April 6, 2023

Jeff Sit

Seneca College

BDB600

Business Intelligence 2

final project report

BDM500 Project MS1

**Decide and describe the problem to be solved/analyzed**:   
To investigate an issue regarding the housing market crisis in Canadian society. After researching for datasets to use, the problem I decided I will be analyzing is the housing and living situation in Canadian society. Specifically, I wish to know how much money the typical Canadian household makes and spends, and most importantly, how these numbers have changed over time. These figures will be compared to assess the difference in wages and expenses and how they have changed over time.

**Data and data sources are available and defined:**There are 2 primary datasets I will be using plus a third one I may use if necessary.  
First, the household spending dataset which contains data on the average amount of money a Canadian household spends throughout the year and how it is spent.  
Second, the household income dataset which contains the average weekly income per household depending on the industry they work in.  
Finally, there is the consumer price index dataset which provides data in the form of a reference point with regards to how much the same product/service has adjusted as a percentage.

**Business/challenge questions are clearly defined:**

1. For the amount of money each industry is making, are there any industries that make enough money to pay for the average household expenditure of that same year? Explain why or why this might not be the case.
2. Comparing 2010 and 2019, on average how much more money is being made in income for every industry? How might this answer change if inflation were to be considered?

**Solution strategy described:**The 2 datasets I will be using cover different timeframes. Therefore, I will be comparing the period of 2010 to 2019 as that is common across all datasets. Through these changes in time, I will be assessing the degree of impact using spending factors like food, transportation, or education, while also considering income factors like one’s field of occupation. This will all be done to determine what the same product/service today cost before some years ago. A similar process will follow for the income dataset. If possible, I also want to see what can be done about minimizing spending while maximizing income.

BDM500 Project MS2

**Data preprocessing completed:**

1. Done in Jupyter Notebook and in Power BI (just in case)
2. In the case of the spending dataset, there is missing data in the tobacco categories due to how before 2019, the category was slightly narrower in scope, only factoring tobacco, and alcohol. This category would eventually be succeeded by the tobacco, alcohol, and non-medicinal cannabis category. To streamline this data preprocessing stage, the results for these 2 separate categories will be merged into a single column. There is also a gap between 2017 and 2019 since spending data for 2018 is completely missing. Thus, 2018 will be avoided when attempting to look for comparisons.
3. For the income by industry dataset, there is missing data in a few categories here. However, the missing data is found in years that are outside the scope of the comparison outlined for this project (which is 2010 to 2019). Therefore, as these rows are for the early 2000s or towards 2020 and beyond, they can be dropped.

**Solution alternatives evaluated:**

1. Given that the main problem to be studied involves analyzing the typical Canadian household’s income and spending practices, it goes without saying that I will need to at minimum study these 2 datasets along the same period. However, as there are over 20 columns for the industry income dataset and close to twice that number of columns for the household spending dataset at nearly 40 columns, clearly there are more specific ways I can achieve this:
   1. One option would be to home in on a specific industry or two like construction or healthcare, and then to compare these industries head-to-head. This can be done with extremely similar or different professions with similarity measured by education required, level of income, or some other way they could be comparable.
   2. A second option would be to perform the same thing from above but on the household spending dataset. Here, I would isolate a cluster of categories for a type of spending and assess how that cumulative expenditure has changed. For example, looking at the essential categories such as grocery costs, rent, and private/public transit will likely shed some insight on what is heavily contributing to the increase in spending.

**Data visualization techniques applied:**

1. Examples of both solutions from the previous section were visualized using different techniques: 1 bar chart, 1 line chart. Both can be found below and in the Power BI file.

Chart, bar chart

Description automatically generatedChart, line chart

Description automatically generated

**Business/challenge questions reviewed or refined:**

1. Initial question: For the amount of money each industry is making, are there any industries that make enough money to pay for the average household expenditure of that same year? Explain why this might or might not be the case.
   1. Revised question: Assume for each industry’s weekly earnings, that they are salaried and are paid for every week in the year (52 weeks). With this assumption, are there any industries that make enough money to pay for the average household expenditure of that same year? Explain why this might or might not be the case.
   2. This needed to be clarified since the industry income dataset displayed all values in weekly amounts earned while the household spending dataset dealt in yearly costs. Therefore, a standard method to convert the weekly amount to a yearly amount needed to be implemented and detailed in the question.
2. Initial question: Comparing 2010 and 2019, on average how much more money is being made in income for every industry? How might this answer change if inflation were to be considered?
   1. Revised question: Comparing 2010 and 2019, on average how much more money is being made in income for every industry? Does your answer appear drastically different when illustrated as a percentage of the base income? How might this answer change if inflation were to be considered?
   2. For the most part, this question is still the same, but I wanted to add the percentage component since I felt this could be a meaningful detail. This is because a 100-dollar increase may appear insignificant for an industry that makes 1000 dollars weekly but could be very impactful for jobs that make around 300 dollars weekly.

BDM500 Project MS3

**Chosen solution alternative applied:**

1. Done in jupyter notebook.
2. The original plan was to study the typical Canadian household’s income and spending practices, meaning I would need to study the industry income dataset and average house spending dataset along the same period. With regards to only choosing one of the solution alternatives, I have decided that I will be focusing on the amount of money earned per occupational industry. Details surrounding the procedure and this decision can be found in the step below.

**Actions and processes to support implementation of alternative detailed:**

1. Given that the data for the income dataset was more complete with a larger range of years to work with, I felt that choosing this dataset was the right call. To begin comparing the columns of this dataset, I decided to isolate only a handful of columns based on 3 set cases that I felt would bring out the most meaningful information:
   1. First, I found the most well earning occupation compared against the sum of 3 of some of the lowest earning occupations and industries. This case would highlight how even a married couple working full time could be making less than a single working adult regardless of marital status, based on their occupation. This case compared (mining, quarrying, oil, and gas extraction) versus (food services), (arts, entertainment, recreation), and (administrative and support, waste management and remediation services).
   2. This second case is like the first one but dialed up to exasperate and display the full extent of this issue of income inequality. I picked the lowest weekly wage that made a 4-digit salary and pitted against 3 times the lowest weekly wage occupation. The result was goods produce industry versus accommodation and food services.
   3. Finally, the last case saw the 2 most similar weekly wage industries compared against each other. This was to test how despite nearly identical starts in weekly salary, how would different jobs affect wage changes throughout the years? This case is unique where the difference between the 2 columns is calculated then outputted as its own column. This comparison involved Finance and insurance versus Information and cultural industries, where a positive value indicates Finance is greater while a negative value means Information and cultural industries is greater.

**More data visualization used + Data is explored using different dimensions:**

1. Examples of each of the 3 cases (given this solution alternative) were visualized using different techniques: 1 column chart, 1 line chart, 1 area (line) chart. All 3 of these can be found below and in the Power BI file.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated Chart, waterfall chart

Description automatically generated

**Gap analysis detailed, described, and completed + Methods to reach target goal:**

|  |  |  |  |
| --- | --- | --- | --- |
| GOAL: Increase number of insights generated from income dataset | | | |
| Current State | Future State | Gap Desc. | Potential next steps |
| -3 insights generated from ~1/3 of available columns  -price index dataset is not used | -6 to 10 Insights generated from 70% to 100% of available columns  -streamline process to get insights | -difference of 2 to 3x as many insights  -too many ways to generate insights, leads to inconsistent results. | -use price index dataset for next set of insights  -ensure every column has at least 1 insight first |

BDM500 Project MS4

**Stakeholders involved in project listed, categorized and their roles in the project described:**

* Government
* Students
* Graduates/alumni
* Financial experts
* Business executives
* Parents/homeowners
* Investors
* Society at large

Stakeholders Table

|  |  |
| --- | --- |
| High influence + Low interest | High influence + High interest |
| * Business executives * Investors | * Financial experts * Government * Society at large |
| Low influence + Low interest | Low influence + High interest |
|  | * Students * Graduates/alumni * Parents/homeowners |

***Government:***

This group of stakeholders would be interested in the success of this project since they would want to know the various financial situations of the citizens they govern over. This is also a group of stakeholders with a high influence over the project since the government can always propose new laws/bills to massively alter the economic situation of the nation.

***Students:***

This group of stakeholders would want to know the financial constraints that come with certain career pathways. This group of stakeholders have little to no power/influence over the economic situation they reside in other than to vote for politicians that show promises of delivering the results they seek. The best they can do is excel in a field they are passionate about while hoping it pays an above average wage.

***Graduates/alumni:***

Similar group of stakeholders to “Students” since they are just more experienced students who have finished a post-secondary credential of some kind. Still have low influence since they possess no ability to change the economic landscape on their own. Highly interested in this project as it will give them a starting point for their wages depending on what sector of work they operate in.

***Financial experts:***

This group of stakeholders would likely be the most knowledgeable group on this subject. Naturally, they would be interested in this project since analyzing financial trends is part of their job description. Due to their extensive knowledge in business and finance, this group of stakeholders hold great influence over the project as they are the ones to make reports and articles about economics.

***Business executives:***

This group of stakeholders likely has little interest in the project since they likely make significantly more than average due to their higher positions. Therefore, the financial struggles the average person in any industry faces are not applicable to them. This is unfortunate as this is also a group of stakeholders with a high influence over the project since the private companies can choose how they manage their funds, including how they pay their employees lower along the company ladder.

***Parents/homeowners:***

This group of stakeholders would want to know how much they can expect to pay in expenses every year. This is especially pertinent if there are children being raised within the household as that would contribute to their expenses as well. Low influence in the project since the only method of affecting the economic situation is to vote for the politicians of their choice.

***Investors:***

Similar to “Business executives” group, where these stakeholders would be interested in the success of this project simply so that they could know where to invest their money to get the biggest return of money back. As they are uninterested in changing the economy in any way unless it benefits them, their overall interest level can be labelled as low.

***Society at large:***

As the largest group of stakeholders, society would want to see this project’s results to ensure their living standards are being maintained, if not increasing. Overall influence level would high simply for the fact that Canada is a democratic country where rulings are dependent on the society of people.

**Practices and innovations that ensure integrity and privacy of data described and detailed:**

1. Data access and editing permissions will be limited to ensure only the relevant parties that are authorized have access to the corresponding data.
2. Change logs will be regularly used across every sector to ensure any edits in the data can be easily recognized, traced, and backed up.
3. Data will be validated in multiple stages to ensure it is correct at all times such as when it is initially collected to when it is used or modified.
4. Updates to OS/software will be installed as soon as possible to prevent system interruptions or even data loss.

**Conclusion:**  
Therefore, it can be concluded that the amount of money earned from wages has been slowly outpaced by that of the total average amount of money expended per year. This finding is made clear given that even for the most well-paying industries of work, they still struggle to make an amount greater than that of the yearly average expenditure for the same year. Hence, the situation only gets bleaker as the pay per industry gets worse and worse, as evidenced when occupations were compared head-to-head.

The overarching reason for why studying this dataset was warranted in the first place hinged on the possibility of there being some hope that there would not be that staggering of a difference between a timeline of compared years. Considering that there does exist a massive difference in these yearly average expenditure values alone, it becomes apparent that there is a severe issue of economic inequality as a result of how the weekly wages differ even more between each other given the multitude of industries available to choose from.

Ultimately, the key takeaway from this analysis would be for people to pick fields of work that they not only find enjoyable, but also can make a reasonable amount of money from as well. The definition of reasonable will differ between individuals, since not everyone will have the same spending habits or be subject to the same initial living or financial conditions. Some people may have family or partners to share the expenses with, in which case it will be significantly easier to meet the total average expenditure value when split across two people of varying occupations.

This is not to say that jobs in the fast-food industry for example should be dismissed. In cases where a little bit of money can be made over being unemployed, at minimum, it is probably worth working temporarily to have at least a bit of income. However, there will come a point where society at large will need to band together to demand adequate pay for what is currently considered as low-level hourly work. Recently, increases in the minimum wage at a federal level have been implemented and they are a step in the right direction. So, while the government is making strides to remedy the situation, the private companies and businesses will need to do the same if the pace of wages is to catch up to the speed at which total expenditure per year has increased to.