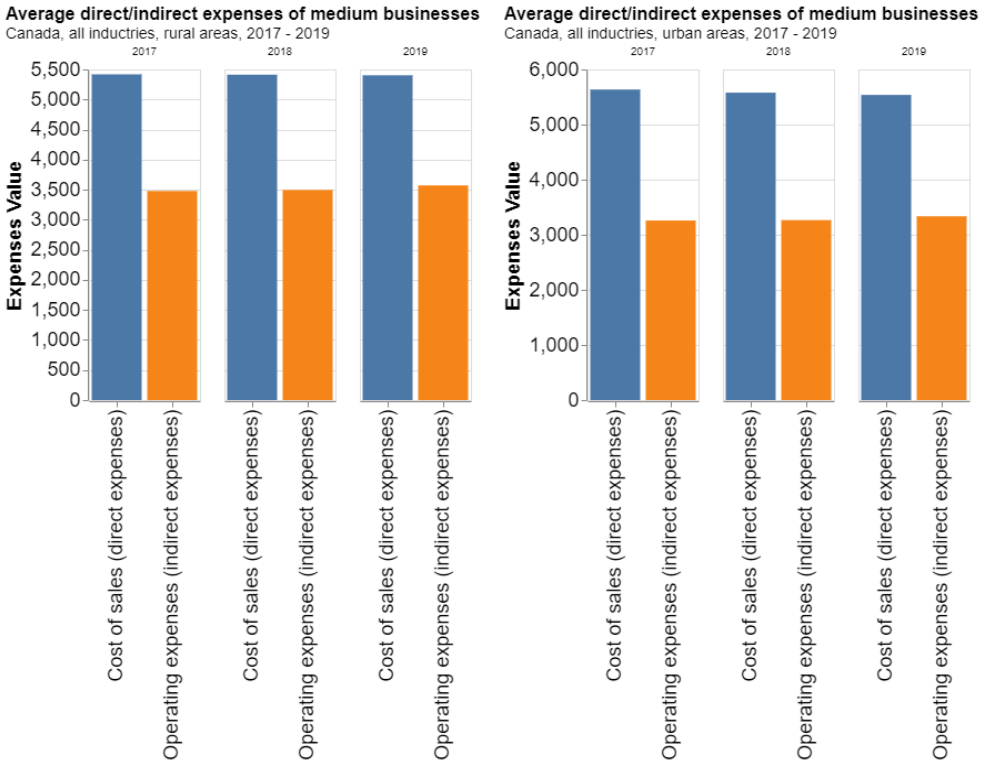


Figure 14: Average direct/indirect expenses of medium businesses

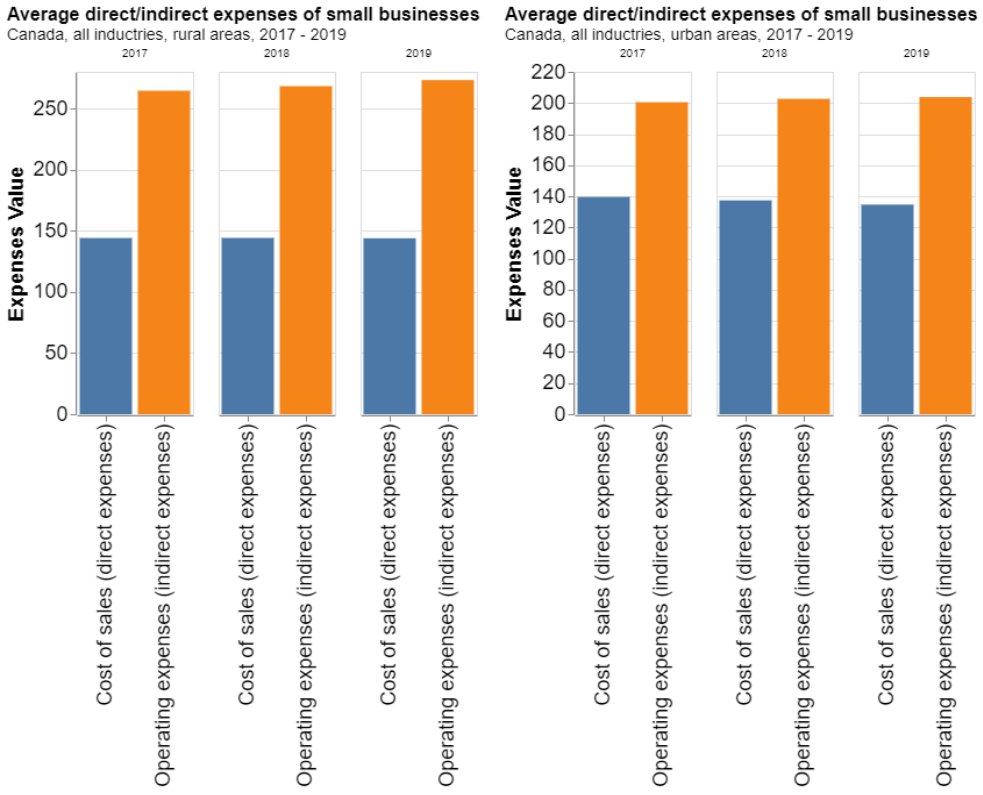


Figure 15: Average direct/indirect expenses of small businesses

When we analyzing the visualizations（Figure 16 and Figure 17） of the expenses distribution of third level expense categories, we found that Regardless of the size of the company and its location, the 'Purchases, materials and subcontracts' was always the main cost of business (43.89% for medium rural business, 44.38% for medium urbanl business, 25.94% for small rural business, 28.93% for small urban business). Therefore, ‘Purchases, materials and subcontracts' expense is an important aspect to consider when a business wants to control their costs.

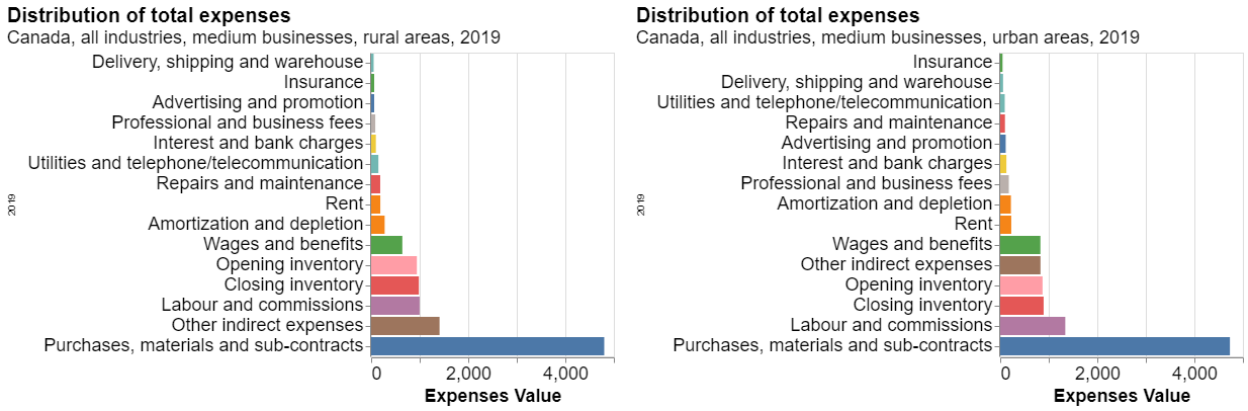


Figure 16: Distribution of total expenses for medium businesses

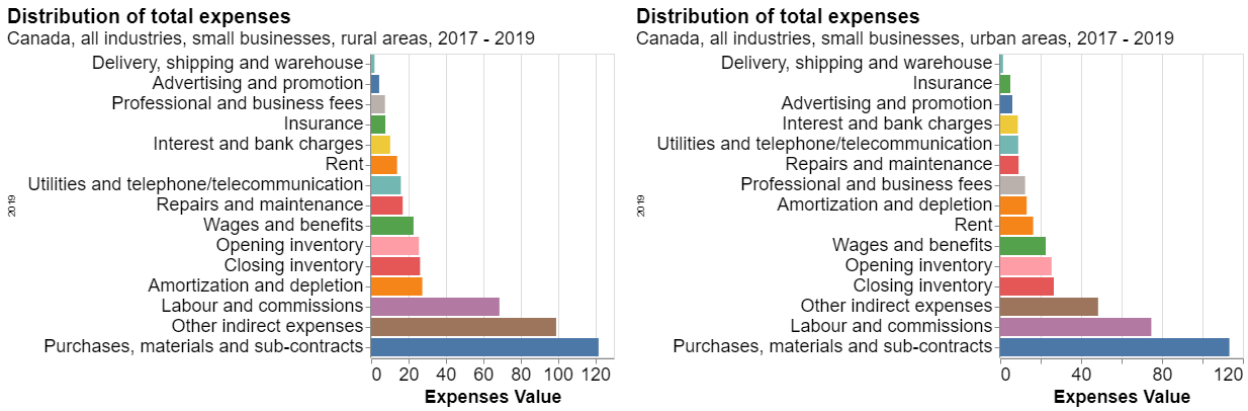


Figure 17: Distribution of total expenses for small businesses

5.1.4 Alberta, Ontario and Quebec are the top three economic active provinces for the small businesses in rural regions

From 2017 to 2019, the average small business number in rural regions are 154,087, 131,292 and 106,843 for Quebec, Ontario and Alberta respectively. These provinces are the top three provinces that have the most small business numbers in rural areas. Moreover, these three provinces are the only three provinces

whose small business numbers are over 100 thousand, which is 33% greater than the fourth province, British Columbia.

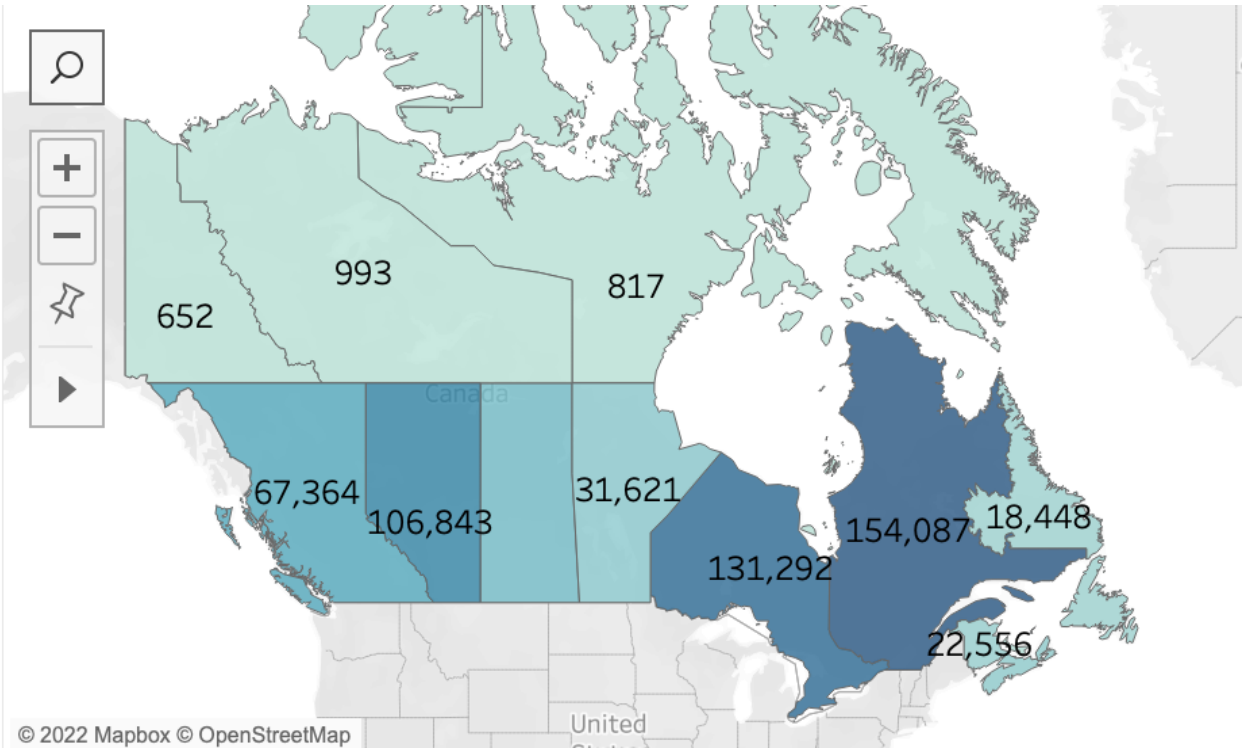


Figure 18: Average total businesses in different provinces from 2017 to 2019, small rural business

The top three industries in Alberta who have the largest small business numbers in rural regions are constructions, algricuture, forestry, fishing and hunting and mining, quarrying and oil and gas extraction.The construction industry has total 52662 small business in rural regions, which is the twice than mining, quarrying and oil and gas extraction.

However, the utility industry which has the least small business amount in the rural region of Alberta has the largest average total revenue from 2017 to 2019.

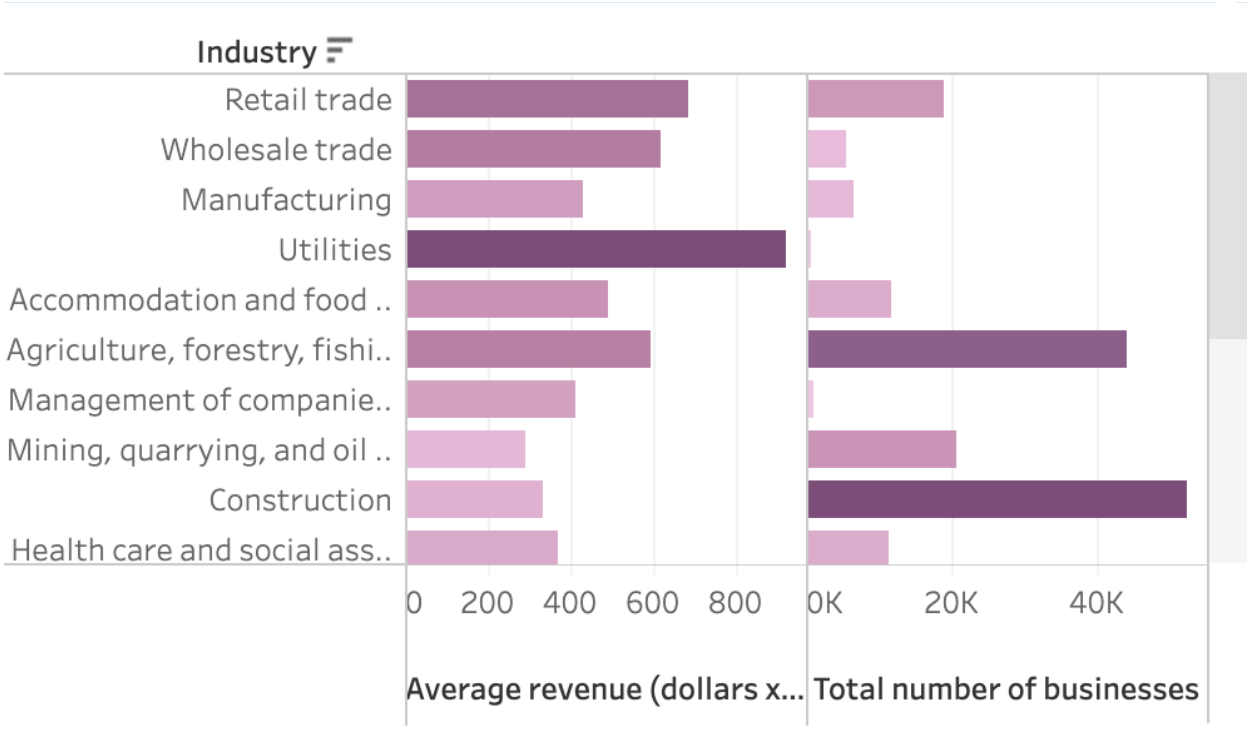


Figure 19: Average total businesses & Average revenue in different Industries in Alberta from 2017 to 2019, small rural business

5.1.5 Nunavut, British Columbia and Northwest Territories are the top three provinces whose small rural businesses are most profitable

In Nunavut, the average net profit of small rural businesses is $664,100, which is the highest in Canada. Although it only has 817 businesses operated from 2017 to 2019, most of the businesses made profits. The second highest province is British Columbia, which had $525,900 in average net profit from 2017 to 2019. The small rural businesses in Northwest Territories have average annual net profit $516,100, which is the third highest in Canada. The rural regions in three provinces are not the most economically active provinces in Canada since they don't have the highest business amount or highest total revenue. However, the small business in rural regions in these three provinces made the most profit. It indicates that a higher amount of business will lead to high competition, which will decrease the net profit.

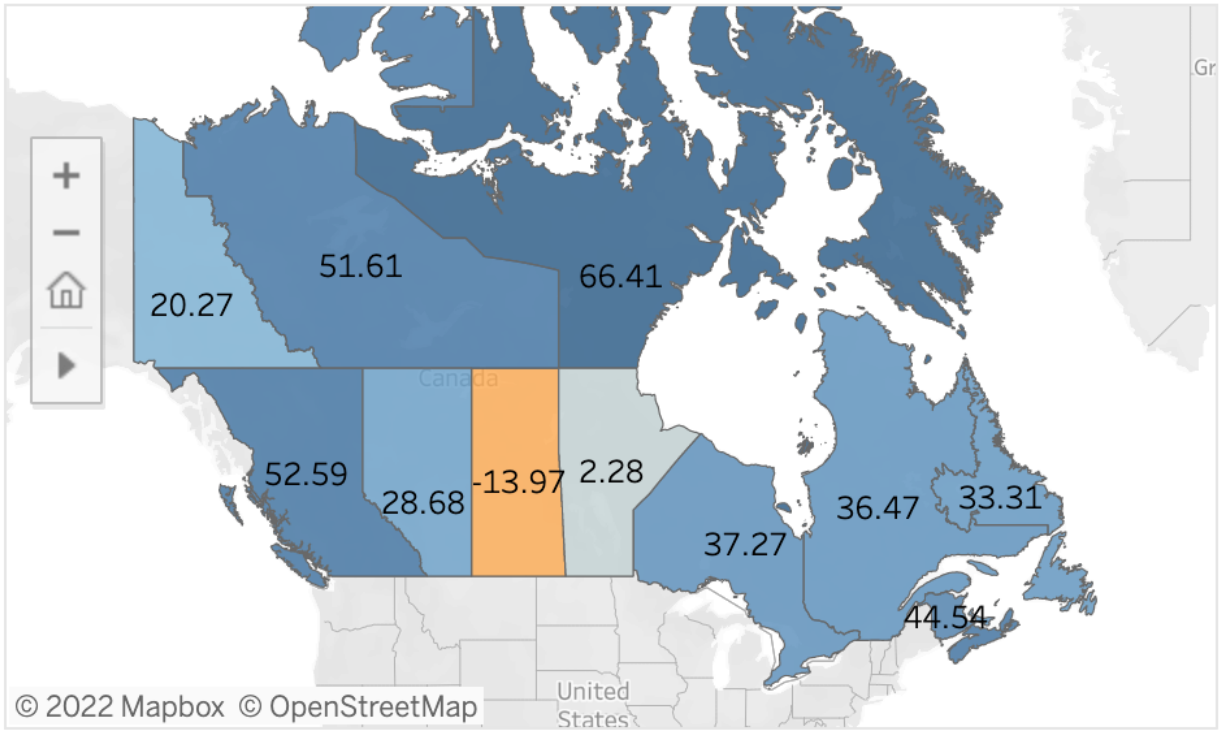


Figure 20: Average Net profit in different provinces from 2017 to 2018, Small rural business

5.2 Net profit Modeling

In order to understand the performance of Canadian rural businesses, we chose the net profit as the response variable in our model and further detected the variables that significantly influence it and the relations between them. We chose 14 relevant factors from the RCBP dataset to build linear regression models. The 14 factors are: 'Total number of businesses', 'Sales of goods and services (percent of total revenue)', 'Cost of sales (direct expenses) (%)', 'Operating expenses (indirect expenses) (%)', 'Total assets', 'Total current assets', 'Total liabilities', 'Total current liabilities', 'Total equity', 'Debt to equity ratio (times)', 'Current debt to equity (%)', 'Revenue to equity ratio (times)', 'Net profit to equity (%)' and 'Gross margin (%)'. We chose rural incorporated businesses and first drew the correlation matrix for medium and small businesses separately. Then we use backward selection to find the most relevant factors for the net profit model.

5.2.1 Rural Medium Businesses Modeling

For the rural incorporated medium businesses, the net profit is relevant to cost of sales (direct expenses) (%), operating expenses (indirect expenses) (%), total assets, total current liabilities and total liabilities.

The cost of sales and operating expenses percentage is the selected expense percentage of total revenue. The linear regression model is approached by R and the following results are (Appendix D) :

Variable	Explanatory Variable name	coefficient
X1	cost of sales (direct expenses) (%)	-87.07
X2	operating expenses (indirect expenses) (%)	-87.6
X3	total assets	0.01711
X4	total current liabilities	0.03335
X5	total liabilities	-0.03062
intercept		8678

Net Profit = -87.07* X1-87.6*X2+0.01711*X3+0.03335*X4-0.03062*X5+8678

The model shows that the total current liabilities and total assets have a positive relation with the net profit, while cost of sales (direct expenses) (%), operating expenses (indirect expenses) (%) and total liability have negative relations with net profit. The interesting part is the total current liability has a positive relation while the total liability has a negative relation with the net profit. That may be because the total current liability stands for short term liability situation, and it stands for the businesses investing more recently and it is a positive signal for the businesses. However, if the businesses have a high long term liability, it becomes a negative signal for the businesses.

5.2.2 Rural Small Businesses Modeling

For the rural incorporated small businesses, the net profit is relevant to cost of sales (direct expenses) (%), operating expenses (indirect expenses) (%), total assets, total current liabilities and total equity. The cost of sales and operating expenses percentage is the selected expense percentage of total revenue. The linear regression model is approached by R and the following results are (Appendix D):

Variable	Explanatory Variable name	coefficient

X1	cost of sales (direct expenses) (%)	-5.043
X2	operating expenses (indirect expenses) (%)	-5.795
X3	total assets	-0.02572
X4	total current liabilities	0.03542
X5	total equity	0.03460
intercept		549.6

Net Profit = -5.043* X1-5.795*X2-0.02572*X3+0.03542*X4+0.03460*X5+549.6

The model shows that total current liabilities and total equity have a positive relation with the net profit loss, while cost of sales (direct expenses) (%), operating expenses (indirect expenses) (%) and total assets have negative relations with net profit loss. A surprising result is that the total assets have a negative relation with the net profit. A possible reason for this result is that the total equity is a factor in the final equation and it has a positive relation with the net profit, so the total assets factor is to adjust back the net profit to improve the accuracy for the model.

5.3 Cross Analysis -- Unemployment Model

The linear regression is a linear approach for modeling the relationship between a scalar response, which is unemployment rate in our case, and one or more explanatory variables, which is the business financial indicators. We use the linear combination of the coefficient times explanatory variables to predict the unemployment rate in Canada.

5.3.1 Small Rural Business Financial Indicators vs Unemployment rate

In this model, we examined the unemployment model in the rural areas. The linear regression model is approached by R and the following results are (Appendix E):

Variable	Explanatory Variable name	coefficient
X1	Percent of profitable businesses	3.978e-01

X2	Total number of businesses	7.548e-04
X3	Total equity (x $1,000)	-2.138e-03
X4	Debt to equity ratio (Times)	-3.045e+00
X1*X5	Total number of profitable businesses	-9.962e-06
intercept		-1.935e+01

Unemployment rate (in rural) = 3.978e-01 * X1 + 7.548e-04 * X2 + (-2.138e-03) * X3 + (-3.045e+00) * X4 + (-9.962e-06) * (X1*X5) + (-1.935e+01)

This model indicates that as the total number of profitable businesses increase, the unemployment will decrease due to the negative coefficient, which matches our belief. A surprising result is that higher the debt-to-equity ratio, lower the unemployment rate. It means that if some companies in a specific region have the difficulty to pay their debt (higher ratio), then the more people will have a job. This may result from the operation debt to expand the business growth. The R-squared of this model is 0.7651, which means about 76.51% of data can be explained by this model. Although it is not a good fit, this is the best R-squared we have for this dataset in terms of the linear regression model.

5.3.2 Small Urban Business Financial Indicators vs Unemployment rate

In this model, we examined the unemployment model in the rural areas. The linear regression model is approached by R and the following results are (Appendix E):

Variable	Variable name	coefficient
X1	Percent of profitable businesses	-2.731e-01
X2	Total number of businesses	3.844e-05
X3	Total equity (x $1,000)	-4.443e-03
X4	Debt to equity ratio (Times)	-2.579e-01

X1*X5	Total number of profitable businesses	-4.918e-07
intercept		3.155e+01

Unemployment rate = (-2.731e-01)* X1 + 3.844e-05* X2 + (-4.443e-03) * X3 + (-2.579e-01) * X4 +(-4.918e-07) * (X1*X5) + 3.155e+01

The negative coefficient of total number of profitable businesses indicates that as we have more businesses making profit in Canada, more people can find jobs which decreases the unemployment rate. The negative coefficient of total equity indicates that if companies tend to have more net value, then it can hire more people to meet its growth. From the conclusion of R, we find that the R-squared is 0.7129, which means about 71% of data can be explained by this model. It is not a very good result, but it can help us to understand the unemployment rate in terms of the financial numbers.

6. Conclusions

After the analysis of rural business performance, several key conclusions are as follows:

- The Agriculture, forestry, fishing and hunting industry had the greatest amount in rural Canada, but it did not make money. Though the gross margin performed very well, the huge operating expenses made this industry lose money.
- Management of companies and enterprises industry had the best debt situation which indicated the industry had good liquidity and solvency. But not all businesses in this industry were in a good financial situation.
- The distribution of direct and indirect expenses in small and medium businesses had the opposite structures. However, regardless of the size and the location of the businesses, the 'Purchases, materials and subcontracts' was always the main cost. Therefore, it is an important aspect to consider when a business wants to control their expenses.
- Different provinces behaved differently according to the number and the net profit of small businesses in rural areas. Provinces which had a large number of small businesses in rural areas tended to have less net profit.

- From the net profit model, we found that the net profit of both medium and small businesses in rural areas had a significant relation with cost of sales(%), operating expenses(%), total assets and total current liabilities.

- The linear regression unemployment model cannot be used to precisely predict the overall unemployment rate in Canada, but it did give us some clues on it. We can use the profitable business number and some financial information to estimate the unemployment rate in a region.

7. Future work

- How did the Agriculture, forestry, fishing and hunting industry keep running when it was not profitable? One can compare with the public policy for the Agriculture, forestry, fishing and hunting industry in rural areas to find more insights.

- More data can be added to the unemployment model to improve training and testing. In our dataset we only had 13 data points of different provinces from 3 years, which was not ideal for training the model. In addition, bringing in more outside data sources, such as government subsidies and economic indicators, will make this model more explainable and more precise.

- The reason for the opposite distribution structures of small and medium businesses in expenses is a good direction to research in the future. More detailed information of small and medium businesses is needed for the differences investigation. The research results may help businesses to develop targeted strategies to reduce expenses.

References:

Statistic Canada. 2022. “Population growth in Canada's rural areas, 2016 to 2021.” Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-x/2021002/98-200-x2021002-eng.cfm.

Statistic Canada. 2022. “Rural Statistics in Canada Rural Canada Business Profiles, 2017 to 2019.” Statistic Canada. https://www150.statcan.gc.ca/n1/pub/45-20-0004/452000042022001-eng.htm.

Statistics Canada. 2022. “Rural Canada statistics.” https://www.statcan.gc.ca/en/subjects-start/society_and_community/rural_canada.

Statistics Canada. “Table 14-10-0375-01 Employment and unemployment rate, annual.” https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410037501.

Statistic Canada. 2022. “Rural Canada Business Profiles Database 2017-2019 Metadata and user guide.” Statistic Canada. https://www150.statcan.gc.ca/n1/pub/45-20-0004/2022001/meta-eng.htm.

Heroku. “Heroku: Cloud Application Platform.” https://www.heroku.com.

Trevor Hastie, Robert Tibshirani, Jerome Friedman. “The elements of statistical learning: Data mining, Inference, and Prediction”. 2nd Edition. Springer. 2008. P43, P59.

Appendix A: Data Manipulation

https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/tree/main/src/code/Data_manipulation.ipynb

Appendix B: Tools, Methodology and Techniques

The detailed information of tools, methodology and techniques used in our project can be found through links listed below:

- Python:https://www.python.org/doc/
- Pandas:https://pandas.pydata.org/docs/
- Numpy:https://numpy.org/doc/stable/
- Altair:https://altair-viz.github.io/
- Tableau:https://help.tableau.com/current/pro/desktop/en-us/gettingstarted_overview.htm
- Dash:https://dash.plotly.com/
    - html.Img:https://dash.plotly.com/dash-html-components/img
    - html.A:https://dash.plotly.com/dash-html-components/a
- dash-bootstrap-components:https://dash-bootstrap-components.opensource.faculty.ai/docs/
    - Row,Col:https://dash-bootstrap-components.opensource.faculty.ai/docs/components/layout/
    - Card:https://dash-bootstrap-components.opensource.faculty.ai/docs/components/card/
    - Button:https://dash-bootstrap-components.opensource.faculty.ai/docs/components/button/
- Heroku:https://devcenter.heroku.com/
- GitHub:https://docs.github.com/en
- Tableau public:https://public.tableau.com/en-us/s/about
- Microsoft Excel:https://support.microsoft.com/en-us/excel

The detailed code/document that described in tools, methodology and techniques section can be found through links listed below:

- [1]:https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/src/code/Data_manipulation.ipynb
- [2]:https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/docs/Analytical%20Paper-First%20Draft.pdf
- [3]:https://dash.plotly.com/deployment
- [4]:https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/src/code/RCBP_Dashboard.html
- [5]:https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/src/code/unemployment.ipynb
- [6]:https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/src/code/unemployment_linear_regression.Rmd

The definition of methods used in tools, methodology and techniques section are provided below:

- **Backward selection:** starts with the full model, and sequentially deletes the predictor that has the least impact on the fit. p59
- ***Linear regression model***: A linear regression model assumes that the regression function E(Y |X) is linear in the inputs X1,...,Xp. p43

Reference:

Trevor Hastie, Robert Tibshirani, Jerome Friedman. “The elements of statistical learning: Data mining, Inference, and Prediction”. 2nd Edition. Springer. 2008. P43, P59.

Appendix C: Data Visualizations for RCBP

RCBP Dashboard for Small businesses:
https://public.tableau.com/app/profile/tingwen7851/viz/SmallBusinessDashboard/Dashboard1?publish=yes

RCBP Dashboard for Medium businesses:
https://public.tableau.com/app/profile/tingwen7851/viz/MediumBusinessProfile/Dashboard2?publish=yes

Appendix D: Code and Result for Net Profit Modeling

Medium Rural Businesses:

```
Call:
lm(formula = Net.Profit.Loss ~ ., data = mydata_mb9)

Residuals:
    Min      1Q  Median      3Q     Max
-320.55  -31.98    5.44   43.42  408.58

Coefficients:
                                        Estimate Std. Error  t value Pr(>|t|)
(Intercept)                            8.678e+03  7.870e+01  110.265  < 2e-16 ***
Cost.of.sales..direct.expenses.....   -8.707e+01  8.627e-01 -100.928  < 2e-16 ***
Operating.expenses..indirect.expenses..... -8.760e+01  8.124e-01 -107.830  < 2e-16 ***
Total.assets                           1.711e-02  1.620e-03   10.564  < 2e-16 ***
Total.liabilities                     -3.062e-02  4.545e-03   -6.736 1.93e-10 ***
Total.current.liabilities              3.335e-02  8.959e-03    3.723  0.00026 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 86.95 on 188 degrees of freedom
  (82 observations deleted due to missingness)
Multiple R-squared:  0.9923,    Adjusted R-squared:  0.9921
F-statistic:  4830 on 5 and 188 DF,  p-value: < 2.2e-16
```

Small Rural Businesses:

```
Call:
lm(formula = Net.Profit.Loss ~ ., data = mydata_sb7)

Residuals:
    Min      1Q  Median      3Q     Max
-351.51  -14.24   -2.03   15.18  467.03

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                             5.496e+02  6.556e+00   83.83   <2e-16 ***
Cost.of.sales..direct.expenses.....    -5.043e+00  7.216e-02  -69.88   <2e-16 ***
Operating.expenses..indirect.expenses..... -5.795e+00  7.552e-02  -76.73   <2e-16 ***
Total.assets                           -2.572e-02  1.007e-03  -25.53   <2e-16 ***
Total.current.liabilities               3.542e-02  1.786e-03   19.84   <2e-16 ***
Total.equity                            3.460e-02  9.112e-04   37.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.98 on 2124 degrees of freedom
  (2562 observations deleted due to missingness)
Multiple R-squared:  0.8787,    Adjusted R-squared:  0.8784
F-statistic:  3077 on 5 and 2124 DF,  p-value: < 2.2e-16
```

Code:

https://github.com/ubco-mds-2021-labs/capstone-project-RCBP_StatsCan/blob/main/src/code/First_RQ_Modeling.Rmd

Appendix E: Code and Result for Cross Analysis–Unemployment Model

Small rural businesses model:

R code:

```r
# load data
new_rural_small_rm <- read.csv('sb_merged_unemployment_rate_rm_rural_new.csv')

# linear regression model

unemployment_model2 <- lm(unemployment_rate ~ Percent.of.profitable.businesses*Total.number.of.businesses + Total.equity
+ Debt.to.equity.ratio..times., data= new_rural_small_rm)
summary(unemployment_model2)
```

R result:

```
Call:
lm(formula = unemployment_rate ~ Percent.of.profitable.businesses *
    Total.number.of.businesses + Total.equity + Debt.to.equity.ratio..times.,
    data = new_rural_small_rm)

Residuals:
     Min       1Q   Median       3Q      Max
-1.38642 -0.27186  0.05894  0.33291  1.13014

Coefficients:
                                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                             -1.935e+01  1.926e+01  -1.005  0.33096
Percent.of.profitable.businesses                         3.978e-01  2.261e-01   1.759  0.09887 .
Total.number.of.businesses                               7.548e-04  2.619e-04   2.882  0.01141 *
Total.equity                                            -2.138e-03  4.807e-03  -0.445  0.66282
Debt.to.equity.ratio..times.                            -3.045e+00  1.975e+00  -1.542  0.14389
Percent.of.profitable.businesses:Total.number.of.businesses -9.962e-06  3.330e-06  -2.992  0.00912 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7392 on 15 degrees of freedom
Multiple R-squared:  0.7651,    Adjusted R-squared:  0.6868
F-statistic:  9.77 on 5 and 15 DF,  p-value: 0.0002629
```

Small urban business Model:

R code:

```r
# load data
new_urban_small <- read.csv('sb_merged_unemployment_rate_urban_new.csv')

# linear regression model

unemployment_model3 <- lm(unemployment_rate ~ Percent.of.profitable.businesses*Total.number.of.businesses + Total.equity
+ Debt.to.equity.ratio..times., data= new_urban_small)
summary(unemployment_model3)


```

R result:

```
Call:
lm(formula = unemployment_rate ~ Percent.of.profitable.businesses *
    Total.number.of.businesses + Total.equity + Debt.to.equity.ratio..times.,
    data = new_urban_small)

Residuals:
     Min       1Q   Median       3Q      Max
-1.07083 -0.47806 -0.09972  0.26081  1.78974

Coefficients:
                                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                              3.155e+01  5.644e+00   5.590 9.41e-06 ***
Percent.of.profitable.businesses                        -2.731e-01  6.768e-02  -4.036 0.000481 ***
Total.number.of.businesses                               3.844e-05  4.521e-05   0.850 0.403529
Total.equity                                            -4.443e-03  2.899e-03  -1.533 0.138436
Debt.to.equity.ratio..times.                            -2.579e-01  1.134e+00  -0.227 0.822020
Percent.of.profitable.businesses:Total.number.of.businesses -4.918e-07  5.545e-07  -0.887 0.383910
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7442 on 24 degrees of freedom
Multiple R-squared:  0.7129,    Adjusted R-squared:  0.6531
F-statistic: 11.92 on 5 and 24 DF,  p-value: 7.171e-06
```