ENSF 592
Final Project Report

Team #34, Two Sum

Abstract

Goal of the project is to provide and analyse Canadian economic data collected from 2011 to 2021. The targeted users for this program are those who are planning on relocation and job search. This program allows for targeted analysis based on province, year and industry as well as general statistics across the provinces of Canada.

Introduction

The concept for this project was derived from a conversation my partner and I had regarding economic outlooks for the information sector in Canada and what we perceived as an apparent lack of consolidated credible data, with a present reliance on subjective sources like payscale.com. The conversation culminated in the concept of our project, designed to engineer a source to consolidate and report economic trends across Canada.

Early stages of the project focused on identifying a target demographic who would likely interact with and leverage our solution. The framework was ultimately refined to two target demographics: young adults choosing potential careers of interest, and established professionals interested in relocation or divergence in career.

To further refine our concept we identified critical areas of interest for our selected target demographics. Fortunately, we, and a myriad of our peers and elders fall within this group. Based on our conversations and a thought analysis we ascertained our chosen demographics areas of interest when professionally planning. The consensus areas of interest were salaries, unemployment rates, sector size, average work hours and rates of change.

Task Distribution and Management

This program is broken into 2 sections:

- Data Handling
- User Interface and Execution

Isaiah Williams focussed on the Data Handling and David Cheng focussed on the User Input/Execution, however, much of the project was done in conjunction.

The outcome of the planning phase was to create small, modular functions that can be extended throughout the project. This will allow for high flexibility as new features can be added as modules to be called upon. Managing version control was completed through git where the sections were broken out into different branches to be merged into main upon completion of testing.

Data Handling

The data required to meet our key parameters was fortunately, readily available through statistics Canada (*Labour force characteristics by industry, 2022), (Employment by Industry Annual*, Statistics Canada, 2022), (*Total compensation per job, by NAICS industry, 2022)*. This source was chosen due to its inherent credibility, abundance of relevant information and structural linearity across reports; three datasets were collected altogether.

However, through closer inspection both online and in the extracted back end csv data, the information within the tables did not align easily for connective mapping, *Figure 1*. The primary challenge when using the datasets was the non-standardised nature of the "North American Industry Classification System (NAICS)" column, which was critical to providing sector related insights and to our eventual hierarchical indexing. Additionally, it was identified that the collected reports contained no information regarding the territories, as a result, the report specifications were further refined to an analysis of solely provinces.

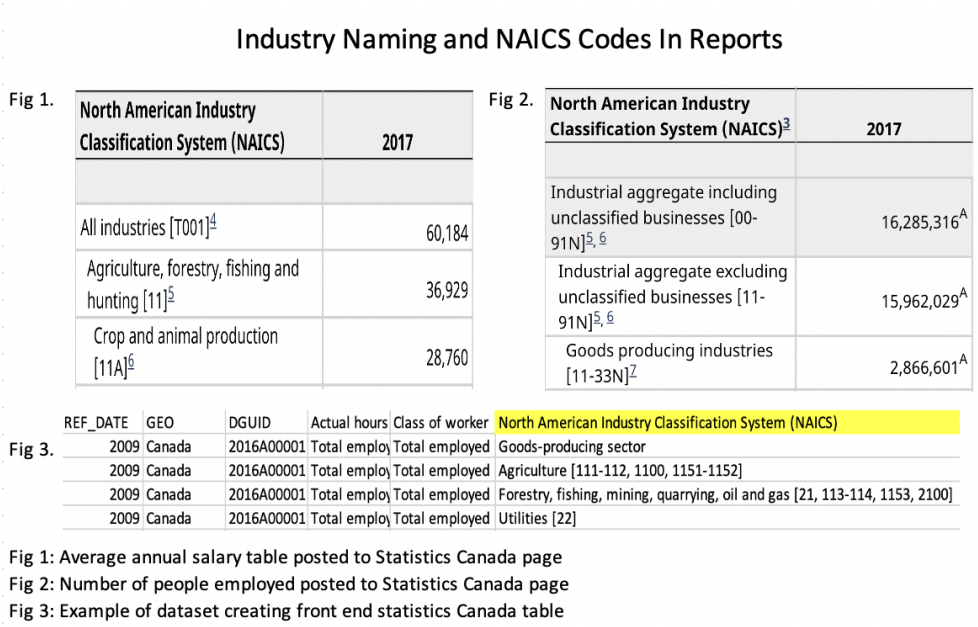


Figure 1: Displaying variance in industry referencing between online representation, background data csv and discrepancy in convention between sources

As displayed in Figure 1, the variable nature of the NAICS industry naming created a quandary as to reporting capability. The primary barriers to success were: inconsistent naming and numbering conventions, with some numbers appearing as single digits [11], others as ranges [44-45] and further rows with no code references altogether.

To resolve this issue, research was conducted into the origin of the NAICS Code information referenced across the tables. Through this analysis a coding structure was identified on the United States Census Bureau webpage, *Figure 2*.

The organization of NAICS is as follows:

- **Sector**: 2-digit code
  - **Subsector**: 3-digit code
    - **Industry Group**: 4-digit code
      - **NAICS Industry**: 5-digit code
        - **National Industry**: 6-digit code

Note: Three sectors are represented by a range of 2-digit codes. These include Manufacturing (31-33), Retail Trade (44-45) and Transportation and Warehousing (48-49).

Figure 2: NAICS coding hierarchy structure

The codes follow a co-opted standard for Canada and the United States regarding the classification of industries through a hierarchical structure. Due to the scope of the project and the challenge of reporting to all levels simultaneously, we refined our focus to reporting statistics at the “Sector” level, two digit codes.

In parallel to the identification of a structure, a data dictionary was extracted from the United States Census Bureau adding the final component for joining and linearization of industry naming across datasets (*2022 NAICS*, North American Industry Classification System, 2022).

The hierarchy chosen for the data was based on province, sector and year. However, as the “North American Industry Classification System (NAICS).xlsx” extract lacked reference to year and province, a cross join was performed with a list of the provinces excluding territories, and subsequently the 10 years in scope of the report. With the addition of this information the NAICS excel sheet would function as an anchor for merging purposes containing all variations of sector, year, and province.

Concurrently, the extracts pertaining to employment information were filtered for pertinent information and had their respective “North American Industry Classification System (NAICS)” columns stripped for NAICS code referencing using regex. Subsequently, duplicate entries and rows with null values in key columns were removed from each of the tables. Additionally, any rows which did not contain a code reference were dropped from the table and all undefined range codings were converted to their respective arrays of values: ex: 22-25, becoming, 22, 23, 24, 25; while, defined ranges ex: 44-45 (retail), remained the same. Finally internal cross joins were used to align each row's extracted list of “Sector” codes with its own line item, as information on each row related to each “Sector” listed.

Lastly, hierarchical joining was conducted using pandas merge functions based on province, sector and year, linking the three economic datasets to the NAICS lookup sheet serving as anchor of connectivity.

Leveraging the finalised dataset, calculations were performed on the columns to infer additional information regarding, hourly wage, number of people unemployed and total sector size.

To convey findings to the end user tables were constructed using the pivot table function *(Figure 3)* as well as, the groupby function in congruence with aggregate, apply and lambda functions *(Figure 4)*. The tables were each designed to address the key component criteria expressed during the information gathering stage of the report. NaN values were not extracted from output tables as it was deemed valuable for users to identify where information was not available in the report to prevent the potential for abstracted conclusions regarding absences.

```
    print(
        "\n\nThis is a pivot table outlining stats for job demand based on your selections:"
    )
    pvt1 = data_subset.pivot_table(
        ["Employed", "Unemployed",  "Total Sector Job Demand",],
        index=["Provinces", "Job_Title"],
        columns="Years",
        aggfunc=sum,
    )

    print(pvt1.replace([0], np.nan))
```

Figure 3: Example code used to create pivot table

```
    # Displaying relative salary for each sector and province comparative to the yearly average salary for that sector
    # Average salary across Canada for user selected sector by year
    selected_province_salary = (
        user_dataframe.groupby(["Provinces", "Job_Title", "Years"], as_index=True)
        .agg({"Salary": "mean"})
        .reset_index(drop=False)
    )
    # Average salary by year for user selected sector and province
    salary_across_canada = (
        collective_dataset.groupby(["Job_Title", "Years"], as_index=True)
        .agg({"Salary": "mean"})
        .reset_index(drop=False)
    )
    # Joining all information
    group = pd.merge(
        selected_province_salary,
        salary_across_canada,
        left_on=["Job_Title", "Years"],
        right_on=["Job_Title", "Years"],
        how="left",
    )
    # Normalizing average salary by year for user selected sector and province by the standard salary for the sector across all provinces
    group["Centered Salary"] = group["Salary_x"] - group["Salary_y"]
    group = group.drop(columns=["Salary_x", "Salary_y"]) # Dropping frivelous columns
    centered_salary_data = group.copy() # Creating a copy of data to be used in following table
    group = group.reset_index(drop=True) # Removing default numerical index
    group = group.set_index(["Provinces", "Job_Title", "Years"]).unstack(2) # Creating a hierarchial index
    group["Net mean"] = group.mean(axis=1) # Creating row total mean
    group = format_currency(group) # Changing table data into currency format

    print(
        "This table displays the salaries for provinces and sectors based on your selection, relative to the average for the sector split by year:"
    )
    print(group, end="\n\n") # Printing data
    print("\n\n")
```

Figure 4: Example code used to create centred data table using groupby function

Salaries were addressed through a variety of reports both at macro and micro scales. In general, tables pertaining to salary provided a comparison of industries to one another and expected earning potentials by province. Additionally, comparison tables were created to trend salaries within a sector against the standard rate across Canada, and year over year rate of change comparison charts *(Figure 5)*. Finally hours worked was weighed against

expected hourly wage to identify the industries with the highest pay off in respect to time investment.

This table displays the salaries for provinces and sectors based on your selection, relative to the average for the sector split by year:

		Centered Salary											Net mean
Years		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
Provinces	Job_Title												
Alberta	Information	$12,230.70	$12,409.80	$16,909.70	$20,761.30	$16,803.30	$11,115.50	$8,626.50	$7,239.50	$7,802.00	$6,860.90	$5,451.00	$11,473.65
British Columbia	Information	$2,185.70	$512.80	$206.70	$-814.70	$1,500.30	$1,974.50	$-691.50	$980.50	$550.00	$3,635.90	$1,056.00	$1,008.75
Manitoba	Information	$-5,019.30	$-5,023.20	$-4,317.30	$-5,901.70	$-7,053.70	$-7,374.50	$-4,398.50	$-3,961.50	$-7,725.00	$-10,562.10	$-9,895.00	$-6,474.89
New Brunswick	Information	$116.70	$-2,312.20	$-5,594.30	$-7,497.70	$-6,901.70	$-7,328.50	$-5,184.50	$-5,777.50	$-3,622.00	$-6,986.10	$-6,509.00	$-5,236.07
Newfoundland and Labrador	Information	$7,129.70	$10,923.80	$17,751.70	$15,502.30	$18,126.30	$17,049.50	$15,024.50	$12,499.50	$7,315.00	$11,788.90	$11,795.00	$13,173.29
Nova Scotia	Information	$-1,815.30	$-3,728.20	$-5,024.30	$-5,137.70	$-5,459.70	$-7,339.50	$-2,956.50	$-2,916.50	$-996.00	$1,847.90	$-1,305.00	$-3,166.44
Ontario	Information	$3,766.70	$4,396.80	$3,089.70	$4,748.30	$4,798.30	$6,930.50	$8,896.50	$13,233.50	$16,232.00	$19,609.90	$23,896.00	$9,963.47
Prince Edward Island	Information	$-11,686.30	$-11,894.20	$-14,148.30	$-14,354.70	$-14,154.70	$-11,584.50	$-17,064.50	$-19,786.50	$-20,439.00	$-24,584.10	$-23,557.00	$-16,659.44
Quebec	Information	$-4,203.30	$-4,429.20	$-4,226.30	$-3,952.70	$-5,008.70	$-950.50	$-646.50	$302.50	$2,877.00	$3,632.90	$4,706.00	$-1,081.71
Saskatchewan	Information	$-2,705.30	$-856.20	$-4,647.30	$-3,352.70	$-2,649.70	$-2,492.50	$-1,613.50	$-1,813.50	$-1,994.00	$-5,244.10	$-5,638.00	$-3,000.62

This table displays the best and worst salary disparities, relative to the average for the sector by year:

Provinces	Job_Title	Years	min_max	
Alberta	Information	2014	max	$20,761.30
		2021	min	$5,451.00
British Columbia	Information	2014	min	$-814.70
		2020	max	$3,635.90
Manitoba	Information	2018	max	$-3,961.50
		2020	min	$-10,562.10
New Brunswick	Information	2011	max	$116.70
		2014	min	$-7,497.70
Newfoundland and Labrador	Information	2011	min	$7,129.70
		2015	max	$18,126.30
Nova Scotia	Information	2016	min	$-7,339.50
		2020	max	$1,847.90
Ontario	Information	2013	min	$3,089.70
		2021	max	$23,896.00
Prince Edward Island	Information	2016	max	$-11,584.50
		2020	min	$-24,584.10
Quebec	Information	2015	min	$-5,008.70
		2021	max	$4,706.00
Saskatchewan	Information	2012	max	$-856.20
		2021	min	$-5,638.00

Name: Centered Salary, dtype: object

Figure 5: Example of salary table displaying salaries within a sector against the standard rate across Canada

Sector size and employment information was conveyed through a pivot table outlining the number of people employed, unemployed and the total sector size within a given province. As well as a variety of tables outlining unemployment rates across Canada, and unemployment rates of change *(Figure 6)*.

10 provinces with highest unemplyment rate changes by sectors and year

Years		2011–2012	2012–2013	2013–2014	2014–2015	2015–2016	2016–2017	2017–2018	2018–2019	2019–2020	2020–2021	Overall Trend
Provinces	Job_Title											
New Brunswick	Information	18.83%	-8.20%	59.13%	-40.90%	1.27%	5.00%	-32.14%	1.75%	139.08%	14.66%	15.85%
Alberta	Information	0.54%	2.16%	-44.44%	14.29%	104.17%	13.47%	-31.47%	33.86%	75.59%	-28.08%	14.01%
Manitoba	Information	-11.65%	1.37%	13.82%	3.81%	-8.49%	-13.28%	58.67%	-39.53%	153.01%	-32.38%	12.54%
Saskatchewan	Information	4.17%	-22.00%	10.26%	-13.95%	32.43%	71.43%	-10.12%	6.62%	86.75%	-50.78%	11.48%
Ontario	Information	-13.85%	5.10%	-13.81%	0.87%	-3.43%	-4.00%	3.70%	6.25%	168.07%	-45.92%	10.30%
Nova Scotia	Information	-13.62%	18.08%	1.95%	6.71%	-45.66%	98.90%	-34.90%	24.15%	38.47%	4.89%	9.90%
British Columbia	Information	-20.80%	13.97%	-33.82%	32.59%	-25.14%	-21.64%	36.19%	18.88%	142.65%	-44.91%	9.80%
Newfoundland and Labrador	Information	-11.43%	7.83%	2.56%	-24.17%	2.20%	21.51%	22.12%	12.32%	107.96%	-60.29%	8.06%
Prince Edward Island	Information	9.58%	38.99%	-46.16%	23.83%	0.38%	22.56%	-23.31%	-4.00%	103.33%	-57.79%	6.74%
Quebec	Information	4.46%	-20.23%	1.43%	-3.87%	-0.00%	-12.82%	-10.92%	22.64%	124.23%	-48.37%	5.65%

Figure 6: Example table displaying the rate of change of unemployment rates yearly. Sorted by highest to lowest average unemployment rates by province over the last 10 years

User Interface/Output

This program allows the user to interact with both a command line interface and graphical user interface. For the purposes of the project grading, **only the command line interface is to be considered.** Through the command line interface, the user will be guided through the phases of data import, general statistics, user selected statistics and plot display.

The user first initiates the program through terminal by calling “python Final_Project.py” and selects the preferred interface by either entering ‘cli’ or ‘gui’ into the command prompt to activate the different interfaces.

```
(base) ➜  final-project-two-sum git:(main) ✗ python Final_Project.py


Enter 'cli' or 'gui' to enter program from command line interface or graphical user interface
```

Figure 7: Program script initiation from terminal

```
if __name__ == "__main__":
    while True:
        try:
            user_input = input(
                "\n\nEnter 'cli' or 'gui' to enter program from command line interface or graphical user interface\n"
            )
            if user_input not in ["cli", "gui"]:
                raise ValueError(f"Please try again")
            else:
                if user_input == "cli":
                    CLI()
                    break
                else:
                    webbrowser.open("http://127.0.0.1:5000/")
                    gui.app.run()
                    break

        except ValueError:
            print(f"Input not valid.\n")
            continue
```

Figure 8: Code snippet of python script logic

Command Line Interface

Data Import

The user is prompted to enter ‘merge’ to start the merge sequence of the raw excel files or the user can continue without merging. This option was introduced to leverage the required excel file of the final dataframe in the project folder which can act as a cache. If this program was to go into production, the underlying data would only need to be updated annually so there would not be a need to duplicate the merge each time the program starts up. If the user chooses to merge, Excel file would be updated in the project folder or added if it does not exist.

```
(base) ➜  final-project-two-sum git:(main) ✗ python Final_Project.py


Enter 'cli' or 'gui' to enter program from command line interface or graphical user interface
cli

******************************************* CANADA ECONOMIC TRENDS *******************************************


Enter 'merge' to get the most current information or any other character to get cached storage
```

Figure 9: Program initiation of data import

General Statistics

The user is prompted to enter 'general' to start the general statistics of economic data across Canada.

```
********************************************* IMPORT END *********************************************

Enter 'general' for general information or any other character to continue to selection
general

******************************************** GENERAL STATS ********************************************

*********** Below are an array of tables to assist you in your market analysis ***********

The overall metrics for the datasets are:
        Years     Salary   Employed  Standard_Hours  Unemployment_Rate  Earnings per hour  Total Sector Job Demand  Unemployed
count  2200.00    1870.00    1870.00         1429.00            1671.00            1429.00                  1561.00     1561.00
mean   2016.00   59837.72   95267.76           36.42               7.12              32.17                116296.43     6089.54
std       3.16   27046.52  145955.75            5.22               5.10              12.45                162734.61     9264.91
min    2011.00   19087.00      35.00           25.10               1.10              12.30                    45.01       10.01
25%    2013.00   37502.75    9273.75           32.60               3.55              21.50                 16667.91      851.25
50%    2016.00   56245.00   29835.00           35.00               5.70              31.66                 41557.29     2313.32
75%    2019.00   72140.00  114265.00           39.60               9.20              38.80                153289.47     6760.23
max    2021.00  162465.00  787390.00           60.60              33.95              75.98                819770.95    71870.37



The 10 top salarys by province for the last 10 years are:
                               Salary                                                                                                                     Sum Total
Years                            2011        2012        2013        2014        2015        2016        2017        2018        2019        2020        2021
Provinces
Alberta                    $64,546.88  $68,579.65  $72,168.29  $76,219.06  $77,161.53  $72,791.00  $73,154.59  $73,482.65  $74,137.00  $77,058.53  $75,729.06  $805,028.24
Newfoundland and Labrador  $56,531.47  $59,070.35  $63,199.71  $66,753.65  $69,631.82  $69,634.24  $71,807.65  $71,014.59  $69,777.00  $76,216.82  $75,692.18  $749,329.47
Ontario                    $55,225.82  $57,039.53  $57,464.71  $59,910.18  $62,246.12  $62,409.76  $63,923.94  $66,746.41  $67,927.41  $73,158.94  $73,346.29  $699,399.12
British Columbia           $55,453.29  $55,766.82  $57,848.35  $58,694.41  $60,979.41  $60,899.76  $62,146.12  $64,596.71  $66,194.82  $72,375.24  $72,721.24  $687,676.18
Saskatchewan               $54,600.18  $57,810.29  $58,703.00  $61,263.82  $62,425.35  $60,871.94  $61,107.94  $62,079.71  $60,891.82  $63,680.12  $64,285.59  $667,719.76
Quebec                     $49,439.18  $51,853.59  $52,764.41  $54,522.65  $55,075.53  $56,094.88  $57,031.88  $59,845.06  $62,201.24  $67,201.06  $68,482.53  $634,512.00
Manitoba                   $49,337.47  $51,168.53  $53,205.65  $54,801.82  $56,528.12  $57,081.12  $59,031.18  $60,218.47  $60,390.00  $62,891.65  $64,512.53  $629,166.53
Nova Scotia                $47,520.94  $48,060.24  $49,742.65  $52,593.82  $54,924.53  $54,670.71  $56,587.12  $57,371.18  $58,190.47  $63,885.53  $63,266.12  $606,813.29
New Brunswick              $50,283.82  $50,227.24  $50,252.41  $51,501.88  $53,854.59  $54,875.71  $54,846.06  $58,209.29  $59,755.76  $60,527.71  $61,134.47  $605,468.94
Prince Edward Island       $40,071.59  $41,107.71  $41,727.82  $43,171.59  $44,278.00  $46,027.82  $45,703.76  $45,972.35  $47,009.82  $50,225.71  $51,739.94  $497,036.12



The 10 top salarys for all sectors based on province from highest to lowest are:
                                                   Salary                                                                                                                                              Sum Total
Provinces                                         Alberta British Columbia     Manitoba New Brunswick Newfoundland and Labrador  Nova Scotia      Ontario Prince Edward Island       Quebec Saskatchewan
Job_Title
Utilities                                     $142,573.64      $118,192.64  $115,581.73   $116,691.64               $112,520.45  $103,737.36  $116,779.00           $67,799.55  $104,015.00  $108,724.00  $1,106,615.00
Mining, Quarrying, and Oil and Gas Extraction $150,618.64      $120,487.73  $102,739.36    $85,860.73               $141,859.00  $108,277.00  $118,475.36           $47,194.09  $105,845.45  $122,924.45  $1,104,281.82
Finance and Insurance                          $88,691.91       $80,843.73   $80,665.27    $64,986.27                $78,698.91   $73,931.18   $86,573.09           $64,525.73   $75,481.73   $73,021.91    $767,419.73
Information                                    $85,340.45       $74,875.55   $67,391.91    $68,630.73                $87,040.09   $70,700.36   $83,830.27           $57,207.36   $72,785.09   $70,866.18    $738,668.00
Manufacturing                                  $89,010.91       $72,448.18   $65,554.82    $57,339.73                $80,559.82   $63,238.55   $80,758.45           $54,516.36   $73,007.09   $73,380.18    $709,814.09
Construction                                   $90,358.00       $68,237.45   $66,094.09    $56,494.36                $80,921.55   $58,352.55   $74,020.36           $50,480.45   $74,541.55   $70,624.45    $690,124.82
Professional, Scientific, and Technical Services $85,563.55     $71,431.45   $61,292.36    $59,731.64                $77,745.36   $62,954.73   $76,009.91           $54,193.55   $66,820.91   $65,573.18    $681,316.64
Educational Services                           $65,913.64       $58,551.82   $58,080.36    $69,186.27                $71,052.45   $52,490.36   $65,993.00           $65,827.45   $54,350.45   $65,416.27    $626,862.09
Transportation and Warehousing                 $76,263.55       $68,325.55   $59,711.27    $55,122.55                $70,323.45   $57,103.27   $55,875.82           $47,676.00   $58,753.09   $63,114.82    $612,269.36
Health Care and Social Assistance              $60,324.36       $56,892.91   $50,416.00    $56,354.45                $60,297.73   $53,076.36   $58,454.55           $46,682.45   $51,293.36   $54,446.00    $548,238.18
```

Figure 10: Screenshot of general statistics

User Selected Statistics

The user is prompted to enter 'select' to filter the economic data across Canada. This section allows the user to choose a single province, job industry and year. The user can also select to filter for all. During the selection, if the user does not select a valid entry, the program will request the user to enter a valid entry and ask if the user would like a list of valid entries. This addresses the requirement for user re-entry of selections as a invalid entry produces a "VALUE EXCEPTION" which will be caught and the selection process initiated once again.

```
****************************************** GENERAL STATS END ******************************************

Enter 'select' to select specific provinces, job titles and year or any other character to exit
select

****************************************** USER STATS ******************************************


Enter province or 'all' for all province data, if you would like to see options press 'o'
Apples
Input for province is not valid.

Enter 'y' to display choices for province or any other key to try again!
y
province choices

New Brunswick
British Columbia
Saskatchewan
Quebec
Ontario
Newfoundland and Labrador
Prince Edward Island
Nova Scotia
Manitoba
Alberta
all

Enter province or 'all' for all province data, if you would like to see options press 'o'
```

Figure 11: Command Line Interface of User input initiation

```python
def get_valid_cli_input(valid_array, category):
    """Prompts user for input through the command line interface, validates input if in array and returns input if valid
    Args:
        valid_array (array): String array containing valid input
        category (string): Category of input to validate ie. "province", 'job outlook"

    Raises:
        ValueError: user_input is not in array

    Returns:      David Cheng, 6 days ago • Completed user input validation method …
        user_input (str): User inputted
    """

    while True:
        try:
            user_input = input(
                f"\nEnter {category} or 'all' for all {category} data, if you would like to see options press 'o'\n"
            )
            if user_input == "o":
                print(f"{category} choices\n")
                print(*valid_array, sep="\n")
            elif user_input not in valid_array:
                raise ValueError(f"{category} not found")
            else:
                break

        except ValueError:
            print(f"Input for {category} is not valid.\n")
            if (
                input(
                    f"Enter 'y' to display choices for {category} or any other key to try again!\n"
                )
                == "y"
            ):
                print(f"{category} choices\n")
                print(*valid_array, sep="\n")
            continue
    return user_input
```

Figure 12: Code snippet of error catching for invalid input and restart of user input procedure

Once the user has selected valid selections, the program will produce tables specific to the user's selection. See below for example of Alberta, Information and 2021.

```
Enter province or 'all' for all province data, if you would like to see options press 'o'
Alberta

Enter job title or 'all' for all job title data, if you would like to see options press 'o'
Information

Enter year or 'all' for all year data, if you would like to see options press 'o'
2021


Below is all the data corresponding to your selections:
Sector                                                                  51
Labour statistics                                  Total compensation per job
UOM                                                                 Dollars
Salary                                                              90115.0
Labour statistics_Jobs                                  Total number of jobs
UOM_Jobs                                                              Jobs
Employed                                                            32680.0
Actual hours worked            Average actual hours (worked in reference week...
UOM_Hours                                                            Hours
Standard_Hours                                                        32.9
Labour force characteristics                              Unemployment rate
UOM_Unemployment                                                 Percentage
Unemployment_Rate                                                 10.733333
Earnings per hour                                                 52.674188
Total Sector Job Demand                                       36609.410007
Unemployed                                                     3929.410007
Name: (Alberta, Information, 2021), dtype: object

Statistics for job demand based on selections

Sector                                                                  51
Labour statistics                                  Total compensation per job
UOM                                                                 Dollars
Salary                                                              90115.0
Labour statistics_Jobs                                  Total number of jobs
UOM_Jobs                                                              Jobs
Employed                                                            32680.0
Actual hours worked            Average actual hours (worked in reference week...
UOM_Hours                                                            Hours
Standard_Hours                                                        32.9
Labour force characteristics                              Unemployment rate
UOM_Unemployment                                                 Percentage
Unemployment_Rate                                                 10.733333
Earnings per hour                                                 52.674188
Total Sector Job Demand                                       36609.410007
Unemployed                                                     3929.410007
Name: (Alberta, Information, 2021), dtype: object

****************************************** USER STATS END ******************************************
```

Figure 13: User specified search for "Alberta, Information, 2021"

Plot display

The user is prompted to enter ‘plot’ to display various plots of the data. The plots are saved under the following path ‘/static/img/../’ which will then be referenced by the graphical user interface. For ease of access,, a separate folder named ‘plots’ was created with the plots copied into and is stored in the project directory. Plotting is the final step of the program and upon completion of the plotting phase, the program will exit.

```
****************************************** USER STATS END ******************************************

Enter 'plot' to save and display interesting plots!
plot

****************************************** PLOTS INITIATED ******************************************


****************************************** PLOTS END ******************************************


********************************************** EXITING **********************************************
```

Figure 14: Completion of plotting and program exit

See below for example of a plot depicting the growth of the average of the top 20 earners from 2011 to 2021. See folder of plots for more plots of analysis.

Figure 15: Top 20 earners from 2011 to 2021

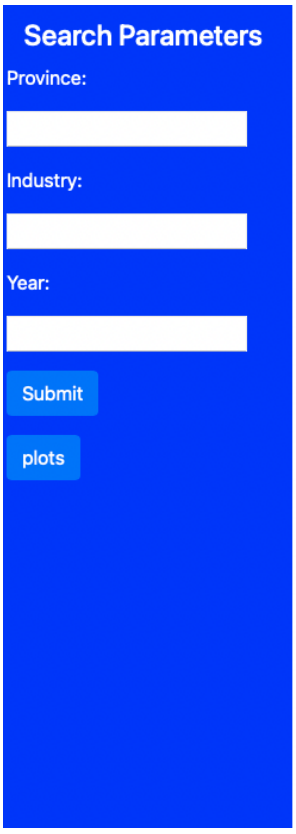
Graphical User Interface

As an added addition to the project scope, the graphical user interface was developed as a simple proof of concept to test the extensibility of our program's modular functions. This interface acts as a web search for our final DataFrame of Canadian Economic Data and displays the generated plots.

```
(base) ➜  final-project-two-sum git:(main) ✗ python Final_Project.py


Enter 'cli' or 'gui' to enter program from command line interface or graphical user interface
gui
 * Serving Flask app "gui" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [15/Jun/2022 19:58:21] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [15/Jun/2022 19:58:21] "GET /static/main.css HTTP/1.1" 200 -
```

Figure 16: Graphical User Interface initiation



Alberta, Information Jobs in 2020

	Alberta
	Information
	2020
Sector	51
Labour statistics	Total compensation per job
UOM	Dollars
Salary	89600.0
Labour statistics_Jobs	Total number of jobs
UOM_Jobs	Jobs
Employed	29545.0
Actual hours worked	Average actual hours (worked in reference week, main job)
UOM_Hours	Hours
Standard_Hours	29.4
Labour force characteristics	Unemployment rate
UOM_Unemployment	Percentage
Unemployment_Rate	14.925
Earnings per hour	58.608059
Total Sector Job Demand	34728.181017
Unemployed	5183.181017

Figure 17: Graphical User Interface with search terms "Alberta, Information, 2020"

Conclusion

This project successfully completed the requirements of the specifications. An opportunity for improvement would be to put our group our modularized functions together by scope and put them into separate python modules which can then be imported and called upon in our user interface class. For example, grouping our input validation and data handling functions into two separate python files that would then be imported into the python file containing our command line interface. This would keep our program's code details hidden and have our code attain a higher level of logic. Functions would also be able to be reused into different programs like seen in our graphical user interface implementation.

References

Labour force characteristics by industry, annual (x 1,000), Statistics Canada, 2022-01-07 [Online] Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301

Employment by Industry Annual, Statistics Canada, 2022-03-29 [Online] Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301

Total compensation per job, by NAICS industry, Statistics Canada, 2022-05-20 [Online] Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301

2022 NAICS, North American Industry Classification System, 2022-06-10 [Online] Available: https://www.census.gov/naics/#:~:text=The%20North%20American%20Industry%20Classification,to%20the%20U.S.%20business%20economy?58967?yearbck=2022