# Module\_7\_Project

Project Title:

The impact of socioeconomics on EV market penetration.

Hypothesis to be tested:

We are testing if education, income or both play a major factor in the willingness of Australians to purchase and own electric vehicles, moving away from petrol/diesel cars.

Data acquisition:

Data used in this project was acquired from the ABS website, using their new table builder interface. Income and education data were taken from the “2021 Census – employment, income and education” database, and were manually formatted to provide data split by postcode.

Vehicle Fuel type data was also obtained using the same method, from the vehicle statistics database.

Initial data cleaning process:

Once data was received exported in CSV’ format, some minor manual cleaning was done to remove redundant information and to ensure it was formatted in a way that would be compatible with the pandas library for python. This involved removing additional government information from the top and bottom of the csv, and having data formatted with simple headers for columns. Rows containing additional data for “totals” which didn’t reflect any individual postcode, but a count of postcodes, for example were also removed.

Data preparation:

* CSV data was imported into a jupyter notebook as pandas dataframes, and functions were created for cleaning headers into a consistent format

def clean\_headers(df):

    df.columns = df.columns.str.replace(" ", "\_")

    return df

def reshape\_df(df):

    return df.pivot\_table(index = ['Vehicle\_type', 'Postcode'],

                          columns = 'Fuel\_type',

                          values = 'Count').reset\_index().rename\_axis(None, axis=1)

* The Education, and Income dataframes were merged with Fuel type information on Postcode.
* CSV data was obtained from MatthewProctor.com regarding the classification of postcodes on a basis of Rural and Metro, so that rural postcodes can be analysed separately if needed. This is also since many rural postcodes have so few datapoints, that they have a tendency to produce misleading results when interpreting the data.
* In order to make a fair comparison between postcodes with vastly different populations, income, duration and fuel type data, where converted into percentages of the total number of residents.
* Median and mean weren’t used, for the analysis with regard to income or as due to the data being grouped in bins, a large number of suburbs plot along the same value, so a weighted score was determined to produce a better visual representation of the data.
* The data for both income and education come from the ABS split into multiple different bins, so in order to rank suburbs based on Education and Income, rather than just taking the highest percentage in the highest bracket which would provide potentially misleading conclusions, each bracket was given a weighting, and the percentages of residents in each bin were multiplied by these weightings. Once done the sum of the weighted Education and Income data were added together to provide an Income and Education “score” for each postcode.
* These scores were then graphed alongside the percentage of electric vehicles located in each postcode, to see if there are any correlations to support our hypothesis.

Income analysis:

Analysing income data for all postcodes (rural and metro) produced an r-value of 0.06, showing little if any correlation at all. Rural and metro data were then analysed separately for suburbs with over 5000 vehicles, in order to remove small postcodes with very little data, which was found to skew the data due to many values having 0 EV’s present.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing calculated income score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles for all postcodes.

Doing this yielded a correlation oof 0.23 and 0.2 for metro and rural respectively, suggesting that although slightly correlated, income doesn’t appear to be a major factor in people’s willingness to purchase electric vehicles as opposed to petrol or diesel.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing calculated income score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles in each postcode for Metro area.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing calculated income score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles in each postcode for Metro area.

Education analysis:

Analysing the income of postcodes, produced a correlation of 0.27 when looking at all postcodes, however, like with income, many values were plotting along 0% on the y axis, due to the number of small postcodes with very few total vehicles which generally had no EV’s at all. Once data was filtered for postcodes with >5000 vehicles, metro areas showed a correlation of 0.63, and 0.67 for rural. Interestingly this seems to suggest that the overall education of a given area seems to be quite a good predictor for people’s willingness to purchase an EV.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing calculated education score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles for all postcodes.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing education score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles in each postcode for Metro area.

Chart, scatter chart

Description automatically generated

Figure : Scatter plot showing calculated education score on the x axis, plotted with the percentage of Electric vehicles out of total vehicles in each postcode for Metro area.

Conclusions:

* Income tends to be a poor indicator of EV’s suggesting that it may take more than falling prices to encourage residents to make the switch to EV’s.
* Education is quite well correlated, leading to the conclusion that it may be peoples understanding of the impact of fossil fuel emissions on the environment that lead to purchasing EV’s.
* Availability of infrastructure and dealerships also play an important role. This can be seen when analysing the outlier of Laverton North, which has over 10% of its vehicles as electric. Laverton north which despite having 4600 Vehicles, is located near a former Toyota manufacturing plant, now a hydrogen refuelling facility. Toyota Mirai sedans were distributed locally as part of the company’s trial program, with easy access to EV refuelling at Toyota’s nearby station