Neighbourhood Cyclability Analysis

Group 152

1. Dataset Description

Data Sources

Five datasets were included to compute the Sydney cyclability score. Four were provided - population density, dwelling density (from Neighbourhoods.csv), service balance (BusinessStats.csv) and bike pod density (BikeSharingPods.csv). The fifth was sourced from the NSW Crime Tool API, which mapped the crime counts, population and boundaries of all suburbs in NSW. Spatial data of each neighbourhood was obtained from the Australian Bureau of Statistics (ABS) under ASGS Statistical Area Level 2 (SA2) 2016. Median income and average monthly rent used to perform correlation analysis with the cyclability score was sourced from CensusStats.csv.

Obtaining and Pre-processing Data

In order to obtain the data for the crime counts in each area, the API url for the NSW Crime Tool was found using the Network tab of developer tools (Chrome). The website used 2 API calls, one to obtain the data such as crime count and population, the other to obtain the geospatial data for each suburb, which can be queried just using the RegionId. Parameters for the API calls were found in the Network tab (Appendix 1). The HTML page contained all possible ids for the parameters, and we filter for all attributes labelled suburb (Appendix 2). The suburb list was scrapped from Wikipedia's list of Sydney suburbs to filter for queries to only include suburbs in Sydney. https://en.wikipedia.org/wiki/List_of_Sydney_suburbs. The crime count was calculated by aggregating the 2018_count of all offences in every suburb. All of the geodata and crime counts were converted into a geojson.FeatureCollection (see data2901_assignment-webscrapping.ipynb).

In the four provided datasets, missing values were replaced with the corresponding mean value of all neighbourhoods. In addition, 2 entries with a longitude and latitude of (0, 0) in BikeSharingPods.csv was omitted.

Moreover, Machine Learning was applied to enrich the crime data scrapped from the web. The data is not as detailed as required, because some areas are divided into smaller pieces in this analysis. Hence, a decision tree regression model was used to fill the data. Population, number of businesses, median annual income and average monthly rent were chosen as features of the model. Data for areas with existing crime counts were used to train the model. After training, the model made predictions for the crime count of that neighbourhood.

2. Database Description

Database Schema

The five datasets and other data produced in the analysis phrase are stored in the database (y19s1d2x01_ysun8114), whose schema is shown below:

crimefeatures

regionid	int
count	int
pop	int
name	varchar
geom	MultiPolygon

bikesharingpods

station_id	int
name	int
num_bikes	int
num_scooters	int
description	text
geom	point

aust

sa2_main16	int
sa2_5dig16	int
sa2_name16	varchar
sa3_code16	int
sa3_name16	varchar
sa4_code16	int
sa4_name16	varchar
gcc_code16	varchar
gcc_name16	varchar
ste_code6	int
ste_name6	varchar
areasqkm16	int
geom	polygon

censusstats

area_id	int
median_annual_household_income	int
avg_monthly_rent	int

businessstats

area_id	int
num_buinesses	int
retail_trade	int
accommodation_and_food_services	int
health_care_and_social_assistance	varchar
art_and_recreation_services	varchar

cyclabilityscore

area_id	int
cyclability_score	double

predictedresult

area_id	int
population	int
crime_count	int

statisticalareas

area_id	int
area_name	varchar
parent_area_id	int

neighbourhoods

area_id	int
area_name	varchar
land_area	double
population	int
number_of_dwellings	int
number_of_businesses	int

zbikepods

area_id	int
z_bikepods_density	double

zcrime

area_id	int
z_crime	double

zdwelling

area_id	int
z_dwelling_density	double

zpopulation

area_id	int
z_population_density	double

zservice

area_id	int
z_service_balance	double

Latitude and longitude values were converted into a point data type in Postgres, using the WGS84 spatial reference system (SRID 4326).

Indexing

Indexes for BikeSharingPods.geom and Aust.geom is indexed using GiST. GiST is chosen as due to its suitability in indexing multi-dimensional spatial data (such as the polygons in Aust.geom). The index will perform spatial join queries more quickly by filtering and refining bounding boxes of Aust.geom to determine which neighbourhood the bike pod point is contained within. Adding the two indexes improved the query time from 18 seconds to 8 seconds in our database.

3. Cyclability Analysis

Formula

Five attributes were used to measure the cyclability:

1. Crime rate inversed

As crime rates have an inverse relationship with cyclability, we used the formula crime rate (inv) $= \frac{\text{population}}{\text{count}}$. As there were more suburbs than neighbourhoods, the suburbs were joined using the suburb's centroid.

2. Bike pod density

   The number of bikes within each neighbourhood is calculated using a spatial join with aust.geom grouped by area_id. Bike density $= \frac{\text{numbikes}}{\text{Aust.areasqkm16}}$.

3. *Population density*

   The population density was calculated by dividing the population by the area of the neighbourhood.

4. Dwelling density

   The calculation of the dwelling density is similar to the population density, that is dividing the number of dwellings by the area of the neighbourhood.

5. Service balance

   Service balance is calculated by first calculating the weight each type of business.

   $$\text{sum} = \text{avg(num businesses)} + \text{avg(retail)} + \text{avg(accomodation/food)} + \text{avg(health care)} + \text{avg(education)} + \text{avg(arts)}$$

   $$\text{weight} = \frac{\text{avg(type\_of\_business)}}{sum}$$

   Then the balance of services can be combined into one value:

   $$\Sigma(\frac{\text{type\_of\_business} \times \text{weight}}{sum})$$

The formula used to calculate the z-score is: $\frac{\text{x}-\text{mean}}{\text{stddev}}$ for each attribute.

Each of the z-scores were stored in a separate table in the database and the cyclability score computed according to the formula:

$$\text{cyclability} = \text{z(population density)} + \text{z(dwelling density)} + \text{z(service balance)} + \text{z(bikepod density)} + \text{z(crime rate inversed)}$$

Results

As seen in our cyclability score, the majority of the scores seems to cluster at -3.8 to 3 (which is 8.4 and 19.4 when mapped into 100-scale), with some extremely high values likely influenced by the population and dwelling densities near the city. The areas with the highest cyclability score seem to cluster around near the city centre of Sydney, gradually reaching lower scores in the outer suburbs. This is to be expected as the population, dwelling and service balance is likely to have much lower z-scores the further away from the inner suburbs. Many outer suburbs exhibit high crime rates, lowering the cyclability score. It can be seen Sydney is most cyclable in the inner west and northern suburbs. Some suburbs further away from the city centre with low population and dwelling densities has very low crime rates (such as St Ives and Menai-Lucas Heights -Woronora) which significantly increased the cyclability score in our analysis.

Cyclabilty Score in Sydney
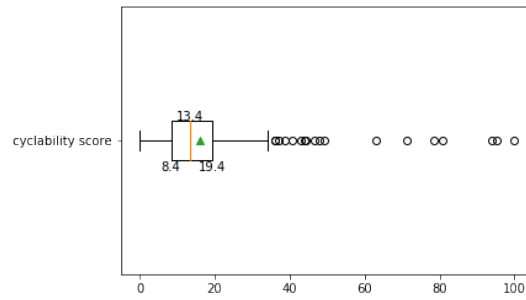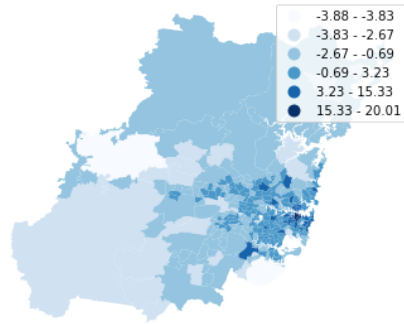
4. Correlation Analysis

Th cyclability score is not highly correlated to neither of the income and the rent. The correlation coefficient of income and cyclability score is 0.24, meaning that there is just a weak correlation between them. Meanwhile, the correlation coefficient of rent and cyclability score is 0.4, which is slightly higher and indicates a moderate correlation. Thus, to a certain degree, the higher the rent is, the more cyclable the area is.

The scatter plots for the income and the rent are shown below:
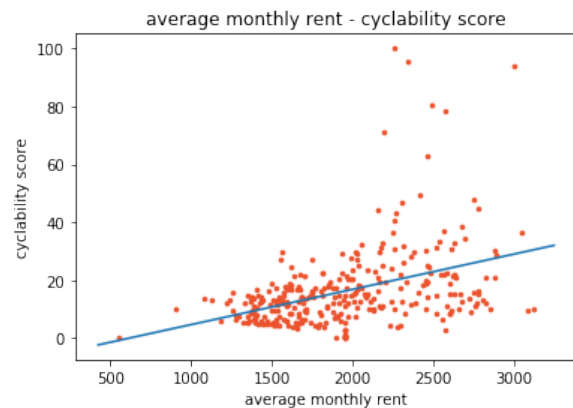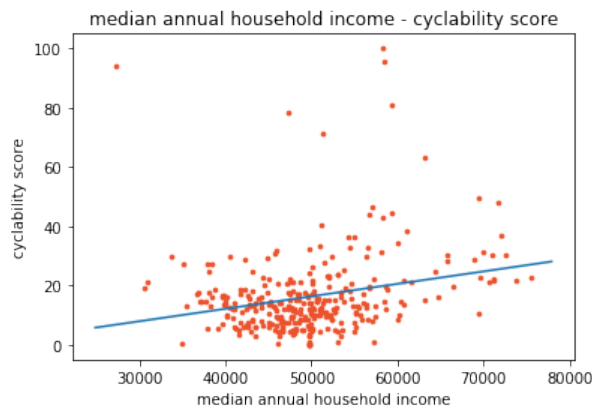
The correlation coefficient matrix is shown below:
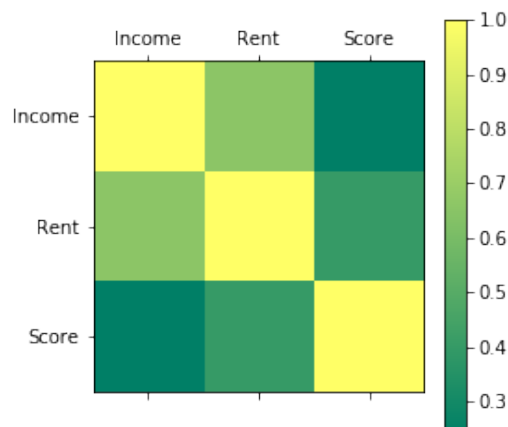
Appendix

Appendix 1

▼ Query String Parameters

    RegionTypeId: 1

    End_Year: 2018

    Start_Year: 2017

    Month: 12

    DataType: I

    OffenceId: 1400011

    RegionId: 13800

Appendix 2

```
<option value="34385" title="3">4385 (Postcode)</option>

<option value="20001" title="2">AARONS PASS (Suburb)</option>

<option value="20002" title="2">ABBOTSBURY (Suburb)</option>
```