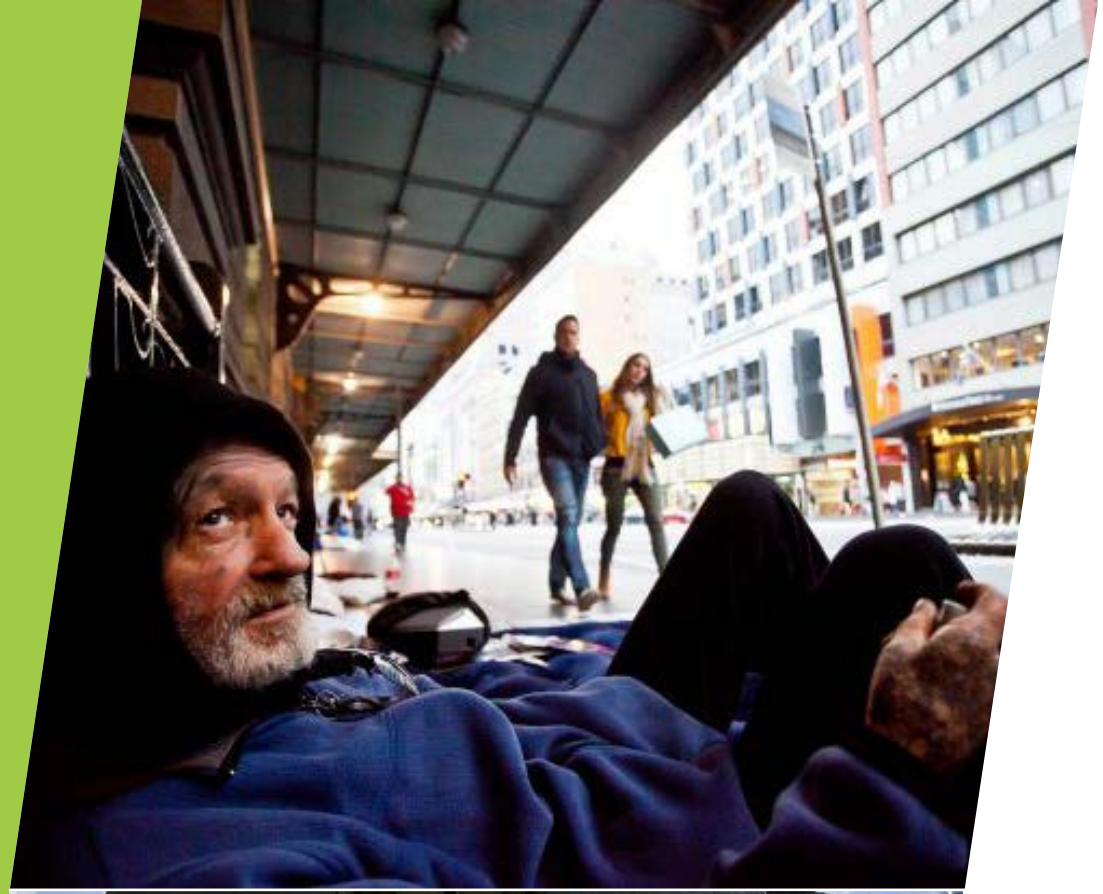


Homelessness in Victoria

By Armaan Dhaliwal-McLeod

What is my research question?

- Are socioeconomic and demographic factors causing homelessness in Victoria's Local Government Areas?

- Why is it worth tackling?

  Homelessness is a growing epidemic in Victoria, and this study seeks to use open data to discover a variety of reasons causing this issue. Not only will this study find correlations with common reasons, it will discover non-intuitive reasons which potentially contribute to this problem

- What features will be linked with homelessness?

  Socioeconomic and demographic issues such as: social housing, drug usage, affordable rental housing, unemployment, cultural diversity, education, social support and population.

What datasets did I use?

- Dataset 1: Local Government Area(LGA) profiles data 2015 –csv
- Dataset covers a wide range of information concerning LGAs, including topics such as diversity, population and social engagement.
- Only relevant information concerning the features to be investigated was extracted from this dataset.

lga_code	lga_name	ppl_born_overseas_perc	social_housing_dwellings	homeless_ppl_est_per_1000_pop
20110	Alpine (S)	15.7	98	3.5
20260	Ararat (RC)	9.2	155	2.5
20570	Ballarat (C)	9.1	2445	4.1
20660	Banyule (C)	23.1	2336	3.7
20740	Bass Coast (S)	15.8	399	1.8

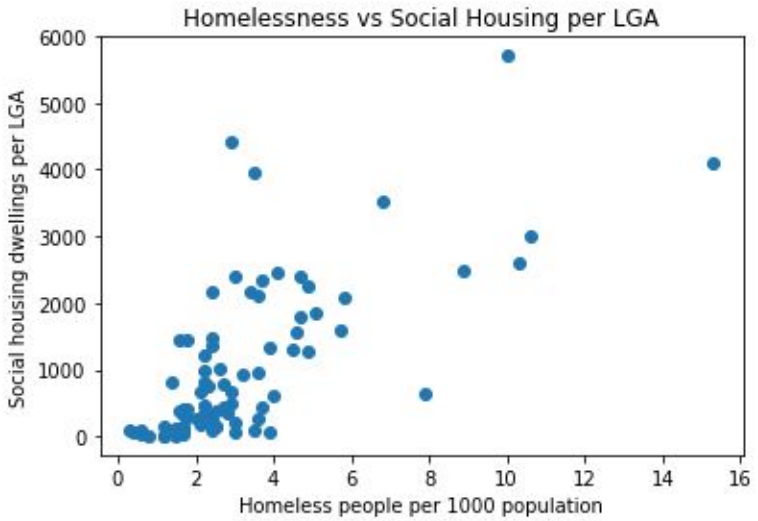
- Dataset 2: VIF 2015 – LGA Yearly Estimated Population 2011-2031
- Dataset gives the yearly estimated populations per LGA.
- Only 2015 was taken, in order to make years between both datasets consistent

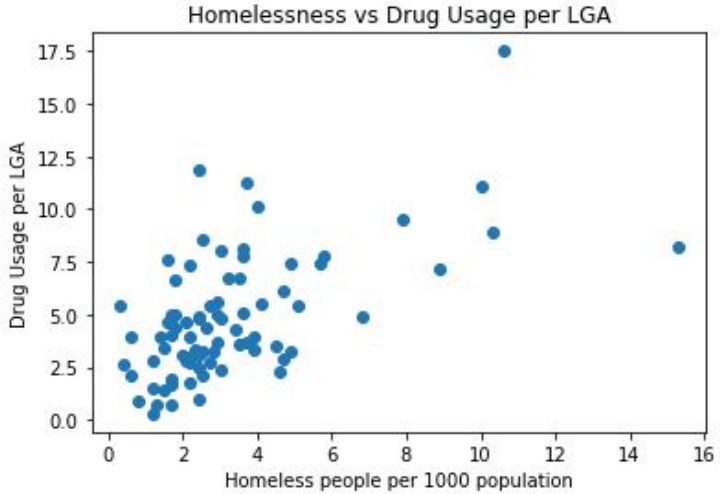
lga_code	lga_name	tpop_2015
20110	Alpine (S)	11997
20260	Ararat (RC)	11236
20570	Ballarat (C)	102260
20660	Banyule (C)	126578

Data Wrangling Methodologies

- Firstly, many scatter plots were used to find correlations between homelessness and the features.

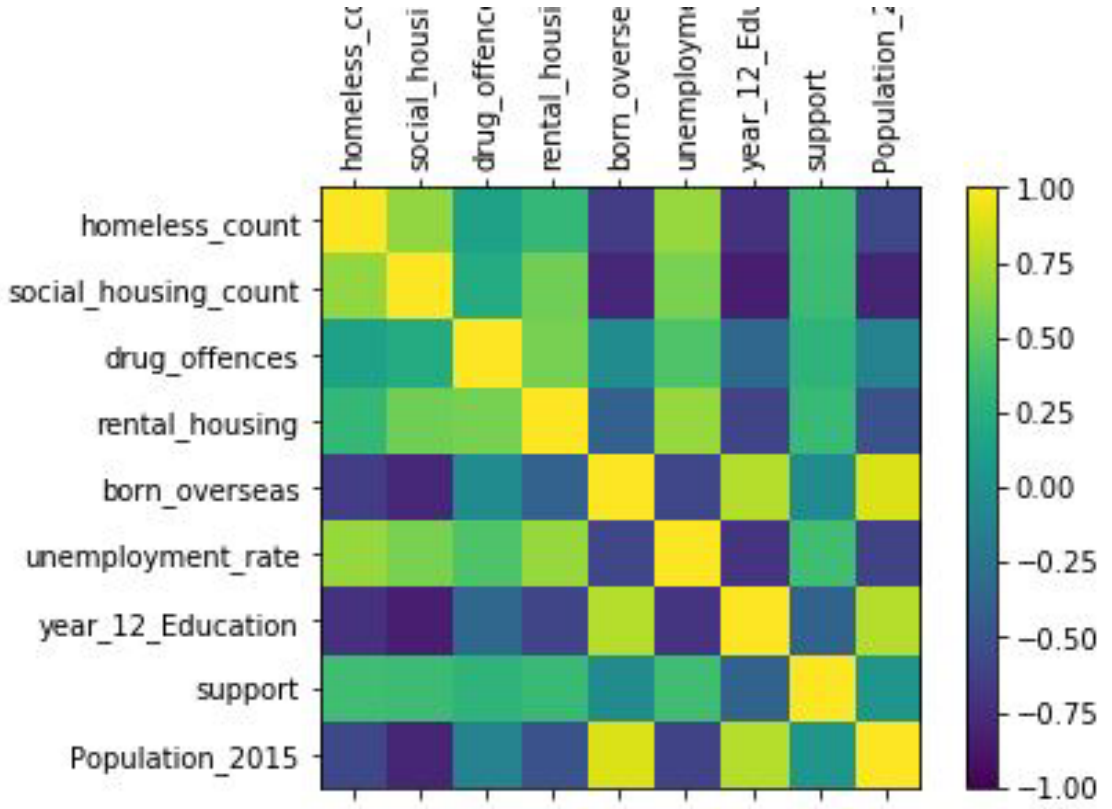
- These were helpful with making sure the project was feasible.





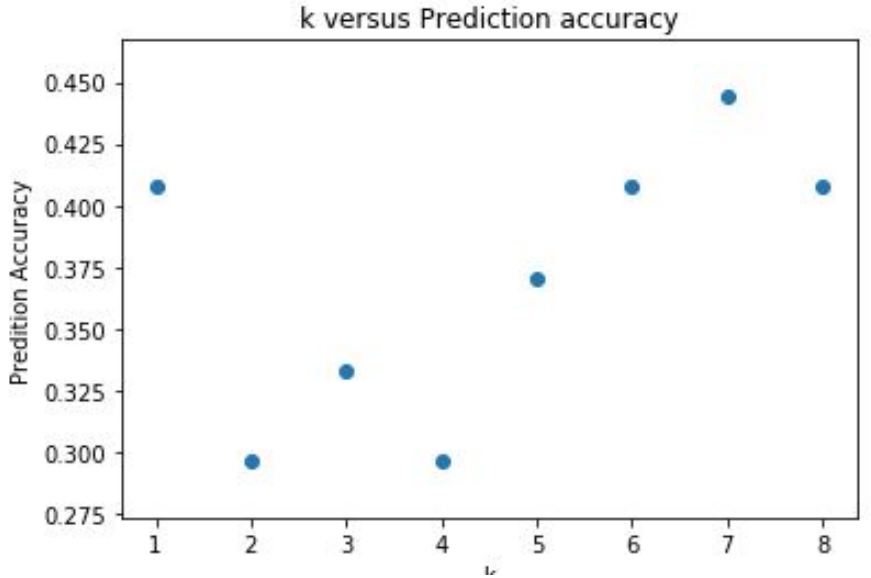
Correlation Matrix

- A correlation matrix was also used to see the correlations between two variables.
- The line of interest here is the top, where moderate to strong correlations are shown with homelessness.



K Nearest Neighbors

- Also used K-NN to make predictions about the data found, and examine how accurate the correlations are, in terms of prediction accuracy
- To the side is a graph measuring the most optimal k value for a K-NN model.



What did I find?

- I found many correlations between homelessness and socioeconomic factors. Not only were their correlations between mainstream causes of homelessness, such as drug usage and unemployment, other factors such as social support and ethnicity also had high correlations. This was also backed up with the K-NN modelling, whereby the highest prediction accuracy found was 0.48, demonstrating that there is correlation in the data.
- This was interesting, as it is clear that the data is showing a trend in homelessness, and that there are many factors contributing to this problem.
- I learnt that data wrangling is very helpful for looking deeper into problems. This project has shown me that you can't rule out any possibilities until you analyze the data, and apply useful methods such as preprocessing, integration and visualizations to bring out the correlations in data.

What have been my Challenges?

- My biggest challenge was creating a clear, formulated research question that was is in scope and beneficial to the Victorian people. I had many problems with this, and it lead to my study being misleading and confusing at times.
- Using higher level models such as K-NN means to show correlations in the data. Even now, I could have extended the k-NN model to include decision trees to help reinforce my answer to the question.
- I could have included more methods on outlier detection, such as removing LGAs like Melbourne and Yarra, as they always have very high results in any correlation, and could possibly create more accurate results.