Economic Factors affecting Electric Vehicle Ownership in Ontario

Final Results and Project Report

by: Irfan Ahmad
CIND820 D1H
November 2023

Toronto Metropolitan University

Table of Contents

Abstract rev 2.0

Electric vehicles have immense environmental and health benefits. In addition to unwanted climate change impacts, respiratory illnesses such as asthma, wheezing, inflammation, and shortness of breath are also adversely affected by bad air quality. Absence of tailpipe emissions in EVs reduces the presence of greenhouse gases in the air which leads to less pollution and better air for breathing.

Since 2022, the Government of Canada is pursuing an ambitious plan to reduce emissions by 40% below 2005 levels by 2030. As such, the public awareness has gained considerable momentum in realizing the benefits of the products offered by the nascent EV industry.

Since purchasing an electric vehicle is relatively expensive in its early years, this project looks at the economic factors that may contribute towards affordability.

The purpose of this project is to come up with a formula that utilizes economic factors in determining the pace at which electric vehicles are registered in Ontario's Forward Sortation Areas (first 3 characters of a postal code). This information when utilized with economic factors such as inflation, interest rates, incentives and GDP growth will provide insight into the areas where more robust infrastructure must be built in order to accommodate increasing EV needs such as charging stations as well as make visible those forward sortation areas where more economic investments are needed in order to improve EV affordability.

The datasets for this study are extracted from the CRA, and the Government of Ontario.

The CRA provides Individual Tax Statistics (group by Forward Sortation Area) that includes Income from multiple sources such as employment, pension, investment, self employment, benefits and all other sources.

The Government of Ontario provides quarterly reports (starting January 2022) on the cumulative number of electric vehicle registrations in Ontario by Forward Sortation Areas. Using these quarterly reports, simple regression will be used to calculate the rate of change of EV registrations in Ontario. Then this information will be integrated with the data on income source for each forward sortation area in Ontario, and finally part of this integrated data will be trained to predict future rate of change of EV registrations in Ontario. To accomplish this, three prediction models will be evaluated; these models are: Multiple Linear Regression, Logistic Regression, and K-Nearest Neighbour. The models will be built using Python/Jupyter Notebook.

After satisfactory testing on the test dataset, the model can be used for provinces other than Ontario.

Note: Since the CRA data is not yet available for the years 2021 and beyond, the prediction models are based on income sources collected for the tax year 2020.

Dataset Links

Individual Tax Statistics from CRA:

https://www.canada.ca/en/revenue-agency/programs/about-canada-revenue-agency-cra/income-statistics-gst-hst-statistics/individual-tax-statistics-fsa.html

Electric Vehicle Registration Count in Ontario from the Government of Ontario:

https://data.ontario.ca/dataset/electric-vehicles-in-ontario-by-forward-sortation-area

Present Knowledge

Due to environmental and health benefits offered by decarbonization of transportation, the government of Canada is aggressively pursuing its ambitious plan to reduce emissions by 40% below 2005 levels by 2030, and to achieve 100% new light-duty zero-emission vehicle sales by 2035.In doing so, a number of action plans are in the works such as:

- Purchase incentives – rebate from federal and provincial governments
- Infrastructure development – charging stations
- Benefit awareness – media advertisements
- Research and Development – alternative fuel efficiency

A number of programs are in place to address the above initiatives such as:

- Up to $5000 incentive from Transportation Canada
  Source: https://tc.canada.ca/en/road-transportation/innovative-technologies/zero-emission-vehicles/light-duty-zero-emission-vehicles

- Electric Vehicle ChargeON Program
  Source: https://www.ontario.ca/page/ev-chargeon-program

- Promote transition to electric transportation through education and advocacy
  Source: https://evsociety.ca/about-ev-society/

- Electric Vehicle Discovery Centre
  Source: https://www.plugndrive.ca/electric-vehicle-discovery-centre/

- Fuel cell research and development by Ballard Power Systems in Burnaby, BC
  Info: https://www.ballard.com/about-ballard/our-vision

Implementation of the above programs have produced encouraging results such as:

- Zero-emission vehicles hit 10.5% market share in Canada in Q2-2023, an all-time high
  Source: https://electricautonomy.ca/2023/08/25/zev-market-share-canada-q2/

- PowerON, Billy Bishop Airport team-up for EV charging solution
  Source: https://electricautonomy.ca/2023/10/26/poweron-billy-bishop-airport-ev-charging/

- Governments of Canada and Ontario finalize agreement with Umicore Rechargeable Battery Materials Canada Inc. for new plant in Loyalist Township
  Source: https://www.canada.ca/en/innovation-science-economic-development/news/2023/10/governments-of-canada-and-ontario-finalize-agreement-with-umicore-rechargeable-battery-materials-canada-inc-for-new-plant-in-loyalist-township.html

Critical Analysis of Present Knowledge

Existing literature on zero-emission electric vehicles (ZEV) effectively provides a wealth of information regarding programs in place for public education and awareness, progress towards uptick in sales volume, infrastructure development, and research facilities for improvements in fuel efficiencies. Existing ZEV registration count is used to provide outlook for future demand in broader geographical areas. However, this approach lacks finer details with respect to economic factors contributing towards ZEV registrations at a smaller postal district level.

The current literature provides reactive information based on observations from broad geographical areas. For example, Electric Autonomy Canada (ZEV registrations hold steady in Q2 2023, according to StatsCan (electricautonomy.ca) discusses ZEV registration at a national and provincial levels. It documents the facts but does not provide action plan for improvements where sales are sluggish.

An important economic factor missing from the current literature is the role of income source on ZEV registrations. Distribution of income source at smaller geographical areas such as forward sortation area (FSA – first three characters of a postal code) can provide a valuable information on the type of income sources contributing most in ZEV purchases.

Other Identical Work

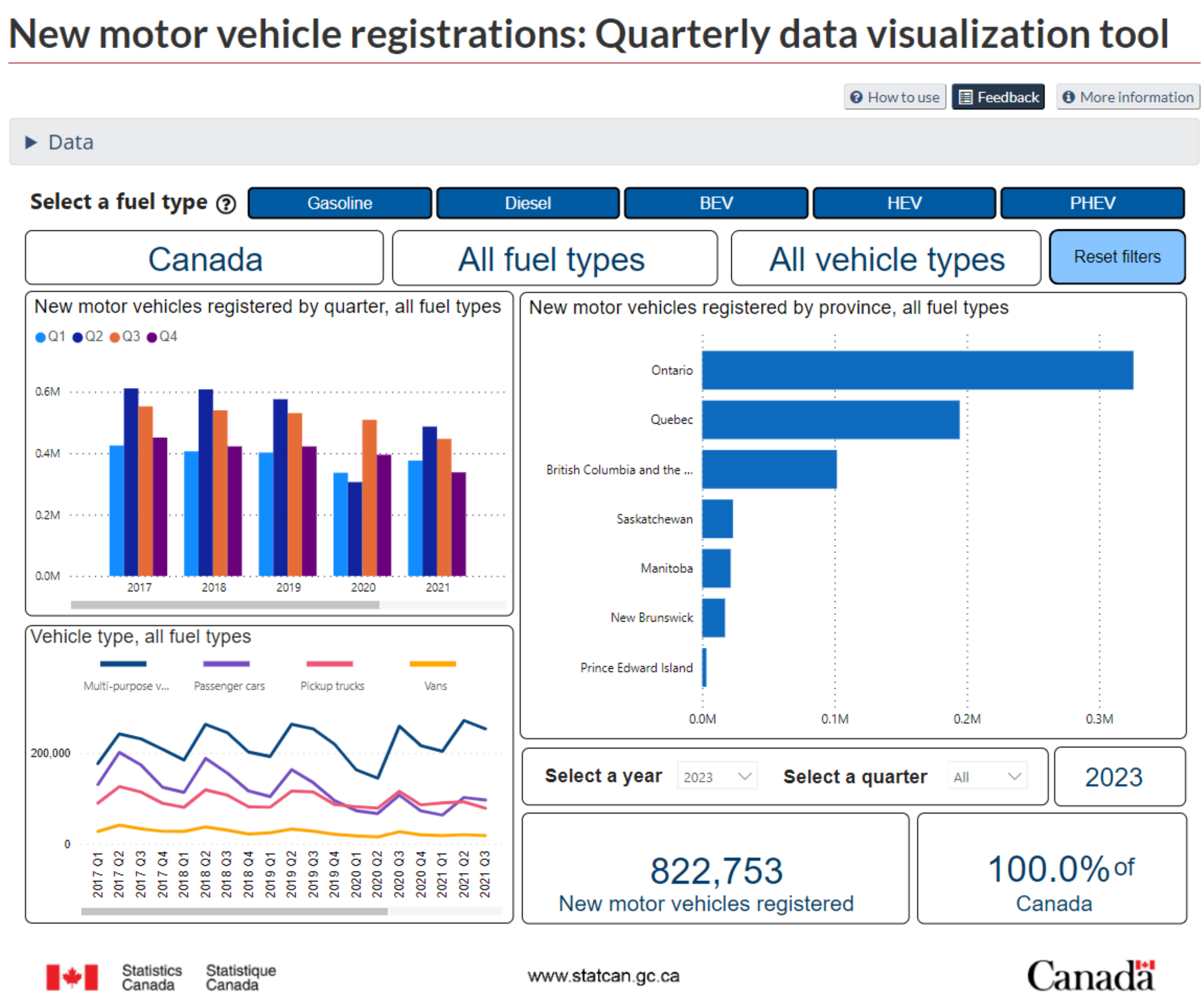
No studies found on comparative impact of income source on electric vehicle ownership.

Other Related Work

The current literature on electric vehicle trends is observational where statistics are displayed for collected ZEV information. The information collection was done by empirical means, such as retrieving ZEV registration counts from relevant departments and collection of household income by surveys. Two such works are as follows:

New Motor Vehicle Registrations by Statistics Canada:

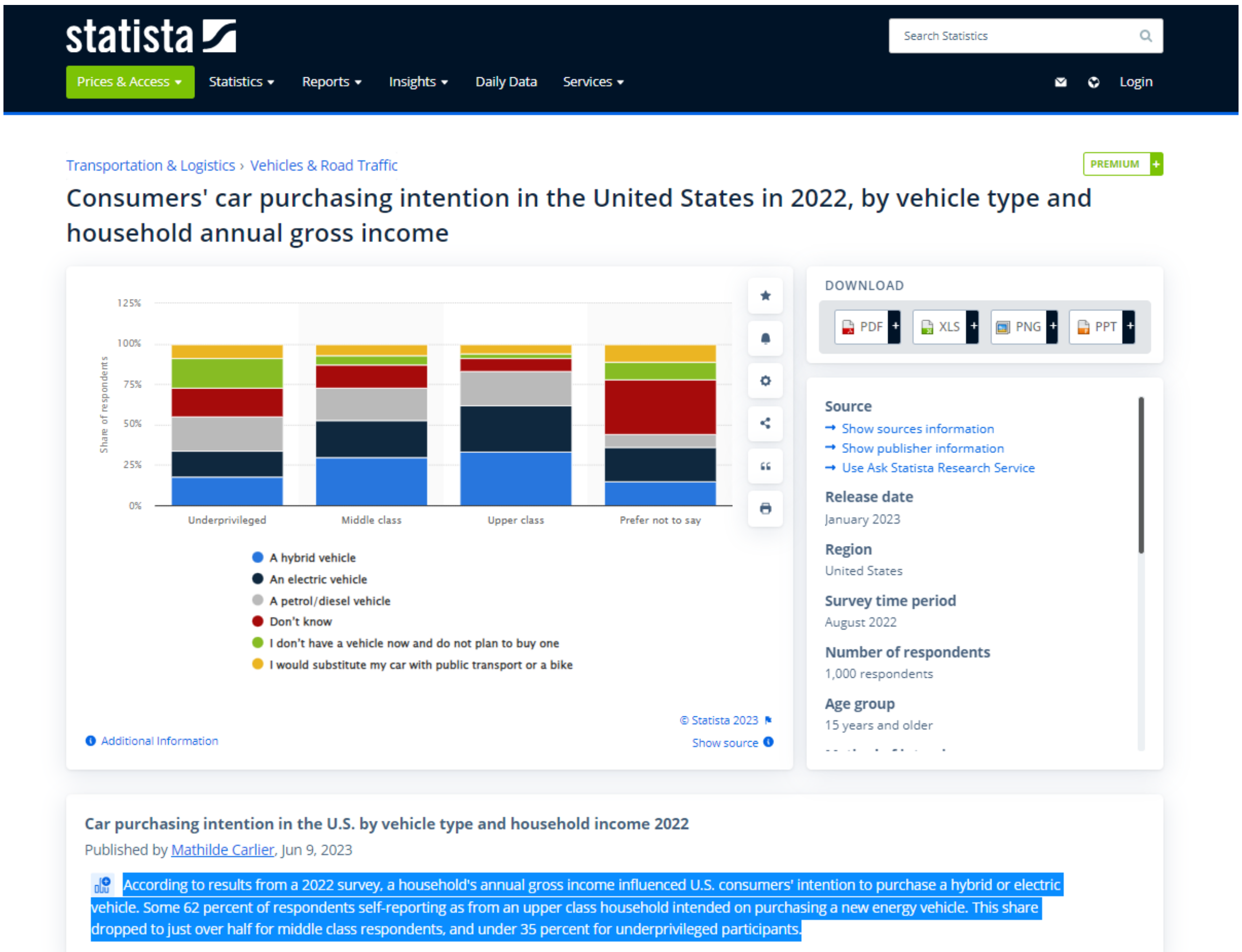
https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2021019-eng.htm



Other Related Work (cont)

Consumers' car purchasing intention in the United States in 2022, by vehicle type and household annual gross income:

https://www.statista.com/statistics/1296900/us-car-purchasing-intention-by-vehicle-type-and-household-income/



Relevance of My Work

My work in this field does not rely on surveys to determine future outlook of ZEV registration nor does it depend solely on simple regression for prediction.

My work combines two sources of consumer data—ZEV registration count and weighted income amount per source for each forward sortation area (FSA) in Ontario.

Analysis of ZEV registration data combined with information on income source will provide a much clearer picture about the existing inventory of zero-emission electric vehicles in each FSA. Information regarding contribution of various income sources on ZEV purchase will allow for a more targeted approach towards awareness, infrastructure, and in implementation of future economic policies.

Among many answers, my work in this field will also provide answers to the following specific questions:

- What parts of Ontario have the strongest growth in ZEV registration?
- What type of income contributes most in ZEV registration?
- Given economic outlook, what FSAs will be affected most in terms of new ZEV registrations?
- Based on demographic outlook, what impact will retirement income have on new ZEV registrations?
- What income source is most common in areas (FSA) where new ZEV registrations are lowest?
- What type of targeted programs are needed to improve ZEV affordability in those forward sortation areas where new ZEV registrations are sluggish?

Value of My Work

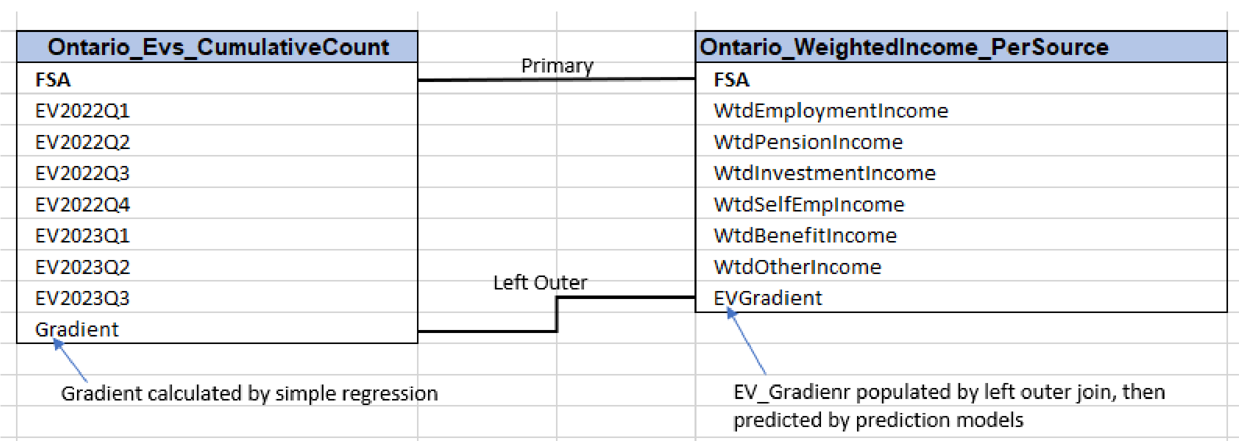
My research work is worthwhile because it identifies the relative importance of a source of income in determining whether or not a zero-emission vehicle will be purchased in a given forward sortation area.

Approach

The Government of Ontario provides quarterly reports (starting January 2022) on the cumulative number of electric vehicle registrations in Ontario by Forward Sortation Areas. Using these quarterly reports, simple regression will be used to calculate the rate of change of EV registrations in Ontario. Then this information will be integrated with the data on income source for each forward sortation area in Ontario, and finally part of this integrated data will be trained to predict future rate of change of EV registrations in Ontario. To accomplish this, three prediction models will be evaluated; these models are: Multiple Linear Regression, Logistic Regression, and K-Nearest Neighbour. The models will be built using Python/Jupyter Notebook.

The steps involved in realizing above objectives are as follows:

- Combine quarterly reports on ZEV registration counts into one table (ZEV registrations for each quarter as a separate filed having FSA as unique ID)
- Using simple regression, determine the gradient of ZEV registration count for each forward sortation area (FSA)
- Link the gradient to the income amount/person/source dataset (StatsCan provides income information for each forward sortation area)
- In the augmented income amount/person/source dataset, the gradient is the dependent variable.
- Divide the income amount/person/source dataset into train and test.
- Train the model using multiple regression to determine gradient as a weighted sum of income source.
- Test the model.
- Predict the gradient of ZEV registration after applying government’s economic forecasts on income type.
- Apply the model to other jurisdictions.



Datasets

GitHub Link: https://github.com/iresearch23/cind820

Ontario_EVs_CumulativeCount

Column Name	Description
FSA	Forward Sortation Area (First 3 characters of a postal code)
EV2022Q1	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 1 of 2022
EV2022Q2	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 2 of 2022
EV2022Q3	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 3 of 2022
EV2022Q4	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 4 of 2022
EV2023Q1	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 1 of 2023
EV2023Q2	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 2 of 2023
EV2023Q2	Cumulative Number of zero-emission electric vehicle registrations at the end of quarter 3 of 2023
Gradient	**Rate of change of zero-emission electric vehicle registrations – to be determined by simple regression**

Number of columns: 9
Number of records: 569

Datasets (cont)

GitHub Link: https://github.com/iresearch23/cind820

Ontario_WeightedIncome_PerSource

Column Name	Description
FSA	Forward Sortation Area (First 3 characters of a postal code)
WtdEmploymentIncome	(Employment income) * (Employment Income # / Total Income #)
WtdPensionIncome	(Pension income) * (Pension Income # / Total Income #)
WtdInvestmentIncome	(Investment income) * (Investment Income # / Total Income #)
WtdSelfEmpIncome	(Self Emp income) * (Self Emp Income # / Total Income #)
WtdBenefitIncome	(Benefit income) * (Benefit Income # / Total Income #)
WtdOtherIncome	(Other income) * (Other Income # / Total Income #)
EVGradient	**Rate of change of zero-emission electric vehicle registrations – populated from Ontario_ZEV_Cumulatine_Count table, then predicted by multiple regression model**

Number of columns: 8
Number of records: 525

Data Preparation (part 1 of 2)

The data preparation is done in two stages: The first stage prepares the data for determining gradient for EV registration count. The data preparation in the second stage integrates this gradient with the income data provided by the CRA.

Analysis of the Government of Ontario's quarterly reports on the cumulative number of electric vehicle registrations in Ontario by Forward Sortation Areas indicate that there are 132 missing values. The missing values are replaced with 0. See the Python screenshot below:

```
In [4]: dfEVRegistration = pd.read_csv('C:\Tmu\CIND820\Ontario_EVs_CumulativeCount.csv', encoding='utf-8')
        print(dfEVRegistration.shape)
        dfEVRegistration.head(5)

(569, 8)

Out[4]:
```

	FSA	EV2022Q1	EV2022Q2	EV2022Q3	EV2022Q4	EV2023Q1	EV2023Q2	EV2023Q3
0	K0A	645.0	745.0	871.0	928.0	987.0	1088.0	1195.0
1	K0B	95.0	103.0	107.0	118.0	119.0	131.0	145.0
2	K0C	167.0	193.0	215.0	234.0	242.0	269.0	302.0
3	K0E	128.0	140.0	155.0	162.0	173.0	182.0	208.0
4	K0G	197.0	223.0	254.0	272.0	294.0	326.0	352.0

Check for missing values:

```
In [5]: dfEVRegistration.isna().sum()

Out[5]: FSA          0
        EV2022Q1    28
        EV2022Q2    26
        EV2022Q3    22
        EV2022Q4    20
        EV2023Q1    19
        EV2023Q2    11
        EV2023Q3     6
        dtype: int64
```

Replace missing values with 0:

```
In [6]: dfEVRegistration["EV2022Q1"].fillna(0, inplace = True)
        dfEVRegistration["EV2022Q2"].fillna(0, inplace = True)
        dfEVRegistration["EV2022Q3"].fillna(0, inplace = True)
        dfEVRegistration["EV2022Q4"].fillna(0, inplace = True)
        dfEVRegistration["EV2023Q1"].fillna(0, inplace = True)
        dfEVRegistration["EV2023Q2"].fillna(0, inplace = True)
        dfEVRegistration["EV2023Q3"].fillna(0, inplace = True)
```

Note: Outliers are not expected in this data as this data is an accurate account of EV registrations.

Data Processing (part 1 of 2)

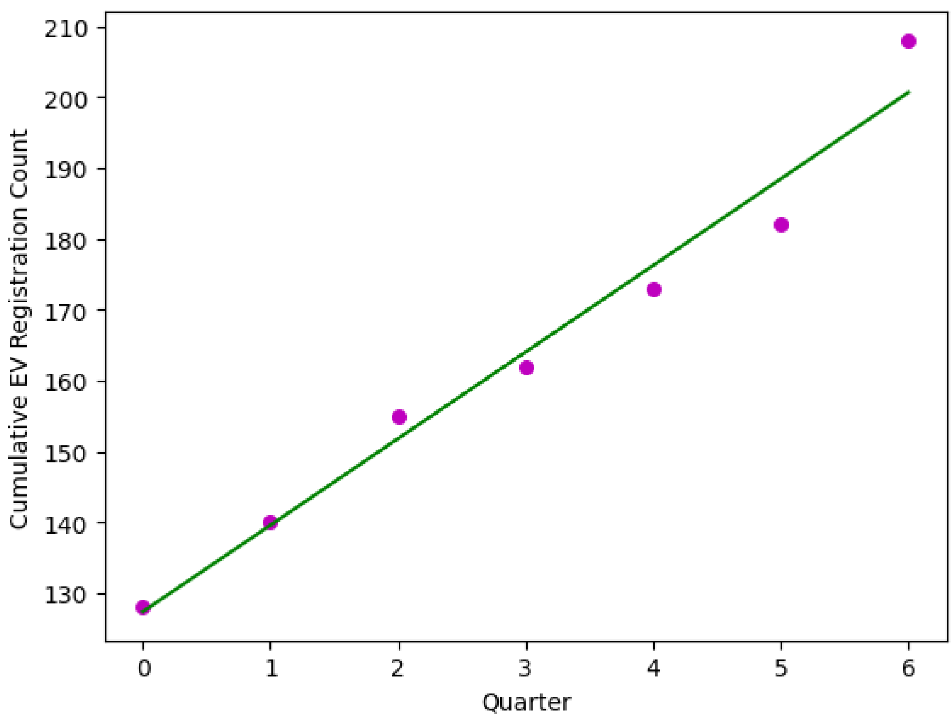
After running simple regression, the gradient is assigned to each row corresponding to forward sortation area:

	FSA	EV2022Q1	EV2022Q2	EV2022Q3	EV2022Q4	EV2023Q1	EV2023Q2	EV2023Q3	\
0	K0A	645.0	745.0	871.0	928.0	987.0	1088.0	1195.0	
1	K0B	95.0	103.0	107.0	118.0	119.0	131.0	145.0	
2	K0C	167.0	193.0	215.0	234.0	242.0	269.0	302.0	
3	K0E	128.0	140.0	155.0	162.0	173.0	182.0	208.0	
4	K0G	197.0	223.0	254.0	272.0	294.0	326.0	352.0	

	Yintercept	Gradient
0	87.571429	660.000000
1	7.785714	93.500000
2	20.857143	169.142857
3	12.214286	127.357143
4	25.392857	197.821429

Example - Graph for record 3:

```
EV Registrations = [128. 140. 155. 162. 173. 182. 208.]
Coefficients = [127.35714286  12.21428571]
```



Data Preparation (part 2 of 2)

After establishing gradient values for the quarterly EV registration count, we are ready to prepare income data and link it to appropriate EV gradient. The second stage of data preparation handles income data extracted from the CRA public data repository.

After extraction from CRA, the weighted income per source is calculated as follows:

Weighted income for source X
= Total income from source X * (# individuals income source X / # individuals all income sources)

The above calculations are directly done in extracted CSV, then the CSV is loaded to the data frame in Python:

```
In [82]: dfWtdIncomePerSource = pd.read_csv('C:\Tmu\CIND820\Ontario_WeightedIncome_PerSource.csv', encoding='utf-8')
print(dfWtdIncomePerSource.shape)
dfWtdIncomePerSource.head(5)
```

(525, 7)

Out[82]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdInvestmentIncome	WtdSelfEmpIncome	WtdBenefitIncome	WtdOtherIncome
0	K0A	1180583638	174107115	67101015	9991241	19249315	41094462
1	K0B	131846461	35705964	12872775	1741128	7779233	5982520
2	K0C	374023803	91756996	25979482	3541081	16401083	15257693
3	K0E	261467549	78028866	15588684	2235807	12750040	14100813
4	K0G	288977631	93333286	18809557	2755332	8808412	15779711

Check for missing values:

```
In [83]: dfWtdIncomePerSource.isna().sum()
```

```
Out[83]: FSA                    0
WtdEmploymentIncome    0
WtdPensionIncome       0
WtdInvestmentIncome    0
WtdSelfEmpIncome       0
WtdBenefitIncome       0
WtdOtherIncome         0
dtype: int64
```

Data Preparation (part 2 of 2) cont…

Look for strong correlation:

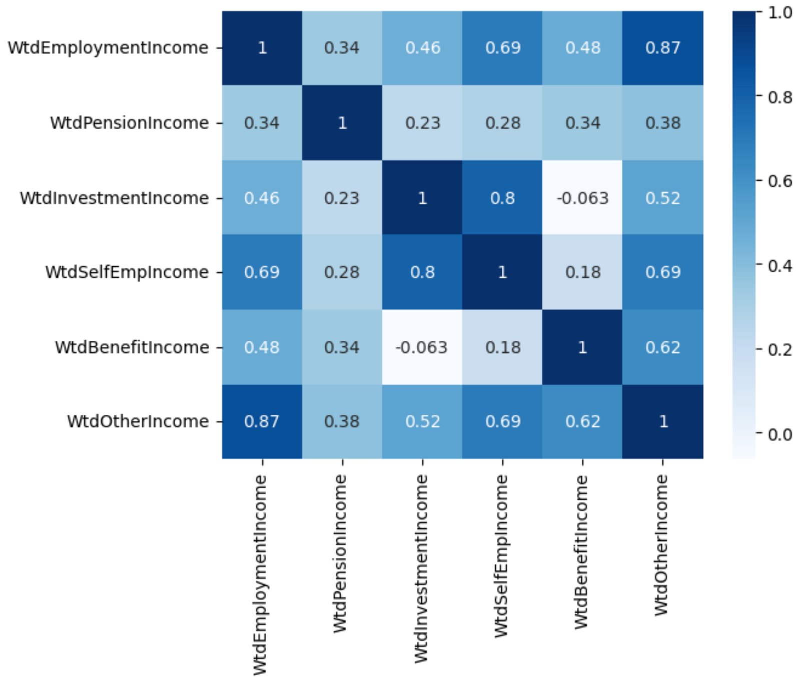
In [86]:
```
dfWtdIncomePerSource.drop('FSA', axis=1).corr()
```

Out[86]:

	WtdEmploymentIncome	WtdPensionIncome	WtdInvestmentIncome	WtdSelfEmpIncome	WtdBenefitIncome	WtdOtherIncome
WtdEmploymentIncome	1.000000	0.336690	0.464325	0.693096	0.477752	0.874617
WtdPensionIncome	0.336690	1.000000	0.230955	0.276162	0.338046	0.378981
WtdInvestmentIncome	0.464325	0.230955	1.000000	0.795886	-0.063307	0.519549
WtdSelfEmpIncome	0.693096	0.276162	0.795886	1.000000	0.179042	0.693825
WtdBenefitIncome	0.477752	0.338046	-0.063307	0.179042	1.000000	0.622184
WtdOtherIncome	0.874617	0.378981	0.519549	0.693825	0.622184	1.000000

Plot correlation heatmap:

In [87]:
```
dataplot = sb.heatmap(dfWtdIncomePerSource.drop('FSA', axis=1).corr(), cmap="Blues", annot=True)
plt.show()
```



Consider a correlation >= 0.70 as strong, then WtdOtherIncome (0.87) and WtdInvestmentIncome (0.80) can be removed.

Data Preparation (part 2 of 2) cont…

Look for string correlation: cont…

After removing strongly correlated variables: WtdOtherIncome and WtdInvestmentIncome

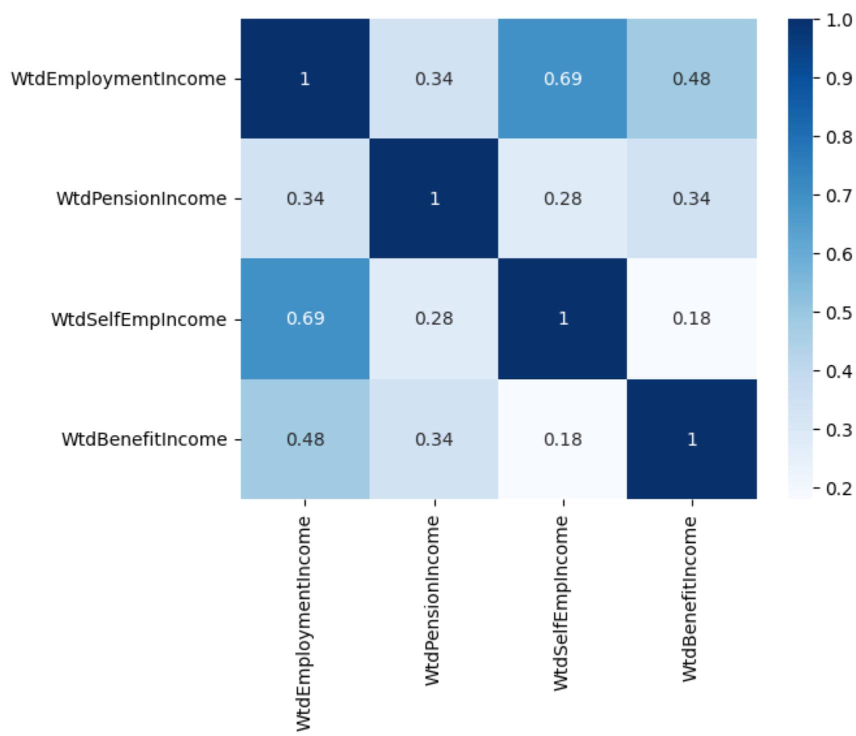
In [88]:
```
del dfWtdIncomePerSource['WtdOtherIncome']
del dfWtdIncomePerSource['WtdInvestmentIncome']
dfWtdIncomePerSource.head(5)
```

Out[88]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmplncome	WtdBenefitIncome
0	K0A	1180583638	174107115	9991241	19249315
1	K0B	131846461	35705964	1741128	7779233
2	K0C	374023803	91756996	3541081	16401083
3	K0E	261467549	78028866	2235807	12750040
4	K0G	288977631	93333286	2755332	8808412

Plot correlation heatmap to verify correlation < 0.7

In [89]:
```
dataplot = sb.heatmap(dfWtdIncomePerSource.drop('FSA', axis=1).corr(), cmap="Blues", annot=True)
plt.show()
```



Data Preparation (part 2 of 2) cont…

Look for outliers > 3 SD away:

```
In [90]: dfWtdIncomePerSource.describe()
```

Out[90]:

	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmpIncome	WtdBenefitIncome
count	5.250000e+02	5.250000e+02	5.250000e+02	5.250000e+02
mean	2.395274e+08	2.852788e+07	2.503401e+06	9.058794e+06
std	2.038361e+08	2.608942e+07	2.885245e+06	7.458949e+06
min	1.013330e+05	0.000000e+00	0.000000e+00	0.000000e+00
25%	1.072442e+08	1.219278e+07	6.697590e+05	3.404927e+06
50%	1.868204e+08	2.182338e+07	1.515815e+06	7.779233e+06
75%	3.198277e+08	3.736638e+07	2.981993e+06	1.270416e+07
max	1.330043e+09	2.525722e+08	1.763455e+07	4.220975e+07

Remove Outliers more than 3 SD away:

```
In [92]: dfWtdIncomePerSource.shape

Out[92]: (525, 5)

In [93]: EmploymentIncome_mean = dfWtdIncomePerSource["WtdEmploymentIncome"].mean()
         EmploymentIncome_std  = dfWtdIncomePerSource["WtdEmploymentIncome"].std()

         PensionIncome_mean    = dfWtdIncomePerSource["WtdPensionIncome"].mean()
         PensionIncome_std     = dfWtdIncomePerSource["WtdPensionIncome"].std()

         SelfEmpIncome_mean    = dfWtdIncomePerSource["WtdSelfEmpIncome"].mean()
         SelfEmpIncome_std     = dfWtdIncomePerSource["WtdSelfEmpIncome"].std()

         BenefitIncome_mean    = dfWtdIncomePerSource["WtdBenefitIncome"].mean()
         BenefitIncome_std     = dfWtdIncomePerSource["WtdBenefitIncome"].std()

         dfWtdIncomePerSource = dfWtdIncomePerSource.loc[dfWtdIncomePerSource["WtdEmploymentIncome"] < EmploymentIncome_mean + 3 * EmploymentIncome_std]
         print(dfWtdIncomePerSource.shape)

         dfWtdIncomePerSource = dfWtdIncomePerSource.loc[dfWtdIncomePerSource["WtdPensionIncome"] < PensionIncome_mean + 3 * PensionIncome_std]
         print(dfWtdIncomePerSource.shape)

         dfWtdIncomePerSource = dfWtdIncomePerSource.loc[dfWtdIncomePerSource["WtdSelfEmpIncome"] < SelfEmpIncome_mean + 3 * SelfEmpIncome_std]
         print(dfWtdIncomePerSource.shape)

         dfWtdIncomePerSource = dfWtdIncomePerSource.loc[dfWtdIncomePerSource["WtdBenefitIncome"] < BenefitIncome_mean + 3 * BenefitIncome_std]
         print(dfWtdIncomePerSource.shape)

(511, 5)
(504, 5)
(494, 5)
(486, 5)
```

Data Preparation (part 2 of 2) cont…

Normalize the numeric attributes then add column 'EVGradient':

Out[94]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmpIncome	WtdBenefitIncome
1	K0B	0.157169	0.346620	0.160442	0.255584
2	K0C	0.446082	0.890743	0.326305	0.538852
3	K0E	0.311805	0.757475	0.206026	0.418898
4	K0G	0.344624	0.906045	0.253899	0.289397
6	K0J	0.224904	0.840682	0.183582	0.404024

Add column 'EVGradient' with default value of zero to the dataframe:

In [95]:

```
dfWtdIncomePerSource['EVGradient']   = 0
dfWtdIncomePerSource.head(5)
```

Out[95]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmpIncome	WtdBenefitIncome	EVGradient
1	K0B	0.157169	0.346620	0.160442	0.255584	0
2	K0C	0.446082	0.890743	0.326305	0.538852	0
3	K0E	0.311805	0.757475	0.206026	0.418898	0
4	K0G	0.344624	0.906045	0.253899	0.289397	0
6	K0J	0.224904	0.840682	0.183582	0.404024	0

Data Preparation (part 2 of 2) cont…

Populate ‘EVGradient’ taken from part 1:

Populate EVGradient from the Gradient values in dfEVRegistration:

In [96]:
```
nRecords = len(dfEVRegistration)

for i in range(nRecords):
        cFSA = dfEVRegistration['FSA'].values[i]
        cFSA = cFSA.strip()
        nGradient = dfEVRegistration['Gradient'].values[i]


        dfWtdIncomePerSource.loc[dfWtdIncomePerSource['FSA'].str.contains(cFSA), 'EVGradient'] = nGradient

dfWtdIncomePerSource.head(5)
```

Out[96]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmplIncome	WtdBenefitIncome	EVGradient
1	K0B	0.157169	0.346620	0.160442	0.255584	93.500000
2	K0C	0.446082	0.890743	0.326305	0.538852	169.142857
3	K0E	0.311805	0.757475	0.206026	0.418898	127.357143
4	K0G	0.344624	0.906045	0.253899	0.289397	197.821429
6	K0J	0.224904	0.840682	0.183582	0.404024	93.357143

Data Processing (part 2 of 2)

Multiple Linear Regression:

After splitting the data into train and test with ratio 75:25, the multiple linear regression has produced the following result:

Mean Squared Error = 3409

Mean Absolute Error = 34

Intercept = 4

Coefficients:

WtdEmploymentIncome: 439

WtdPensionIncome: 40

WtdSelfEmpIncome: 116

WtdBenefitIncome: -96

Therefore, predicted value of quarterly EVGradient is as follows:

EVGradient = 4 + 439 (WtdEmploymentIncome) + 40 (WtdPensionIncome) + 116 (WtdSelfEmpIncome) -96 (WtdBenefitIncome)

The coefficient values suggest that *Employment Income* is the most significant factor in promoting electric vehicle sales, followed by *Self Employment* and *Pension Income*. Districts with large benefit recipients tend to lower the rate at which electric vehicles are purchased.

Data Processing (part 2 of 2)

Logistic Regression and K-Nearest Neighbour Models:

In order to evaluate *Logistic Regression* and *K-Nearest Model*, EVGradient is transformed from continuous to categorical variable with value: 'Low' and 'High'. 50^{th} percentile of EVGradient is used to split the values—EVGradient with values less than 88 are considered 'Low' and values above 87 are considered 'High'.

Performance under *Logistic Regression*:

```
Logistic Regression Confusion Matrix :
 [[49 15]
 [ 4 54]]
```

Performance under Logistic Regression:

Accuracy = (TP+TN)/(TP+TN+FP+FN) = (49+54) / (49+54+15+4) = 103 / 122 = 0.8443

Precision = TP/(TP+FP) = 49 / (49+15) = 49 / 64 = 0.7656

Recall = TP/(TP+FN) = 49 / (49+4) = 49 / 53 = 0.9245

Performance under K-Nearest Neighbour:

```
KNN Confusion Matrix :
 [[50 14]
 [ 3 55]]
```

Performance under K-Nearest Neighbour:

Accuracy = (TP+TN)/(TP+TN+FP+FN) = (50+55) / (50+55+14+3) = 105 / 122 = 0.8607

Precision = TP/(TP+FP) = 50 / (50+14) = 50 / 64 = 0.7813

Recall = TP/(TP+FN) = 50 / (50+3) = 50 / 53 = 0.9434

Better values in all three categories: Accuracy, Precision, and Recall for the KNN model suggests KNN model outperforms Logistic Regression model.

Research Questions and Findings

Q1. What parts of Ontario have the strongest growth in ZEV registration?

Query of the WtdIncomePerSource dataset suggests the strongest EV growth was in L3*, L4*, and L6* districts. These districts are the suburbs of the Greater Toronto Area.

What parts of Ontario have the strongest growth in ZEV registration?

```
In [43]: dfWtdIncomePerSource.nlargest(5, ['EVGradient'])
Out[43]:
```

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmpIncome	WtdBenefitIncome	EVGradient
183	L6C	0.553367	0.116929	0.701627	0.298212	719.214286
137	L3R	0.516518	0.465038	0.662505	0.478577	623.357143
148	L4G	0.871128	0.383687	0.683059	0.347929	616.035714
147	L4E	0.685461	0.105755	0.831202	0.322312	603.571429
145	L4B	0.397504	0.204829	0.617393	0.232644	561.392857

Q2. What type of income contributes most in ZEV registration?

Coefficients of Multiple Linear Regression indicate that employment income contributes most in ZEV registration.

EVGradient = 4 + 439 (WtdEmploymentIncome) + 40 (WtdPensionIncome) + 116 (WtdSelfEmpIncome) -96 (WtdBenefitIncome)

Research Questions and Findings (cont…)

Q3. Given economic outlook, what FSAs will be affected most in terms of new ZEV registrations?

Assuming employment trend is towards contract and self employment work, the FSAs with the higher self employment gradient will be affected most. Top 5 of these FSAs are: M4*, M6*, M9*.

These FSAs represent North York, Forest Hill, and Etobicoke areas of Toronto.

Given economic outlook, what FSAs will be affected most in terms of new ZEV registrations?

```
In [44]: dfWtdIncomePerSource.nlargest(5, ['WtdSelfEmpIncome'])
```

Out[44]:

	FSA	WtdEmploymentIncome	WtdPensionIncome	WtdSelfEmpIncome	WtdBenefitIncome	EVGradient
298	M4V	0.440921	0.347221	1.000000	0.047806	216.214286
322	M6C	0.332768	0.145816	0.965665	0.228830	171.607143
324	M6G	0.395183	0.190317	0.959617	0.275810	152.035714
339	M9A	0.531873	0.386054	0.949265	0.261767	298.321429
333	M6S	0.592496	0.277846	0.928016	0.167655	278.250000

Research Questions and Findings (cont…)

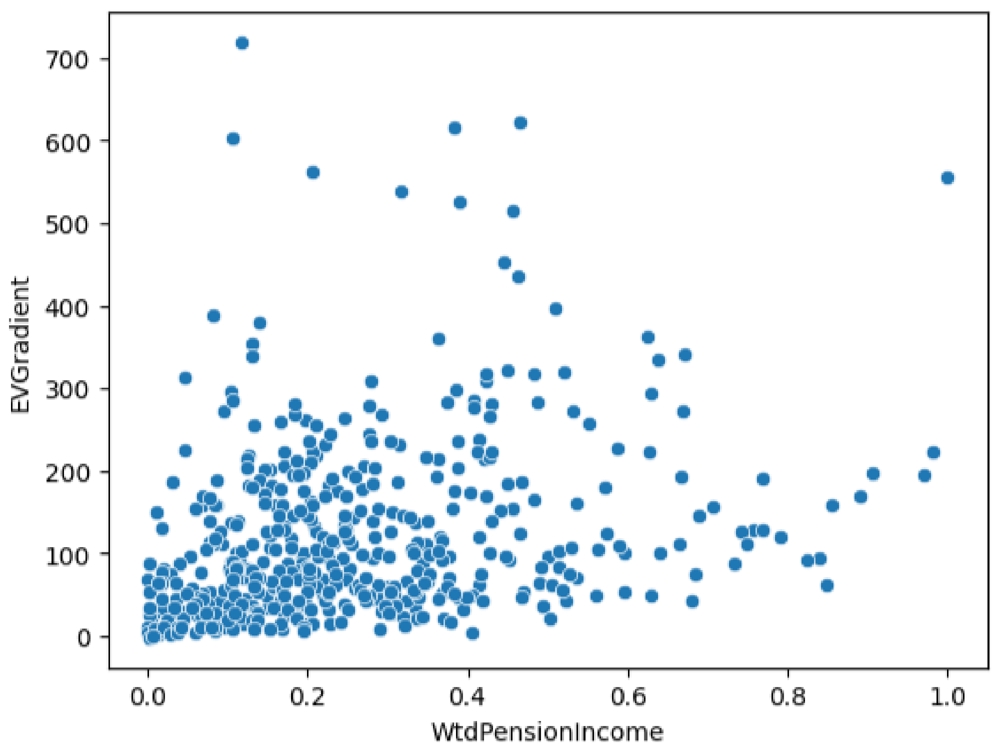
Q4. Based on demographic outlook, what impact will retirement income have on new ZEV registrations?

The scatter plot of Pension Income vs. EVGradient suggests that an increase in pension income does not significantly increases EVGradient. On the contrary, higher EVGradient values exist for low to mid pension income.

Retirement of middle-class workers will likely increase the sale of EVs (perhaps due to cashing out of RRSP or a lump sum payment from an employer). Retirees of high-income brackets are not contributing as much in the sales of EVs (most likely because they already own an EV).

```
In [25]: sb.scatterplot(x='WtdPensionIncome',
                        y='EVGradient', data=dfWtdIncomePerSource)

Out[25]: <Axes: xlabel='WtdPensionIncome', ylabel='EVGradient'>
```



Research Questions and Findings (cont…)

Q5. What income source is most common in areas (FSA) where new ZEV registrations are lowest?

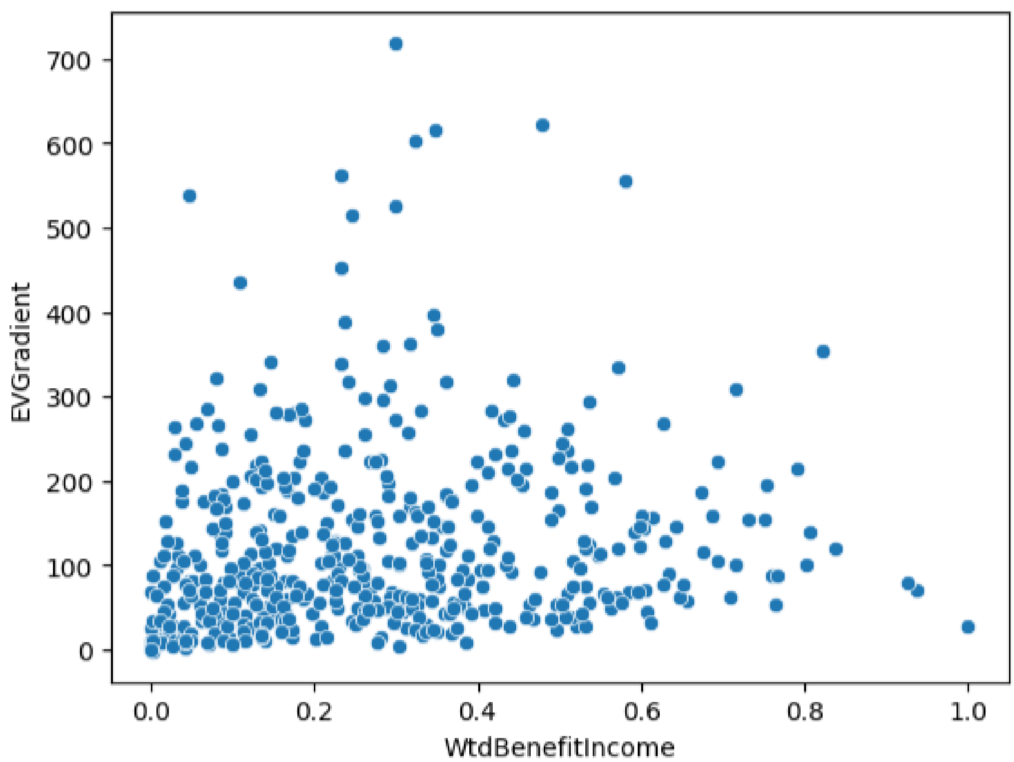
Coefficients of Multiple Linear Regression indicate that the benefit income contributes negatively in new ZEV registration.

EVGradient = 4 + 439 (WtdEmploymentIncome) + 40 (WtdPensionIncome) + 116 (WtdSelfEmpIncome) -96 (WtdBenefitIncome)

This observation is also confirmed by the scatter plot of Benefit Income vs. EVGradient as shown below:

```
In [27]: sb.scatterplot(x='WtdBenefitIncome',
                        y='EVGradient', data=dfWtdIncomePerSource)
```

```
Out[27]: <Axes: xlabel='WtdBenefitIncome', ylabel='EVGradient'>
```



The plot above shows downward trend in EVGradient as the benefit income goes up.

Research Questions and Findings (cont…)

Q6. What type of targeted programs are needed to improve ZEV affordability in those forward sortation areas where new ZEV registrations are sluggish?

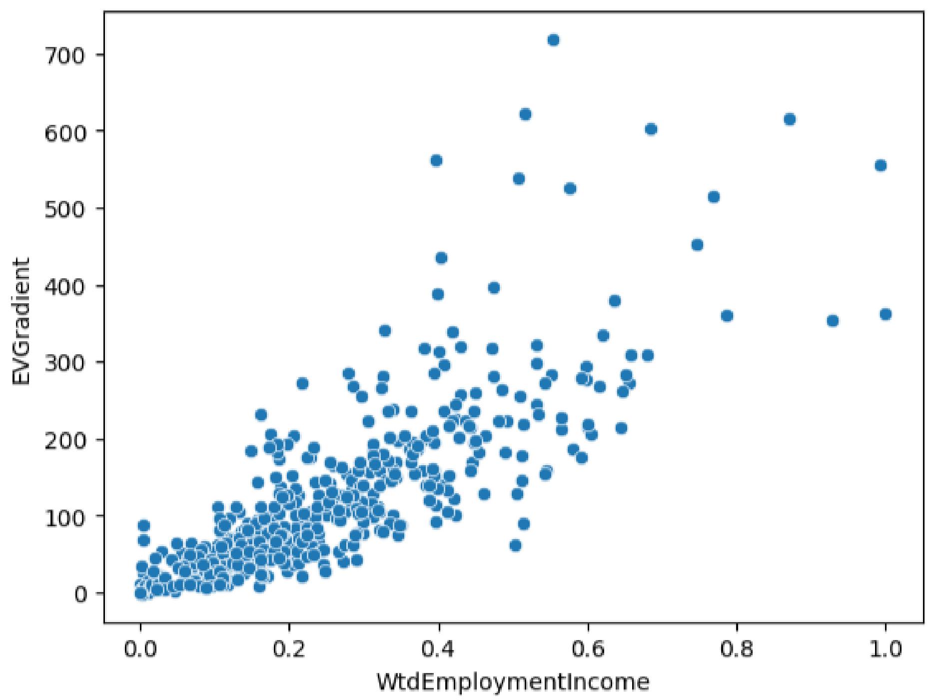
More job opportunities for individuals who rely mostly on social benefits.

Also, as the retired population tend to move to smaller towns, more EV infrastructure should be built in smaller communities in order to encourage purchasing of EVs.

Since higher Employment and Self-employment income contributes more towards EVGradient, more emphasis should be placed on the availability and affordability of higher education. See the scatter plot below showing fewer dots for higher pay but large EVGradient:

```
In [24]: sb.scatterplot(x='WtdEmploymentIncome',
                        y='EVGradient', data=dfWtdIncomePerSource)

Out[24]: <Axes: xlabel='WtdEmploymentIncome', ylabel='EVGradient'>
```



Shortcomings and Conclusion

Shortcomings:

The models studied in this project rely heavily on the impact of income source on the rate of increase in EV registrations. The models do not consider the age and the marital status of individuals in a given postal district. Knowledge of population age will enhance models' predictions since certain age groups are more likely to embrace new technology. Also, a new housing development in a certain district will attract families with young children, providing more insight into the future demands of EVs.

The models also do not consider the saturation of EV ownership. As more and more people acquire EVs, the EVGradient at some point will become close to zero. However, since this is the beginning of EV transition initiative, I expect the gradient to grow or continue to stay constant for the immediate future.

Conclusion:

In order to increase EV ownerships, the following actions are recommended:

- More emphasis on higher education in order to improve affordability.
- More EV infrastructure in smaller towns to encourage retirees to buy EVs.
- More job opportunities for individuals who rely mostly on social benefits.
- More targeted emphasis for people living in big city suburbs.
- More awareness about the benefits of using EVs.
- More incentives to offset cost of EV purchase.

References

Individual Tax Statistics from CRA:
https://www.canada.ca/en/revenue-agency/programs/about-canada-revenue-agency-cra/income-statistics-gst-hst-statistics/individual-tax-statistics-fsa.html

Electric Vehicle Registration Count in Ontario from the Government of Ontario:
https://data.ontario.ca/dataset/electric-vehicles-in-ontario-by-forward-sortation-area

Up to $5000 incentive from Transportation Canada:
https://tc.canada.ca/en/road-transportation/innovative-technologies/zero-emission-vehicles/light-duty-zero-emission-vehicles

Electric Vehicle ChargeON Program:
https://www.ontario.ca/page/ev-chargeon-program

Promote transition to electric transportation through education and advocacy:
https://evsociety.ca/about-ev-society/

Electric Vehicle Discovery Centre:
https://www.plugndrive.ca/electric-vehicle-discovery-centre/

Fuel cell research and development by Ballard Power Systems in Burnaby, BC:
https://www.ballard.com/about-ballard/our-vision

Zero-emission vehicles hit 10.5% market share in Canada in Q2-2023, an all-time high:
https://electricautonomy.ca/2023/08/25/zev-market-share-canada-q2/

PowerON, Billy Bishop Airport team-up for EV charging solution:
https://electricautonomy.ca/2023/10/26/poweron-billy-bishop-airport-ev-charging/

New Motor Vehicle Registrations by Statistics Canada:
https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2021019-eng.htm