DATA2001 Report

500349199

480361420

490272330

I. Dataset Description

Traffic Data: Traffic data for the state of New South Wales was obtained through the NSW Government body, Transport for NSW. They made data in the form of geoJSON available for public download. We cleaned this file using GeoPandas, dropping columns that weren't needed for our calculation before loading the data frame into our database.

Statistical Area Level 2 (SA2) Data: The geometric data for the various areas that are considered SA2 data was obtained through the Australian Bureau of Statistics (ABS). Before loading into the database, we cleaned the shapefile data using GeoPandas. We created a GeoPandas dataframe and then dropped the columns that would not assist in further analysis.

Test Sites Data: The Covid-19 test sites data was obtained through the University of Sydney's Canvas site. Before loading it into our database, the CSV was cleaned and columns centre name, phone number, and opening hours were removed as they weren't needed further for analysis.

Postcodes Data: The postcodes CSV was obtained through the University of Sydney's canvas site. No cleaning of the file needed to occur beforehand and it was simply read into our database using Pandas.

Health Service Data: The health services data was obtained through the University of Sydney's canvas site. We removed the columns address, state, comment, and website as they weren't needed further. After moving the CSV into a Pandas dataframe, we dropped NA values for num_beds before reading the dataframe to SQL in order to assist future analysis.

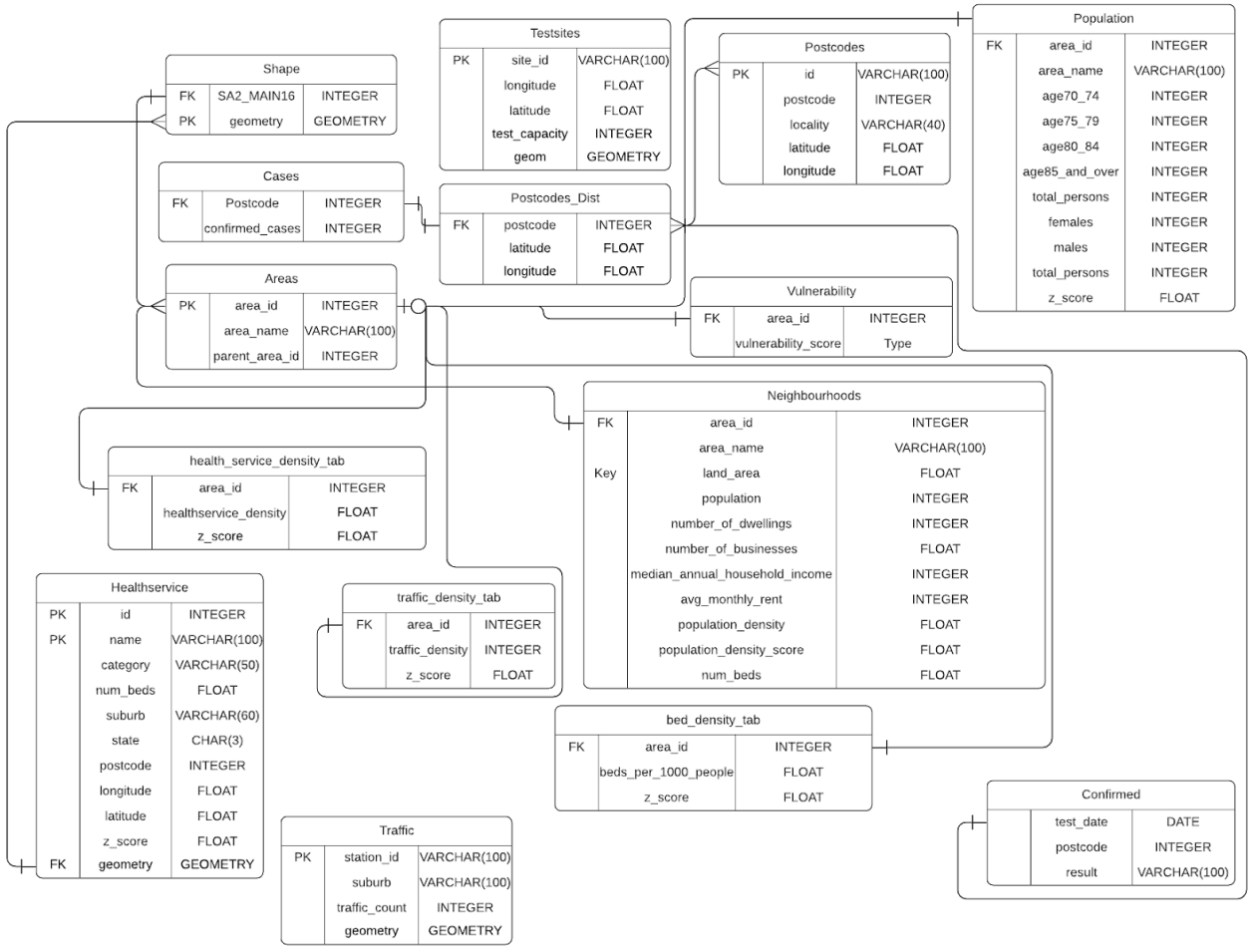
Population Data: The population data was obtained through the University of Sydney's canvas site. Using Pandas, columns for ages 0-69 were dropped as they weren't used further.

Neighbourhoods Data: The neighbourhoods dataset was obtained as a CSV file through the University of Sydney's canvas website. We pre-processed the neighbourhoods dataset by dropping rows null values from our Pandas dataframe before loading it into our table. This

was done due to null values causing continuous issues with the analysis neighbourhoods was required for.

Covid Cases Data: This dataset was obtained through the Australian Government's Open Data collection [1]. As we only need confirmed cases, after loading the CSV file into a Pandas dataframe, we dropped all rows that weren't for a confirmed case. Not only did this make future use of the data simpler, but it optimised the processing of the data also.

II. Database Description



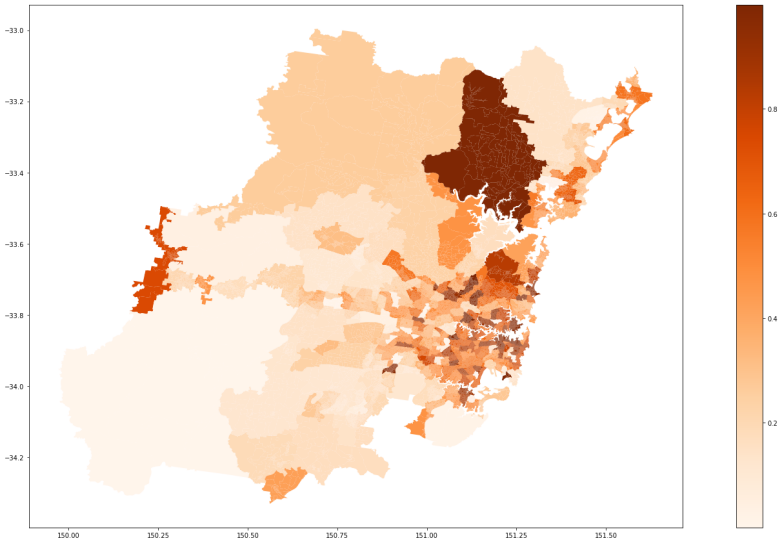
Two spatial indexes were created to assist in speeding up queries requiring a spatial join. One was created for the geometry column of the Shape table and one was created for the geom column (containing a geometric point) for Healthservice.

III. Vulnerability Score Analysis

For our vulnerability score analysis, we used the following equations:

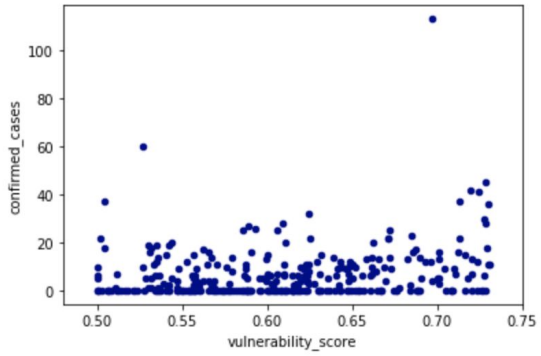
Vulnerability = *S(z(population_density) + z(population_age) – z(healthservice density) – z(hospital_bed density) + z(traffic_density))*

Where ***z(measure, x)*** = $\frac{x - avg_{measure}}{stdev_{measure}}$, **S** = the sigmoid function

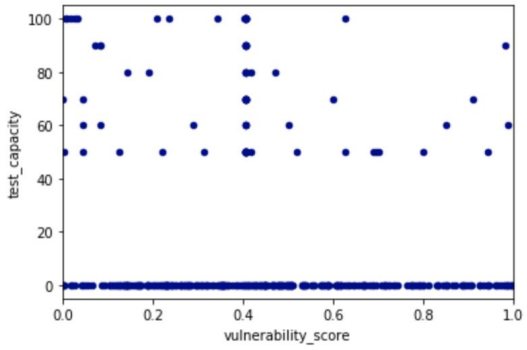


The above heat map was generated for the suburbs of the greater Sydney region. The darker the colour, the higher the calculated vulnerability score. The list of the top 5 most vulnerable suburbs (in descending order) are therefore, Wahroonga, Ryde, Woolloomooloo, Sydney, and Bondi Junction. The least vulnerable suburbs (in ascending order) are St Leonards, Westmead, Kogarah, Randwick, and Liverpool. At first glance, it is possible the high density of health services (and therefore beds) in Westmead and St Leonards may have skewed their vulnerability according to our score.

IV. Correlation Analysis



This graph suggests a potential correlation between our vulnerability score and the more confirmed cases due to the visually apparent slight upward trend. However, the correlation coefficient for our vulnerability score and the confirmed cases is -0.156198. As it is < 0.3, we can consider that our score does not correlate with the actual Covid-19 data in a meaningful way despite the slight trend in the graph. This result is not in accordance with our hypothesis. One possible reason is that our vulnerability equation does not consider criteria importance accurately. It is possible one factor has more of an impact and thus should be considered more strongly in our equation. Another possible reason is that we don't consider the behaviours differences such as propensity for overseas travel which was a central tenant in most of Australia's Covid-19 cases.



When graphed, test capacity and vulnerability score have no apparent trend. The correlation coefficient for our vulnerability score and the confirmed cases is 0.223855. As it is < 0.3, we can consider that our score does not correlate with the test capacity data. In Australia, the majority of Covid-19 cases (62.2%) were acquired overseas [2]. Therefore, a suburb is not necessarily made excessively vulnerable based on traits such as population density but rather through a combination of those factors combined with whether the members of that suburb travel frequently. In the future, it would be worth looking at our vulnerability score when compared to only confirmed cases acquired locally.

Table 1: various combinations of measures and their correlation coefficients

Target1	Target2	Correlation Coefficient
test capacity	vulnerability	-0.154388
test capacity	confirmed cases	0.109034
population density	confirmed cases	0.198109
population age	confirmed cases	-0.052365
healthservice	confirmed cases	-0.056348
hospital_bed	confirmed cases	-0.024608
traffic density	confirmed cases	0.19174

In order to rule out some data correlating but the score being skewed by the measures that didn't, we calculated the correlation between confirmed cases and the individual measures of the equation. What can be seen in table 1 is that none of our measures correlate. This suggests that these measures aren't the core determinants in suburb vulnerability for cases.

V. Conclusions

Overseas travel and close contact with those who travel seem to be the biggest factors. We also can't forget incidents such as the Ruby Princess [3] which led to locally acquired cases not necessarily based on long-term traits of the suburb the infected individual lived in. Additionally, due to the reasonably quick containment of the virus, we dealt less with a widespread outbreak and more with clusters in specific areas. This means the virus had limited freedom to spread based on external factors outside of close contact with an overseas traveller. Perhaps, our vulnerability score would lead to stronger correlations in countries with higher case numbers and worse containment.

Very little is known about Covid-19 at the moment and therefore information about how individual immunity and environmental preference (e.g. hot or cold) may impact case numbers. With regard to individual immunity, it is also possible that those with stronger immune systems did not seek out medical treatment due to mild symptoms. Until we know more medically about Covid-19, it will be challenging to identify most at-risk areas. For now, it is best to focus on protecting the most vulnerable individuals, and once we understand more about the virus, study what made areas vulnerable allowing for a better future pandemic response plan.

VI. References

1. Data.gov.au. 2020. NSW COVID-19 Tests By Location And Result. [online] Available at: <https://data.gov.au/dataset/ds-nsw-5424aa3b-550d-4637-ae50-7f458ce327f4/details?q=>.
2. Australian Government Department of Health. 2020. Coronavirus (COVID-19) Current Situation And Case Numbers. [online] Available at: <https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers#cases-by-source-of-infection>.
3. Zhou, N., 2020. Anatomy Of A Coronavirus Disaster: How 2,700 People Were Let Off The Ruby Princess Cruise Ship By Mistake. [online] the Guardian. Available at: <https://www.theguardian.com/world/2020/mar/24/anatomy-of-a-coronavirus-disaster-how-2700-people-were-let-off-the-ruby-princess-cruise-ship-by-mistake#maincontent>.