

Townsville Suburb Clustering

Introduction

Description and Discussion of Background

Townsville is a major regional centre in Queensland, Australia. It has a population of over 193,000. It is a major government, commercial and defence centre. The Townsville local government area covers 3,736 square kilometres. The area considered in this analysis is about 180 sq kms. Population density in the urban area is just over 1060 persons per sq km. It is generally considered that Townsville is largely the same in its economic and social structure.

This project will seek to cluster the urban suburbs of Townsville to examine this issue. The features used in the cluster analysis can broadly be considered as Lifestyle factors. They are: Socio-Economic Indexes; Access to Public Transport; Animal Complaints; Crime Data; Access to Council Community Facilities; Access to Community Venues.

The variability in a city's population is an important consideration when planning its social and economic development. Town planning may target some programs to certain suburb features, that would not work, or be unnecessary, in other suburbs. For example, the installation of cameras used to monitor for crime. The insights gained from this project would be useful to people moving to or moving within the Townsville District. It can give people an idea of commonalities and differences between suburbs.

Data and Methodology

Data was sourced from several online sites.

Suburb Names

Data Source: https://en.wikipedia.org/wiki/List_of_Townsville_suburbs

Data Preparation: The suburb names were copied from the Wikipedia page into an Excel spreadsheet. This list was then edited to include only the suburbs being considered in this project. Suburbs that were geographically not part of the extended urban area were not included. This data was then read into a pandas dataframe.

```
df.head()
```

Suburb

0	Townsville City
1	Aitkenvale
2	Annandale
3	Belgian Gardens
4	Castle Hill

Geographic Location Data:

Geographic location data was retrieved using the *geopy* client, with the *Nominatim* geocoder for accessing the OpenStreetMap (OSM) dataset. Location data was added to the pandas dataframe.

```
df.head()
```

	Suburb	latitude	longitude
0	Townsville City	-19.26	146.82
1	Aitkenvale	-19.30	146.77
2	Annandale	-19.31	146.78
3	Belgian Gardens	-19.25	146.79
4	Castle Hill	-19.26	146.80

Socio-Economic Data

Data Source: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument>
State Suburb, Indexes, SEIFA 2016

Data Preparation: The data was downloaded as an Excel workbook. The Queensland data (State Suburb Code = 3#####) was copied to a blank spreadsheet. The suburb names from the Townsville Suburbs Excel workbook were copied into the blank spreadsheet. These then became the index keys to extract the four socio-economic indicators from the Queensland data subset lookup table. The formula values were then converted to set values (ie not formula). The table was then read into a pandas dataframe. The suburb name strings were trimmed of trailing and leading spaces.

The 4 Socioeconomic factors included were:

Index of Relative Socio-economic Disadvantage (IRSD)

Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)

Index of Economic Resources (IER)

Index of Education and Occupation (IEO)

A higher number is better.

These indexes loaded into the dataframe as objects, which were converted to numeric values (int64) for processing.

```
se_df.head()
```

	Suburb	rseid	rsead	ier	ieo
0	Townsville City	1055	1059	943	1094
1	Aitkenvale	940	937	915	964
2	Annandale	1077	1068	1069	1051
3	Belgian Gardens	1015	1021	966	1056
4	Castle Hill	1127	1164	1135	1169

This dataframe was then merged with the suburb dataframe (which also contained the geolocation data).

```
df_ec.head()
```

	Suburb	latitude	longitude	rseid	rsead	ier	ieo
0	Townsville City	-19.26	146.82	1055	1059	943	1094
1	Aitkenvale	-19.30	146.77	940	937	915	964
2	Annandale	-19.31	146.78	1077	1068	1069	1051
3	Belgian Gardens	-19.25	146.79	1015	1021	966	1056
4	Castle Hill	-19.26	146.80	1127	1164	1135	1169

Public Transport Availability

Data Source: <https://data.gov.au/dataset/ds-dga-7559d6fa-bb12-4407-85d7-4c42718d4ed7/details?q=bus%20shelter>

Townsville Bus Stops and Shelters (esri shapefile – zipped)

After downloading and unzipping the Bus_Shelters_Townsville.dbf file was loaded into LibreOffice Calc. The data was then copied into an Excel spreadsheet, where unnecessary features were removed.

stop_id	stop_code	stop_name	stop_lat	stop_long	zone_id	Bus_Shelter
890000	890000	Stockland Elizabeth Street	-19.29803300000	146.76437000000	4	Yes
890001	890001	Anne St at Richard Street	-19.29235100000	146.76662600000	4	Yes
890002	890002	Fulham Rd at Vincent State School	-19.28846900000	146.76421800000	4	Yes
890003	890003	Fulham Rd at Vincent Shopping Centre	-19.28818500000	146.76181100000	4	Yes
890004	890004	Fulham Rd at St James Village	-19.28827400000	146.75970300000	4	Yes

This excel file was loaded into a pandas dataframe. Geolocation data was then used to calculate the number of bus stops within 500m of the geolocation point for each suburb.

```
df_ec.head()
```

	Suburb	latitude	longitude	rseed	rseed	ier	ieo	bus_stop
0	Townsville City	-19.26	146.82	1055	1059	943	1094	13.00
1	Aitkenvale	-19.30	146.77	940	937	915	964	1.00
2	Annandale	-19.31	146.78	1077	1068	1069	1051	7.00
3	Belgian Gardens	-19.25	146.79	1015	1021	966	1056	7.00
4	Castle Hill	-19.26	146.80	1127	1164	1135	1169	0.00

Animal (dog) Complaints made to Council

Data Source: <https://data.gov.au/dataset/ds-dga-5a005841-f4f2-4c52-82db-8cce70715d72/details?q=>

The downloaded Excel file listed each incident in chronological order;

Animal Type	Complaint Type	Date Received	Suburb	Electoral Division	Year
dog	Attack	July 2020	Alice River	Division 1	2020
dog	Attack	July 2020	Alice River	Division 1	2020
dog	Attack	July 2020	Alice River	Division 1	2020
dog	Enclosure	July 2020	Alice River	Division 1	2020

This raw data was transformed into a pivot table in Excel. The (financial) year 2020 was selected and dogs were selected.

Animal Type	dog								
Year	2020								
Count of Date Received	Column Labels								
Row Labels	Aggressive Animal	Attack	Enclosure	Noise	Private	Impound	Wandering	Grand Total	
Aitkenvale	13	8	17	31		55	17	141	
Alice River	8	9	10	9		3	6	45	
Alligator Creek	10	4	2				2	18	

This pivot table values were copied to another Excel spreadsheet, which was loaded into a pandas dataframe.

Suburb	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Animal Complaints	Grand Total
Aitkenvale	13	8	17	31	55	17		141
Alice River	8	9	10	9	3	6		45
Alligator Creek	10	4	2			2		18
Annandale	9	7	5	8	12	4		45

Dataset is based on the 6 primary complaint categories of

Aggressive Animal – refers to aggressive dogs

Attack – refers to dog attacks

Enclosure – refers to dogs outside their enclosure or with an inadequate enclosure

Noise – refers to dogs making noise

Private Impound – refers to dogs that need to be picked up and impounded by Council

Wandering – refers to wandering dogs

The animal complaints dataframe was merged with the main dataframe.

```
ac_df.head()
```

	Suburb	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Animal Complaints	Grand Total
0	Aitkenvale	13.00	8.00	17.00	31.00	55.00	17.00		141
1	Alice River	8.00	9.00	10.00	9.00	3.00	6.00		45
2	Alligator Creek	10.00	4.00	2.00	0.00	0.00	2.00		18
3	Annandale	9.00	7.00	5.00	8.00	12.00	4.00		45
4	Arcadia	0.00	3.00	0.00	0.00	0.00	4.00		7

```
df_ec_ac.head()
```

	Suburb	latitude	longitude	rseed	rseed	ier	ieo	bus_stop	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Complaints	Animal Grand Total
0	Townsville City	-19.26	146.82	1055	1059	943	1094	13.00	0.00	2.00	2.00	5.00	4.00	2.00		15
1	Aitkenvale	-19.30	146.77	940	937	915	964	1.00	13.00	8.00	17.00	31.00	55.00	17.00		141
2	Anndale	-19.31	146.78	1077	1068	1069	1051	7.00	9.00	7.00	5.00	8.00	12.00	4.00		45
3	Belgian Gardens	-19.25	146.79	1015	1021	966	1056	7.00	2.00	2.00	4.00	2.00	8.00	2.00		20
4	Castle Hill	-19.26	146.80	1127	1164	1135	1169	0.00	3.00	0.00	0.00	2.00	1.00	1.00		7

Crime Data (July 2019 – June 2020)

Data Source: <https://www.data.qld.gov.au/dataset/crime-locations-2000-present/resource/3a1448e0-649b-4b7c-813b-a5a5bb8ea14e>

Crime data API was used to access Queensland Crime Database by Suburb. Each API request returned a json file with the requested data.

Typical json file result (first few lines): [{ "Type": "Miscellaneous Offences", "Date": "2019-07-01 00:01:00", "Postcode": "4814", "Area of Interest": "Aitkenvale" }, { "Type": "Other Theft (excl. Unlawful Entry)", "Date": "2019-07-01 15:15:00", "Postcode": "4814", "Area of Interest": "Aitkenvale" }

Crime data was extracted from the json file, categorised as a property crime or a person crime, with the respective totals for each added to the suburb record in the main dataframe.

```
df_ec_ac.head()
```

Suburb	latitude	longitude	r sed	r sead	ier	ieo	bus_stop	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Animal Complaints Grand Total	prop_cc	pers_cc
Townsville City	-19.26	146.82	1055	1059	943	1094	13.00	0.00	2.00	2.00	5.00	4.00	2.00	15	476.00	1686.00
Aitkenvale	-19.30	146.77	940	937	915	964	1.00	13.00	8.00	17.00	31.00	55.00	17.00	141	1146.00	589.00
Annandale	-19.31	146.78	1077	1068	1069	1051	7.00	9.00	7.00	5.00	8.00	12.00	4.00	45	230.00	128.00
Belgian Gardens	-19.25	146.79	1015	1021	966	1056	7.00	2.00	2.00	4.00	2.00	8.00	2.00	20	172.00	232.00
Castle Hill	-19.26	146.80	1127	1164	1135	1169	0.00	3.00	0.00	0.00	2.00	1.00	1.00	7	37.00	4.00

Council Community Facilities Were Counted for each Suburb

Data Source: <https://data.gov.au/dataset/ds-dga-efa565ea-68cc-4c1d-a298-d5d81afc4343/details?q=>

The geojson file was used as it contained geolocation information for each venue.

Typical json file result (first few lines):

```
{"type":"FeatureCollection","features":[{"type":"Feature","id":"ckan_efa565ea_68cc_4c1d_a298_d5d81afc4343.1","geometry":{"type":"Point","coordinates":[146.81429389,-19.25824839]},"geometry_name":"geom","properties":{"objectid":1,"facility_i":1,"name":"Townsville City Council Customer Service Centre","category":"Administrative","address":"103 Walker Street, Townsville","disability":"Yes","disabili_1":"Ramp, Lift","accessible":"Onsite","accessib_1":"There is one designated accessible parking space which is located in the rear car park. There are accessible parking spaces available in Walker Street and Wills Street","accessib_2":"There is one designated unisex accessible toilet on the ground floor in the forecourt"}}, {"type":"Feature","id":"ckan_efa565ea_68cc_4c1d_a298_d5d81afc4343.2","geometry":{"type":"Point","coordinates":[146.81786577,-19.25824839]}}
```

The category of facility and the geolocation information for each facility were extracted into a pandas dataframe.

```
cf_df.head()
```

	category	fac_lat	fac_long
0	Administrative	-19.26	146.81
1	Library	-19.26	146.82
2	Theatre	-19.27	146.81
3	Theatre	-19.27	146.81
4	Community Centre	-19.27	146.81

The geolocation information was then used to count the number of Council facilities within 1000m of each Suburb geolocation. This total was added to the main dataframe.

```
df_ec_ac.head()
```

atitude	longitude	rsed	rsead	ier	ieo	bus_stop	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Animal Complaints Grand Total	prop_cc	pers_cc	council_fac
-19.26	146.82	1055	1059	943	1094	13.00	0.00	2.00	2.00	5.00	4.00	2.00	15	476.00	1684.00	4.00
-19.30	146.77	940	937	915	964	1.00	13.00	8.00	17.00	31.00	55.00	17.00	141	1146.00	589.00	1.00
-19.31	146.78	1077	1068	1069	1051	7.00	9.00	7.00	5.00	8.00	12.00	4.00	45	230.00	128.00	0.00
-19.25	146.79	1015	1021	966	1056	7.00	2.00	2.00	4.00	2.00	8.00	2.00	20	172.00	232.00	1.00
-19.26	146.80	1127	1164	1135	1169	0.00	3.00	0.00	0.00	2.00	1.00	1.00	7	37.00	4.00	1.00

Suburb Venue Data

This next section uses Foursquare API to get venue details for each suburb. It counts the number of venues in different categories within 1000m of suburb geolocations.

The defined categories were:

cafe = ['Café', 'Coffee Shop', 'Sandwich Place', 'Pizza Place', 'Juice Bar', 'Ice Cream Shop', 'Fried Chicken Joint', 'Burger Joint']

restaurant = ['Restaurant', 'Steakhouse']

supermarket = ['Supermarket', 'Market', 'Grocery Store']

pub = ['Pub', 'Liquor Store', 'Brewery']

entertainment = ['Theater', 'Multiplex']

recreation = ['Pool', 'Historic Site', 'Football Stadium', 'Bowling Alley', 'Basketball Stadium', 'Art Museum', 'Aquarium', 'Beach']

A json file was returned in response to the API.

Typical json return from FourSquare (first few lines): { "meta": { "code": 200, "requestId": "5f52be68eac69b39120fe129" }, "response": { "headerLocation": "Townsville", "headerFullLocation": "Townsville", "headerLocationGranularity": "city", "totalResults": 25, "suggestedBounds": { "ne": { "lat": -19.292946890999993, "lng": 146.77556542716533 }, "sw": { "lat": -19.310946909000001, "lng": 146.75652897283467 } }, "groups": [{ "type": "Recommended Places", "name": "recommended", "items": [{ "reasons": { "count": 0, "items": [{ "summary": "This spot is popular", "type": "general", "reasonName": "globalInteractionReason" }] }, "venue": { "id": "4d963f3bedc941bd2689925c", "name": "Boost Juice", "location": { "address": "310 Ross River Rd", "lat": -19.299975253741703, "lng": 146.76278951919068, "labeledLatLngs": [{ "label": "display", "lat": -19.299975253741703, "lng": 146.76278951919068 }], "distance": 406, "postalCode": "4814", "cc": "AU", "city": "Aitkenvale", "state": "QLD", "country": "Australia", "formattedAddress": ["310 Ross River Rd", "Aitkenvale QLD 4814", "Australia"] }, "categories": [{ "id": "4bf58dd8d48988d112941735", "name": "Juice Bar", "pluralName": "Juice Bars", "shortName": "Juice Bar", "icon": { "prefix": "https://ss3.4sqi.net/img/categories_v2/food/juicebar", "suffix": ".png" }, "primary": true }], "photos": { "count": 0, "groups": [] }, "referralId": "e-0-4d963f3bedc941bd2689925c-0" }, { "reasons": { "count": 0, "items": [{ "summary": "This spot is popular", "type": "general", "reasonName": "globalInteractionReason" }] },

The venue data was extracted from this json file, categorised and counted. The count data was added to the main dataframe.

```
df_ec_ac.head()
```

ive nal	Attack	...	Animal Complaints Grand Total	prop_cc	pers_cc	council_fac	cafe_count	restaurant_count	supermarket_count	pub_count	entertainment_count	recreation_count
.00	2.00	...	15	476.00	1686.00	4.00	12.00	15.00	1.00	4.00	0.00	2.00
.00	8.00	...	141	1146.00	589.00	1.00	5.00	9.00	1.00	1.00	0.00	0.00
.00	7.00	...	45	230.00	128.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00
.00	2.00	...	20	172.00	232.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00
.00	0.00	...	7	37.00	4.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00

This resulted in the final main dataframe with the following features and data types:

```
df_ec_ac.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 45 entries, 0 to 44
Data columns (total 24 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Suburb                                     45 non-null     object
1   latitude                                  45 non-null     float64
2   longitude                                  45 non-null     float64
3   rsed                                       45 non-null     int64
4   rsead                                      45 non-null     int64
5   ier                                        45 non-null     int64
6   ieo                                        45 non-null     int64
7   bus_stop                                  45 non-null     float64
8   Aggressive Animal                        45 non-null     float64
9   Attack                                    45 non-null     float64
10  Enclosure                                45 non-null     float64
11  Noise                                    45 non-null     float64
12  Private Impound                          45 non-null     float64
13  Wandering                                45 non-null     float64
14  Animal Complaints Grand Total            45 non-null     int64
15  prop_cc                                  45 non-null     float64
16  pers_cc                                  45 non-null     float64
17  council_fac                              45 non-null     float64
18  cafe_count                               45 non-null     float64
19  restaurant_count                        45 non-null     float64
20  supermarket_count                       45 non-null     float64
21  pub_count                               45 non-null     float64
22  entertainment_count                     45 non-null     float64
23  recreation_count                         45 non-null     float64
dtypes: float64(18), int64(5), object(1)
```

Suburb	latitude	longitude	rsed	rsead	ier	ieo	bus_stop	Aggressive Animal	Attack	Enclosure	Noise	Private Impound	Wandering	Grand Total	prop_cc	pers_cc	council_fac	cafe_count	restaurant_count	supermarket_count	pub_count	entertainment_count	recreation_count
Townsville City	-19.2622223	146.8158417	1055	1059	943	1094	13	0	2	2	5	4	2	15	476	1677	4	11	14	1	5	0	2
Aitkenvale	-19.3019469	146.7660472	940	937	915	964	1	13	8	17	31	55	17	141	1146	587	1	5	9	1	1	0	0
Annandale	-19.3122951	146.7790291	1077	1068	1069	1051	7	9	7	5	8	12	4	45	230	128	0	0	1	1	0	0	0
Belgian Gardens	-19.2473194	146.7926118	1015	1021	966	1056	7	2	2	4	2	8	2	20	172	232	1	0	1	0	0	0	1
Castle Hill	-19.2580333	146.8026248	1127	1164	1135	1169	0	3	0	0	2	1	1	7	37	4	1	0	0	0	1	0	0

Data Analysis

Exploratory Data Analysis

No missing values (as per .info() result).

<TO BE COMPLETED>

Cluster Analysis

<TO BE COMPLETED>

Results/Discussion/Conclusion

<TO BE COMPLETED>