Consumer Spending in Canada: A 12-Year Provincial Analysis of Patterns, Key Drivers, Regional Variations, and Future Trends

Shannon School of Business, Cape Breton University

A Capstone Project on Business Analyst Research Paper

In Partial Fulfillment of the Requirements in BUSS-5802-11

Instructor: Enayat Rajabi

Group E

Jereika Aiko Tan- Student ID: 0264312

Luzviminda Deinla- Student ID: 0274950

Shijie Liu - Student ID: 0276875

Simarpreet Kaur- Student ID: 0285168

Table of Contents

1 Introduction

Over the past 12 years, consumer spending in Canada has evolved under the influence of income, inflation, and demographic changes, with significant regional variations. This study examines these trends, explores key drivers, and employs predictive models to provide actionable insights for businesses, policymakers, and researchers navigating future market dynamics.

2 Objectives

- Examine historical spending patterns (2010–2021) in key Canadian provinces (Ontario, British Columbia, Alberta, Quebec, and Saskatchewan), focusing on housing, food, transportation, and discretionary goods.
- Analyze the influence of economic and demographic factors (e.g., income, inflation, household size, aging) on spending behavior.
- Identify and explore the causes of regional spending differences.
- Develop and evaluate predictive models to forecast future demand, supporting businesses and policymakers.

3 Dataset and Methods

3.1 Data Collection and Description

This study uses data from Statistics Canada, the Open Government Portal, and the Bank of Canada. The dataset includes national and provincial expenditures, income, and CPI (2010–2021), with 50,945 data points across key variables: Figure 0_Data Dictionary

3.2 Data Preparation

Data preparation and initial cleaning were done using MS Excel with the following steps:

- Merged five files into a unified dataset with 10 key spending categories, removing duplicates while retaining relevant data.
- Addressed inconsistencies from merging by replacing blanks with zero values to ensure completeness.
- Removed irrelevant rows and columns for consistency.
- Added a 'Data Type' column to group data (e.g., 'Spend: Category,' 'Spend: Age,' 'Income: Age') for easier filtering and targeted analysis.

3.3 Exploratory Data Analysis (EDA)

The data was then loaded into Python Jupyter Notebook for further inspection, cleaning, and exploratory data analysis.

3.3.1 Summary Statistics of Key Variables

Figure 1_Summary Statistics for Numeric Variables

- The dataset shows considerable variability and skewness in key variables. Average Expenditure per Household and Disposable Income per Household highlight significant disparities influenced by high-income households. Missing data, represented as zero values, were used as placeholders to preserve dataset integrity without impacting the analysis.

3.3.2 Trend Analysis: Overview of Spending Over Years

3.3.2.1 Household Spending Trends by Category (2010-2021)

Line charts were used to visualize trends in key expenditure categories (Food, Housing, Transportation, and Discretionary Goods) from 2010 to 2021:

Figure 2_Household Spending Trends by Category (2010-2021)

- Essential spending on food and housing steadily increased, with housing showing the highest growth.
- Transportation spending initially rose but dropped sharply after 2019 due to reduced commuting during the COVID-19 pandemic.
- Discretionary spending peaked in 2019 but slightly declined by 2021, reflecting economic caution during the pandemic.

3.3.2.2 Household Spending Trends by Province (2010-2021)

Line charts were made to show household spending trends in key provinces (Ontario, British Columbia, Alberta, Quebec, and Saskatchewan) from 2010 to 2021:

Figure 3_Household Spending Trends by Province (2010-2021)

- Alberta had the highest average household spending, despite some ups and downs.
- British Columbia and Ontario showed steady growth, while Quebec and Saskatchewan had lower, more stable spending.
- These trends show differences in household spending across Canada during this time.

3.3.3 Provincial Share of Household Expenditure

A pie chart was used to show household spending distribution across provinces:

Figure 4_Provincial Share of Household Expenditure

- Alberta leads with the highest share (23.8%), followed by Ontario (20.1%) and British Columbia (20.0%).
- Saskatchewan (19.6%) and Quebec (16.5%) have smaller shares, highlighting Alberta's higher household spending.

3.3.4 Impact of Economic Factors (Disposable Income and Consumer Price Index)

The analysis examined how Disposable Income and CPI influenced spending patterns across provinces, visualized with line charts:

Figure 5_Impact of Economic Factors (Disposable Income and Consumer Price Index)

- **Alberta:** Consumer spending dropped during the 2016 oil price crash (CAD 113,000 in 2014 to CAD 106,000 in 2016) but recovered to over CAD 120,000 by 2021. CPI steadily rose, reflecting inflation.
- **British Columbia:** Spending grew steadily from CAD 75,000 in 2010 to CAD 101,000 in 2021, with income growth and moderate inflation indicating stability.
- **Ontario:** Spending increased from CAD 76,000 in 2010 to CAD 98,000 in 2021, alongside income growth. Rising CPI pointed to higher living costs.

- **Quebec:** Spending rose modestly from CAD 63,000 in 2010 to CAD 81,000 in 2021, influenced by regional factors. CPI increased steadily, with a softer inflationary impact.
- **Saskatchewan:** Spending was more volatile, especially around 2016 due to agricultural reliance. However, income and CPI steadily grew, showing resilience to economic shifts.

3.3.5 Impact of Demographic Factors (Age Groups, Household Type and Income Quintile)

The analysis shifted to demographic factors, examining age groups, household types, and income quintiles using stacked bar charts:

Figure 6_Impact of Demographic Factors (Age Groups, Household Type and Income Quintile)

- **Age Groups:** Younger households (under 30) spend less, especially on transportation and discretionary goods, due to lower incomes. Middle-aged households (30–54) have the highest spending, driven by higher incomes and family needs, while older households (55+) spend less overall, focusing more on essentials and less on discretionary goods. Housing expenses decline for older households, likely due to downsizing or reduced mortgage payments.
- **Household Types:** Spending increases with household size. Single-person households spend more on discretionary goods, while couples with children allocate the most to housing and transportation. Larger households face the highest costs overall, while discretionary spending decreases with household size.
- **Income Quintiles:** Household spending rises significantly with income. Higher-income households spend more across all categories, especially on discretionary goods, while lower-income households allocate a greater share to essentials like housing and food, reflecting financial constraints.

3.3.6 Regional Comparison of Household Spending by Category

Bar charts were generated for each province to compare average household spending across categories, underscoring regional variations and their causes:

Figure 7_Regional Comparison of Household Spending by Category

- Alberta leads in total spending across categories by 2021, followed by British Columbia, with housing consistently being the top expense in all provinces.
- While Alberta, Ontario, and British Columbia allocate the most to housing, Quebec spends less on housing but prioritizes food more.
- Discretionary spending surged in Alberta, British Columbia, and Saskatchewan, likely driven by higher incomes and lifestyle choices, while Quebec's modest, stable growth reflects its conservative spending culture and lower cost of living.

3.3.7 Box Plot Analysis of Key Spending Categories by Province

In order to further examine differences across provinces, we used box plots to analyze each key spending category:

Figure 8_Box Plot Analysis of Key Spending Categories by Province

- **Food:** Income disparities and regional cost-of-living variances are behind Alberta's and Saskatchewan's higher median spending. In British Columbia and Ontario, food spending

is moderate, whereas in Quebec it is lower and more consistent, reflecting a more conservative lifestyle.

- **Housing:** Ontario and Alberta have the highest median housing costs, followed by British Columbia. Quebec has the lowest median, signifying stable housing costs, with Saskatchewan showing modest spending and a wider range.
- **Transportation:** Alberta spends the most on transportation with significant variability, followed by Saskatchewan. British Columbia, Ontario, and Quebec spend less, with Quebec showing the least variability and few high-cost outliers in some provinces.
- **Discretionary Goods:** Alberta and Saskatchewan have the highest discretionary spending with more variability, showing diverse habits. British Columbia and Ontario spend moderately, while Quebec shows the lowest and most stable spending, indicating a more conservative approach.

3.4 Predictive Model Fitting, Evaluation and Selection

3.4.1 Data Transformation and Pre-processing

To generate detailed predictions that account for economic and demographic factors, as well as yearly and regional variations, the dataset was refined to focus on categories with complete data for all three key numeric variables, while retaining both yearly and regional information. Although Age Group and Household Type also have complete data, they provide only national-level insights. This refinement ensures the dataset remains clean, focused, and suitable for accurate and regionally representative predictions.

- **Input variables:** Year, Region, Income Quintile, Consumer Price Index, and Disposable Income per Household (CAD).
- **Target variable:** Average Expenditure per Household (CAD).

3.4.2 One-Hot Encoding of Categorical Variables

Categorical variables were converted into binary columns to prepare the data for machine learning models.

3.4.3 Outlier Detection and Handling

Figure 9_Box Plots of Capped Data to Confirm Outlier Handling

- Outliers were identified and addressed using the IQR method and a capping system.

3.4.4 Multicollinearity Analysis: Variance Inflation Factor (VIF) and Correlation Matrix Heatmap

Multicollinearity among predictor variables was analyzed to identify highly correlated features that could affect the reliability and interpretability of regression models:

Figure 10_Initial VIF Values of Numeric Variables;

Figure 11_Initial Correlation Heatmap of All the Variables Including the Target Variable

- High VIFs: Year (VIF = 117.99) and Consumer Price Index (VIF = 154.66) indicate severe multicollinearity, which can make regression coefficients unstable and unreliable.
- Correlation Heatmap: Strong correlations between Income Quintiles and Disposable Income per Household (CAD) suggest redundancy, as both variables likely reflect similar aspects of household wealth or economic capacity.

3.4.5 Addressing Multicolliearity and Dimensionality Reduction

Figure 12_VIF Values of Numeric Variables After Addressing Multicollinearity

Figure 13_Correlation Heatmap of All the Variables After Addressing Multicolliearity

- A new variable, CPI_Trend, was created by combining CPI with Year to mitigate multicollinearity.
- Income Quintile was dropped to reduce redundancy with Disposable Income per Household (CAD).
- Post-adjustments, all features have low VIFs (1.006), indicating minimal multicollinearity.
- **Revised input variables after the changes:** Region_British Columbia, Region_Ontario, Region_Quebec, Region_Saskatchewan, Average expenditure per household (CAD), Disposable income per household (CAD) and CPI_Trend (Year * CPI)

3.4.6 Predictive Models Fitting and Evaluation

Four predictive models were evaluated on the dataset (Linear Regression, Random Forest Regressor, XGBoost Regressor, and Neural Network Regressor), with the following results:

Figure 14_Predictive Model Comparison

- Linear Regression emerged as the best overall model, achieving the lowest MSE, RMSE, and MAE, along with a high R^2 score of 0.98.
- XGBoost performed well, particularly excelling in minimizing Mean Percentage Error (MPE) at 7.36%.
- Both Linear Regression and XGBoost achieved the highest R^2 scores, but Linear Regression had better accuracy and lower errors overall.
- For minimizing percentage error, XGBoost is the preferred choice, while Linear Regression excels in overall accuracy and error reduction.

3.4.7 Cross validation across Four Models

Cross-validation was conducted across the four models to compare their performance and robustness, ensuring reliable evaluation through multiple data splits using MSE and R^2 metrics.

Figure 15_Predictive Model Cross Validation Results

- Linear Regression consistently outperformed other models, with the lowest MSE mean and standard deviation, indicating high accuracy and consistency. It also achieved the highest R^2 mean and the lowest R^2 standard deviation, reflecting stable performance and the best explanation of variance in the data.
- Overall, Linear Regression proved to be the most accurate and robust model for predicting household expenditure.

3.4.8 Optimizing Linear Regression Performance with Ridge and Lasso Regularization Techniques

Three regularized linear regression models (Linear, Ridge, and Lasso) were evaluated using cross-validation to reduce overfitting and improve reliability. We also visualized the residuals to ensure the model is well-calibrated and the assumptions of regression analysis are met:

Figure 16_Actual vs Predicted Expenditure (Optimized Linear Regression)

- All models demonstrated strong predictive accuracy ($R^2 = 0.98$), with Linear and Lasso performing similarly, while Ridge showed slightly lower accuracy; Linear Regression was chosen as the best model for its consistency across metrics. Linear and Lasso had similar performance, with identical MAE and percentage error values. Ridge had slightly higher MSE and RMSE, indicating slightly lower accuracy.

3.4.9 Trial Prediction and Accuracy Evaluation

We used Linear Regression to forecast Average Household Expenditure (CAD), accounting for regional differences, disposable income, and inflation. Trained on 2010–2017 data, the model predicted Ontario's 2021 expenditure with a CPI of 141.6 and a Disposable Income of CAD 99,550:

- The predicted expenditure for 2021 is CAD 104,137.66, with an MAE of CAD 5,896.66 and a percentage error of 6.00%.
- While reasonably accurate, the model shows a slight discrepancy compared to the actual value of CAD 98,241, which is expected in predictive modeling.

4 Discussion and Results

The analysis of Canadian consumer spending over the past 12 years highlights steady increases in housing and food expenditures, reflecting their essential role in household budgets. In contrast, transportation and discretionary spending showed greater variability, particularly during the pandemic. Regional differences also emerged, with Alberta and Saskatchewan experiencing spending volatility due to industry cycles, while British Columbia and Ontario displayed steadier trends. Key factors like the Consumer Price Index (CPI), disposable income, household type, age, and income level significantly shaped consumer behavior.

For demand forecasting, Linear Regression demonstrated the highest accuracy for long-term predictions, while XGBoost excelled in minimizing relative errors for short-term forecasts. However, the analysis is limited by data constraints, including the exclusion of factors such as consumer confidence, technological advancements, and government policies, which may also influence spending patterns. Additionally, some data imputation methods and assumptions, such as replacing missing values with placeholders, may have impacted model precision. Integrating real-time data and addressing these limitations can help businesses refine strategies further, allocate resources efficiently, and adapt to evolving consumer needs with greater accuracy.

5 References

Government of Canada. (n.d.). *Detailed household income*. Retrieved September 18, 2024, from https://search.open.canada.ca/data/?sort=metadata_modified+desc&search_text=detailed+household+income&page=1&keywords_en=household+spending+and+savings