Predicting solar installation rates using demographic data

Anthony Liu

February 19, 2020

1 Introduction / Business Problem

The central problem that the analysis in this report will attempt to address is "Can we predict solar installation rates in the different local government areas of Australia using demographic data?". As the effects of global climate change are becoming better understood, the move towards renewable energy sources has become an important focus [1]. Solar energy is emerging as one of the most popular forms of renewable energy for reasons such as decreasing costs, environmental ethics, health, government incentives and accessibility [2, 3].

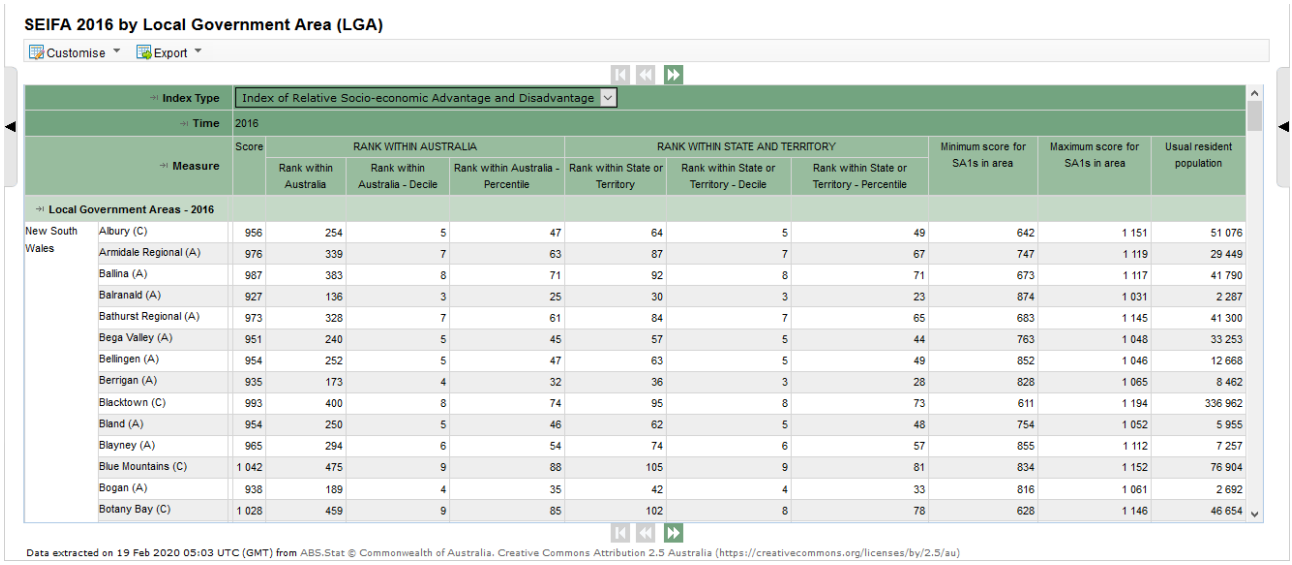
Australia is in a particularly privileged position to have some of the best solar energy resources in the world [4]. Being able to answer the proposed question may lead to insights which guide policy, investment or further research to better utilise this natural resource. Given that installation of solar constitutes a purchase and purchases are made by individuals under their own unique circumstances, there is good reason to believe that demographic factors (such as financial status, dwelling structure, education, etc.) may affect installation rates. In this analysis we limit our scope to prediction of solar installation rates in each local government area (LGA) based on available demographic data. Questions of inference such as which specific demographic factors are related to installation rates are not considered.

Parties who may be interested in this problem include policy makers, investors and researchers. Reliable predictions on installation rates may guide projections and highlight both the level and content of policy intervention justified in affecting solar uptake. Also, investors in the solar industry hold an advantage if they have reliable guidance on which local government areas in the future (accounting for demographic shifts) are likely to install solar. Finally, results from attempting to answer this question may lead to new lines of inquiry among researchers (e.g. if we can indeed predict installation rates then that provides justification to then do further research on which exact demographic factors have an effect).

2 Data

To answer our problem we will consider two separate types of data: data regarding solar installation rates in each LGA and data regarding demographics in each LGA. We obtain our first dataset on solar installations from the Australian Photovoltaic Institute [5]. The Australian Photovoltatic Institute (APVI) is a not-for-profit company whose stated objective is to "Support the increased development and use of PV via research, analysis and information" [6]. The photovoltaic (PV) data by the APVI was compiled from Australian government body sources: PV installation data came from the Clean Energy Regulator [5, 7] and attached LGA data was from the Australian Bureau of Statistics (ABS) [5, 8]. Of particular interest to this analysis is the density variable, an estimate of the percentage of dwellings in each LGA which have installed solar systems.

Detailed demographic data on each LGA including socio-economic indicators, dwelling structure, age, sex, cultural background, education and employment was also obtained from the ABS [9]. In addition, we also attempt to include geographical features of each LGA using data from the Foursquare API. Figures 1 and 2 provide a preview of how the socio-economic indicators and education/employment data look on the ABS website respectively. These entries will later be wrangled into an appropriate format and act as individual features/predictors of a machine learning algorithm to predict density (i.e. solar installation rate in each LGA).



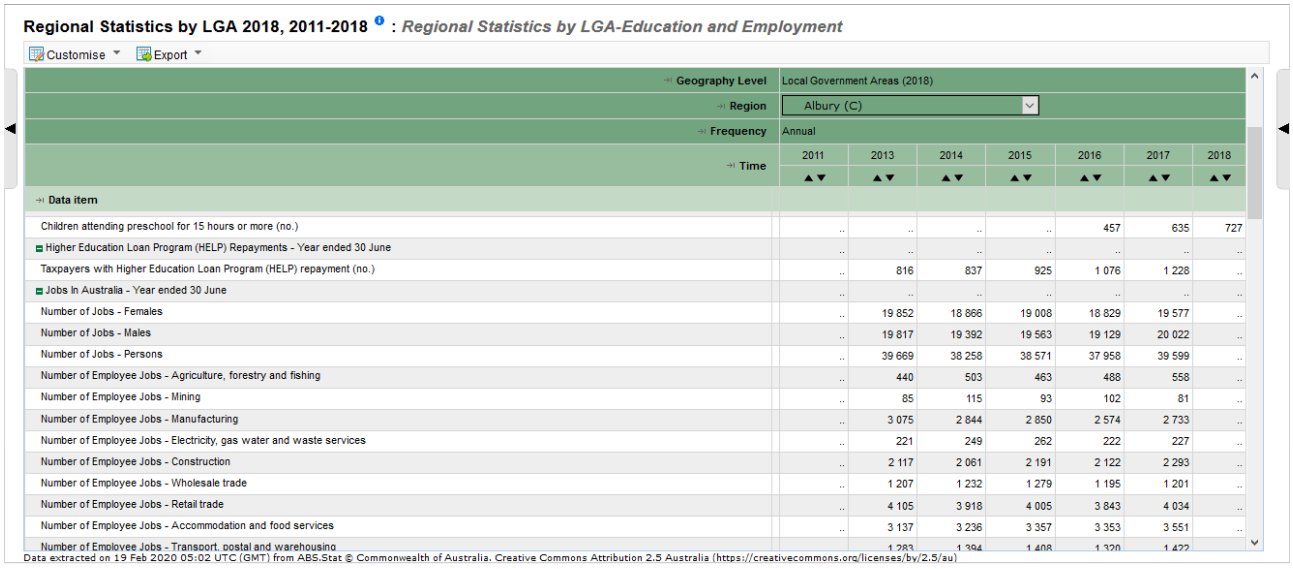
SEIFA 2016 by Local Government Area (LGA)

Customise Export

Index Type	Index of Relative Socio-economic Advantage and Disadvantage									
Time	2016									
Measure	Score	RANK WITHIN AUSTRALIA			RANK WITHIN STATE AND TERRITORY			Minimum score for SA1s in area	Maximum score for SA1s in area	Usual resident population
		Rank within Australia	Rank within Australia - Decile	Rank within Australia - Percentile	Rank within State or Territory	Rank within State or Territory - Decile	Rank within State or Territory - Percentile			
Local Government Areas - 2016										
New South Wales	Albury (C)	956	254	5	47	64	5	49	642	1 151	51 076
	Armidale Regional (A)	976	339	7	63	87	7	67	747	1 119	29 449
	Ballina (A)	987	383	8	71	92	8	71	673	1 117	41 790
	Balranald (A)	927	136	3	25	30	3	23	874	1 031	2 287
	Bathurst Regional (A)	973	328	7	61	84	7	65	683	1 145	41 300
	Bega Valley (A)	951	240	5	45	57	5	44	763	1 048	33 253
	Bellingen (A)	954	252	5	47	63	5	49	852	1 046	12 668
	Berrigan (A)	935	173	4	32	36	3	28	828	1 065	8 462
	Blacktown (C)	993	400	8	74	95	8	73	611	1 194	336 962
	Bland (A)	954	250	5	46	62	5	48	754	1 052	5 955
	Blayney (A)	965	294	6	54	74	6	57	855	1 112	7 257
	Blue Mountains (C)	1 042	475	9	88	105	9	81	834	1 152	76 904
	Bogan (A)	938	189	4	35	42	4	33	816	1 061	2 692
	Botany Bay (C)	1 028	459	9	85	102	8	78	628	1 146	46 654

Data extracted on 19 Feb 2020 05:03 UTC (GMT) from ABS.Stat © Commonwealth of Australia. Creative Commons Attribution 2.5 Australia (https://creativecommons.org/licenses/by/2.5/au)

Figure 1: A preview of the socio-economic indicators dataset on the ABS website, prior to exporting as a CSV file and performing data wrangling



Regional Statistics by LGA 2018, 2011-2018 : *Regional Statistics by LGA-Education and Employment*

Customise Export

Geography Level	Local Government Areas (2018)						
Region	Albury (C)						
Frequency	Annual						
Time	2011	2013	2014	2015	2016	2017	2018
Data item							
Children attending preschool for 15 hours or more (no.)	..	..	..	..	457	635	727
Higher Education Loan Program (HELP) Repayments - Year ended 30 June	..	..	..	..	..	..	..
Taxpayers with Higher Education Loan Program (HELP) repayment (no.)	..	816	837	925	1 076	1 228	..
Jobs In Australia - Year ended 30 June	..	..	..	..	..	..	..
Number of Jobs - Females	..	19 852	18 866	19 008	18 829	19 577	..
Number of Jobs - Males	..	19 817	19 392	19 563	19 129	20 022	..
Number of Jobs - Persons	..	39 669	38 258	38 571	37 958	39 599	..
Number of Employee Jobs - Agriculture, forestry and fishing	..	440	503	463	488	558	..
Number of Employee Jobs - Mining	..	85	115	93	102	81	..
Number of Employee Jobs - Manufacturing	..	3 075	2 844	2 850	2 574	2 733	..
Number of Employee Jobs - Electricity, gas water and waste services	..	221	249	262	222	227	..
Number of Employee Jobs - Construction	..	2 117	2 061	2 191	2 122	2 293	..
Number of Employee Jobs - Wholesale trade	..	1 207	1 232	1 279	1 195	1 201	..
Number of Employee Jobs - Retail trade	..	4 105	3 918	4 005	3 843	4 034	..
Number of Employee Jobs - Accommodation and food services	..	3 137	3 236	3 357	3 353	3 551	..
Number of Employee Jobs - Transport, postal and warehousing	..	1 283	1 394	1 408	1 320	1 422	..

Data extracted on 19 Feb 2020 05:02 UTC (GMT) from ABS.Stat © Commonwealth of Australia. Creative Commons Attribution 2.5 Australia (https://creativecommons.org/licenses/by/2.5/au)

Figure 2: A preview of the education and employment dataset on the ABS website, prior to exporting as a CSV file and performing data wrangling

References

[1] https://www.ipcc.ch/report/renewable-energy-sources-and-climate-change-mitigation/

[2] https://www.pewresearch.org/fact-tank/2016/10/05/americans-strongly-favor-expanding-solar-power-to-he

[3] https://www.seia.org/solar-industry-research-data

[4] https://www.ga.gov.au/scientific-topics/energy/resources/other-renewable-energy-resources/solar-energy

[5] Australian PV Institute (APVI) Solar Map, funded by the Australian Renewable Energy Agency, accessed from pv-map.apvi.org.au on 19 February 2020. Dataset can be downloaded from https://pv-map.apvi.org.au/historical#4/-26.67/134.12

[6] http://apvi.org.au/about-us/

[7] http://www.cleanenergyregulator.gov.au/RET/Forms-and-resources/Postcode-data-for-small-scale-installa

[8] https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.003July%202016?OpenDocument

[9] http://stat.data.abs.gov.au/