CSCI 4146

PROCESS OF DATA SCIENCE

Final Project Report

NSGROW

Prepared by

Kanav Bhardwaj | B00884919
Abhiroop Yerramilli | B00869102
Yakov Fainshtein | B00850311
Matthew Carl | B00864121

Submitted to

Dr. Evangelos E. Milios

Dalhousie University

Halifax, Nova Scotia

04/4/2024

Table of Contents

Introduction

Background Information

Cities, and sub-regions, around the world, serve as hubs of economic activities. The main backbone of this urban landscape is efficient and strategic resource allocation. Strategic investments in various amenities and careful management of available resources help promote sustainable development, economic growth, innovation, and community development, and ensure fiscal sustainability among others. Due to the importance of resource allocation, poor or ineffective resource allocation can hinder the development and well-being of an urban community. The effects include economic stagnation, social inequalities, environmental degradation, educational and health disparities, and an increase in social unrest, just to name a few. Addressing the challenges around efficient and strategic resource allocation requires a data-driven approach that can leverage insights and different approaches from Municipal Fiscal Statistics to aid decision-making. In this situation, a data-driven approach can highlight previous and current trends, patterns, and relationships, which in turn can be used to predict and analyze future needs. This will aid policymakers and stakeholders to make informed decisions, maximize the impact of strategic investments, promote economic and community development, and directly impact environmental sustainability.

Problem Statement

Municipal leaders in Nova Scotia currently face the challenge of efficiently managing their budgets across various sectors. Without accurate forecasting, there is a risk of overspending, which can lead to financial strain and compromised service delivery. The objective of our project is to develop a predictive tool that enables municipal leaders to forecast yearly expenditures in all sectors accurately. By leveraging this tool, leaders can proactively identify sectors at risk of high expenditure, enabling strategic adjustments to spending patterns before fiscal thresholds are exceeded. This tool aims to promote fiscal responsibility, ensure the sustainability of municipal services, and enhance the overall financial health of municipalities in Nova Scotia.

Problem Significance

The significance of our project lies in addressing the critical need for effective resource allocation in Nova Scotia's municipalities. These communities depend on strategic budgeting to ensure economic and social well-being. Ineffective budget management can lead to economic stagnation, social inequality, and environmental issues, which in turn strain healthcare, education, and can escalate social unrest. Municipal leaders face the specific challenge of forecasting expenditures without the risk of overspending that could lead

to fiscal stress and impact service delivery. Our project introduces a predictive tool tailored to the fiscal environment of Nova Scotia. It will allow leaders to anticipate yearly expenses, identify potential financial risks, and adjust spending to maintain economic stability. This initiative concentrates on data from Nova Scotia's counties and contextual economic indicators, aiming to enhance fiscal foresight without delving into sector-specific predictions. It's a strategic decision that streamlines focus and maximizes the tool's effectiveness. By improving fiscal planning, this tool is set to bolster municipal service sustainability and promote the financial health of Nova Scotia's communities, making it an essential step towards long-term prosperity.

Project Objectives

The project's core objective is to leverage historical data to develop predictive models for annual regional expenditure, focusing on identifying high expenditure sectors and their interrelationships. This involves a systematic analysis of historical data across various sectors to detect patterns, trends, and correlations. The predictive models will enable precise forecasting of yearly expenditures for specific regions, with a particular emphasis on sectors with significant spending. To support municipal decision-making, the project will utilize visualizations to elucidate these relationships and expenditure trends. Additionally, it incorporates a continuous evaluation mechanism to enhance the predictive tool's accuracy and efficiency over time. This professional approach is designed to facilitate strategic planning and resource allocation by providing clear, actionable insights into regional spending patterns.

Project Scope

The scope of our project pertains to the province of Nova Scotia as well as a small number of fiscal details about Canada as a whole. The project is aimed at providing a tool to leaders of counties only within Nova Scotia, not other out of province counties or municipalities. The project focuses on fiscal data from 2013 – 2022 for the counties of Nova Scotia, their populations, as well as Canada's annual GDP and annual average inflation rate during these times. The project also includes data about Canadian housing, the land area of all regions, and population breakdowns for each of the target regions. It should be known that the expenditure and profit sectors included in the data are heavily affected by factors not being considered in this project scope. Because of this, the project only aims to predict total expenditures and not by sector. However, the expense breakdown from year to year is more or less constant, thus the project team decided it would be unnecessary to seek additional features pertaining to only specific expenditure sectors.

Methodology

Data Collection & Preparation

Primary Data Set:

Our Primary dataset, “Municipal Fiscal Statistics: Operating Fund Total Revenues and Expenditures by Regional Municipality”, from the Data Nova Scotia Website, features revenue and expense time-series data for Nova Scotia's regional municipalities. Initial analysis revealed that some regions lacked complete data for our period of interest, leading us to exclude these due to their minimal size and data. The refined dataset covered 49 towns and rural municipalities. This prompted us to incorporate external datasets to capture broader fiscal influences beyond yearly activities. Discovering more comprehensive data for Nova Scotia's larger counties than for smaller municipalities, we consolidated the 49 areas into Nova Scotia’s 18 counties for a more effective analysis.

External Data Set #1: Global GDP’s

From Worldbank.org, this dataset details annual GDP growth percentages globally from 1960 to 2022. We focused on this dataset to examine its potential as an indicator for predicting Nova Scotia's expenditures. Specifically looking at Canada's GDP growth/ decline from 2013 to 2020.

External Data Set #2: Canadian Inflation Rate

We sourced a dataset from Statista.com, originally compiled from the Bank of Canada's records, containing Canada's monthly inflation rates from January 1993 to December 2023. We identified inflation as a valuable addition to our study due to its effect on the financial sector. Focusing on the period from 2013 to 2020, we calculated the average annual inflation rates to incorporate into our analysis.

External Data Set #3: Regional Populations, Nova Scotia

From the Government of Canada's website, this dataset provides a comprehensive population breakdown for all counties within Nova Scotia from 2001 to 2022. Population plays a direct role in fund allocation thus we knew it was pivotal to our project. Our focus was specifically on the populations of all counties between 2013 and 2020. Moreover, we calculated the year-over-year (YOY) population change for each district, both as a discrete value and as a percentage.

External Data Set #4: Land Areas of Targeted Regions

We obtained a dataset from the Government of Canada website to inform our analysis, selecting it for its comprehensive coverage of municipal divisions across Canada, based on the 2016 and 2021 national censuses. This dataset provided detailed metrics on population, dwelling counts, land areas, and population densities.

External Data Set #5: County-Wise Average Residential and Commercial Tax Rates

From the Government of Canada's website, this dataset outlines the residential and commercial tax rates for each county in Nova Scotia from 2009 to 2024. Tax rates, much like inflation, significantly influence economic behavior, affecting how individuals and businesses allocate their finances. Recognizing the importance of understanding these dynamics for our analysis, we specifically focused on examining both commercial and residential tax rates from 2013 to 2020.

External Data Set #6: Minimum Wage, Nova Scotia

From the Open Data Nova Scotia website, this dataset provides us information of the minimum wage of the province from year 2002 to 2023 from which we specifically considered the data from year 2013 to 2020. Minimum Wage influences the fiscal dynamics of the province, making it worthwhile to be examined.

During the preparation phase, we performed feature engineering to obtain important statistics which have been listed in the Appendix under *Table 2: Engineered Features Breakdown*. After the data was processed, we created new features to better describe the data:

Next Year Expenses/Revenues: Since we are predicting the expenses/revenues for the next year, these features were added to aid with the modelling.

Per Capita Expenses/Revenues: We divided the expenses/revenues related feature by population to get a per-person statistic, which allowed us to better study the features of each region, by not having to worry about specific populations.

Population Density: Given the variation of land areas in the counties, it makes sense to consider the density of the people when making predictions.

Note: A dissection of our final preprocessed data set can be found in *Table 1: Final Dataset Feature Breakdown*

Data Analysis

Part 1: Grouping

The Municipal Fiscal Statistics dataset encompasses data spanning from 2013 to 2020 across various regions, including Towns, Regional Municipalities, and Rural Municipalities. However, the data for each individual region is limited and, on its own, offers little insight. Additionally, the period under review saw several regions merge into other counties, complicating any analysis. To address these challenges and the sparse data issue, we opted to aggregate the regional data at the county level. This approach allows for a more comprehensive perspective, facilitating broader analyses that would not be possible with the fragmented regional data. A map which outlines these 18 counties can be seen in *Figure 1* below.

Figure 1: County Map of Nova Scotia

After grouping, we both one hot encoded the data and binned it. One hot encoding was performed specifically on the municipality types, (County, District, Regional), as is a nominal categorical feature. Binning was performed in the counties specifically by population size and land area size. A binning breakdown can be seen in the tables below.

Size	***Counties***	***Value***
Population Size		
0 – 11000	Small	1
11000 – 22000	Medium	2
22000 – 46000	Large	3
46000 – 90000	Larger	4
90000 – 150000	Cape Breton	5
150000 – 1000000	Halifax	6
Land Size		
1000 – 2000	Antigonish, Richmond	1
2000 – 3000	Kings, Yarmouth, Queens, Cape Breton, Shelburne, Digby, Victoria, Pictou, Lunenburg	2
3000 – 4000	Hants, Annapolis, Colchester, Inverness	3
4000 – 5000	Guysborough, Cumberland	4
5000 – 6000	Halifax	5

Table 1: Categorization of Population and Land Area

Part 2: Correlation

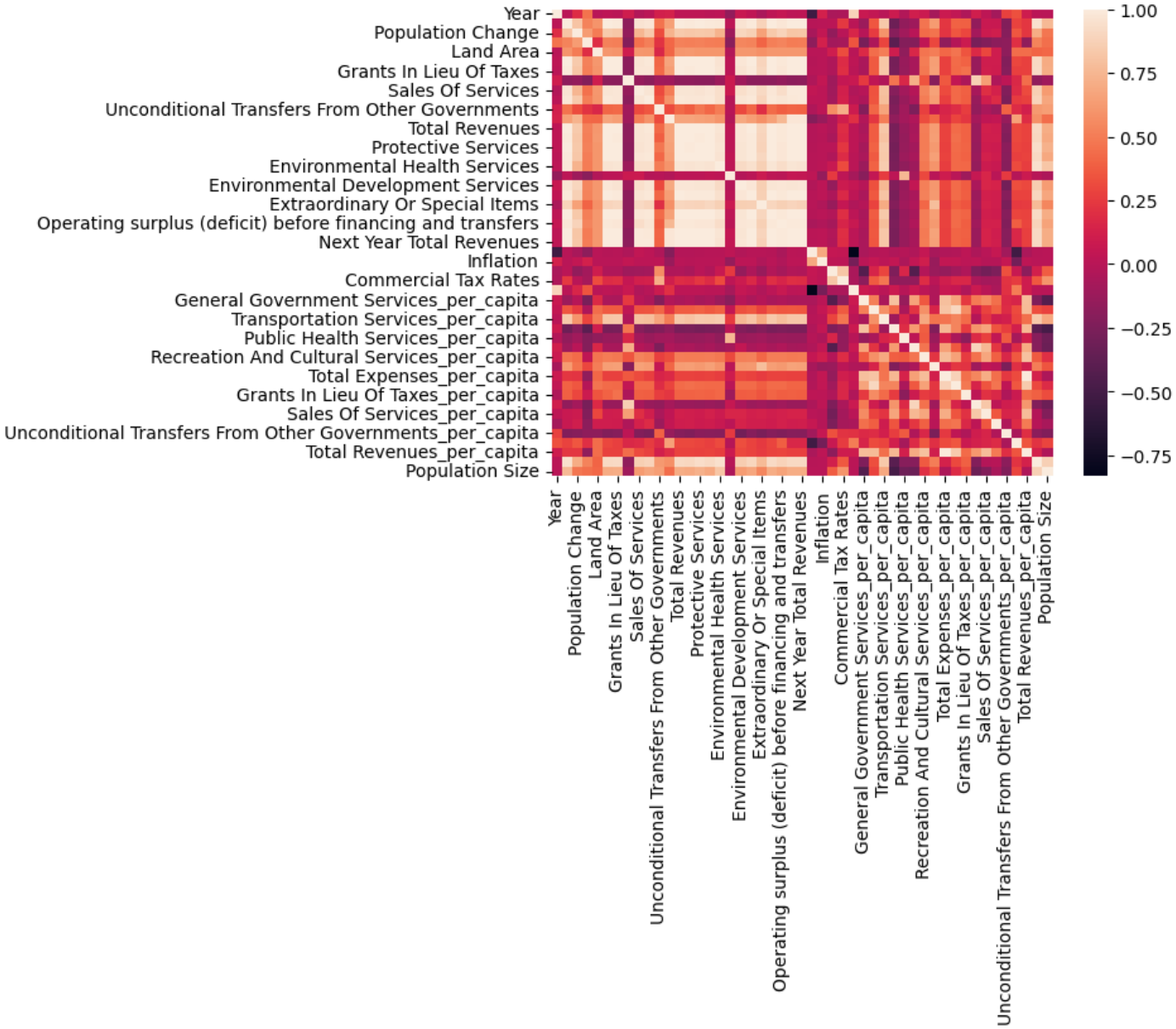

Figure 2: Correlation Heap Map

Part 3: Visualizations

Below are multiple visualizations which are crucial to understanding our problem domain and project goal overall. In *Figure 2* shows the trend of both Nova Scotia's revenues expenses from 2013 to 2020, both seeing a steady overall increase YOY. Continuing, *Figure 3* and *Figure 4* illustrate a breakdown of these revenue and expense totals. YOY, the ratios are very consistent, with only minor fluctuations. *Figure 5* and *Figure 6* show breakdown by county, of Nova Scotia's annual revenue and expenses. These plots show the obvious trend, that population is directly proportional to the total amount of fiscal activity within a county. *Figure 6* and *Figure 7*, highlight a very interesting conclusion, that smaller counties generally get more money per person allocated to them than larger ones. Although we did not expect this to be the case, it makes sense as almost all counties have certain fixed costs, most of which are un effected to the population of the area.

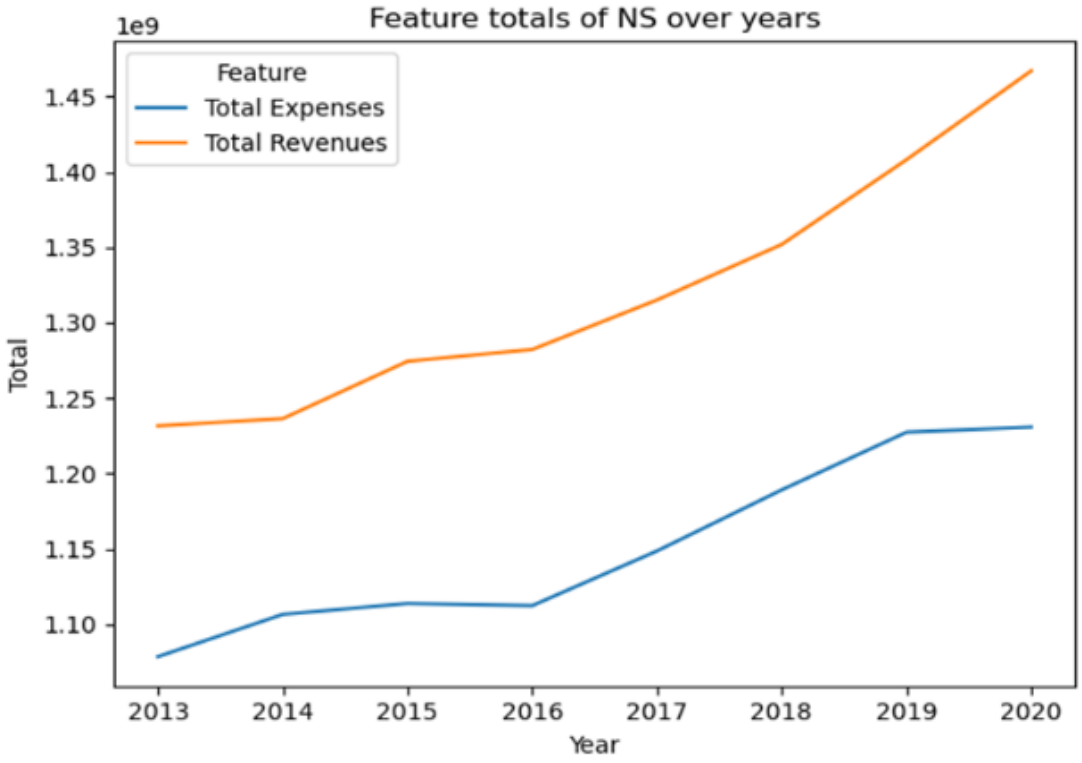

Figure 2: Total Expenses & Revenues of Nova Scotia 2013-2020

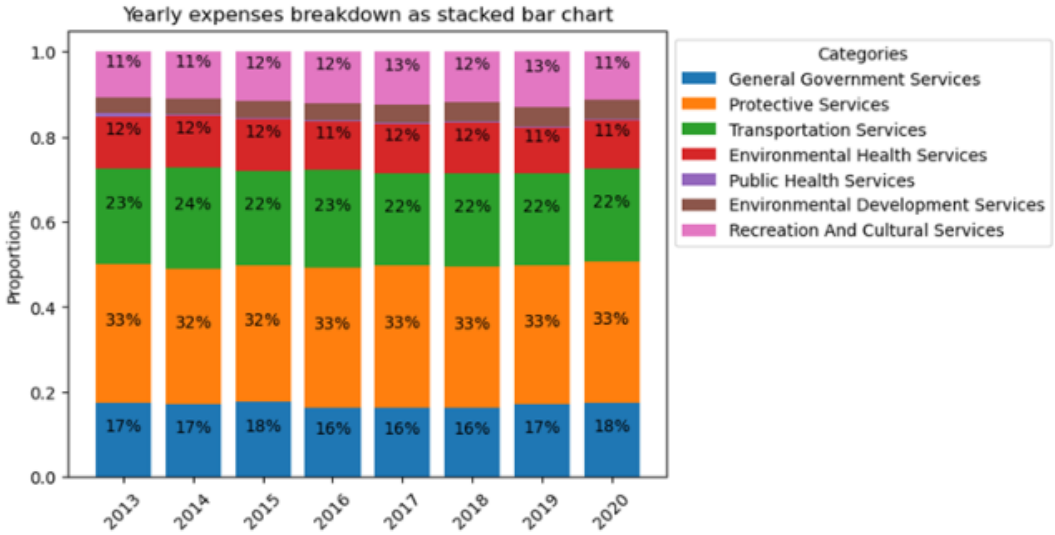

Figure 3: Yearly Expenses Broken down as Percentage of Total

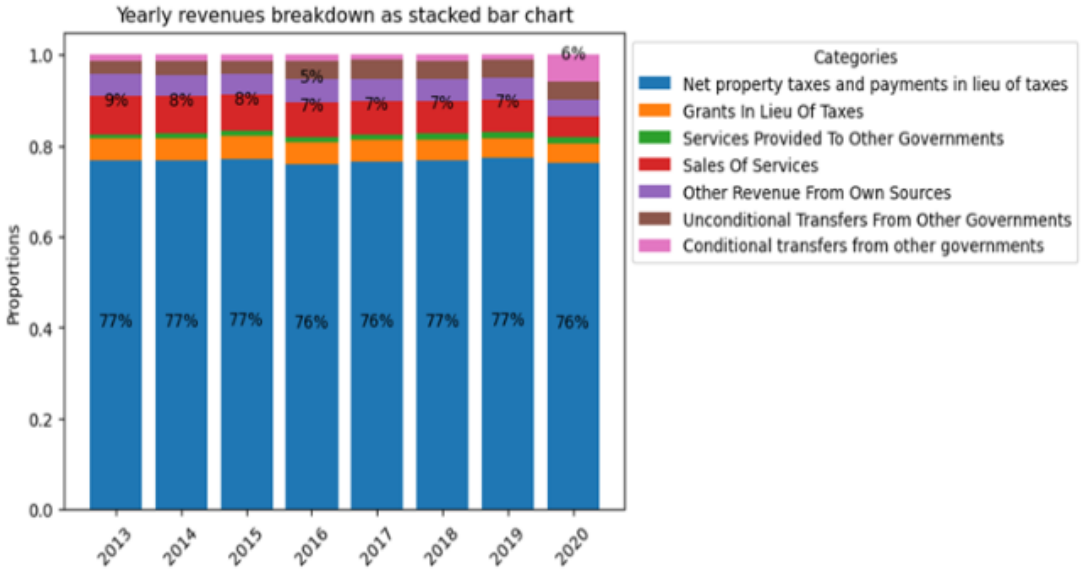

Figure 4: Yearly Revenues Broken down as Percentage of Total

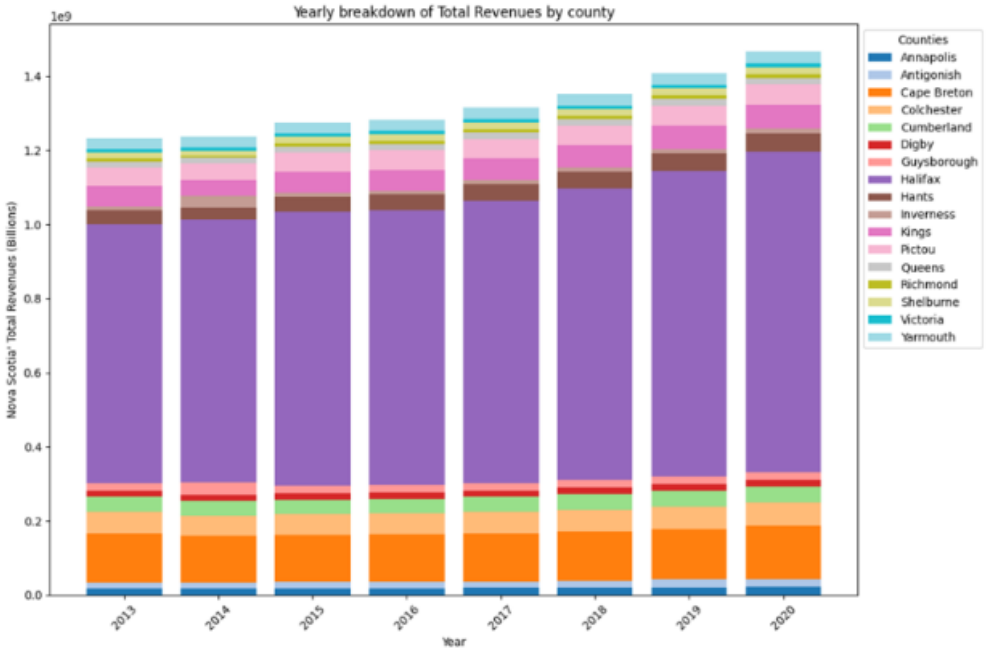

Figure 5: Yearly County Revenues Broken down from Provincial Total

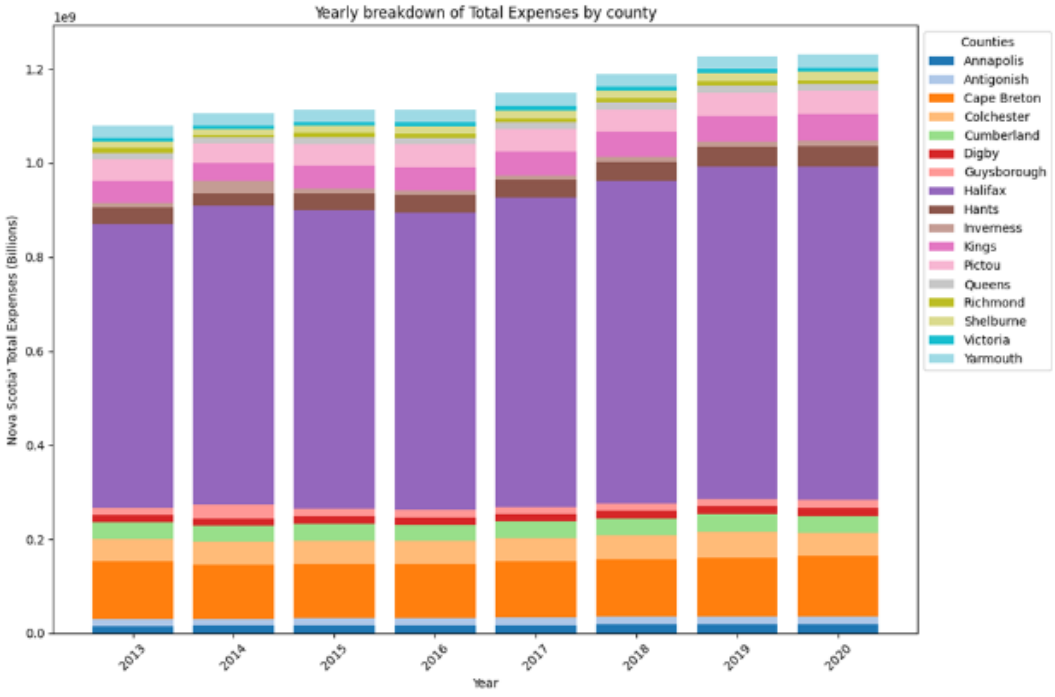

Figure 6: Yearly County Expenses Broken down from Provincial Total

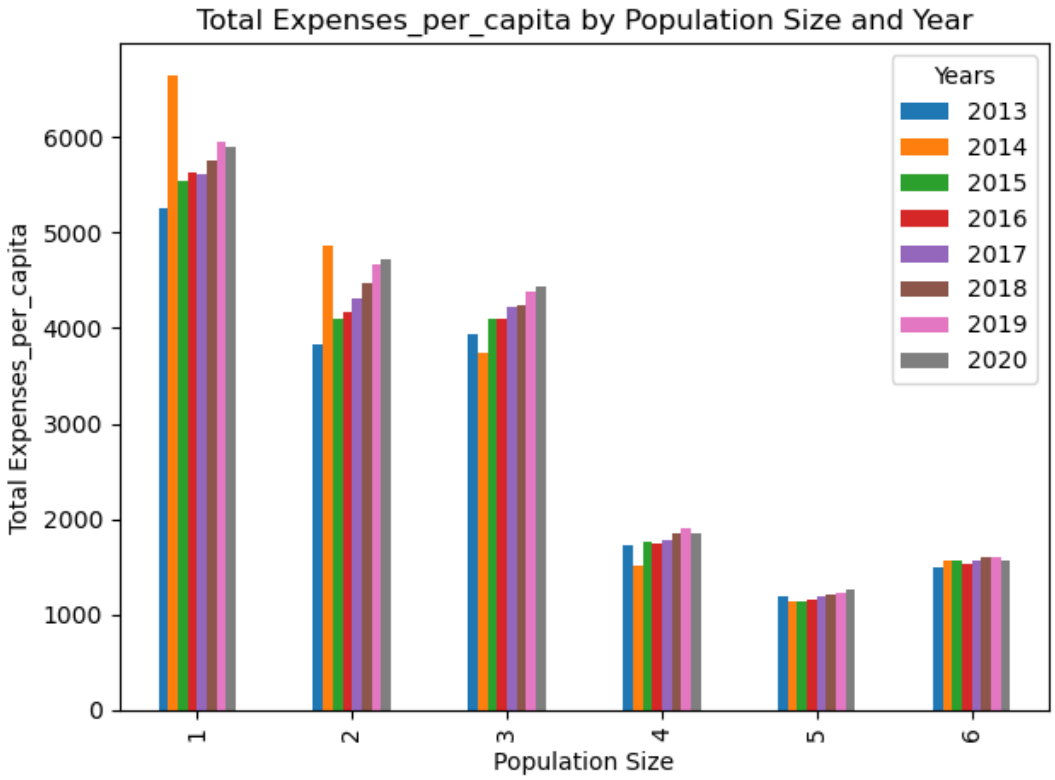

Figure 7: Total Expense per Capita Split based on Population Size and Year

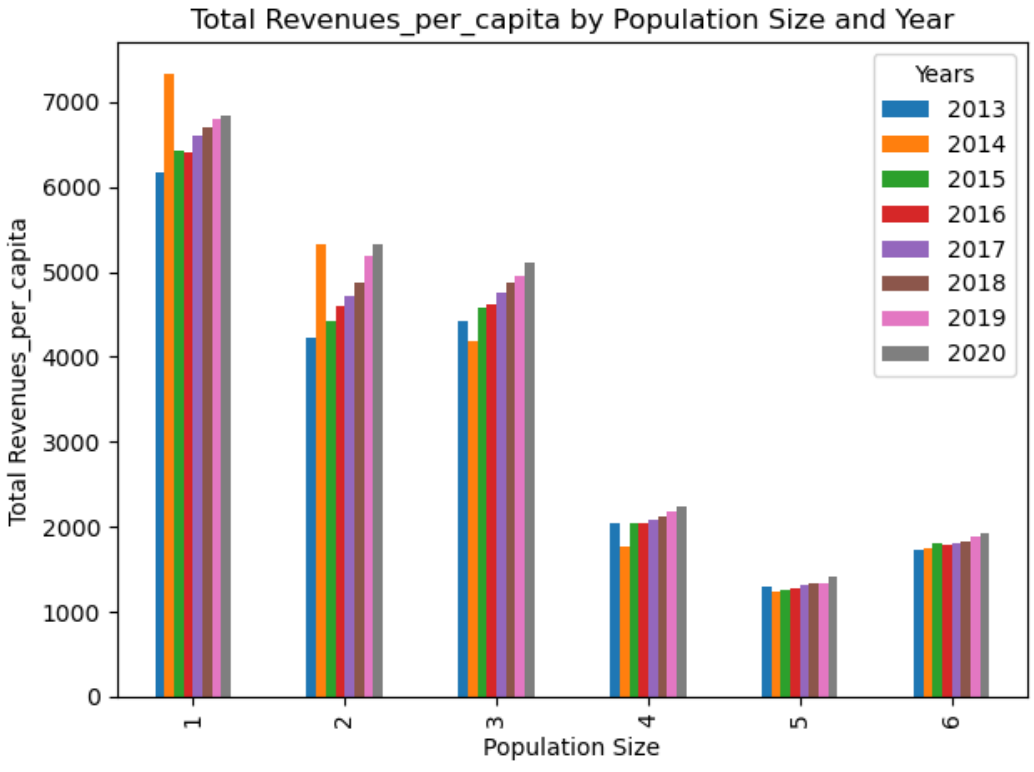

Figure 8: Total Revenues per Capita Split based on Population Size and Year

Modelling

Model 1: Linear Regression

We used Linear Regression Model as a baseline. The goal of the model is to use current years' data (already observed) to predict next year's Expenses and Revenues.

Assumptions

Based on the correlations we discovered in the last section; we picked the top features and assumed a linear relationship. The assumption of linear relationship of total expenses/revenues with features is based on prior domain knowledge. Furthermore, we are only using this model as a baseline.

Feature Selection

We used correlation maps to find the features to use the features that are correlated with the target feature. The following features were considered:

- Population
- Population Change
- Inflation
- GDP
- Minimum Wage
- Land Area of County

The target features:

- Total Next Year Expenses
- Total Next Year Revenues

Model Configuration

The model performs a 3-step calculation:

1) Predict Next Year Total Expenses and Next Year Total Revenues using holistic features:
    - Population
    - Population Change
    - Inflation
    - GDP
    - Minimum Wage
2) Using the prediction of Next Year Total Expenses and Next Year Total Revenues to predict the allocations to each county.
3) After predicting the Expenses and Revenues of each county, predict allocations to each sector.

Training

We sorted the data on year and performed a time series split using *TimeSeriesSplit*. It divides the data into folds, keeping the time series nature of the data. All the four models were trained using *TimeSeriesSplit*: predicting 'Total Next Year Expenses', 'Total Next Year Revenues'. A county-wise prediction of Revenue, and Expenses were also predicted.

Fitted Models:

Feature	*Weights for log NEXT_YEAR_TOTAL_EXPENSES*	*Weights for log NEXT_YEAR_TOTAL_REVENUES*
YEAR	0.011	0.019
logPOPULATION	0.0001	N/A
GDP	0.01	0.003
INFLATION	-0.015	0.008
MINIMUM_WAGE	0.002	0.002
log TOTAL_EXPENSES	0.0002	0.0005
TOTAL_REVENUES	0.0006	0
log TOTAL_REVENUES	0	0.0002
Constant	-2.78	-18.28
	Weights for log TOTAL_EXPENSES	**Weights for log TOTAL_REVENUES**
YEAR	0.03	0.03
logPOPULATION	0.55	0.58
POPULATION CHANGE	-0.07	-0.08
POPULATION DENSITY	0.01	0.008
LAND AREA	0.41	0.45
COUNTY	-0.15	-0.18
DISTRICT	-0.03	-0.03
REGIONAL	0.18	0.21
RESIDENTIAL TAX RATES	-0.50	-0.66
POPULATION SIZE	0.17	0.20
Constant	-61.57	-68.51

Testing:

We used *TimeSeriesSplit* again, to test the model on every iteration. Due to the limitations of the model and the dataset (limited data), the model was tested on data for the year 2020 to predict total expenses, total revenues, and county wise breakdowns for expenses and revenues for the year 2020.

Model 2: Gradient Boosting Machines (GBM)

Given our small data size, GBM is another model we are considering due to its ability to increase accuracy. This belief is based on the linear regressive nature of the problem, which in theory, makes GBM a valid ensemble technique.

Assumptions:

We went with a similar approach to the linear regression model; the topmost correlated features were assumed to yield the best results. The top 5, 10, and 15 features were used in model development. COVID was assumed to play a role on the data for the year 2019, and this was reflected by choosing a split to analyze this time where the pandemic first arose.

Feature Selection:

The top 15 most correlated features were used to predict the target feature. However, only 5 of the top correlated features resulted in the best performance.

This same data is condensed in the following bar chart for easier comprehension:

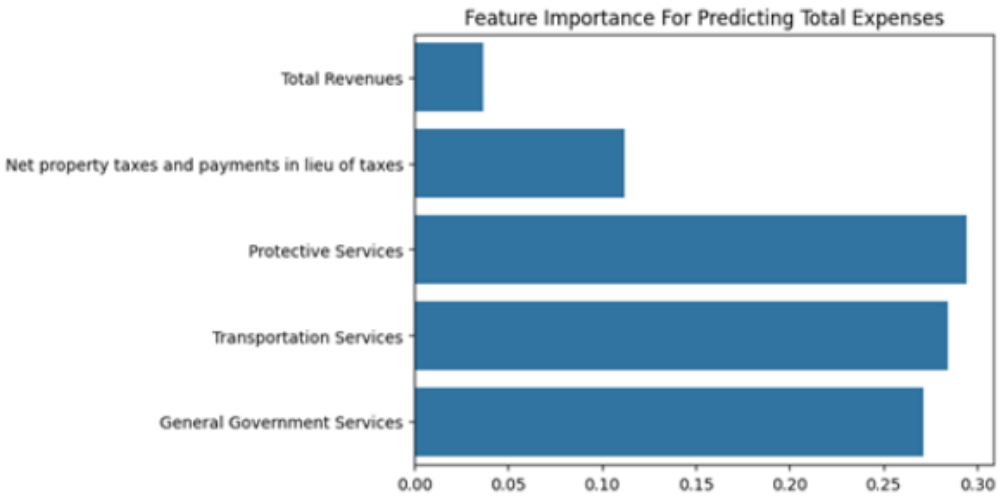

Figure 10: Feature Importance for Predicting Total Expenses using GBM.

Model Configuration:

The model is configured using various splits and hyperparameter tuning grids. The best performer then gets further analyzed. This is elaborated upon in the training section below. The target feature is Total Expenses.

Training:

The training of this model is split into 3 sections: split, initial hyperparameter tuning, and fine hyperparameter tuning.

Splits:

Various splits were created to check various possible configurations which could result in a more accurate model.

- A time series split was conducted with 3 different feature amounts (5, 10, and 15 most correlated features).
- A COVID split was conducted only using 5 features such that training data was pre-COVID and test data was during/post COVID. Pre-COVID was also attempted which took years 2013-2017 into training and used 2018-2019 as testing.

Grids:

- At first an initial wide grid was made capable of equally analyzing all parameters for all different splits.
- The best 2 splits were chosen, and finer tuning was conducted given the results of the initial grid.
- The model with lowest RMSE was chosen and analysis was completed with those parameters.

Testing

The model was then fitted over a selectable number of iterations (50 in this case). From here the average RMSE, R^2, and average percentage error were calculated. Scatterplots were used to help visualize the predicted vs actual, as well as the feature dependencies and residuals.

Model Evaluation:

The performance of the two predictive models was rigorously evaluated. Considering limitations of both the model, and the dataset, the final model was evaluated on the year 2020 alone. The Model Performance and Analysis section below goes into depth on Model Evaluation, Performance, and Analysis.

Results

Model Performance & Analysis

Linear Regression was used to predict Total Expenses, Total Revenue, and county wise breakdowns for expenses and revenue, while GBM was used to predict Total Expenses, and county wise breakdowns for expenses. Both models used the data from the year 2020 for evaluation.

The metrics used for both models are Percentage Difference (Actual value vs Predicted value) and Root Mean Square Error (RMSE). The main intention behind using Percentage Difference and RMSE for evaluating these models is to assess accuracy and performance of the model compared to the actual values. RMSE gives us an idea of the standard deviation of the predictions from the actual values, and relative error (percentage differences) provide an insight into the magnitude of deviation. Both are very useful here.

Below is the performance for both the models using the metrics discussed in this section:

Model Performance for Linear Regression:

Feature	*Actual (in CAD)*	*Predicted (in CAD)*	*%age Difference*
Total Expenses 2020	1230868649	1212541850	1.48 %
Total Revenues 2020	1467053121	1408906860	12.13 %

Year	Region	Total Expenses	Total Revenues	Predicted Next Year Total Expenses	Predicted Next Year Total Revenues	%age Error Expenses	%age Error Revenues
2020	Digby	16891255	18960166	17831784	19967483	-5.568142	-5.312807
2020	Guysborough	16100548	19282783	16287551	19749532	-1.161470	-2.420548
2020	Halifax	708955527	865285000	715513794	822086748	-0.925060	4.992373
2020	Victoria	8786747	10256693	9088064	9859349	-3.429221	3.873997
2020	Hants	43095201	49575049	42554226	50468476	1.255302	-1.802171
2020	Cape Breton	126935138	142219006	124262886	135008891	2.105211	5.069727
2020	Inverness	12735592	13697179	20612907	23565834	-61.852759	-72.048814
2020	Cumberland	37128393	42627666	35681794	41731339	3.896207	2.102688
2020	Shelburne	17512846	18904460	17326148	19121731	1.066063	-1.149311
2020	Kings	56289133	63438378	56816271	63724983	-0.936483	-0.451785
2020	Antigonish	16867362	21567761	15245080	16554337	9.617876	23.244991
2020	Pictou	48267139	55379522	37318892	41124445	22.682610	25.740701
2020	Colchester	49426808	63758102	49292591	57379208	0.271547	10.004837
2020	Richmond	8668365	9581925	9500310	10490969	-9.597485	-9.487071
2020	Queens	15683475	17842769	13445666	14854730	14.268579	16.746498
2020	Annapolis	19590452	22043387	19310326	21117290	1.429911	4.201246
2020	Yarmouth	27934668	32633275	26526494	30603496	5.040955	6.219967

Table 2: Total Revenues and Expenses Actual vs Predicted. Total and by County

The total RMSE for expenses is $3,805,472 and $11,637,542 for revenues.

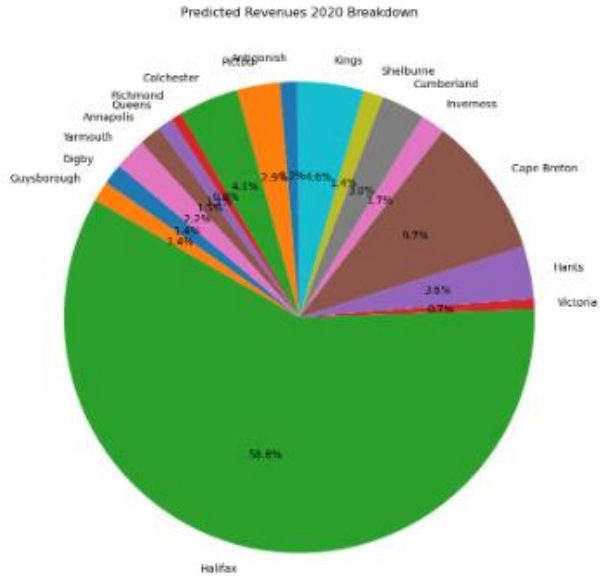

Figure 11: Model Predicted Revenue Breakdown by County

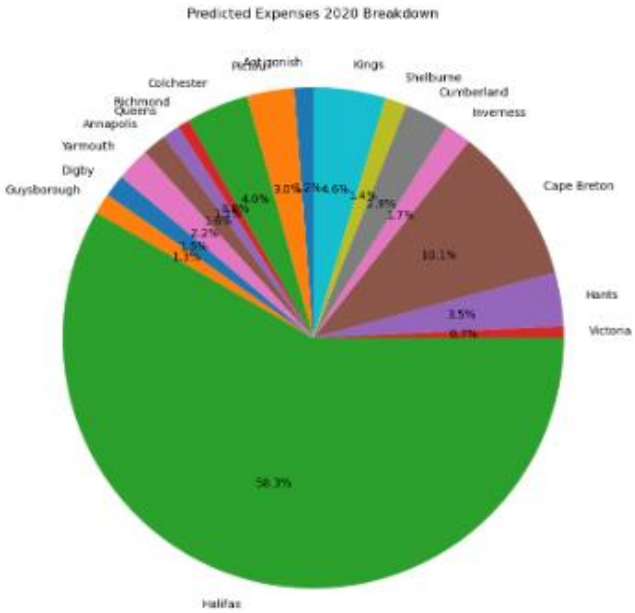

Figure 12: Model Predicted Expense Breakdown by County

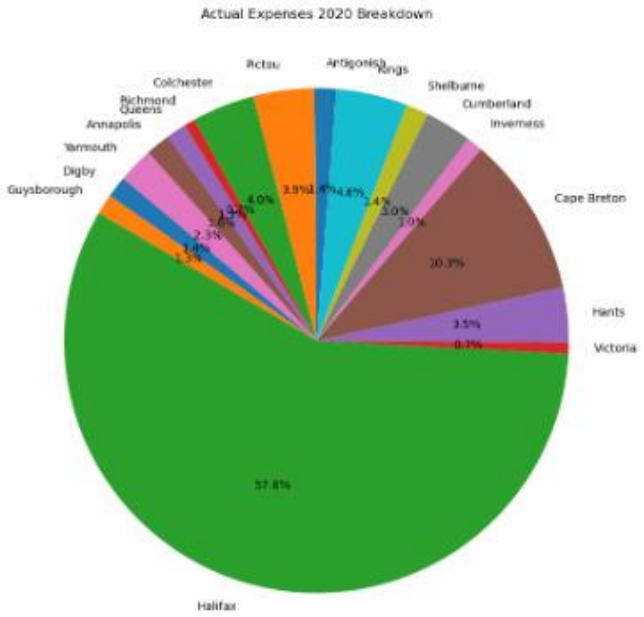

Figure 13: Actual Breakdown of Expenses by County

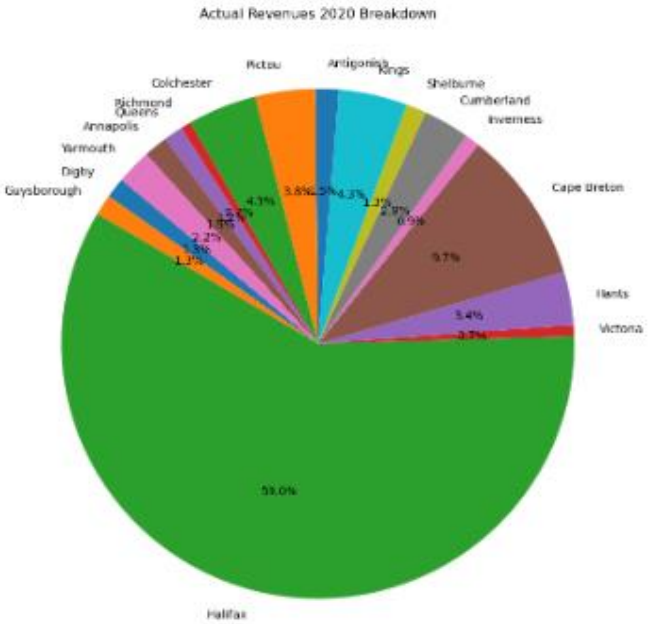

Figure 14: Actual Revenue Breakdown by County

Model Performance for GBM:

Region	*Residuals*	*% Difference*
Annapolis	5.513122e+05	2.814188
Antigonish	8.687442e+05	5.150445
Cape Breton	2.202796e+07	17.353712
Colchester	1.569158e+06	3.174709
Cumberland	5.660330e+05	1.524529
Digby	2.793718e+05	1.185751
Guysborough	1.068636e+06	6.083862
Halifax	2.261080e+07	3.136256
Hants	2.566579e+06	5.955603
Inverness	6.330574e+05	4.970773
Kings	3.039407e+06	5.399633
Pictou	6.365456e+05	1.318797
Queens	6.890957e+05	4.295345
Richmond	9.953213e+04	1.148223
Shelburne	3.291792e+05	1.879644
Victoria	3.099620e+05	3.527608
Yarmouth	2.487281e+05	0.782363

Table 1: Residuals and precent Different between Predicted and Actual

Average RMSE over 50 iterations: $6703723.075

Average R^2 over 50 iterations: 0.9988

Average Percentage Error over 50 iterations: 7.31%

Additional Performance Graphs:

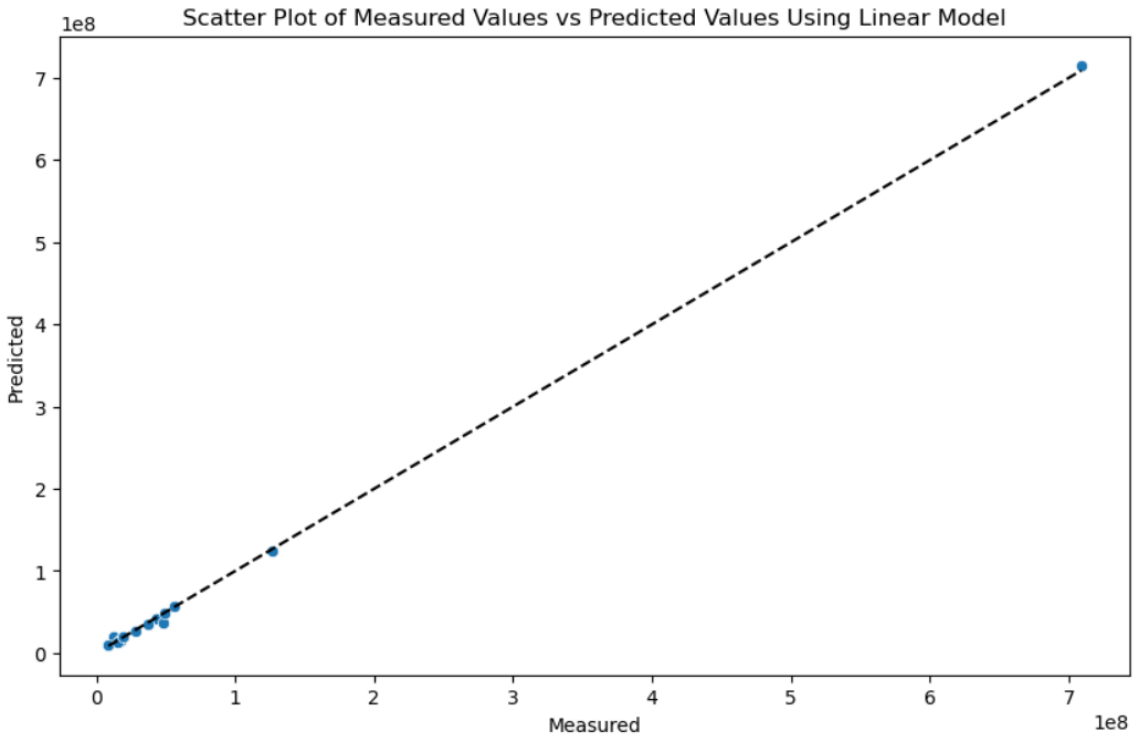

Figure 3: Scatter Plot showing Predicted vs. Actual for Linear Regression

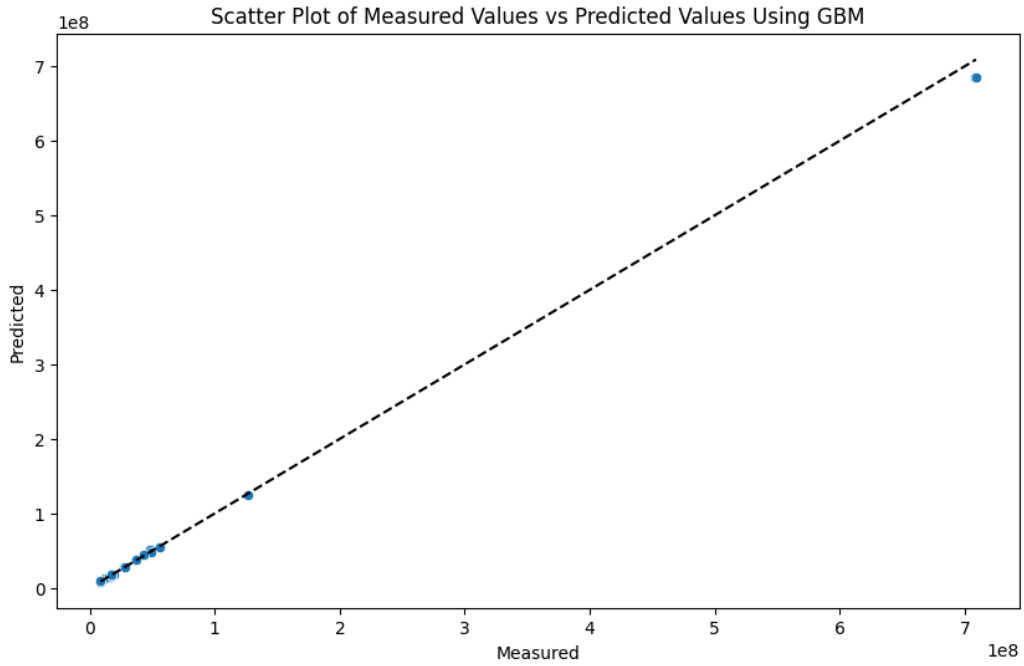

Figure 4: Scatter Plot showing Predicted vs. Actual for Gradient Boosting Machines

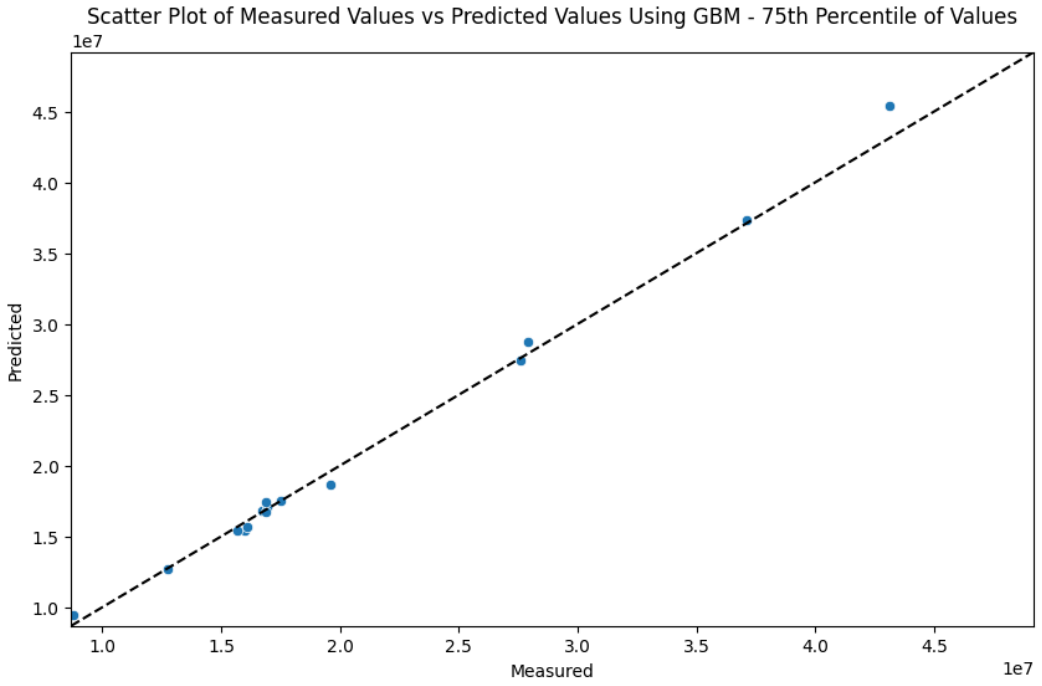

Figure 5: Scatter Plot showing Predicted vs. Actual for Gradient Boosting Machines showing 80the Percentile of Data Points

Further analyzing the feature dependencies, all the features which were introduced into the data played a similar role with Total Revenues having the lowest dependency. When fitting the model many times, these values change which seems to be a common occurrence with ensemble models as they may converge on different solutions differently with each fit.

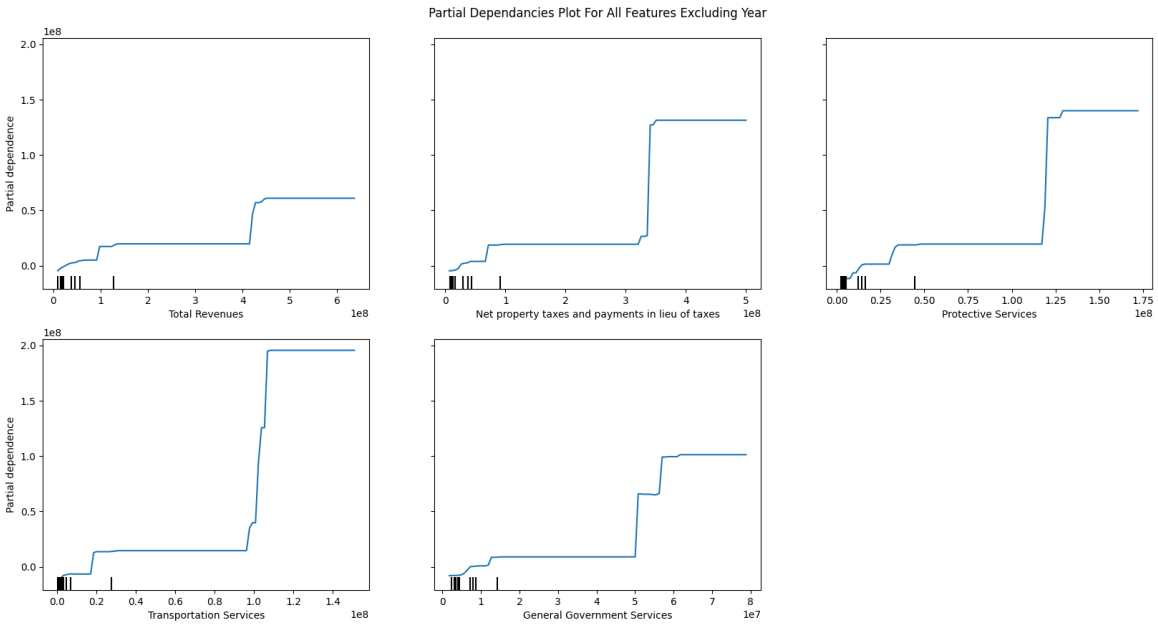

Figure 6: Partial Dependencies for Features Chosen in the Final Model Excluding Year

The final statistical scatterplot used is the residual plot. It can be seen below, for each predicted value the residual (the difference between predicted and actual). There does not appear to be many outliers, most residuals are within ±$1e6. This also includes the outliers from our initial data which have overly high total expenses. Nonetheless, the model was still able to predict the values within that same residual.

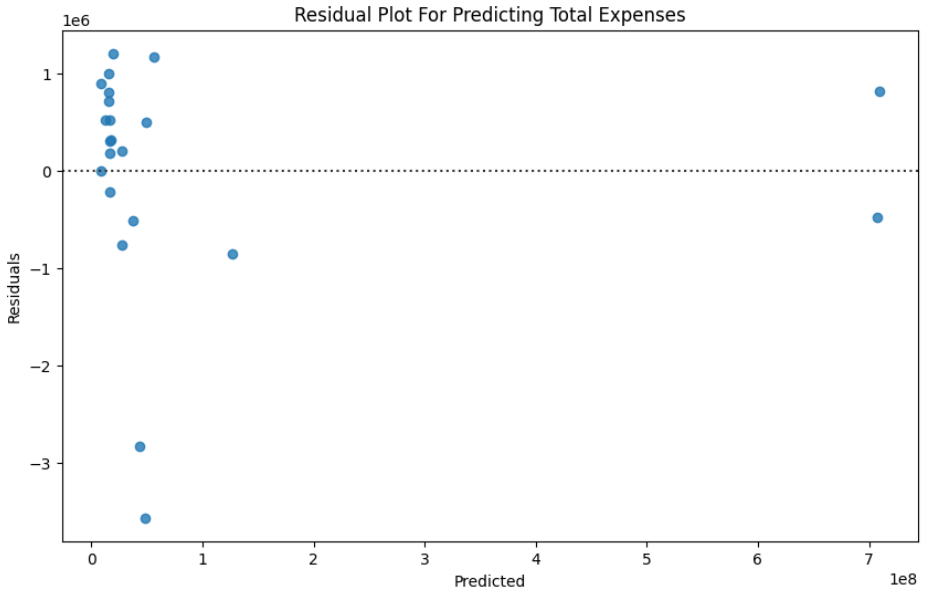

Figure 7: Residuals vs Predicted for Total Expenses using GBM Model

As a bonus some additional data about the training and cross validation score can be found on the right (NMSE), as well as the average evaluation time (training, and scoring) can be found below. The cross validations score is very unstable with a small number of training examples but then settles, while the training score remains close to 0. With the time on the right, it can be noted that the number of examples does not heavily affect the training time.

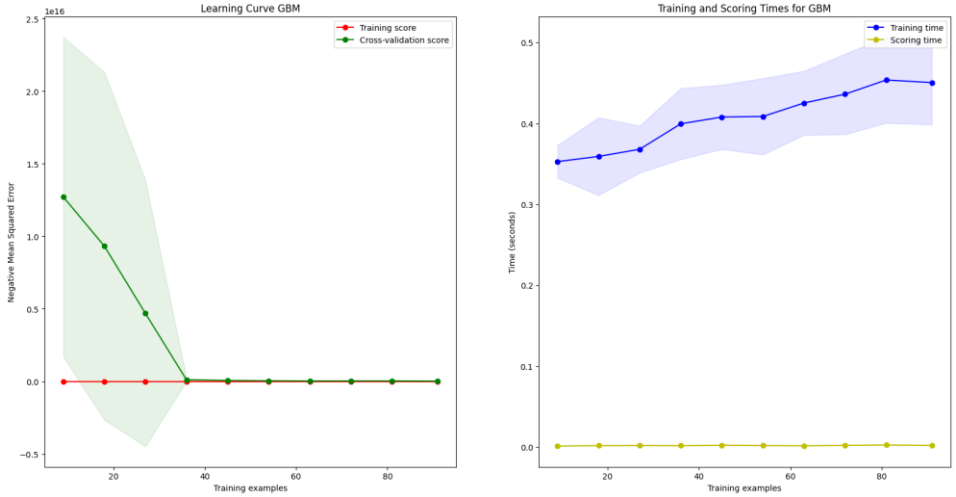

Figure 8: Learning Rate Based on Training Examples for GBM (Both Time and NMSE)

Interpretation of Results

Total Expenses for 2020

Model	*Percentage Difference (Predicted vs Actual)*	*Root Mean Squared Error (RMSE)*
Linear Regression	1.48%	3,805,472
Gradient Boosting Machines	7.32%	6,703,723.08

Table 2: Comparison of GBM and Linear Regression considering RMSE and Average Percent Difference

We can see that comparing the models for forecasting total expenses for 2020, there is a notable difference between the two models. The Linear Regression model demonstrates a smaller percentage difference of 1.48% between predicted and actual values, suggesting relatively accurate predictions. However, its RMSE value of 3,805,472 indicates a moderate level of deviation between predicted and actual expenses. In contrast, the Gradient Boosting Machines model exhibits a higher percentage difference of 7.32%, indicating a greater level of discrepancy between predicted and actual values. Additionally, its RMSE value of 6,703,723.08 suggests a larger magnitude of prediction errors compared to the Linear Regression model. Despite its higher percentage difference and RMSE, the Gradient Boosting Machines model may still offer valuable insights and predictive capabilities, albeit with a greater margin of error.

Total County Wise Expenses for 2020

Region	*% Difference (GBM)*	*% DIFFERENCE (LINEAR REGRESSION)*
Annapolis	2.814188	1.429911
Antigonish	5.150445	9.617876
Cape Breton	17.353712	2.105211
Colchester	3.174709	0.271547
Cumberland	1.524529	3.896207
Digby	1.185751	-5.568142
Guysborough	6.083862	-1.161470
Halifax	3.136256	-0.925060
Hants	5.955603	1.255302
Inverness	4.970773	-61.852759
Kings	5.399633	-0.936483
Pictou	1.318797	22.682610
Queens	4.295345	14.268579
Richmond	1.148223	-9.597485
Shelburne	1.879644	1.066063
Victoria	3.527608	-3.429221
Yarmouth	0.782363	5.040955

Table 3: Analyzing Average Percent Difference by County Using Both Models

Both models show strengths and weaknesses in their predictive capabilities across different regions. The GBM model generally displays higher percentage differences, indicating larger prediction errors in some regions compared to the Linear Regression model. However, the Linear Regression model also exhibits substantial percentage differences in certain regions, suggesting limitations in accurately forecasting expenses. Each model is better (lower % differences) in a few regions and is worse in other regions. Another insight from the table is that GBM shows primarily positive % differences whereas linear regression shows both positive and negative % differences. This is an important factor when deciding which model to select. If only positive differences are required, GBM is more favorable, and when both positive and negative differences are required, Linear Regression is more favorable.

The interpretations of the results directly relate to the problem statement and objectives outlined for the project. The problem statement highlights the challenge faced by municipal leaders in Nova Scotia regarding efficient budget management and the risk of overspending without accurate forecasting. This aligns with the findings indicating that both models have strengths and weaknesses in predicting expenses across different regions, emphasizing the need for accurate forecasting tools.

The objectives of the project, including systemic analysis of historical data, using visualizations to elucidate trends, and incorporating multiple predictive models, are all reflected in the interpretations of the results. The analysis of historical data across various sectors to detect patterns and trends directly corresponds to the evaluation of the models' predictive capabilities. Additionally, the use of visualizations to highlight relationships and trends mirrors the comparison of the models' performance across different regions.

Implications

The results shed light on the importance of accurately forecasting expenses to promote fiscal responsibility and ensure the sustainability of municipal services. They also highlight a few key implications:

- Despite the differences in predictive accuracy between the Linear Regression and Gradient Boosting Machines models, both offer valuable insights that can aid municipal leaders in making informed decisions regarding budget management and resource allocation.
- From a societal perspective, accurate budget forecasting contributes to addressing issues such as economic inequality, social welfare, and environmental sustainability.
- Overall, the transformative potential of accurate budget forecasting tools is highlighted, enabling stakeholders to drive positive change and foster sustainable development for the benefit of all.

Limitations

Based on modelling, interpretation of results, and the implications, we can identify a few limitations:

- Limited Available Data: Our analysis was constrained by limited availability of data, primarily sourced from government resources which restricted how much data we had to work with. This can potentially lead to overfitting. Having more data can reduce the chance of overfitting and would directly impact % difference and RMSE. A potential solution to fix this limitation would be to consider data from a wider range of years.
- Outlier Year: From our data analysis we can conclude that 2020 was an outlier year in terms of source breakdown, revenue and expenditure breakdown, and county wise fiscal data. This adds on to limited data availability as we are only considering the data for the years 2013 to 2020 (each year has significant impact on the prediction).
- The models only consider the top 5 most correlated features to train the model. Additional features either from the dataset or external features that can be added, need to be considered to improve the model. This will decrease the chance of overfitting and potentially improve model performance.
- Since the model uses external features such as Population/Change, Inflation, GDP, etc., it can only be seen one year in the future. We can use population predictors, GDP, and Inflation projections; however, the accuracy of the model would suffer greatly from this.
- Only 2 models have been explored. More relevant models can be explored that can work with this dataset, and potentially provide better results. More bagging and ensemble techniques can be explored in the future that can improve the performance of singular models.

Conclusion

Summary of Key Findings

Overall, this project proved to be a challenging yet rewarding task, which tested our teamwork capabilities and skills as emerging and novice data scientists. In the data analysis phase, we examined the breakdown of Nova Scotia annual expenditures and how it evolved from year to year. This specifically highlighted the constant increase in overall expenditure as well as the variation in make up when dissected by county. These analyses necessitated the need for our model in order to better predict which counties would uncharacteristically high expenses for the coming year. Through data analyses, it was decided binning by land area and population was required, to better understand the data. This allowed us to successfully create our GBM and Linear Regression models. Both Models performed better than expected, given the small number of datapoints for this project. The Linear regression model achieved a percentage difference of 1.48% while the GBM achieved 7.32%. The RMSE for both models were higher than our team would have liked, however we attribute this to the minimal amount of data we were working with overall. In the future, we would have liked to either find or simulate data for years up to 2024, however this data was not available at the time of the project. In conclusion, we feel this model can be a viable tool for municipal leaders in Nova Scotia, to help gauge how certain counties will be performing with regards to expenses in the coming year.

Achievements of Objectives

At the beginning of this project, we outlined a couple objectives which we wanted to achieve:

1) Leverage historical data to develop predictive models for annual regional expenditure.
2) Focus on identifying high expenditure sectors and their interrelationships.
3) Commence a systematic analysis of historical data to detect patterns, trends, and correlations.
4) Enable precise forecasting of yearly expenditures for specific regions, emphasizing significant spending sectors.
5) Utilize visualizations to elucidate relationships and expenditure trends for municipal decision-making.
6) Incorporate a continuous evaluation mechanism to enhance tool accuracy and efficiency.

Of these goals, we achieved all but the last goal, which involves incorporating a continuous mechanism into our tool. Aside from this shortcoming, we were successful in achieving the goals of this project. We dove into a deep analysis of historical financial data from various sectors, searching for patterns and connections that could help us predict future spending in Nova Scotia's municipalities. These efforts culminated in the creation of two

predictive models: a Gradient Boosting Machine (GBM) and a Linear Regression model. Despite the hurdle of working with a limited dataset, both models surpassed our expectations. The Linear Regression model impressed us with its ability to predict 2020's total expenditures with only a 1.48% error margin. Though the GBM model had a larger error margin, it still offered us valuable foresight overall and the ability of choice with respect to modelling options.

Appendix

Code Repository

https://git.cs.dal.ca/kbhardwaj/nsgrow

Member Contributions

Kanav Bhardwaj:

- Data Collection
- Data Preprocessing
- Feature Engineering
- Data Visualization
- Modelling
- Final Report and Presentation

Abhiroop Yerramilli:

- Data Collection
- Data Preprocessing
- Data Visualizations
- Final Report and Presentation

Yakov Fainshtein:

- Data Collection
- Data Preprocessing
- Modelling
- Final Report and Presentation

Matthew Carl:

- Data Collection
- Data Preprocessing
- Data Visualizations
- Final Report and Presentation

Data Quality Report

Continuous Feature	*Min*	*1st Quartile*	*Mean*
Year	2013	2014.75	2016.5
Population	7023	14008.75	53196.49265
Population Change	-901	-142.5	277.2352941
Population Change Percentage	-2.23	-0.74	-0.233676471
Land Area	1246	2393	2936
Net property taxes and payments in lieu of taxes	3867261	9920940.75	59590477.22
Grants In Lieu Of Taxes	115617	445682.5	3624079.154
Services Provided To Other Governments	0	158901.5	906702.3603
Sales Of Services	0	519732	5649442.934
Other Revenue From Own Sources	207926	663086.75	3621576.926
Unconditional Transfers From Other Governments	0	845281	2851094.926
Conditional transfers from other governments	0	135140.5	1456731.691
Total Revenues	4455895	16642229.25	77700105.2
General Government Services	1226424	3372318.25	11335911.63
Protective Services	1184790	3494020.25	21909404.3
Transportation Services	268619	1232045.25	14881137.25
Environmental Health Services	743357	2933374.25	7828582.346
Public Health Services	0	0	251684
Environmental Development Services	97155	790524.25	2744555.728
Recreation And Cultural Services	311772	1445998	7892757.875
Extraordinary Or Special Items	-2380241	0	863207.4632
Total Expenses	3834232	15058259.25	67707240.57
Operating surplus (deficit) before financing and transfers	-170959	1395767.5	9992864.625
Next Year Total Expenses	3834232	15304985	68315024.8
Next Year Total Revenues	4455895	16848068	78449733.17
GDP	-5.038233441	0.94140595	1.191849887
Inflation	0.725	1.079166667	1.49375
Residential Tax Rates	0.571433333	1.2	1.3515344
Commercial Tax Rates	1.526666667	2.69	3.125657782
Minimum Wage	10.05	10.3	10.775
General Government Services_per_capita	112.5182284	170.1558324	239.8992937
Protective Services_per_capita	129.527714	265.4971995	335.868983
Transportation Services_per_capita	29.36689625	74.32977096	137.9175021
Environmental Health Services_per_capita	81.26784738	149.7391405	203.9036311
Public Health Services_per_capita	0	0	13.39194955
Environmental Development Services_per_capita	10.46921174	34.12307085	65.06957236
Recreation And Cultural Services_per_capita	34.07663267	79.83337278	108.231776
Extraordinary Or Special Items_per_capita	-23.71537458	0	3.741268368
Total Expenses_per_capita	419.1791844	875.0060585	1108.023974
Net property taxes and payments in lieu of taxes_per_capita	422.7900951	736.6547553	902.094454

Grants In Lieu Of Taxes_per_capita	10.13317822	25.28687941	48.23329192
Services Provided To Other Governments_per_capita	0	6.324124284	57.07433499
Sales Of Services_per_capita	0	23.15615967	90.24807627
Other Revenue From Own Sources_per_capita	15.97035414	38.43292436	62.97790008
Unconditional Transfers From Other Governments_per_capita	0	26.30454946	77.76210917
Conditional transfers from other governments_per_capita	0	7.209765275	18.57978485
Total Revenues_per_capita	487.1427791	973.8172024	1256.969951
Population Density	1.823879118	5.689987815	15.52748379
Population Size	1	2	2.647058824
Continuous Feature	**Median**	**3rd Quartile**	**Max**
Year	2016.5	2018.25	2020
Population	21049	44875.75	450893
Population Change	-41	57	11076
Population Change Percentage	-0.2	0.2325	2.52
Land Area	2836	3627	5477
Net property taxes and payments in lieu of taxes	16106060.5	41979217.5	673795000
Grants In Lieu Of Taxes	731383.5	1630358.5	40760000
Services Provided To Other Governments	574339	1181345	5363653
Sales Of Services	1151186.5	2814447.25	76128650
Other Revenue From Own Sources	1523009.5	2551843	42553000
Unconditional Transfers From Other Governments	1390471	2598629.75	19193672
Conditional transfers from other governments	346946.5	789809	63394000
Total Revenues	21210988	52963129.25	865285000
General Government Services	4752373	8369226.75	113117000
Protective Services	5402950	15007023.5	227684000
Transportation Services	2489106	5581036.75	190023000
Environmental Health Services	3694841	7126141.5	50543000
Public Health Services	87899.5	260186.75	2568551
Environmental Development Services	1132294	2158332.5	30858000
Recreation And Cultural Services	2068375	4220929.75	107211000
Extraordinary Or Special Items	0	0	17653500
Total Expenses	18652275	47259009.25	708955527
Operating surplus (deficit) before financing and transfers	3436017.5	6282748.75	156329473
Next Year Total Expenses	18977929	47495333.5	708955527
Next Year Total Revenues	21567761	52967794.5	865285000
GDP	2.11712275	2.775589266	3.033834903
Inflation	1.520833333	1.916666667	2.266666667
Residential Tax Rates	1.4	1.495	2.06
Commercial Tax Rates	3.083333333	3.413333333	5.23875
Minimum Wage	10.525	10.8875	12.55
General Government Services_per_capita	199.0123742	251.2903349	1076.001781
Protective Services_per_capita	315.6437075	417.6423822	1158.055725

Transportation Services_per_capita	109.9114426	157.0977872	455.8457865
Environmental Health Services_per_capita	173.613541	231.9335873	791.4936387
Public Health Services_per_capita	2.660529616	10.62729538	144.3600884
Environmental Development Services_per_capita	48.8376675	67.6092888	357.6367882
Recreation And Cultural Services_per_capita	97.21384016	126.3741042	264.9288804
Extraordinary Or Special Items_per_capita	0	0	52.92571429
Total Expenses_per_capita	1050.067184	1216.98622	3761.102036
Net property taxes and payments in lieu of taxes_per_capita	888.4151461	963.1847915	3502.085751
Grants In Lieu Of Taxes_per_capita	40.44774276	71.49447635	171.4155667
Services Provided To Other Governments_per_capita	16.35681963	34.89894616	728.4602743
Sales Of Services_per_capita	39.80742947	91.91700212	629.272342
Other Revenue From Own Sources_per_capita	55.8841599	76.63282976	411.1260127
Unconditional Transfers From Other Governments_per_capita	67.36468415	106.3797351	221.8027908
Conditional transfers from other governments_per_capita	14.42464059	23.93996926	140.5965495
Total Revenues_per_capita	1175.668969	1323.538279	4174.749618
Population Density	7.300963082	14.30900853	82.32481285
Population Size	2	3	6
Continuous Feature	**STD DEV**	**# Instances**	**% Missing Values**
Year	2.299758441	136	0
Population	95639.39496	136	0
Population Change	1691.515123	136	0
Population Change Percentage	0.820167832	136	0
Land Area	1036.032604	136	0
Net property taxes and payments in lieu of taxes	137786179.5	136	0
Grants In Lieu Of Taxes	8894099.796	136	0
Services Provided To Other Governments	1161144.508	136	0
Sales Of Services	15993118.33	136	0
Other Revenue From Own Sources	8235960.304	136	0
Unconditional Transfers From Other Governments	4292134.512	136	0
Conditional transfers from other governments	5713627.323	136	0
Total Revenues	175624912.6	136	0
General Government Services	22610788.8	136	0
Protective Services	47728214.88	136	0
Transportation Services	42215105.95	136	0
Environmental Health Services	10384046.41	136	0
Public Health Services	426745.687	136	0
Environmental Development Services	5671158.316	136	0
Recreation And Cultural Services	19897271.89	136	0
Extraordinary Or Special Items	3621084.586	136	0
Total Expenses	150900291.7	136	0
Operating surplus (deficit) before financing and transfers	25150365.09	136	0

Next Year Total Expenses	152664065.7	119	12.5
Next Year Total Revenues	177864673.2	119	12.5
GDP	2.498104839	136	0
Inflation	0.503809306	136	0
Residential Tax Rates	0.30619659	136	0
Commercial Tax Rates	0.791876409	136	0
Minimum Wage	0.76603839	136	0
General Government Services_per_capita	125.9437504	136	0
Protective Services_per_capita	120.6039664	136	0
Transportation Services_per_capita	95.78458424	136	0
Environmental Health Services_per_capita	99.64032285	136	0
Public Health Services_per_capita	30.37351786	136	0
Environmental Development Services_per_capita	65.58899318	136	0
Recreation And Cultural Services_per_capita	48.20079094	136	0
Extraordinary Or Special Items_per_capita	11.19450257	136	0
Total Expenses_per_capita	401.7141486	136	0
Net property taxes and payments in lieu of taxes_per_capita	308.3633001	136	0
Grants In Lieu Of Taxes_per_capita	28.83623645	136	0
Services Provided To Other Governments_per_capita	139.1607612	136	0
Sales Of Services_per_capita	127.5461409	136	0
Other Revenue From Own Sources_per_capita	42.78562678	136	0
Unconditional Transfers From Other Governments_per_capita	59.29268392	136	0
Conditional transfers from other governments_per_capita	17.55659583	136	0
Total Revenues_per_capita	493.2468907	136	0
Population Density	18.27899486	136	0
Population Size	1.416983817	136	0
Continuous Feature	**Cardinality**		
Year	8		
Population	135		
Population Change	122		
Population Change Percentage	107		
Land Area	17		
Net property taxes and payments in lieu of taxes	136		
Grants In Lieu Of Taxes	136		
Services Provided To Other Governments	119		
Sales Of Services	135		
Other Revenue From Own Sources	136		
Unconditional Transfers From Other Governments	135		
Conditional transfers from other governments	129		
Total Revenues	136		
General Government Services	136		
Protective Services	136		

Transportation Services	136
Environmental Health Services	136
Public Health Services	91
Environmental Development Services	136
Recreation And Cultural Services	136
Extraordinary Or Special Items	37
Total Expenses	136
Operating surplus (deficit) before financing and transfers	136
Next Year Total Expenses	119
Next Year Total Revenues	119
GDP	8
Inflation	8
Residential Tax Rates	86
Commercial Tax Rates	90
Minimum Wage	8
General Government Services_per_capita	136
Protective Services_per_capita	136
Transportation Services_per_capita	136
Environmental Health Services_per_capita	136
Public Health Services_per_capita	94
Environmental Development Services_per_capita	136
Recreation And Cultural Services_per_capita	136
Extraordinary Or Special Items_per_capita	37
Total Expenses_per_capita	136
Net property taxes and payments in lieu of taxes_per_capita	136
Grants In Lieu Of Taxes_per_capita	136
Services Provided To Other Governments_per_capita	119
Sales Of Services_per_capita	135
Other Revenue From Own Sources_per_capita	136
Unconditional Transfers From Other Governments_per_capita	135
Conditional transfers from other governments_per_capita	129
Total Revenues_per_capita	136
Population Density	135
Population Size	6

Categorical Features	***Mode***	***2nd Mode***
Region	Annapolis	Inverness
Land Area Size	0-5000	5000-50000
Categorical Features	**Proportion of Mode**	**Frequency of 2nd Mode**
Region	5.882352941	8
Land Area Size	94.11764706	8
Categorical Features	**% Missing Values**	**Cardinality**
Region	0	17
Land Area Size	0	2
Categorical Features	**Frequency of Mode**	**Proportion2nd Mode**
Region	8	5.882352941
Land Area Size	128	5.882352941

References

Data

[1] Municipal Fiscal Statistics: Operating Fund Total Revenues and Expenditures by Regional Municipality:

https://data.novascotia.ca/Municipalities/Municipal-Fiscal-Statistics-Operating-Fund-Total-R/thwb-cfp5

[2] Canadian GDP:

https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=CA

[3] Canadian Inflation Rate:

https://www.statista.com/statistics/271247/inflation-rate-in-canada/

[4] Population data: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810000203&geocode=A000212

[5] More Population Data:
https://www150.statcan.gc.ca/t1/tbl1/en/cv!recreate.action?pid=1710013901&selectedNodeIds=1D18,1D19,1D20,1D21,1D22,1D23,1D24,1D25,1D26,1D27,1D28,1D29,1D30,1D31,1D32,1D33,1D34,1D35,3D111,3D112,3D113,3D114,3D115&checkedLevels=1D2&refPeriods=20160101,20220101&dimensionLayouts=layout3,layout2,layout2,layout2&vectorDisplay=false

[6] Land Area:

https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=9810000202 [1]

https://www.youtube.com/watch?v=h1BnRBzYjYY

Images

https://i.pinimg.com/originals/ed/e9/6f/ede96fc77a612da3efc514ee7462dadf.jpg

Tables

Feature	***Description***
Region	County in Nova Scotia
Year	Year of data
Population	Population of the region
Population Change	Population change in numbers
Population Change Percentage	Percentage of population change
Land Area	Area of the region in square kilometers
Net property taxes and payments in lieu of taxes	Total property taxes collected, including payments instead of taxes
Grants In Lieu Of Taxes	Payments by government entities not subject to regular taxation
Services Provided To Other Governments	Revenues from services provided to other governmental units
Sales Of Services	Revenue from services sold to the public or other organizations
Other Revenue From Own Sources	Miscellaneous revenues not categorized elsewhere
Unconditional Transfers From Other Governments	Funds received from higher levels of government without conditions
Conditional transfers from other governments	Funds from higher governments with specified use conditions
Total Revenues	Sum of all revenue sources
General Government Services	Expenses for administrative services of the government
Protective Services	Expenses for police, fire, and emergency services
Transportation Services	Expenses related to transportation infrastructure and services
Environmental Health Services	Expenses for waste management and sanitation services
Public Health Services	Expenses for public health programs and initiatives
Environmental Development Services	Expenses for land use and environmental protection initiatives
Recreation And Cultural Services	Expenses for recreational and cultural facilities and services
Extraordinary Or Special Items	Unusual or one-time revenues or expenses
Total Expenses	Sum of all government expenditures
Operating surplus (deficit) before financing and transfers	Net of revenues over expenses before accounting for debts and transfers
Next Year Total Expenses	Forecasted total expenses for the following year
Next Year Total Revenues	Forecasted total revenues for the following year
County	A governmental division within a province in Canada
District	A subdivision of a county, often for administrative or political purposes
Regional	Pertaining to the entire region or encompassing several districts
GDP	Gross Domestic Product of the region
Inflation	Rate at which the general level of prices for goods and services is rising
Residential Tax Rates	Tax rates applicable to residential properties
Commercial Tax Rates	Tax rates applicable to commercial properties
Minimum Wage	The minimum legal wage for workers in the region

Table 2: Engineered Features Breakdown

Feature	***Description***
Next Year Total Expenses	The expenses for the next year of the region
Next Year Total Revenues	The revenues for the next year of the region
General Government Services_per_capita	General government services expenses divided by the population
Protective Services_per_capita	Protective services expenses divided by the population
Transportation Services_per_capita	Transportation services expenses divided by the population
Environmental Health Services_per_capita	Environmental health services expenses divided by the population
Public Health Services_per_capita	Public health services expenses divided by the population
Environmental Development Services_per_capita	Environmental development services expenses divided by the population
Recreation And Cultural Services_per_capita	Recreation and cultural services expenses divided by the population
Extraordinary Or Special Items_per_capita	Extraordinary or special items expenses divided by the population
Total Expenses_per_capita	Total expenses divided by the population
Net property taxes and payments in lieu of taxes_per_capita	Net property taxes and payments in lieu of taxes divided by the population
Grants In Lieu Of Taxes_per_capita	Grants in lieu of taxes divided by the population
Services Provided To Other Governments_per_capita	Revenue from services provided to other governments divided by the population
Sales Of Services_per_capita	Revenue from sales of services divided by the population
Other Revenue From Own Sources_per_capita	Other revenue from own sources divided by the population
Unconditional Transfers From Other Governments_per_capita	Unconditional transfers from other governments divided by the population
Conditional transfers from other governments_per_capita	Conditional transfers from other governments divided by the population
Total Revenues_per_capita	Total revenues divided by the population
Population Density	Population density is calculated as population divided by the land area of the region
Population Size	A categorization of population into bins for analysis purposes
Land Area Size	A categorization of land area into bins for analysis purposes

Figures

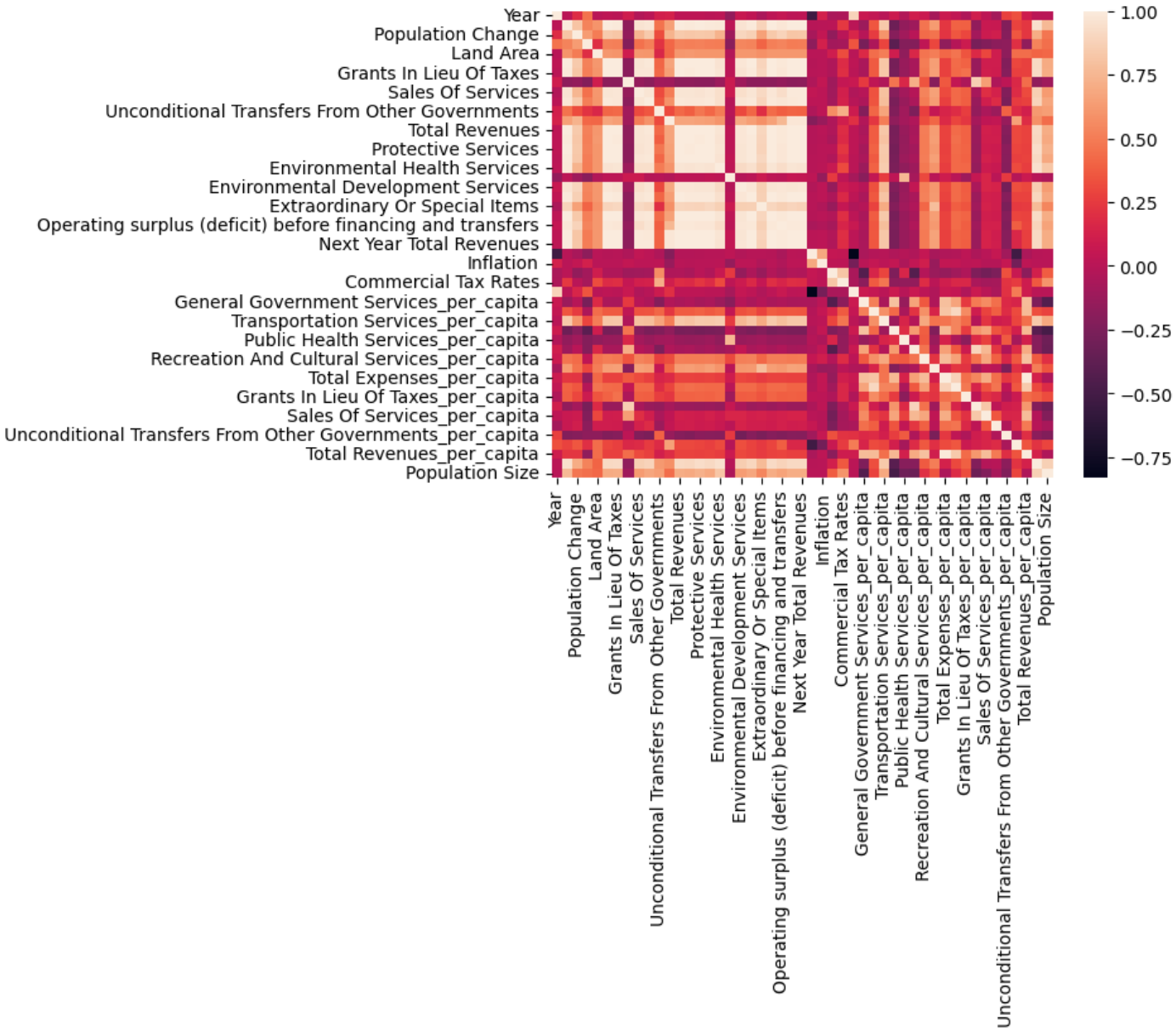

Figure 9: Correlation Heap Map

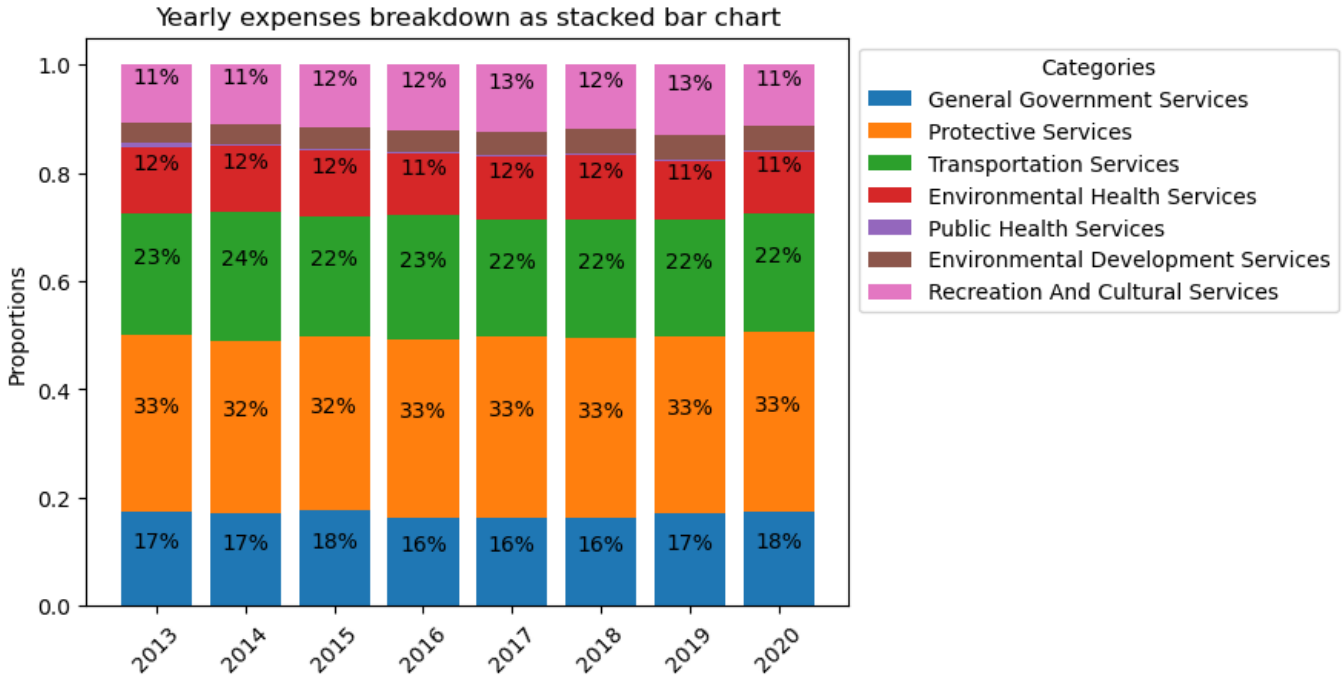

Figure 10: Percentage of Yearly Expenses Spent on Various Categories

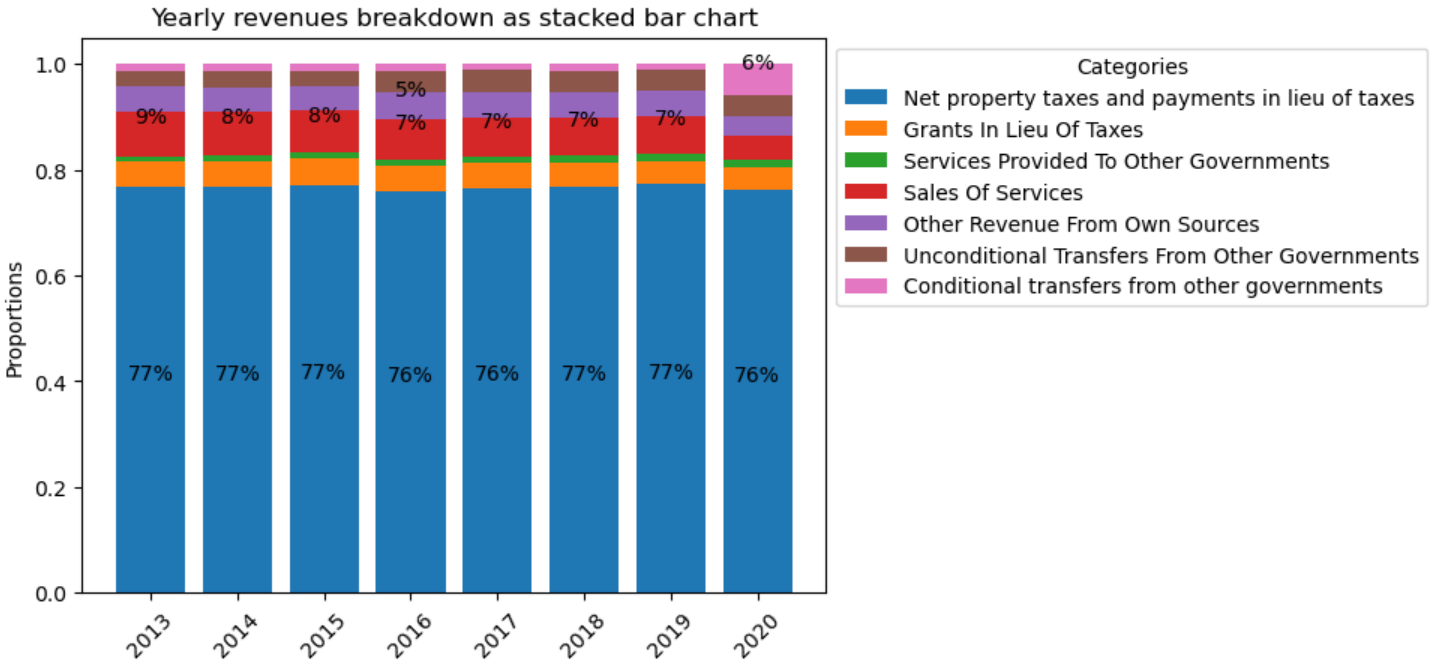

Figure 11:Percentage of Yearly Revenues Earned from Various Categories

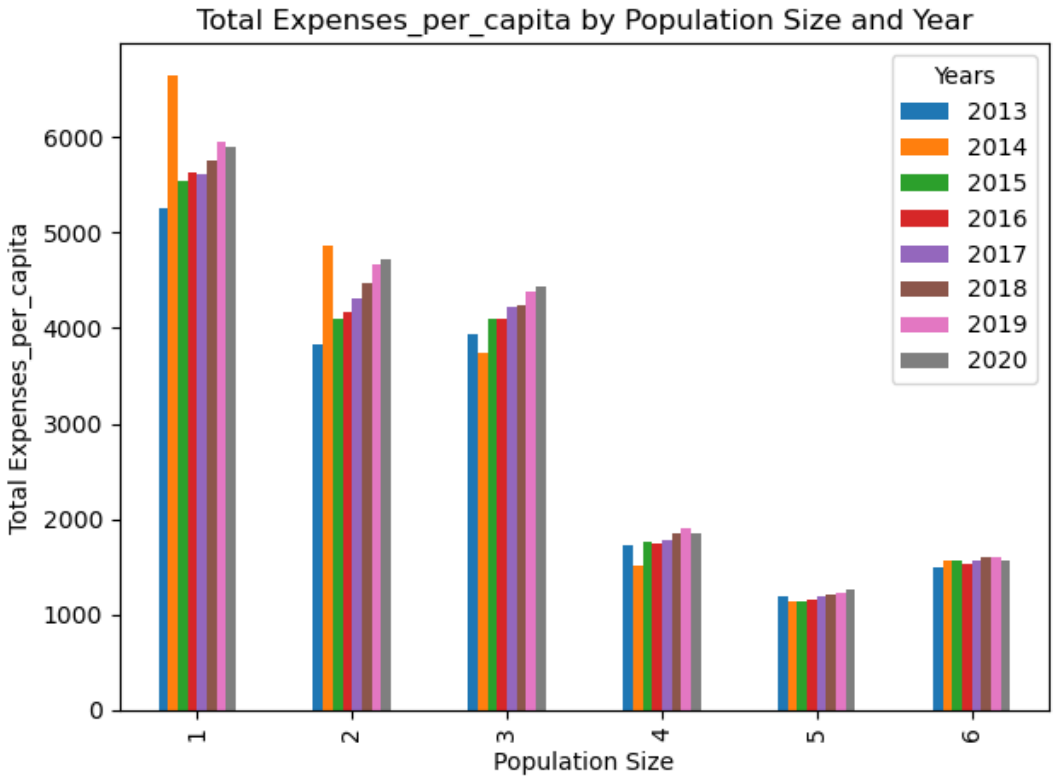

Figure 12:: Total Expense per Capita Split based on Population Size and Year

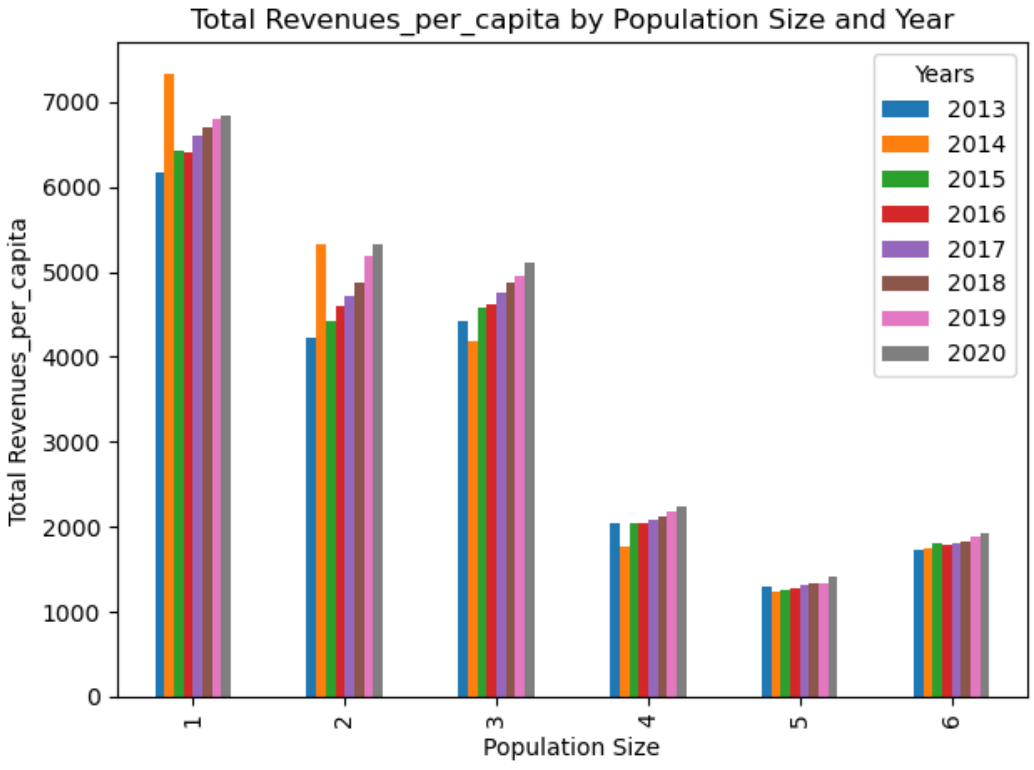

Figure 13: Total Revenues per Capita Split based on Population Size and Year

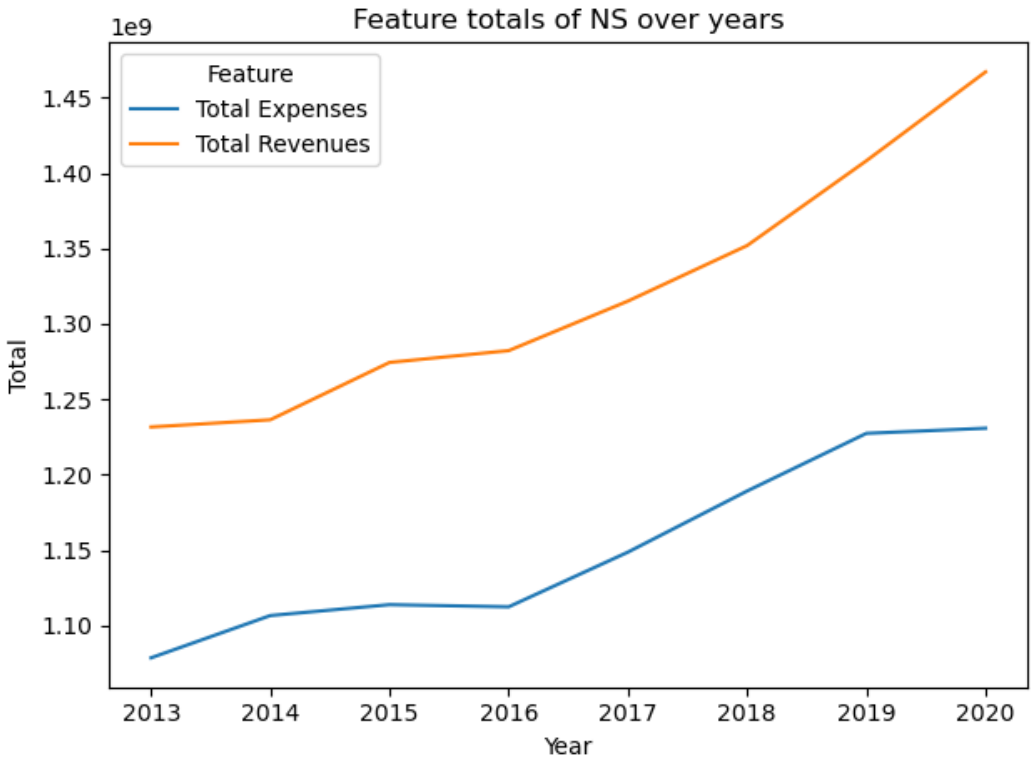

Figure 14: Total Revenues and Expenses for All of Nova Scotia each Year.

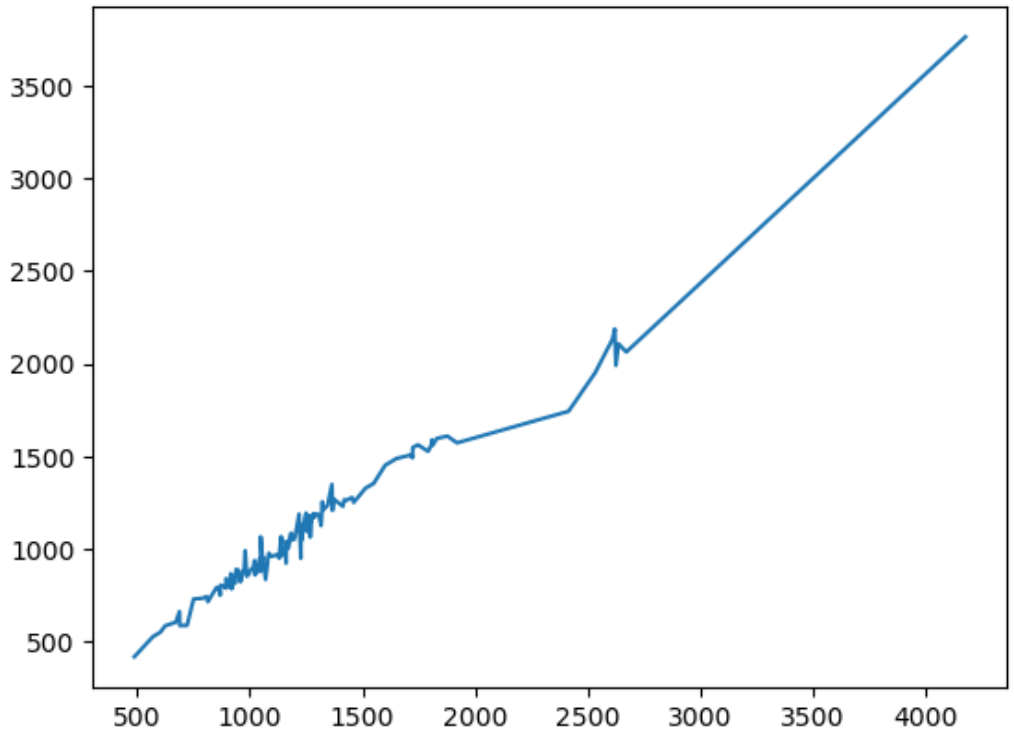

Figure 15: Total Expenses per Capita vs Total Revenues per Capita.

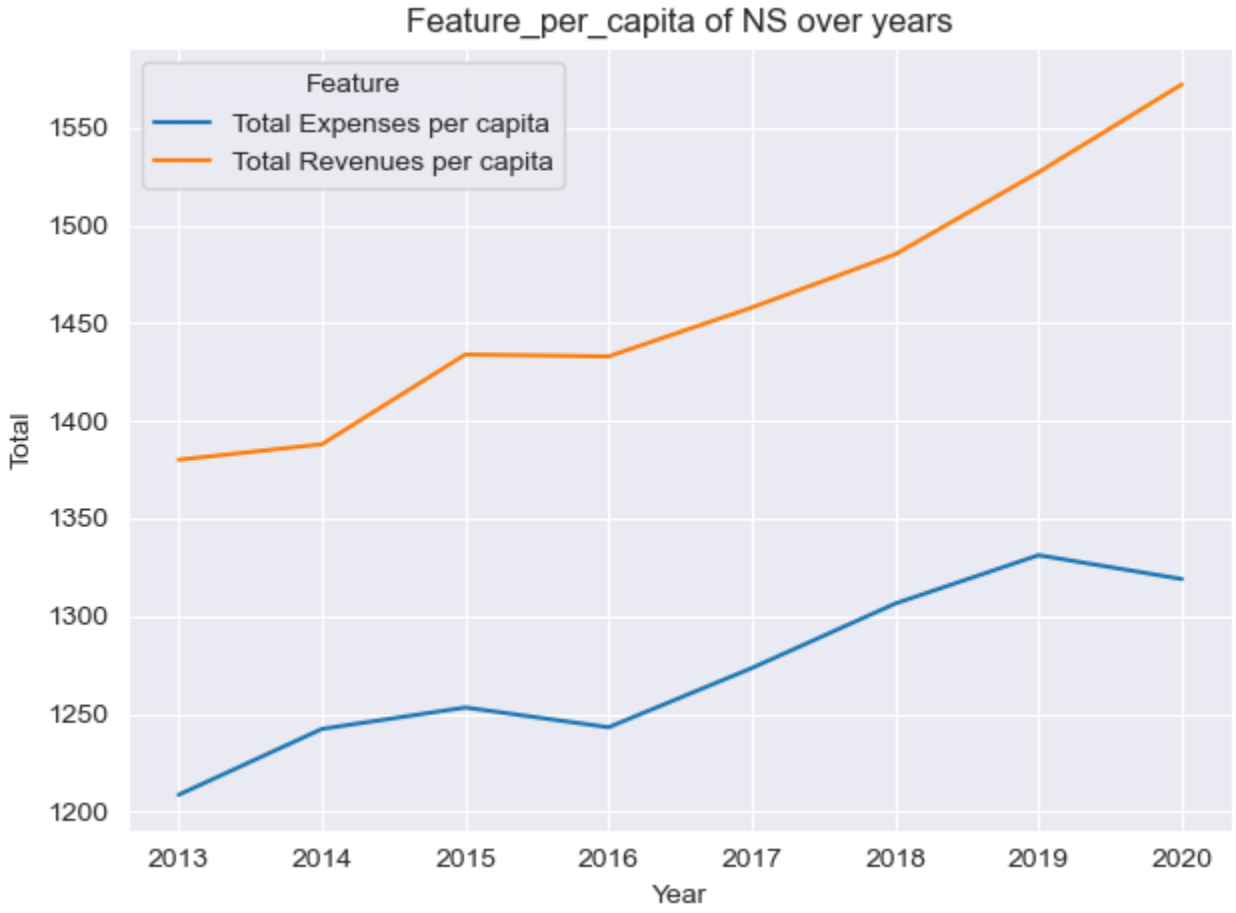

Figure 16: Revenues and Expenses Per Capita for All of Nova Scotia each Year.

Feature	Weights for log NEXT_YEAR_TOTAL_EXPENSES	Weights for log NEXT_YEAR_TOTAL_REVENUES
YEAR	0.011	0.019
logPOPULATION	0.0001	N/A
GDP	0.01	0.003
INFLATION	-0.015	0.008
MINIMUM_WAGE	0.002	0.002
log TOTAL_EXPENSES	0.0002	0.0005
TOTAL_REVENUES	0.0006	0
log TOTAL_REVENUES	0	0.0002
Constant	-2.78	-18.28

Feature	Weights for log TOTAL_EXPENSES	Weights for log TOTAL_REVENUES
YEAR	0.03	0.03
logPOPULATION	0.55	0.58
POPULATION CHANGE	-0.07	-0.08
POPULATION DENSITY	0.01	0.008
LAND AREA	0.41	0.45
COUNTY	-0.15	-0.18
DISTRICT	-0.03	-0.03
REGIONAL	0.18	0.21
RESIDENTIAL TAX RATES	-0.50	-0.66
POPULATION SIZE	0.17	0.20
Constant	-61.57	-68.51

Feature	*Actual (in CAD)*	*Predicted (in CAD)*	*%age Difference*
Total Expenses 2020	1230868649	1212541850	1.48 %
Total Revenues 2020	1467053121	1408906860	12.13 %

Year	Region	Total Expenses	Total Revenues	Predicted Next Year Total Expenses	Predicted Next Year Total Revenues	%age Error Expenses	%age Error Revenues
2020	Digby	16891255	18960166	17831784	19967483	-5.568142	-5.312807
2020	Guysborough	16100548	19282783	16287551	19749532	-1.161470	-2.420548
2020	Halifax	708955527	865285000	715513794	822086748	-0.925060	4.992373
2020	Victoria	8786747	10256693	9088064	9859349	-3.429221	3.873997
2020	Hants	43095201	49575049	42554226	50468476	1.255302	-1.802171
2020	Cape Breton	126935138	142219006	124262886	135008891	2.105211	5.069727
2020	Inverness	12735592	13697179	20612907	23565834	-61.852759	-72.048814
2020	Cumberland	37128393	42627666	35681794	41731339	3.896207	2.102688

2020	Shelburne	17512846	18904460	17326148	19121731	1.066063	-1.149311
2020	Kings	56289133	63438378	56816271	63724983	-0.936483	-0.451785
2020	Antigonish	16867362	21567761	15245080	16554337	9.617876	23.244991
2020	Pictou	48267139	55379522	37318892	41124445	22.682610	25.740701
2020	Colchester	49426808	63758102	49292591	57379208	0.271547	10.004837
2020	Richmond	8668365	9581925	9500310	10490969	-9.597485	-9.487071
2020	Queens	15683475	17842769	13445666	14854730	14.268579	16.746498
2020	Annapolis	19590452	22043387	19310326	21117290	1.429911	4.201246
2020	Yarmouth	27934668	32633275	26526494	30603496	5.040955	6.219967

Table 4: Total Revenues and Expenses Actual vs Predicted. Total and by County

Size	***Counties***	***Value***
Population Size		
0 – 11000	Small	1
11000 – 22000	Medium	2
22000 – 46000	Large	3
46000 – 90000	Larger	4
90000 – 150000	Cape Breton	5
150000 – 1000000	Halifax	6
Land Size		
1000 – 2000	Antigonish, Richmond	1
2000 – 3000	Kings, Yarmouth, Queens, Cape Breton, Shelburne, Digby, Victoria, Pictou, Lunenburg	2
3000 – 4000	Hants, Annapolis, Colchester, Inverness	3
4000 – 5000	Guysborough, Cumberland	4
5000 – 6000	Halifax	5

Table 5: Categorization of Population and Land Area