Understanding the Housing Crisis in Nova Scotia (with comparisons to the rest of Canada)

# Objective

The growing housing crisis in Canada has been making news headlines for almost a year. The cost of buying a house has skyrocketed, rental rates are soaring, and there is low availability of both, leading to increasing numbers of unhoused people. As governments attempt to address the crisis, the question of what Canadians can afford to pay for housing becomes very important, especially in Nova Scotia, where income taxes are among the highest in the country while minimum wage is the lowest.

This project will examine the housing crisis, comparing incomes to housing costs across the country, and then diving deeper by comparing this to the housing market (i.e. housing supply) in Nova Scotia in particular.

# Data Sources

The data sourced for this project comes from two (external) open sources: Statistics Canada and the Canada Mortgage and Housing Corporation (CMHC). Statistics Canada is the Canadian federal government's statistical office. The Canadian Income Survey (CIS) is conducted by Statistics Canada. CMHC is a crown corporation (a government organization structured like an independent company). CMHC is the national housing agency. More information can be found at the organizations' websites: <https://www.statcan.gc.ca/en/start> and <https://www.cmhc-schl.gc.ca/en/about-us>. Two datasets will be used for the analysis. These datasets are summarized in Table 1 and more detailed data profiles are found in the following section.

Table : DataSet Overview

|  |  |  |
| --- | --- | --- |
| Original Table Name | CIS2018 | Housing-market-indicators-nova-scotia-1990-2016 |
| Simple Description | Canada Income Survey 2018 | Nova Scotia housing market overview |
| Data Source | Stats Can | CMHC |
| Data Scope | Canada | Nova Scotia |
| Time Range | 2018 | 1990-2016 |
| Geographic Categories | provinces | None |
| Number of Rows in cleaned data | 94336 | 27 |
| Number of Columns in cleaned data | 19 | 35 |
| Number of Continuous Variables | 5 | 35 |
| Number of Categorical Variables | 28 | 0 |

# Data Cleaning

##### Canadian Income Survey 2018

The original CIS 2018 data file had 94,336 rows and 192 columns. "Valid Skip" was used to indicate entries where the survey question was not applicable to the respondent, based on their answer to a previous question. Cleaning included the following steps:

* removing the year column, since this dataset is for a single year
* creating a subset with only the relevant columns
* renaming columns for clarity and consistent formatting
* replacing "Valid Skip" with NaN and converting the following object columns to numerical
  + after-tax income
  + monthly mortgage payment
  + monthly condo fee
  + monthly rent.

No mixed-type columns were identified in the dataframe. The only missing values identified in the data set were the responses that were a "valid skip." The only action taken for these was to replace 'valid skip' with NaN. The final subset contains 94336 rows and 24 columns.

##### Nova Scotia Housing Market Indicators 1990-2016

The original data file was a formatted Excel spreadsheet. Some formatting was required in Excel to convert it to a csv file, namely consolidating headings and removing footnote rows. The import shape of the file was 35 rows and 28 columns. Cleaning involved the following steps:

* transposing the data
* renaming columns for clarity and consistent formatting
* changing the data types for all columns from object to float.

Some columns contained missing data, however preliminary examination of the file in Excel indicated that not all variables were measured every year. Since this accounts for the missing values, no action was taken.

# Data Profiles

##### Canada Income Survey 2018

The CIS is collected by the Centre for Income and Socioeconomic Well-being Statistics within Statistics Canada. It is distributed to a subset of respondents at the same time as the Labour Force Survey (LFS). Data from the LFS is used to supplement the CIS questions, and tax returns are used to complete many of the quantitative fields. The CIS data includes a weighting factor which must be applied for the data to be representative of the Canadian population and not just this sample. The survey provides detailed information about the respondents' income sources, government benefits, household and housing, and poverty indicators (e.g. food security, core housing needs, low income). A subset of anonymized data is available as a PUMF (public use microdata file) with both a data dictionary and a user guide. The user guide is the source of most of the information for this profile.

The public use data for the 2020 Canadian Income Survey has not been released yet, but the public use data for the 2018 Canadian Income Survey was released January 5, 2022. While there is a lag, this is the most recent dataset available from Statistics Canada.

The data is organized by survey respondents (people), but does contain flags to extract data for entire households. This data was cleaned and imputed by Statistics Canada before being released. All personally identifying data was stripped before the data was released. Statistics Canada's data security is legislated by the Statistics Act. Privacy concerns have been addressed by Statistics Canada to ensure that the data is in compliance with confidentiality requirements of the Statistics Act. Among the measures taken, personally Identifiable Information (PII) has been removed, some data has been aggregated, quantitative data outliers, which could be PII, have been replaced, noise has been added to the data, and some data has been swapped.

The quality and trustworthiness of this data are high. The original data was released in separate micro- and macrodata files, but a consolidated copy was available from the Abacus Data Network (a collaborative data repository maintained by four university libraries in British Columbia. The data was downloaded in SPSS format, opened in SPSS software and exported to a csv for import to Python.

The CIS was chosen as a dataset was chosen because it has continuous and categorical variables related to income and housing for all of the provinces and territories in Canada. A subset of the PUMF data has been selected for analysis, and a data profile is found in Table 2.

File Citation: Statistics Canada, 2022, "CIS2018 SPSS sav.zip", Canadian Income Survey, 2018, https://hdl.handle.net/11272.1/AB2/G6T0LC/RAD63D, Abacus Data Network, V1

Dataset Citation: Statistics Canada, 2022, "Canadian Income Survey, 2018", https://hdl.handle.net/11272.1/AB2/G6T0LC, Abacus Data Network, V1, UNF:6:RlzI4LxHQ+ZRmY8Hn85cuw== [fileUNF]

Table : Data Profile for CIS 2018 Data Selected for Analysis

| Column  (description for non-obvious columns) | Type | | | Structured | | Time-variant (V)  or invariant (I) |
| --- | --- | --- | --- | --- | --- | --- |
| household id  (identifier for the household respondent belongs to) | qualitative | ordinal | structured | | I | | |
| person id  ( identifier for the individual) | qualitative | ordinal | structured | | I | | |
| sample weight  (weight to apply sample value to entire population) | quantitative | continuous | structured | | I | | |
| Province | qualitative | nominal | structured | | V | | |
| age group  (as of December 31, 2018) | qualitative | ordinal | structured | | V | | |
| Gender | qualitative | binary | structured | | V | | |
| marital status | qualitative | nominal | structured | | V | | |
| after tax income | quantitative | continuous | structured | | V | | |
| household size | qualitative | nominal | structured | | V | | |
| household composition | qualitative | nominal | structured | | V | | |
| household major income flag  (flags the major income earner of the household) | qualitative | binary | structured | | V | | |
| after tax low income flag  (flags households earning less than low income cutoff) | qualitative | binary | structured | | V | | |
| before tax low income flag  (flags households earning less than low income cutoff) | qualitative | binary | structured | | V | | |
| after tax below lim flag  (flags households below low-income measure threshold) | qualitative | binary | structured | | V | | |
| disposable income below mbm flag  (flags households that can't afford market basket measure) | qualitative | binary | structured | | V | | |
| ownership of dwelling  (owner vs renter) | qualitative | binary | structured | | V | | |
| monthly mortgage payment  (excludes property taxes) | quantitative | continuous | structured | | V | | |
| monthly condo fee | quantitative | continuous | structured | | V | | |
| monthly rent | quantitative | continuous | structured | | V | | |
| rent subsidy flag  (flags rent subsidized by government, employer, or relative) | qualitative | nominal | structured | | V | | |
| core housing need indicator  (flags households with unsuitable, inadequate, or unaffordable housing) | qualitative | nominal | structured | | V | | |
| adult food security status | qualitative | nominal | structured | | V | | |
| child food security status | qualitative | nominal | structured | | V | | |
| household food security status | qualitative | nominal | structured | | V | | |

##### Nova Scotia Housing Market Indicators (1990-2016)

The housing market indicators data is compiled by the CMHC using a combination of surveys they have conducted and data from both the Bank of Canada, and Statistics Canada. The data table was last updated in July 2017, and contains historic data for every year between 1990 and 2016. While this data is two years behind the CIS data sourced for this analysis, time-series analysis and forecasting will be possible with the housing market indicators data. This data is considered trustworthy, since the CMHC is a crown corporation that represents the government.

The housing market indicators table includes survey data from the CMHC's starts and completions survey, market absorption survey, rental market survey, seniors' housing survey. Mortgage data is from the Bank of Canada, and socioeconomic data is from Statistics Canada.

This dataset was chosen to look specifically at housing supply and costs in Nova Scotia. By comparing the forecast from this data with the 2018 CIS data, we can look at what housing affordability looked like to Canadians in 2018. The first column contains the specific housing market indicator (e.g. housing starts, housing completed, housing supply) followed by columns for each year of data. The units for each row are included with the indicator description in the first column. The data was transposed as part of cleaning.

Citation: Canada Mortgage and Housing Corporation (CMHC), Housing Market Indicators for Nova Scotia, 1990-2016, accessed on June 16, 2022. This information is reproduced and distributed on an “as is” basis with the permission of CMHC.

Table : Data Profile for CMHC Nova Scotia Housing Market Indicators (1990-2016)

| Column | Type | | Structured | Time-variant (V)  or invariant (I) |
| --- | --- | --- | --- | --- |
| dwelling | quantitative | continuous | structured | I |
| Total starts | quantitative | continuous | structured | I |
| Single detached starts | quantitative | continuous | structured | I |
| multiple starts | quantitative | continuous | structured | I |
| semi detached starts | quantitative | continuous | structured | I |
| row starts | quantitative | continuous | structured | I |
| apartment starts | quantitative | continuous | structured | I |
| starts by market | quantitative | continuous | structured | I |
| freehold homeownership starts | quantitative | continuous | structured | I |
| rental starts | quantitative | continuous | structured | I |
| condominium ownership starts | quantitative | continuous | structured | I |
| co op and unknown starts | quantitative | continuous | structured | I |
| total completions | quantitative | continuous | structured | I |
| residential building permits | quantitative | continuous | structured | I |
| res bldg permit value thousands | quantitative | continuous | structured | I |
| new completed and unabsorbed home supply | quantitative | continuous | structured | I |
| single and semi detached supply | quantitative | continuous | structured | I |
| row and apartment supply | quantitative | continuous | structured | I |
| rental vacancy percent | quantitative | continuous | structured | I |
| rental availabilty rate percent | quantitative | continuous | structured | I |
| standard space seniors vacancy rate | quantitative | continuous | structured | I |
| pct change new housing price index | quantitative | continuous | structured | I |
| pct change consumer price index | quantitative | continuous | structured | I |
| pct change construction wage rate index | quantitative | continuous | structured | I |
| pct change owned accommodation costs | quantitative | continuous | structured | I |
| pct change rental accommodation costs | quantitative | continuous | structured | I |
| bachelor average rent dollars | quantitative | continuous | structured | I |
| one bedroom average rent | quantitative | continuous | structured | I |
| two bed average rent dollars | quantitative | continuous | structured | I |
| three plus bed average rent dollars | quantitative | continuous | structured | I |
| july 1 population thousands | quantitative | continuous | structured | I |
| labour force participation rate percent | quantitative | continuous | structured | I |
| pct change employment | quantitative | continuous | structured | I |
| pct change unemployment | quantitative | continuous | structured | I |
| pct change real disposable income | quantitative | continuous | structured | I |
| net migration | quantitative | continuous | structured | I |

# Limitations and ethical considerations

No PII exists in either dataset sourced for this project. The CIS 2018 had some data on immigration; however this data was not included in the subset for analysis in order to avoid cultural bias in the interpretation.

The most recent year included in either dataset is 2018. The data will not give a picture of the current housing crisis, but it may show how it developed. More recent data may have been included in reports from Statistics Canada and the CMHC, even if the data is not openly available.

Although the CIS is survey data, income data reported in the CIS has been checked against respondents' tax returns, so the accuracy of the income data will be better than the accuracy of the housing costs, which were self-reported.

No detailed information about the data collection and cleaning methods used for the Housing Market Indicators Data is available.

# Preliminary descriptive Statistics

Since the Housing Market Data is all time series data, no preliminary statistics have been included here; however they were checked as a quality control tool during data cleaning. Preliminary descriptive statistics from the 2018 CIS are included in Table 4.

Table : Preliminary Descriptive Statistics from the 2018 Canadian Income Survey

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Sample weight (not yet applied) | After tax income (CAD) | Household size | Monthly mortgage payment | Monthly condo fee | Monthly rent |
| count | 94336 | 77148 | 94336 | 45296 | 3818 | 22157 |
| mean | 385 | 40811 | 3 | 1309 | 370 | 994 |
| std | 435 | 34574 | 1 | 730 | 237 | 513 |
| min | 10 | -133295 | 1 | 0 | 0 | 0 |
| 25% | 126 | 19025 | 2 | 800 | 250 | 700 |
| 50% | 212 | 34355 | 3 | 1200 | 300 | 900 |
| 75% | 463 | 54435 | 4 | 1700 | 450 | 1300 |
| max | 4809 | 922020 | 7 | 4500 | 1600 | 3200 |

# Key Questions to Explore

* Where do Canadians live?
* Where do low-income Canadians live?
* What is the income distribution in Canada (i.e. is there skew)?
* What is the income distribution in Nova Scotia?
* What is the proportion of low-income homeowners to renters?
* What is the proportion of non-low-income homeowners to renters?
* How many Canadians were paying more than 30-35% of their after-tax income for housing?
* How did the Nova Scotia housing market change between 1990 and 2016?
* Does the Nova Scotia housing supply keep pace with population?
* Does the forecast of the Housing Market Indicators Data match currently reported (i.e. in reports, not open data sets) values? If not, what changed?