Apr 2020

IBM Applied Data Science Capstone Report

EVALUATE METROPOLITAN ADELAIDE HOTEL LOCATIONS BY NET GAMING REVENUE, POPULATION DENSITY AND TRAFFIC VOLUME

PETER BUCHANAN

Contents

Introduction

Background

South Australia (abbreviated as SA) is the southern, central state of mainland Australia. It has a total land area of 983,482 square kilometres. The state has a population of more than 1.75 million people, 77% of whom live in Adelaide and surrounding metropolitan areas.

The Australian Bureau of Statistics (ABS), classify hotels as operating in 'accommodation services' industry, rather than the 'pub, tavern and bar' industry. Hotels with gaming machines tend to be larger, earning more revenue and employing more people. In addition to EGMs, these venues tend to offer a variety of bar, bistro/restaurant, retail bottle shop and gaming facilities including TAB and Keno, as well as other forms of entertainment.

To own and operate gaming machines in South Australia you must have a gaming machine licence and hold gaming machine entitlements. Each gaming machine entitlement gives the licence holder the right to own and operate a single gaming machine at their venue.

When a new gaming machine licence is granted, no entitlements will be issued. To operate gaming machines, a licence holder will need to purchase entitlements either through the government's approved trading system or under a contractual agreement with Club One.

Each gaming machine licence specifies the maximum number of gaming machines approved for a venue. The most any venue can have is 40 machines.

A gaming machine licence can't be granted to premises located under the same roof as shops or within shopping complexes.

Business Understanding

A hotelier owns and operates a one-hundred-year-old freehold heritage-listed hotel on a significant combined land holding of 6,833sqm

It is located in a quiet suburb next to a park and picturesque views of the city skyline.

The venue has experienced a decline in EGM expenditure and patronage in recent years. A recently refurbished gaming room and new machines failed to reverse a decline in revenue.

Considerable renovation and replacement of heating, ventilation, and air conditioning (HVAC) would be required to increase gaming machine approval from 20 to 40 machines.

SWOT Analysis of the hotel identified an opportunity to develop the site into a retirement village.

Presale of the residential development has raised capital for the construction of a new purpose-built Hotel in a more suitable location.

The new Hotel will utilise smart technology including

- Cashless gaming, to reduce cash handling, eliminate hopper float and hard count labour
- Self Service Cash Redemption Terminals to reduce labour, fraud and automate reconciliation
- Cloud integrated POS to facilitate online food and alcohol order and delivery
- General-Purpose EFT based reloadable reward card.

  Transfers from the prepaid card to competitive accounts can be restricted. Supported by Apple, Google Pay any EFTPOS capable cash register and EFT payment Gateway for Online food and liquor orders

Breakdown of Business units, projected Income and Expenditure

- Focus on increased Net Gaming Revenue and reduced labour costs through automation
- Discontinuance Hotel Accommodation due to high labour costs and low revenue
- Aggressive pricing strategy, Meals and Food sales, Take away liquor and other beverages

Business Unit	Income Actual	Income Desired
Sale of liquor and other beverages (on-premises)	14%	10%
Sale of liquor and other beverages (take away)	31%	25%
Accommodation	14%	0%
Net Gaming Revenue (NGR)	18%	52%
Takings from meals and food sales	19%	9%
Other gambling (TAB, Keno, etc)	4%	4%
Total	100%	100%

Hotel Expenditure	Proportion
Labour costs	21.20%
Purchases	43.30%
Gambling taxes and levies	10.50%
Rent, leasing and hiring	6.00%
Other expenses	19.00%
Total	**100.00%**

Project Scope

1. Provide locations in Metropolitan Adelaide for a new hotel based on the following criteria:
    a. High gaming revenue area
    b. Competitor location, avoid locations where the competitor hotel has a presence
    c. Located on a main arterial road with medium to high traffic volume
    d. Located in proximity to shopping complex or transport hub
    e. High population density
    f. Provide Geodemographic classification of city amenities by applying a clustering algorithm
    g. Customer demographics for customer insights and segmentation

Audience & Stakeholders

The audience for this project is the hotelier, marketers and advisors who make decisions based on data insights. This project explores some of those data insights specific to venue location.

Data

This section describes the data sourced for this project.

Data Sources

South Australian Government Data Directory (Data.SA)

- #### Liquor and Gaming Licences
  A data set containing publicly accessible licence details under Liquor and Gaming legislations.
  This information is provided from Consumer and Business Services

- #### City and Traffic Volume ESRI Shapefile (.shp)
  A shapefile stores non-topological geometry for the spatial features in a data set

Australian Bureau of Statistics (ABS)

- #### State Suburbs (SSC)
  ABS approximation of localities gazetted by the Geographical Place Name authority in each State. Gazetted Localities are the officially recognised boundaries of suburbs (in cities and larger towns) and localities (outside cities and larger towns)

- #### Local Government Areas (LGAs)
  The Australian Statistical Geography Standard (ASGS) Local Government Areas (LGAs) are an ABS approximation of officially gazetted Local Government Areas as defined by each State and Territory Local Government Department

- #### Local Government Areas ESRI Shapefile (.shp)
  A shapefile stores non-topological geometry and attribute information for the spatial features in a data set. The geometry for a feature is stored as a shape comprising a set of vector coordinates

- #### Census Data QuickStats
  QuickStats makes a high level, at a glance Census data available for most areas, from small areas to state, territory and Australia level

Australian Tax Office (ATO)

- #### Taxation Statistics
  An overview of the income and tax status of Australian individuals, companies, partnerships, trusts and funds for 2016-17

The South Australian Centre for Economic Studies (SACES)

- #### Gambling Database
  The South Australian Centre for Economic Studies has developed a Gambling Database that provides regional level data on gaming machine activity in South Australian licensed venues

Foursquare

- #### Places API
  Returns a list of venues by latitude, longitude, search radius and venue category
  Foursquare is one of the largest sources of location-based venue data with over 105 million venues

Google

- #### Geocoding and Places API
  The Geocoding API is a service that provides geocoding and reverse geocoding of addresses
  Place API search returns a list of places along with summary information about each place.

Data Collection

Most files were programmatically downloaded and saved to a local data repository with specific columns extracted into Pandas Dataframes.
Other data sources originating from WebAPI or Web Scrape were imported directly into Dataframes

Pandas Dataframe	Shape	Filetype	Contains
hotel_df	(533, 16)	Zipped CSV	Logistics and Geocoding and Statistics for all hotels in South Australia
abs_city_df	(358122, 5)	Zipped CSV	Name and unique code for all Australian Cities
abs_lga_name_df	(28205, 3)	Zipped CSV	Name and unique code for all Australian Local Government Areas
abs_city_shape_df	(1891, 7)	Zipped SHP	Shapefile for all South Australian Cities
gov_traffic_shape_df	(2577, 17)	Zipped SHP	Shapefile for all South Australian Roads
abs_lga_shape_df	(562, 6)	Zipped SHP	Shapefile for all South Australian Local Government Areas
ato_income_df	(2691, 4)	Zipped CSV	Australian population and average income by Postal Code
gaming_revenue_df	(71, 24)	Excel	South Australian gaming database, statistics by Exploded Local Government Areas
top_ngr_citycode_df	(23, 31)	Zipped CSV	Top Gaming Revenue LGA Cities
shopping_centre_df	(35, 5)	Goggle API	Shopping malls within Top Gaming Revenue LGA Cities
census_df	(2369, 12)	Web Scrape	Australian census data by City
foursquare_category_df	(943, 8)	Foursquare API	Foursquare Venue Category Hierarchy flattened, there are five hierarchical levels
foursquare_venue_explore_df	(1446, 44)	Foursquare API	Foursquare Venue data within Top Gaming Revenue LGA Cities

Data Preparation

Data Transformation

Data sets sourced across Australian government agencies follow a consistent naming schema, statistical areas or regions are built up from or, approximated by whole Mesh Blocks

Mesh Blocks are the smallest geographical area defined by the Australian Bureau of Statistics (ABS) and form the building blocks for the larger regions of the Australian Statistical Geography Standard (ASGS).
Where possible, one Mesh Block will have a dwelling count of between 30 and 60

Foursquare and Government Geo location transformed to Google Geocoding API format

```
# data correction convert ABS City to match GoogleAPI format
abs_city_lga_name_df.City.replace(to_replace = "St Kilda", value ="Saint Kilda", inplace=True)
abs_city_lga_name_df.City.replace(to_replace = "St Agnes", value ="Saint Agnes", inplace=True)
abs_city_lga_name_df.City.replace(to_replace = "Clovelly Park", value ="Tonsley", inplace=True)
abs_city_lga_name_df.City.replace(to_replace = "Kingston Se", value ="Kingston SE", inplace=True)
abs_city_lga_name_df.LgaName.replace(to_replace = "Campbelltown (C) (SA)", value ="Campbelltown (C)", inplace=True)
abs_city_lga_name_df.LgaName.replace(to_replace = "Kingston (DC) (SA)", value ="Kingston (DC)", inplace=True)
```

Where possible transformation, filtering, column names and data types occur at the point of import

```
# Import to Dataframe: ato_income_df
ato_income_df = pd.read_excel(ATO_INCOME_FILE, header = 0, sheet_name ='Snapshot Table 7D',
                              names =   ["PostalCode","PostalCodePopulation", "PostalCodeAvgTaxableIncome","PostalCodeRankTaxableIncome"],
                              usecols=  "B,C,F,I",
                              skiprows = [0,1,2],
                              dtype= {"PostalCode": str, "PostalCodeAvgTaxableIncome": float, "PostalCodeRankTaxableIncome": int})
```

Data Cleaning

Google Geocoding API was used extensively to replace inaccurate or inconsistent geodata
Both forward and reverse geocoding were used to clean up address, city and postal codes

Data Testing

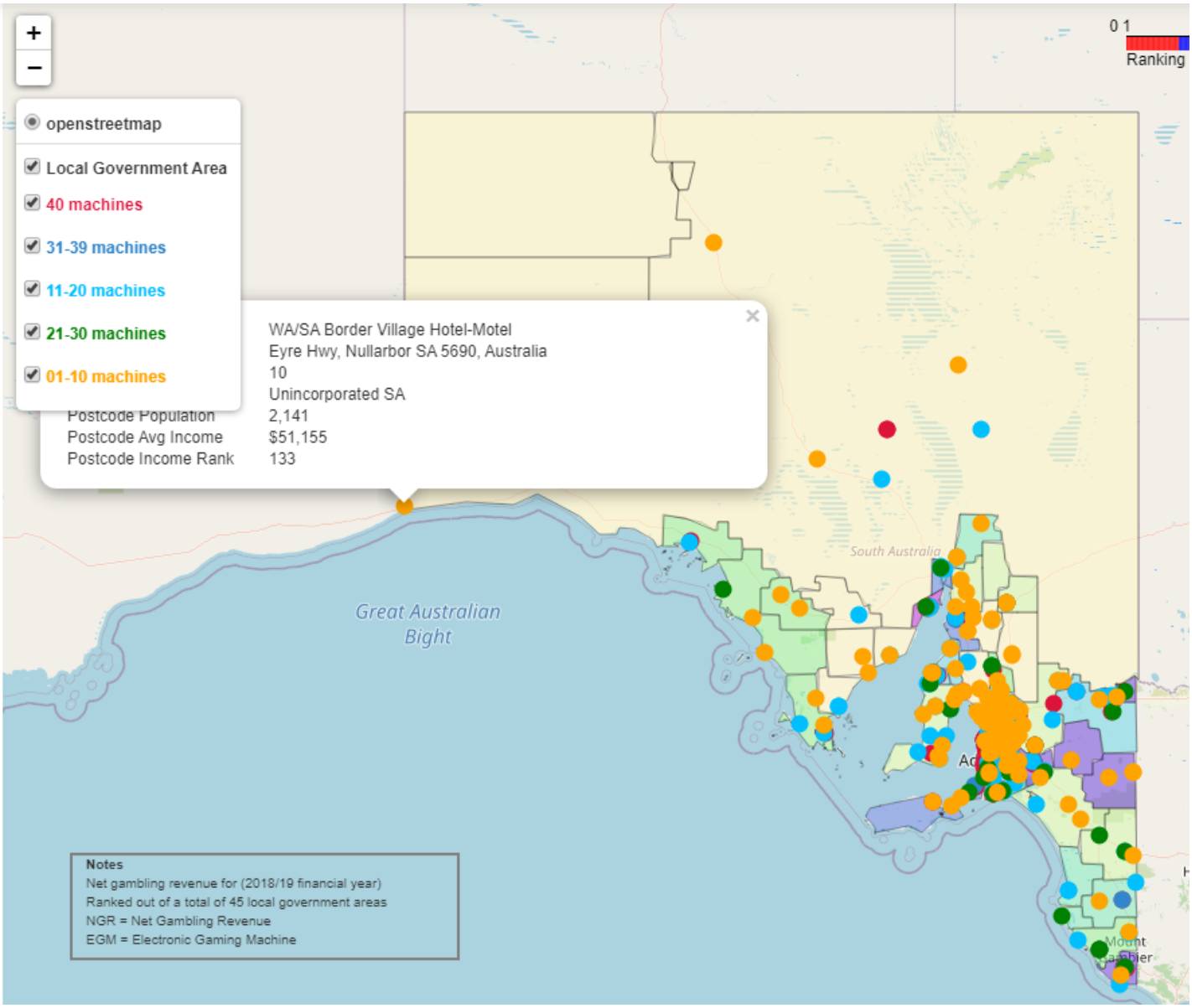
a. Throughout Python code, are checkpoints where Dataframes are exported to CSV for inspection in Excel

b. Hotel locations are plotted on a choropleth map of Local Government Areas in the state of South Australia. Initial testing revealed hotel locations outside the region of 'South Australia', requiring data correction of Hotel name and address.

```
# data correction
if(hotel_df.at[index, 'PremisesName'] == 'Avenues Hotel'):
    hotel_df.at[index, 'Address'] = '114 Payneham Road Stepney SA 5069'

elif(hotel_df.at[index, 'PremisesName'] == 'WA/SA Border Village Hotel-Motel'):
    hotel_df.at[index, 'Address'] = 'Border Village Roadhouse, Eyre Hwy, Nullarbor SA 5690'

elif(hotel_df.at[index, 'PremisesName'] == 'The Pier Hotel (SA)'):
    hotel_df.at[index, 'Address'] = 'Glenelg Pier Hotel, 18 Holdfast Promenade, Glenelg SA 5045'

elif(hotel_df.at[index, 'PremisesName'] == 'Park Hotel'):
    hotel_df.at[index, 'Address'] = 'Park Hotel Mount Gambier, 163 Commercial St W, Mount Gambier SA 5290'
```

c. The following check was implemented to raise an exception on subsequent imports

```
# raise exception if city not GEOCODE_REGION_SHORT (South Australia)
if GEOCODE_REGION != (logistics['region']):
    raise ValueError(GEOCODE_REGION + ' not returned in region_long: {}'.format(hotel_df.at[index, 'Address']))
```



Gaming Revenue Dataframe

The Gaming Revenue Dataframe contains Gambling data compiled by the South Australian Centre for Economic Studies (SACES)

a. Net Gaming Revenue by Local Government Area was obtained from Consumer and Business Services.
b. Per adult, population estimates were obtained from Local Government Areas published by the Australian Bureau of Statistics and housing

```
gaming_revenue_df shape is  (71, 24)
```

	Key	LgaName	LgaGroup	VenueCount	EgmCount	TotalNgr	TotalNgrPerAdult	MachinesPer1000Adult	AvgNgrPerMachine	AverageNgrPerVenue	AggregateNgrPerLgaStr	AverageNgrPerVenueStr
0	1	Adelaide (C)	Adelaide (C)	37	790	22031.14	939.63	33.27	27976.05	595436.28	$22,031,142	$595,436
1	2	Adelaide Hills (DC)	Adelaide Hills (DC)	15	211	3600.18	117.74	6.90	17022.15	240012.32	$3,600,185	$240,012
2	3	Alexandrina (DC)	Alexandrina (DC)	13	250	7569.47	343.41	11.29	30522.06	582266.93	$7,569,470	$582,267
3	4	Barossa (DC)	Barossa (DC)	14	256	6329.95	327.96	13.18	24345.97	452139.35	$6,329,951	$452,139
4	5	Barunga West (DC)	Barunga West (DC), Copper Coast (DC)	16	273	8929.39	625.48	19.02	33194.74	558086.58	$8,929,385	$558,087
5	5	Copper Coast (DC)	Barunga West (DC), Copper Coast (DC)	16	273	8929.39	625.48	19.02	33194.74	558086.58	$8,929,385	$558,087
6	6	Berri and Barmera (DC)	Berri and Barmera (DC)	6	164	6007.86	708.37	19.33	35975.22	1001310.36	$6,007,862	$1,001,310

TotalNgrPerAdultStr	AvgNgrPerMachineStr	RankTotalNgr	RankTotalNgrPerAdult	RankEgmsPer1000Adult	RankAvgNgrPerMachine	MeshBlockCode	CityCode	City	StateCode	State	LgaCode
$940	$27,976	10	3	1	31	40000010000	40002	Adelaide	4	South Australia	40070
$118	$17,022	34	42	36	40	40003800000	40008	Aldgate	4	South Australia	40120
$343	$30,522	25	31	23	28	40008440000	41180	Port Elliot	4	South Australia	40220
$328	$24,346	26	32	16	34	40012970000	40441	Flaxman Valley	4	South Australia	40310
$625	$33,195	22	13	5	25	40015810000	41520	Ward Hill	4	South Australia	40430
$625	$33,195	22	13	5	25	40046150000	41502	Wallaroo	4	South Australia	41560
$708	$35,975	27	8	4	24	40016731000	40072	Barmera	4	South Australia	40520

```
gaming_revenue_df.dtypes
```

```
Key                      int64
LgaName                 object
LgaGroup                object
VenueCount               int64
EgmCount                 int64
TotalNgr               float64
TotalNgrPerAdult       float64
MachinesPer1000Adult   float64
AvgNgrPerMachine       float64
AverageNgrPerVenue     float64
AggregateNgrPerLgaStr   object
AverageNgrPerVenueStr   object
TotalNgrPerAdultStr     object
AvgNgrPerMachineStr     object
RankTotalNgr             int64
RankTotalNgrPerAdult     int64
RankEgmsPer1000Adult     int64
RankAvgNgrPerMachine     int64
MeshBlockCode           object
CityCode                object
City                    object
StateCode               object
State                   object
LgaCode                 object
dtype: object
```

Hotel Dataframe

The hotel Dataframe contains a list of all Hotels issued with a liquor and/or gaming licence.
Only active hotels with one or more gaming machines are imported

The process as follows:

1. Import hotel licence details from Consumer and Business Services into Pandas
2. Merge statistical data from the Australian Bureau of Statistics (ABS) and Australian Tax Office (ATO)
3. Google Geocoding API to obtain Latitude, Longitude, City and formatted address

```
hotel_df shape is  (533, 7)
```

LicenceNumber	PremisesName	Address	City	PostalCode	Latitude	Longitude	MachineCount
51201413	AAMI Stadium	1 Turner Dr, West Lakes SA 5021, Australia	West Lakes	5021	-34.879935	138.493982	40
50106296	Aberfoyle Tavern	The Hub Shopping Centre, Christie Way, Aberfoyle Park SA 5159, Australia	Aberfoyle Park	5159	-35.075343	138.591324	40
50106694	Aces Bar and Bistro	62-72 Gouger St, Adelaide SA 5000, Australia	Adelaide	5000	-34.930112	138.596767	36
50100012	Alberton Hotel	124 Port Rd, Alberton SA 5014, Australia	Alberton	5014	-34.861083	138.514249	40
50100020	Albion Hotel	423 Churchill Rd, Kilburn SA 5084, Australia	Kilburn	5084	-34.860351	138.583392	40
50100038	Aldgate Pump Hotel	1 Strathalbyn Rd, Aldgate SA 5154, Australia	Aldgate	5154	-35.015222	138.736707	40
50100046	Aldinga Hotel	Old Coach Rd, Aldinga SA 5173, Australia	Aldinga	5173	-35.267102	138.482776	40

Top Net Gaming Revenue Cities Dataframe

The Gaming Revenue Dataframe ranks Average Net Gaming Revenue per machine by Local Government Area.
LGA Group (Campbelltown, Tea Tree Gully) ranks highest at $100,625 per machine per year.

The process as follows:

1. Merge Gaming Revenue Dataframe with South Australian Local Government Area LGA Dataset
2. Drop all but the highest-ranking LGA Group, leaving **23** cities of interest

```
top_ngr_citycode_df shape is  (23, 31)
```

	MeshBlockCode	CityCode	City	StateCode	State	LgaCode	LgaName	Key	LgaGroup	VenueCount	EgmCount
0	40023510000	40048	Athelstone	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
1	40027041000	40211	Campbelltown	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
2	40027190000	40576	Hectorville	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
3	40023600000	41020	Newton	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
4	40023650000	41090	Paradise	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
5	40027202000	41455	Tranmere	4	South Australia	40910	Campbelltown (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
6	40184040000	40069	Banksia Park	4	South Australia	47700	Tea Tree Gully (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438
7	40184140000	40424	Fairview Park	4	South Australia	47700	Tea Tree Gully (C)	7	Campbelltown (C), Tea Tree Gully (C)	12	438

Shopping Malls within Top Net Gaming Revenue Cities Dataframe

Shopping Mall dataset obtained via Google Place API Nearby Search
The dataset contains shopping malls within the Top Net Gaming Revenue Cities Dataframe

The process as follows:

3. Obtain all Shopping Malls, Plazas and Centres within 5000 metres of city centroid
4. Parameter Type, restricts the results to places matching the specified type, in this case, 'shopping_mall'

```
shopping_centre_df shape is  (35, 5)
```

	Id	Premises	Latitude	Longitude	Address
0	12e844fe8d106a4452fee040478299f29dbc8d53	Westfield Tea Tree Plaza	-34.830731	138.692442	976 North East Road, Modbury
1	791966865f09a661584c45ead005fc6b1623bea0	Newton Village	-34.885876	138.689731	299 Montacute Road, Newton
2	390d135eadda064ed581f3b956961339ca374c6d	Newton Central	-34.877316	138.674570	84 Gorge Road, Newton
4	1540af67b045a09e8c739030cc75b963b8274c41	Dernancourt Village Shopping Centre	-34.862114	138.682789	832-840 Lower North East Road, Dernancourt
7	736a4320dc2073c58e522bd4c715e9515f8876a8	Mercury Plaza	-34.876607	138.675860	94-96 Gorge Road, Newton

Census Dataframe

Following columns scraped from the Australian Bureau of Statistics website and imported into Dataframe via Beautiful Soup Python library for extracting data from HTML and XML files

- Population
- Age
- Registered marital status
- Social marital status
- Education
- Level of highest educational attainment
- Employment
- Median weekly incomes
- Family composition
- Single (or lone) parents
- Employment status of parents
- Dwelling count
- Dwelling structure
- Number of bedrooms
- Tenure
- Household income
- Number of registered motor vehicles
- Dwelling internet connection

```
census_df shape is  (2369, 12)
```

	CityCode	City	Title	Label	CityCount	CityPercent	RegionCount	RegionPercent	CountryCount	CountryPercent	CityCountNum	CityPercentNum
0	40048	Athelstone	People	Male	4,548	49.7	825,997	49.3	11,546,638	49.3	4548.0	49.7
1	40048	Athelstone	People	Female	4,601	50.3	850,652	50.7	11,855,248	50.7	4601.0	50.3
2	40048	Athelstone	People	Aboriginal and/or Torres Strait Islander people	46	0.5	34,184	2.0	649,171	2.8	46.0	0.5
3	40048	Athelstone	Age	Median age	43	--	40	--	38	--	43.0	0.0
4	40048	Athelstone	Age	0-4 years	444	4.9	97,072	5.8	1,464,779	6.3	444.0	4.9
5	40048	Athelstone	Age	5-9 years	501	5.5	100,466	6.0	1,502,646	6.4	501.0	5.5
6	40048	Athelstone	Age	10-14 years	541	5.9	95,461	5.7	1,397,183	6.0	541.0	5.9
7	40048	Athelstone	Age	15-19 years	622	6.8	100,686	6.0	1,421,595	6.1	622.0	6.8
8	40048	Athelstone	Age	20-24 years	621	6.8	107,986	6.4	1,566,793	6.7	621.0	6.8
9	40048	Athelstone	Age	25-29 years	480	5.2	107,361	6.4	1,664,602	7.1	480.0	5.2
10	40048	Athelstone	Age	30-34 years	451	4.9	110,751	6.6	1,703,847	7.3	451.0	4.9
11	40048	Athelstone	Age	35-39 years	533	5.8	102,818	6.1	1,561,679	6.7	533.0	5.8
12	40048	Athelstone	Age	40-44 years	611	6.7	106,650	6.4	1,583,257	6.8	611.0	6.7
13	40048	Athelstone	Age	45-49 years	701	7.7	113,619	6.8	1,581,455	6.8	701.0	7.7
14	40048	Athelstone	Age	50-54 years	728	8.0	113,272	6.8	1,523,551	6.5	728.0	8.0
15	40048	Athelstone	Age	55-59 years	699	7.6	111,896	6.7	1,454,332	6.2	699.0	7.6
16	40048	Athelstone	Age	60-64 years	614	6.7	102,028	6.1	1,299,397	5.6	614.0	6.7
17	40048	Athelstone	Age	65-69 years	569	6.2	96,286	5.7	1,188,999	5.1	569.0	6.2
18	40048	Athelstone	Age	70-74 years	428	4.7	72,563	4.3	887,716	3.8	428.0	4.7
19	40048	Athelstone	Age	75-79 years	324	3.5	53,885	3.2	652,657	2.8	324.0	3.5
20	40048	Athelstone	Age	80-84 years	174	1.9	39,374	2.3	460,549	2.0	174.0	1.9
21	40048	Athelstone	Age	85 years and over	102	1.1	44,479	2.7	486,842	2.1	102.0	1.1
22	40048	Athelstone	Registered marital status	Married	4,469	58.4	659,638	47.7	9,148,218	48.1	4469.0	58.4
23	40048	Athelstone	Registered marital status	Separated	144	1.9	44,199	3.2	608,059	3.2	144.0	1.9
24	40048	Athelstone	Registered marital status	Divorced	517	6.8	127,916	9.2	1,626,890	8.5	517.0	6.8
25	40048	Athelstone	Registered marital status	Widowed	303	4.0	83,281	6.0	985,204	5.2	303.0	4.0
26	40048	Athelstone	Registered marital status	Never married	2,216	29.0	468,621	33.9	6,668,910	35.0	2216.0	29.0
27	40048	Athelstone	Social marital status	Registered marriage	4,227	57.9	582,236	47.4	8,001,141	47.7	4227.0	57.9
28	40048	Athelstone	Social marital status	De facto marriage	404	5.5	126,020	10.3	1,751,731	10.4	404.0	5.5
29	40048	Athelstone	Social marital status	Not married	2,664	36.5	520,539	42.4	7,024,973	41.9	2664.0	36.5
30	40048	Athelstone	Education	Preschool	103	4.2	20,292	4.2	347,621	4.8	103.0	4.2
31	40048	Athelstone	Education	Primary - Government	453	18.3	94,526	19.5	1,314,787	18.2	453.0	18.3

Foursquare Venue Categories Dataframe

The Foursquare application programming interface (API) was used to obtain Venue Category.
The five hierarchical levels were flattened and indexed by CategoryId

```
foursquare_category_df shape is  (943, 8)
```

	HierarchyLevel	CategoryName	CategoryId	Level1	Level2	Level3	Level4	Level5
0	1	Arts & Entertainment	4d4b7104d754a06370d81259	Arts & Entertainment				
1	2	Amphitheater	56aa371be4b08b9a8d5734db	Arts & Entertainment	Amphitheater			
2	2	Aquarium	4fceea171983d5d06c3e9823	Arts & Entertainment	Aquarium			
3	2	Arcade	4bf58dd8d48988d1e1931735	Arts & Entertainment	Arcade			
4	2	Art Gallery	4bf58dd8d48988d1e2931735	Arts & Entertainment	Art Gallery			
5	2	Bowling Alley	4bf58dd8d48988d1e4931735	Arts & Entertainment	Bowling Alley			
6	2	Casino	4bf58dd8d48988d17c941735	Arts & Entertainment	Casino			
7	2	Circus	52e81612bcbc57f1066b79e7	Arts & Entertainment	Circus			
8	2	Comedy Club	4bf58dd8d48988d18e941735	Arts & Entertainment	Comedy Club			
9	2	Concert Hall	5032792091d4c4b30a586d5c	Arts & Entertainment	Concert Hall			
10	2	Country Dance Club	52e81612bcbc57f1066b79ef	Arts & Entertainment	Country Dance Club			
11	2	Disc Golf	52e81612bcbc57f1066b79e8	Arts & Entertainment	Disc Golf			
12	2	Exhibit	56aa371be4b08b9a8d573532	Arts & Entertainment	Exhibit			
13	2	General Entertainment	4bf58dd8d48988d1f1931735	Arts & Entertainment	General Entertainment			
14	2	Go Kart Track	52e81612bcbc57f1066b79ea	Arts & Entertainment	Go Kart Track			
15	2	Historic Site	4deefb944765f83613cdba6e	Arts & Entertainment	Historic Site			
16	2	Karaoke Box	5744ccdfe4b0c0459246b4bb	Arts & Entertainment	Karaoke Box			
17	2	Laser Tag	52e81612bcbc57f1066b79e6	Arts & Entertainment	Laser Tag			
18	2	Memorial Site	5642206c498e4bfca532186c	Arts & Entertainment	Memorial Site			
19	2	Mini Golf	52e81612bcbc57f1066b79eb	Arts & Entertainment	Mini Golf			

Foursquare Venue Dataframe

The Foursquare API was used to obtain venue information for Top Net Gaming Revenue Cities

The process as follows:

5. Obtain all venues and venue categories within 5000 metres of each city centroid:
6. Iterate over the **23** cities in the Top Net Gaming Revenue Cities Dataframe
   Iterate over each of the **943** Categories (Hierarchy Level 1 to 5)

   Note: Chunked 10 Categories per request and added 30-second delay between city iteration
   Approx. 46,000 venues returned
7. Drop duplicate venue Id
8. Reverse Google Geocode the Foursquare Venue Latitude, Longitude coordinates to determine Venue City and formatted address.
9. Merge Dataframe with flattened Venue Category Hierarchy
10. Merge Dataframe with Top Net Gaming Revenue Cities Dataframe and drop where city not mapped

```
foursquare_venue_explore_df shape is  (1446, 44)
```

	VenueId	VenueName	VenueLatitude	VenueLongitude	CategoryId	VenueAddress	City	HierarchyLevel	CategoryName	Level1	Level2	Level3	Level4
0	4c874f439c808cfad73e131e	Turramurra	-34.853200	138.700771	4bf58dd8d48988d18b941735	1000 Lower North East Rd, Highbury SA 5089, Australia	Highbury	3	Basketball Stadium	Arts & Entertainment	Stadium	Basketball Stadium	
2	4de7344eb0fb2165e1a08def	Golden Grove Rec Centre	-34.792939	138.694200	4bf58dd8d48988d18b941735	2 Tenison Pl, Golden Grove SA 5125, Australia	Golden Grove	3	Basketball Stadium	Arts & Entertainment	Stadium	Basketball Stadium	

Level5	MeshBlockCode	CityCode	StateCode	State	LgaCode	LgaName	Key	LgaGroup	VenueCount	EgmCount	TotalNgr	TotalNgrPerAdult	MachinesPer1000Adult	AvgNgrPerMachine
	40184360000	40581	4	South Australia	47700	Tea Tree Gully (C)	7.0	Campbelltown (C), Tea Tree Gully (C)	12.0	438.0	43922.81	366.6	3.65	100625.0
	40184101000	40511	4	South Australia	47700	Tea Tree Gully (C)	7.0	Campbelltown (C), Tea Tree Gully (C)	12.0	438.0	43922.81	366.6	3.65	100625.0

AggregateNgrPerLgaStr	AverageNgrPerVenueStr	TotalNgrPerAdultStr	AvgNgrPerMachineStr	RankTotalNgr	RankTotalNgrPerAdult	RankEgmsPer1000Adult	RankAvgNgrPerMachine	Address	PostalCode	
$43,922,813	$3,660,234	$367	$100,625	5.0	27.0	42.0	1.0	Highbury SA 5089, Australia	5089	-
$43,922,813	$3,660,234	$367	$100,625	5.0	27.0	42.0	1.0	Golden Grove SA 5125, Australia	5125	-

Latitude	Longitude	People	MedianAge	HouseholdIncome	Dwellings
-34.854000	138.697000	6588.0	42	1,705	2,334
-34.783889	138.706111	10237.0	42	1,519	3,763

```
foursquare_venue_explore_df.dtypes
```

```
VenueId                  object
VenueName                object
VenueLatitude           float64
VenueLongitude          float64
CategoryId               object
VenueAddress             object
City                     object
HierarchyLevel            int64
CategoryName             object
Level1                   object
Level2                   object
Level3                   object
Level4                   object
Level5                   object
MeshBlockCode            object
CityCode                 object
StateCode                object
State                    object
LgaCode                  object
LgaName                  object
Key                     float64
LgaGroup                 object
VenueCount              float64
EgmCount                float64
TotalNgr                float64
TotalNgrPerAdult        float64
MachinesPer1000Adult    float64
AvgNgrPerMachine        float64
AggregateNgrPerLgaStr    object
AverageNgrPerVenueStr    object
TotalNgrPerAdultStr      object
AvgNgrPerMachineStr      object
RankTotalNgr            float64
RankTotalNgrPerAdult    float64
RankEgmsPer1000Adult    float64
RankAvgNgrPerMachine    float64
Address                  object
PostalCode               object
Latitude                float64
Longitude               float64
People                  float64
MedianAge                object
HouseholdIncome          object
Dwellings                object
dtype: object
```

Methodology

This section describes the data exploration that were conducted and how they relate to the original business problem of providing two or more locations in Metropolitan Adelaide for a new hotel based on criteria:

a. High gaming revenue area
b. Competitor location, avoid locations where the competitor hotel has a presence
c. Located on a main arterial road with medium to high traffic volume
d. Located in proximity to shopping complex or transport hub
e. High population density
f. Provide Geodemographic classification of city amenities by applying a clustering algorithm
g. Customer demographics for customer insights and segmentation

Data Preparation Choices

High Gaming Revenue (a)

The Independent Gaming Corporation Ltd (IGC) was awarded the sole SA Gaming Machine Monitor Licence on the 10th of November 1993 under Section 14, (1) (d) of the Gaming Machines Act 1992.

IGC provides venues access to gaming machine data through its Web Portal. The portal provides all gaming venues with a comprehensive suite of reports with analysis of their gaming machine financial performance and weekly Local Government Area, Net Gaming Revenues (data origin)

a. The general public does not have access to the portal, it would have been the preferred source
b. IGC release an annual PDF report which is publicly available via Government Data Directory PDF was scraped in the first iteration of code until I discovered the SACES gambling database
c. SACES gambling database is a clean and validated version of (b) (I used this option)

The Australian Bureau of Statistics (ABS) provides Local Government Area (LGA) geometric location and attribute information (data origin)

Google Geocoding API provides location geocoding for choropleth mapping (data origin)

Competitor Location (b)

Liquor and Gaming licencing is provided from Consumer and Business Services (data origin)

Traffic volumes (c)

The Department of Planning, Transport and Infrastructure is the (data origin)

Proximity to shopping complex (d)

Foursquare vs Google Places (Adelaide Australia)

a. Google Places accurately returned all shopping malls including Geocoded address (I used this option)
b. Foursquare missed a few and required a second step of reverse Geocoding

High population density (e)

Provided by the Australian Bureau of Statistics (ABS) (data origin)

Geodemographic classification of city amenities by applying a clustering algorithm (f)

The Foursquare Places API is used to explore and categorise venues in each city (mandate)

Customer Demographics (f)

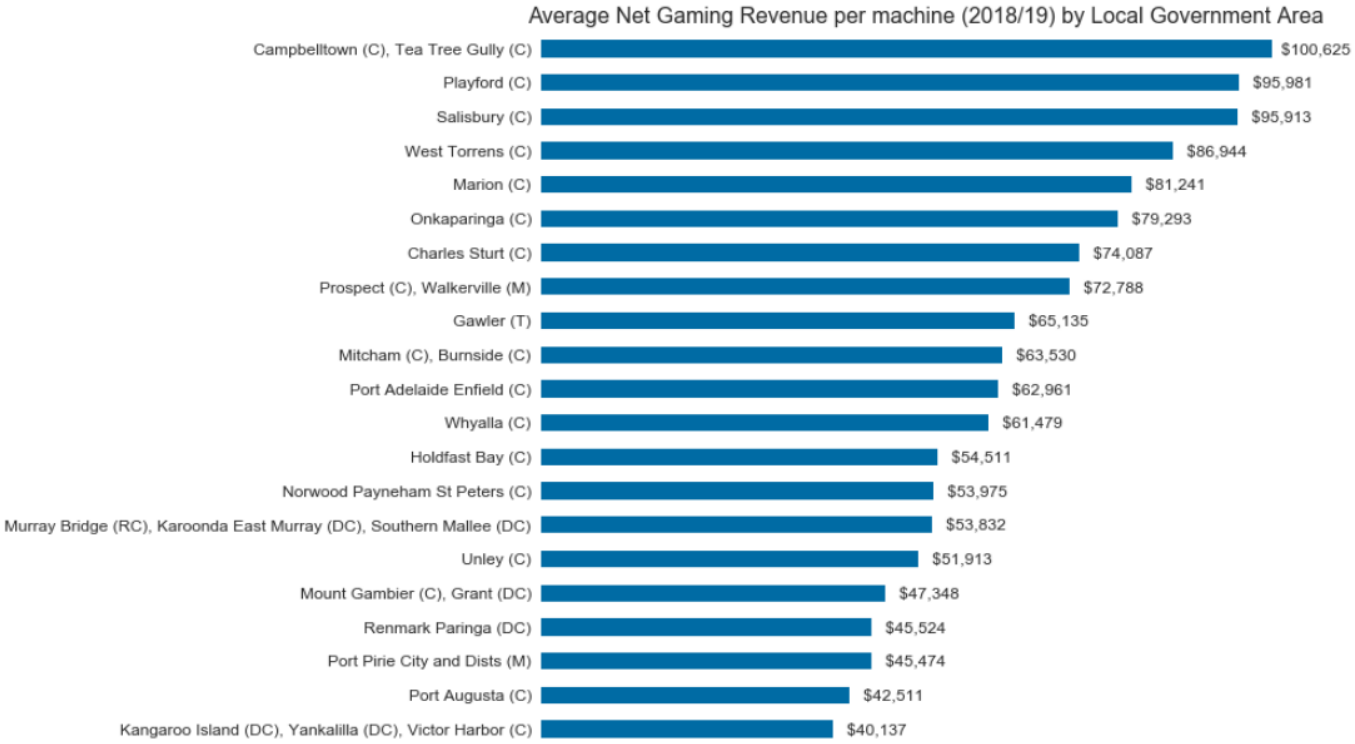
Provided by the Australian Bureau of Statistics (ABS) (data origin)

Histograms

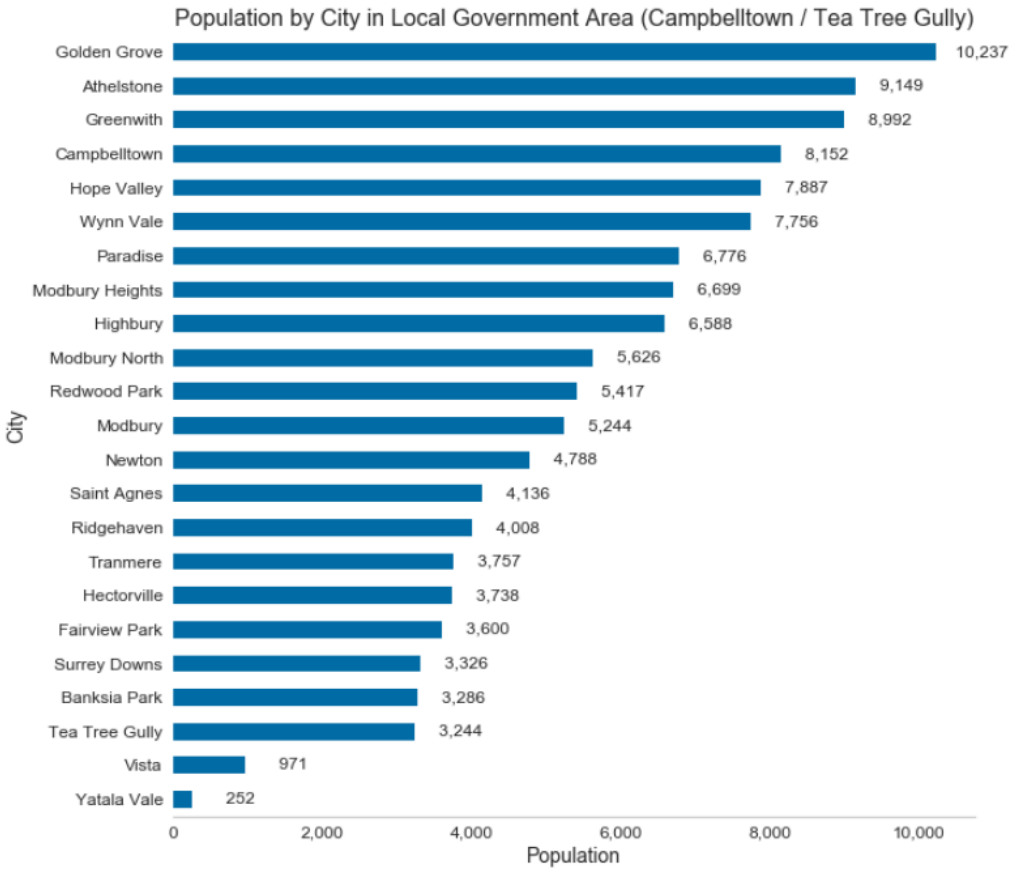
Average Net Gaming Revenue per machine by Local Government Areas (LGA)

Observations:

- 86% revenue increase between rank 1st (desired) and rank 10th (actual hotel income)
- The highest-ranking LGA group is (Campbelltown, Tea Tree Gully)



Population by City in Local Government Area (Campbelltown / Tea Tree Gully)



Choropleth Maps

Map 1. South Australian Local Government Areas (LGA) ranked by Average NGR per machine

Presents all Local Government Areas (LGA) in South Australia, regions colour coded by Average Net Gaming Revenue per machine Rank (1 to 44)

Observations:

- Most high ranking LGA hotels have 40 EGMs installed, lower-ranking have fewer EGMs
- The more disadvantaged a region is in socioeconomic terms (lower average income, higher unemployment, lower educational attainment) the higher is their EGM expenditure. This is associated with the prevalence of gaming venues and machines

Map 2. Population by City in Local Government Area (Campbelltown / Tea Tree Gully)

Observations:

- A belt of low population density surrounding the city of Tea tree gully

Map 3. Foursquare venue and Google Place API search area

Visualizes API Search radius

Observations:

- A search radius of 5000 meters is sufficient to capture all cities in the target location

Map 4. Traffic Volume by City in Local Government Area (Campbelltown / Tea Tree Gully)

Presents traffic volume information for arterial roads through the City

Observations:

- North East Road, passing through Modbury has high traffic volume
- Followed by McIntyre and Kings Road, through Modbury North and Modbury Heights

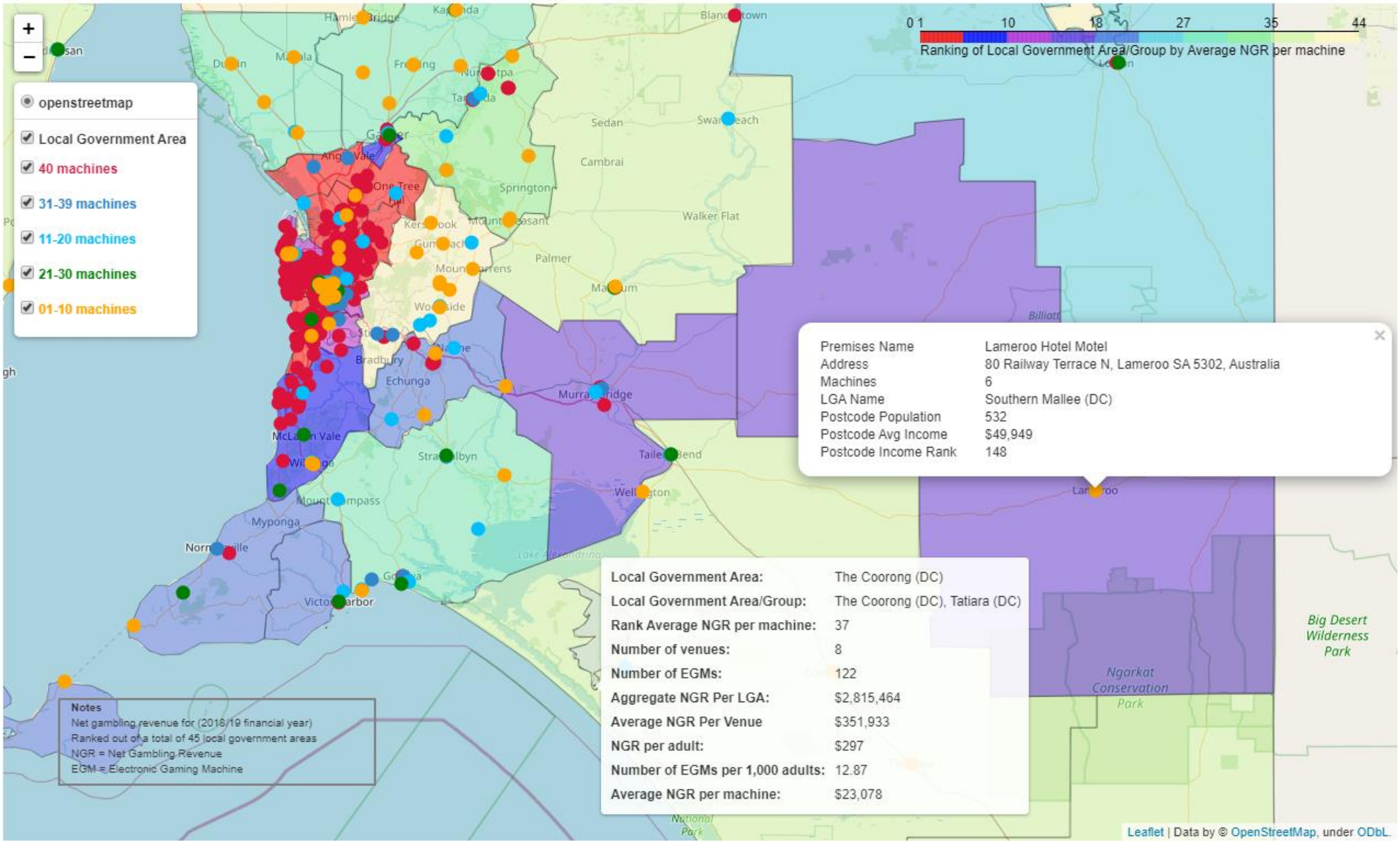
Map 5. Shopping Malls, Traffic Volume and Hotel by City

Presents Shopping Malls, Hotels and traffic volume information for arterial roads through the City

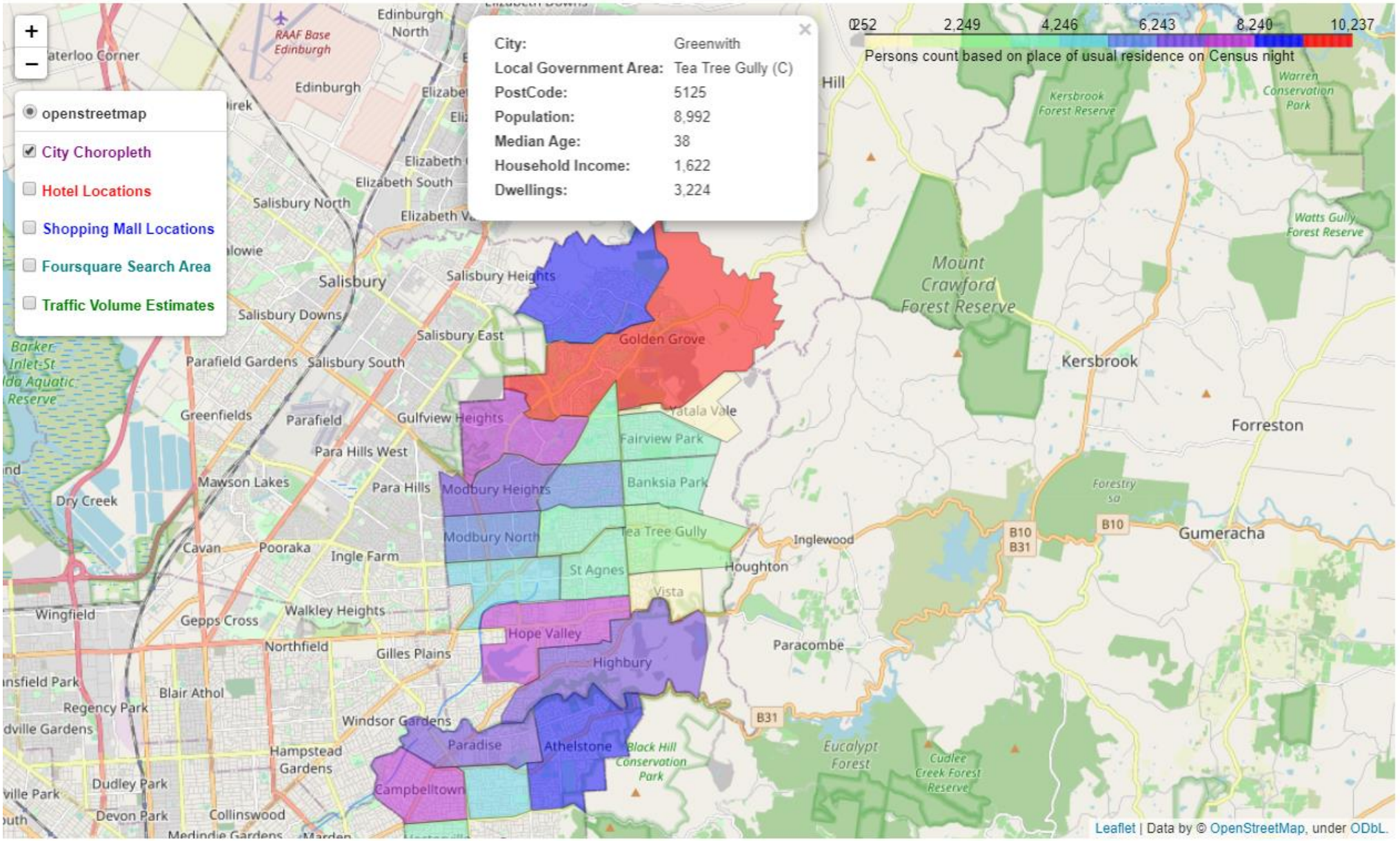
Observations:

- A correlation between the size of the shopping mall and Traffic Volume
- A correlation between the proximity of Hotels to shopping malls

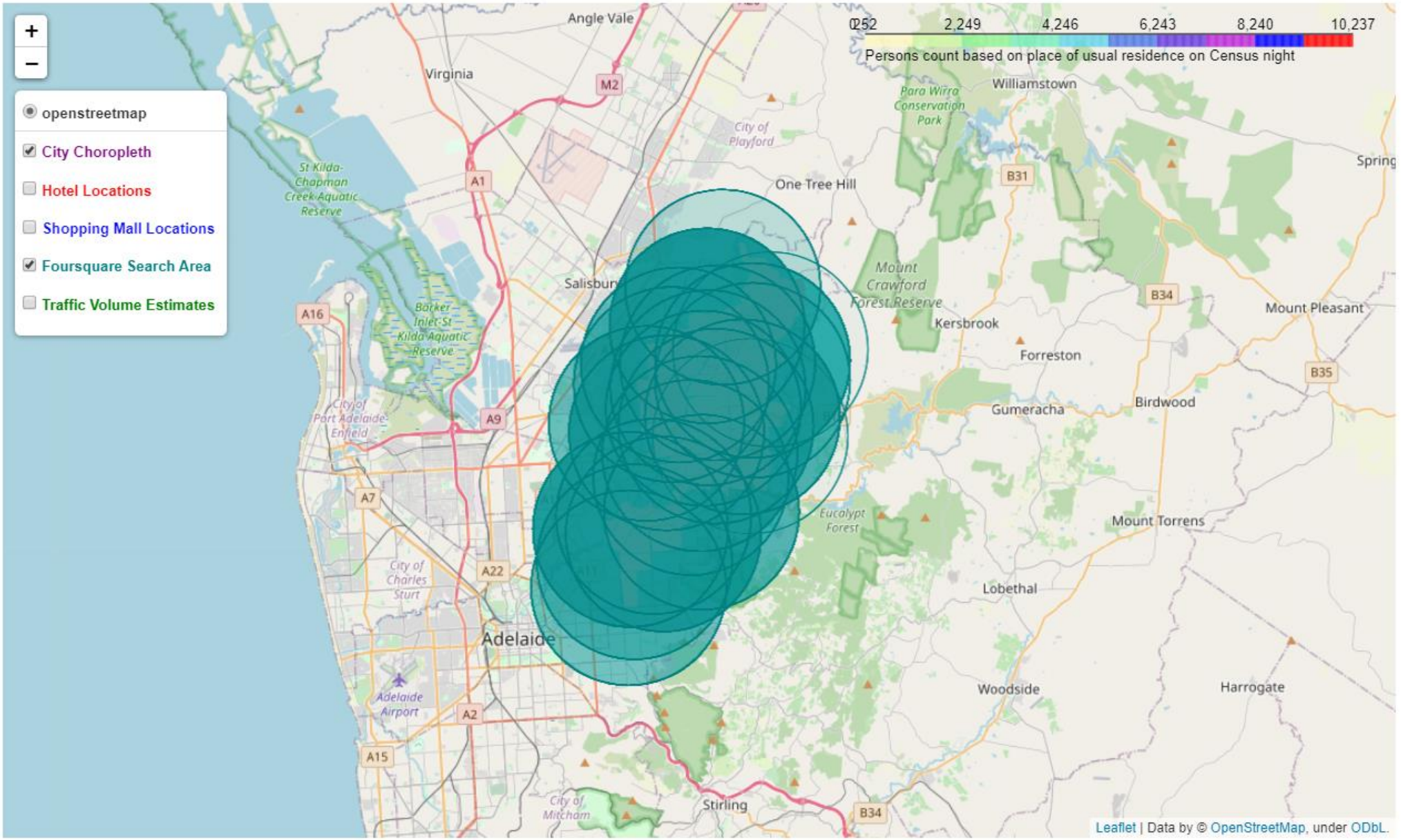
Map 1. South Australian Local Government Areas (LGA) ranked by Average NGR per machine



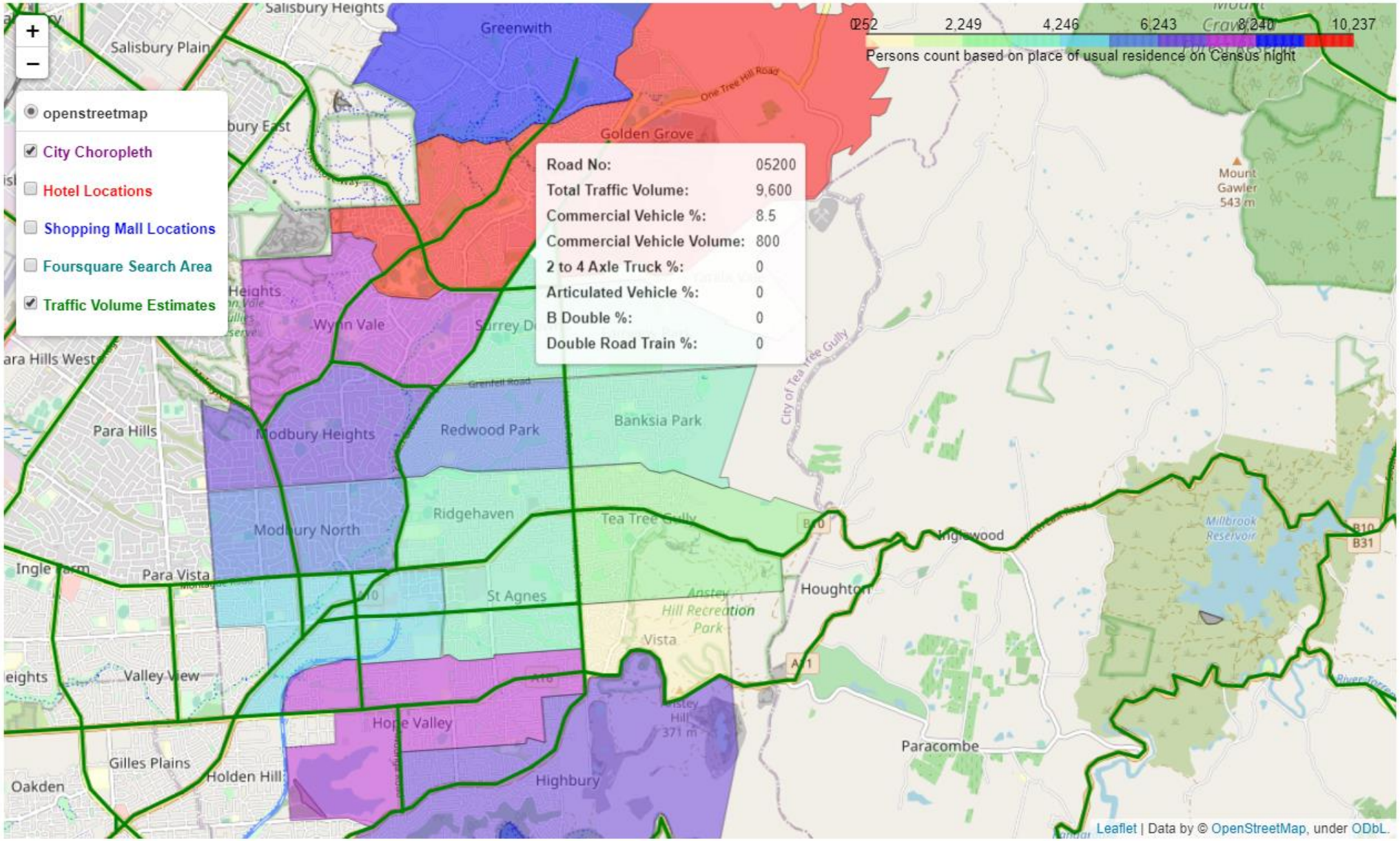
Map 2. Population by City in Local Government Area (Campbelltown / Tea Tree Gully)



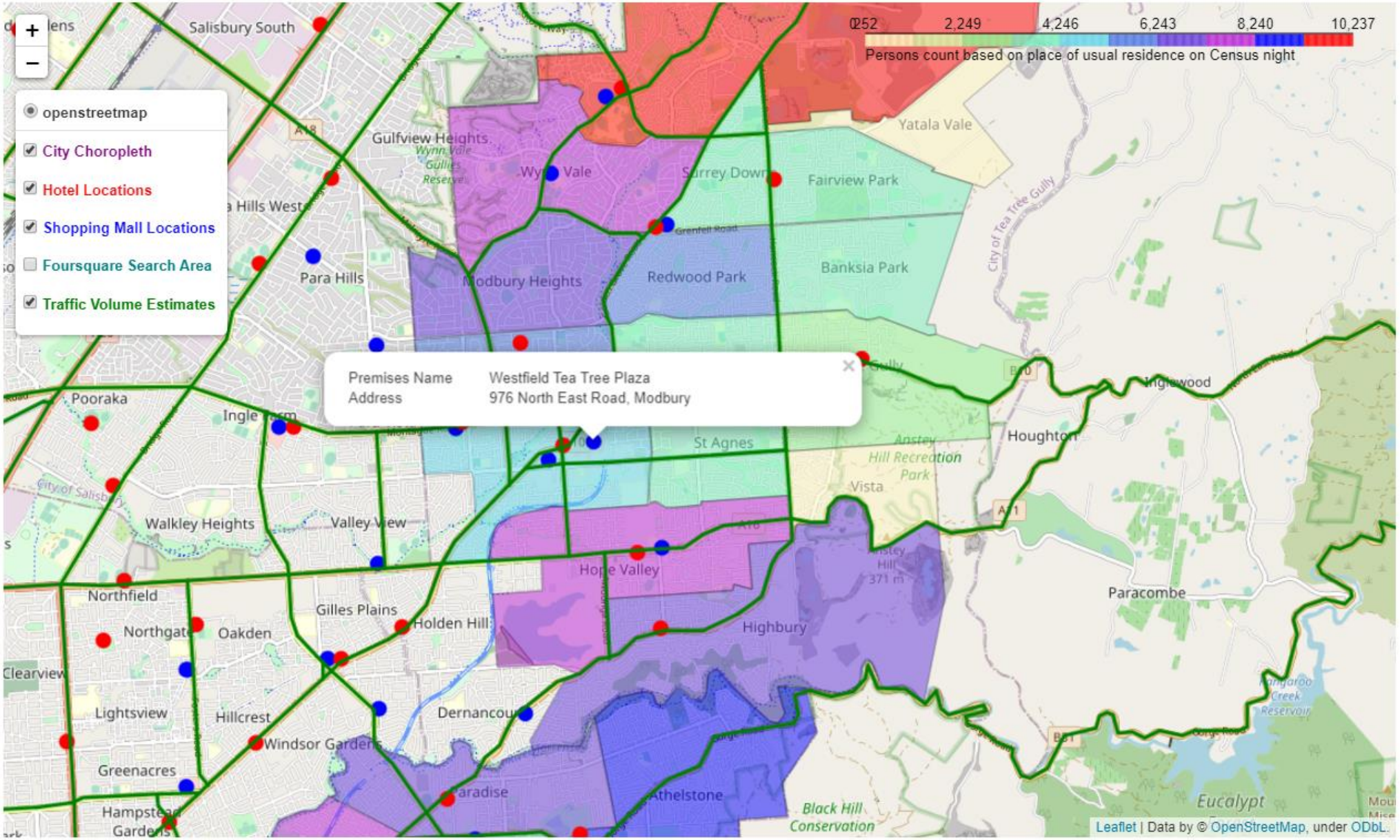
Map 3. Foursquare venue and Google Place API search area



Map 4. Traffic Volume by City in Local Government Area (Campbelltown / Tea Tree Gully)



Map 5. Shopping Malls, Traffic Volume and Hotel by City in Local Government Area (Campbelltown / Tea Tree Gully)



Cluster Analysis of City Amenities

Categorical Data

Foursquare Places API was used to obtain venue information for each of the Top Net Gaming Revenue Cities. There are five Hierarchy Level categories, with Level 2 the most predominate.
The most granular venue Hierarchy containing data is used in the Analysis.
Venues were reverse geocoded to accurately locate them within the bounds of each City

```
274 unique categories in 23 cities
foursquare_venue_explore_df shape is  (1446, 44)
```

	VenueId	VenueName	VenueLatitude	VenueLongitude	CategoryId	VenueAddress	City	HierarchyLevel	CategoryName	Level1	Level2	Level3
0	4c874f439c808cfad73e131e	Turramurra	-34.853200	138.700771	4bf58dd8d48988d18b941735	1000 Lower North East Rd, Highbury SA 5089, Australia	Highbury	3	Basketball Stadium	Arts & Entertainment	Stadium	Basketball Stadium
2	4de7344eb0fb2165e1a08def	Golden Grove Rec Centre	-34.792939	138.694200	4bf58dd8d48988d18b941735	2 Tenison Pl, Golden Grove SA 5125, Australia	Golden Grove	3	Basketball Stadium	Arts & Entertainment	Stadium	Basketball Stadium
3	4dfd624f483b96a3aaa974aa	Foxfield Football Oval	-34.880648	138.701467	4bf58dd8d48988d189941735	84 Maryvale Rd, Athelstone SA 5076, Australia	Athelstone	3	Football Stadium	Arts & Entertainment	Stadium	Football Stadium

One-Hot Encoding

Machine Learning algorithms require categorical data to be converted to a numerical form.

1. Each Venue category value is converted into a new column and assigned a 1 or 0
2. The rows are then Grouped by City with the mean of frequency of each venue category

```
foursquare_venue_one_shot_grouped_df shape is  (23, 275)
```

	City	Accessories Store	American Restaurant	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Assisted Living	Athletics & Sports	Australian Restaurant	Auto Garage	Automotive Shop	BBQ Joint
0	Athelstone	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Banksia Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Campbelltown	0.000000	0.000000	0.009709	0.000000	0.000000	0.000000	0.009709	0.000000	0.009709	0.019417	0.000000	0.000000	0.000000
3	Fairview Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.012821
4	Golden Grove	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.022472	0.000000	0.000000	0.005618	0.000000	0.022472	0.005618
5	Greenwith	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Hectorville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000
7	Highbury	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.000000	0.000000	0.000000
8	Hope Valley	0.000000	0.019608	0.000000	0.019608	0.000000	0.000000	0.019608	0.019608	0.000000	0.000000	0.000000	0.019608	0.000000
9	Modbury	0.008021	0.002674	0.000000	0.005348	0.000000	0.005348	0.008021	0.000000	0.000000	0.002674	0.002674	0.018717	0.002674
10	Modbury Heights	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000
11	Modbury North	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023810	0.000000	0.000000	0.023810	0.000000	0.000000	0.000000
12	Newton	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.011628	0.000000	0.000000	0.011628	0.000000	0.000000	0.011628
13	Paradise	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	Redwood Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	Ridgehaven	0.000000	0.000000	0.000000	0.028571	0.000000	0.000000	0.014286	0.000000	0.014286	0.014286	0.014286	0.071429	0.000000
16	Saint Agnes	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.015152	0.000000	0.000000	0.000000	0.000000	0.045455	0.000000
17	Surrey Downs	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.031250	0.000000	0.000000	0.000000	0.000000	0.031250	0.000000
18	Tea Tree Gully	0.000000	0.000000	0.000000	0.000000	0.017857	0.000000	0.035714	0.000000	0.000000	0.000000	0.000000	0.000000	0.017857
19	Tranmere	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	Vista	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	Wynn Vale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	Yatala Vale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Determine the optimal number of clusters

Clustering is the process of making a group of abstract objects into classes of similar objects.

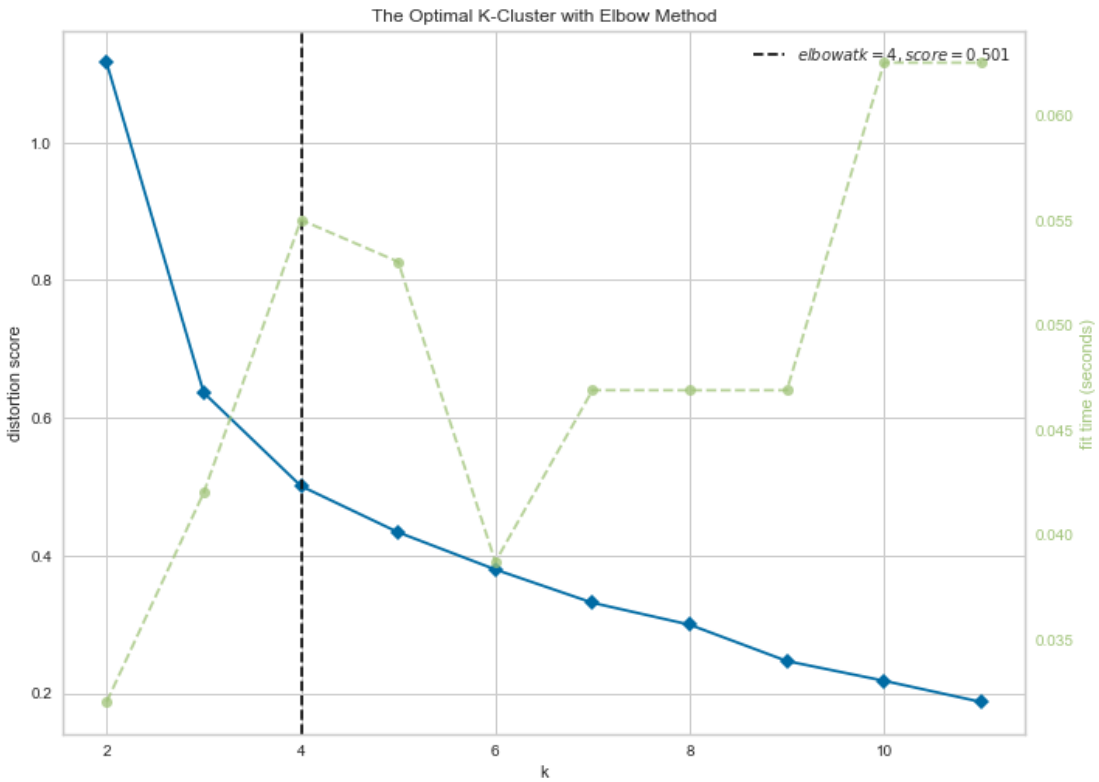
For a certain class of clustering algorithms (in particular k-means), there is a parameter commonly referred to as K that specifies the number of clusters to detect.
The correct choice of K is often ambiguous, with interpretations depending on the shape and scale of the distribution of points in a data set and the desired clustering resolution of the user.

In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use.

If K increases, average distortion will decrease, each cluster will have fewer constituent instances, and the instances will be closer to their respective centroids. However, the improvements in average distortion will decline as K increases. The value of K at which improvement in distortion declines the most is called the elbow, at which we should stop dividing the data into further clusters.

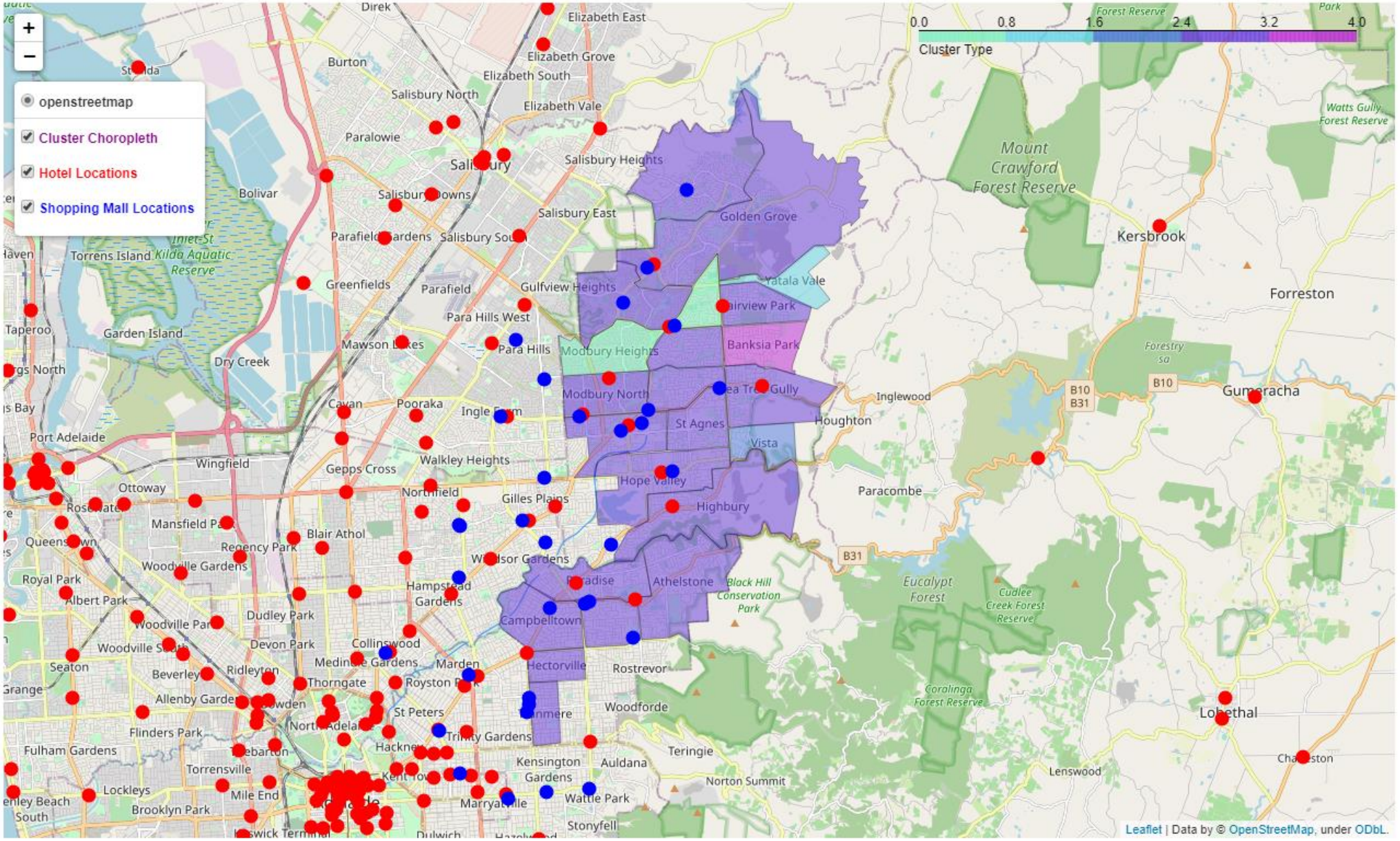
Yellowbrick's KElbowVisualizer implements the “elbow” method of selecting the optimal number of clusters by fitting the K-Means model with a range of values for K.

The KElbowVisualizer recommends K=4 but after examining results K=5 provided a better outcome



	City	ClusterLabel	ClusterDescription	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Athelstone	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Gas Station	Soccer Field	Pizza Place	Bus Line	Church	Playground	Cemetery	Liquor Store	Bus Station	Nursery School
1	Banksia Park	4	Cluster 4: Recreational Area (Golfcourse, Gyms, Parks)	Middle School	Playground	Football Stadium	Church	Public Art	Salon / Barbershop	Bus Station	Financial or Legal Service	Fire Station	Fish & Chips Shop
2	Campbelltown	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Pizza Place	Salon / Barbershop	Grocery Store	Office	Bakery	Miscellaneous Shop	Housing Development	Bowling Alley	Park	Food Truck
3	Fairview Park	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Park	Trail	Butcher	Bus Station	Gas Station	Playground	Pharmacy	Bus Line	Tennis Court	Dog Run
4	Golden Grove	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Miscellaneous Shop	Salon / Barbershop	Bank	Bus Station	Cafe	Automotive Shop	Gas Station	Asian Restaurant	Stables	Electronics Store
5	Greenwith	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Pizza Place	Community Center	College & University	Liquor Store	Dessert Shop	Building	Italian Restaurant	Grocery Store	Medical Center	Fish & Chips Shop
6	Hectorville	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Deli / Bodega	Gas Station	Pharmacy	Electronics Store	Furniture / Home Store	College Academic Building	Field	Bus Station	Church	Salon / Barbershop
7	Highbury	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Bus Station	Lake	Yoga Studio	Office	Cafe	Chinese Restaurant	College Technology Building	Pub	Post Office	Medical Lab
8	Hope Valley	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Playground	Pizza Place	Park	Bakery	Church	Grocery Store	Pub	Daycare	Miscellaneous Shop	Sports Bar
9	Modbury	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Cafe	Salon / Barbershop	Women's Store	Bank	Shoe Store	Jewelry Store	Department Store	Doctor's Office	Electronics Store	Cosmetics Shop
10	Modbury Heights	0	Cluster 0: Quiet Suburban	Bus Line	Playground	Park	Bus Station	Shopping Mall	Liquor Store	Dog Run	Salon / Barbershop	Gaming Cafe	Sporting Goods Shop
11	Modbury North	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Church	Bus Line	Indian Restaurant	Salon / Barbershop	Gas Station	Gym	Bookstore	Bus Station	Soccer Field	Office
12	Newton	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Bakery	Miscellaneous Shop	Doctor's Office	Pharmacy	Cafe	Shopping Mall	Office	Supermarket	Fruit & Vegetable Store	Pizza Place
13	Paradise	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Bus Stop	Church	School	Bus Station	Park	Cafe	Playground	Rock Club	College Academic Building	Sports Bar
14	Redwood Park	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Park	Coworking Space	Cafe	Chinese Restaurant	Office	Bakery	Fast Food Restaurant	Bus Station	Fried Chicken Joint	Student Center
15	Ridgehaven	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Automotive Shop	Medical Center	Pizza Place	Spa	Building	Sports Club	Shopping Mall	Playground	Gas Station	Bank
16	Saint Agnes	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Automotive Shop	Factory	Gas Station	Shopping Mall	School	Salon / Barbershop	Supermarket	Medical Center	Bus Line	Bakery
17	Surrey Downs	0	Cluster 0: Quiet Suburban	Bus Line	Post Office	Supermarket	Bus Station	Soccer Field	Cafe	School	Restaurant	Residential Building (Apartment / Condo)	Coffee Shop
18	Tea Tree Gully	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Salon / Barbershop	Church	Bus Station	Medical Center	Park	Doctor's Office	Pub	Coffee Shop	Real Estate Office	Asian Restaurant
19	Tranmere	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Lounge	Gas Station	Pizza Place	Korean Restaurant	Park	Gym / Fitness Center	Jewelry Store	Fast Food Restaurant	Medical Center	Shopping Mall
20	Vista	2	Cluster 2: Recreation Park	School	Comedy Club	Yoga Studio	Fast Food Restaurant	Elementary School	Event Space	Factory	Farm	Farmers Market	Field
21	Wynn Vale	3	Cluster 3: Busy Area (Shopping Malls, Densely Populated, High Traffic Volume)	Pizza Place	Grocery Store	Playground	Doctor's Office	Dog Run	Pharmacy	Bus Station	Cafe	Fish & Chips Shop	Event Space
22	Yatala Vale	1	Cluster 1: Rural	Garden Center	Beach	Yoga Studio	Fast Food Restaurant	Elementary School	Event Space	Factory	Farm	Farmers Market	Field

Map 6. Venue Category Clusters



Results

Results section states the findings, without bias or interpretation, and arranged in a logical sequence

Histograms

Observations:

- 86% revenue increase between rank 1st (desired) and rank 10th (actual hotel income)
- The highest-ranking LGA group is (Campbelltown, Tea Tree Gully)
- Population by City ranges from approx. four to ten thousand

Choropleth Maps

Observations:

- The more disadvantaged a region is in socioeconomic terms (lower average income, higher unemployment, lower educational attainment) the higher EGM expenditure
- Most high ranking LGA hotels have 40 EGMs installed, lower-ranking have fewer EGMs
- A belt of low population density surrounding the city of Tea tree gully
- A Venue search radius of 5km is sufficient to capture all cities in the target area
- North East Road, passing through Modbury has a high traffic volume
  Followed by McIntyre and Kings Road, through Modbury North and Modbury Heights

Cluster Analysis of City Amenities

The following table summarizes the result of each cluster

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Properties	Quiet Suburban	Rural	Recreation Park	Busy Area Shopping Malls, Densely Populated High Traffic Volume	Recreational Area Golf course, Gyms, Parks
Venues	58	2	2	1376	8
Cities	Modbury Heights Surrey Downs	Yatala Vale	Vista	Saint Agnes Ridgehaven Newton Highbury Paradise Modbury Modbury North Hectorville Athelstone Tranmere Golden Grove Redwood Park Fairview Park Greenwith Wynn Vale Campbelltown Hope Valley Tea Tree Gully	Banksia Park

Hotels within the Top Net Gaming Revenue per Machine Cities

Presents all Hotels within Local Government Areas of (Campbelltown, Tea Tree Gully)
This area has the highest Average NGR per machine at $100,625 per machine per year.

Observations:

- One Club has been included in the table
- Most hotels are located within walking distance of major shopping malls
- Most hotels are located on main arterial roads
- Most traffic volume between Medium (10-20K) to High (20-50K)

Hotel	City	Shopping Mall	Traffic per Day	EGMs	Avg Annual Income	Accommodation	Local Government
Blue Gums Hotel	Fairview Park	100 meters	18,100	40	$51,125	Yes	Tea Tree Gully (C)
Campania Sports and Social Club	Modbury North	1.4 km	n/a	20	$47,216	No	Tea Tree Gully (C)
Clovercrest Hotel Motel	Modbury North	1.6 km	16,000	40	$47,216	Yes	Tea Tree Gully (C)
Golden Grove Tavern	Surrey Downs	1.2 km	16,700	40	$51,125	No	Tea Tree Gully (C)
Highbury Hotel	Highbury	2.1 km	14,000	40	$54,869	No	Tea Tree Gully (C)
Modbury Plaza Hotel	Modbury	100 meters	30,100	40	$47,216	No	Tea Tree Gully (C)
Paradise Hotel	Paradise	50 meters	32,100	40	$49,424	Yes	Campbelltown (C)
Rezz Hotel	Newton	1.1 km	13,400	40	$48,149	Yes	Campbelltown (C)
Tea Tree Gully Hotel	Tea Tree Gully	1.2 km	7,000	40	$51,939	Yes	Tea Tree Gully (C)
The Valley Inn	Hope Valley	1.3 km	16,300	40	$45,713	No	Tea Tree Gully (C)
Village Tavern	Golden Grove	130 meters	22,500	40	$56,373	No	Tea Tree Gully (C)

Hotel	Address	Web Site
Blue Gums Hotel	345 Hancock Rd, Fairview Park SA 5126, Australia	https://www.bluegumshotel.com.au/
Campania Sports and Social Club	5 Famechon Cres, Modbury North SA 5092, Australia	https://www.facebook.com/campania.clubsa/
Clovercrest Hotel Motel	450 Montague Rd, Modbury North SA 5092, Australia	http://www.clovercresthotel.com.au/
Golden Grove Tavern	Golden Grove Rd, Surrey Downs SA 5126, Australia	http://ggtavern.com.au/
Highbury Hotel	1017 Lower North East Rd, Highbury SA 5089, Australia	https://www.highburyhotel.com.au/
Modbury Plaza Hotel	989 North East Road, Modbury SA 5092, Australia	https://www.modburyplazahotel.com.au/
Paradise Hotel	700 Lower North East Rd, Paradise SA 5075, Australia	http://www.paradisehotel.com.au/
Rezz Hotel	20 Hamilton Terrace, Newton SA 5074, Australia	https://www.rezz.com.au/
Tea Tree Gully Hotel	1349 North East Road, Tea Tree Gully SA 5091, Australia	https://www.thegullyphg.com.au/
The Valley Inn	1210 Grand Jct Rd, Hope Valley SA 5090, Australia	https://www.thevalleyinn.com.au/contact.html
Village Tavern	The Golden Way &, The Grove Way, Golden Grove SA 5125, Australia	https://www.villagetavern.com.au/

Gambling Indicator	Local Government Area	2018/19
Number of venues 30 June	Campbelltown (C), Tea Tree Gully(C)	12
Number of EGMs 30 June	Campbelltown (C), Tea Tree Gully(C)	438
Net gambling revenue	Campbelltown (C), Tea Tree Gully(C)	$43,922,813
NGR per adult	Campbelltown (C), Tea Tree Gully(C)	$367
Number of EGMs per 1,000 adults at 30 June	Campbelltown (C), Tea Tree Gully(C)	3.65
Average NGR per machine	Campbelltown (C), Tea Tree Gully(C)	$100,625

Discussion

Explanation and interpretation of results or findings

Determinants of Hotel Location

Spatial location is one of the most important factors for a new hotel establishment.
As a service industry, hotels are keen on locations that are proximate to their potential markets.

Agglomeration effect refers to benefits the hotel can receive from the clustering of economic activity. Generally, one incentive for choosing to locate close to shopping malls and transport hubs is the positive spill over effects. Traffic is one of the primary factors that can impact business operations. For hospitality establishments, high foot and highway traffic areas are generally the best places to start a business.

Sale of liquor and other beverages (take away) account for 25-35% hotel revenue
If the hotel bottle shop is one of many in a three-mile radius, then it may be forced to compete solely on price. Doing so makes reaching profitability considerably more difficult. But provided the hotel offers, a wide selection of products and offers like lottery and tobacco, location will matter less

A more detailed survey of the twelve Hotels within Local Government Areas of (Campbelltown, Tea Tree Gully) is recommended to determine approximate NGR revenue rank by City by hourly gaming floor headcount.

Access to the Independent Gaming Corporation Ltd (IGC) web portal would provide weekly Local Government Area, Net Gaming Revenues rather than the annual dataset used in this report.

Proposed Hotel Locations

City	Recommended	Comment	Recommended Road	Population	Traffic per Day
Athelstone	No	Residential; Small shopping mall	Gorge Rd	9,149	2,900
Banksia Park	No	Recreational Area; Small shopping mall	Hancock Rd	3,286	16,500
Campbelltown	Yes	Residential; Small shopping mall	Lower North East Rd	8,152	37,700
Fairview Park	No	Residential; Blue Gums Hotel	Hancock Rd	4,375	16,500
Golden Grove	Yes	High Population; Large shopping mall; Village Tavern	The Golden Way	10,237	11,400
Greenwith	No	Residential; Small shopping mall	Target Hill Road	8,992	n/a
Hectorville	No	Residential; To close to Glynde Hotel	Glynburn Rd	3,738	22,100
Highbury	No	Residential; Too close to Highbury Hotel	Lower North East Rd	6,588	19,000
Hope Valley	No	Residential; Shopping; Too close to The Valley Inn	Grand Junction Rd	7,887	16,300
Modbury	Yes	Busy Area; Shopping Malls, Very High Traffic Volume	North East Rd	5,244	44,300
Modbury Heights	Yes	Quiet Suburban; Next to A18 Highway, Check Zoning	McIntyre Rd	6,699	34,600
Modbury North	Yes	Quiet Suburban; Next to A18 Highway, Check Zoning	McIntyre Rd	5,626	32,200
Newton	Yes	Residential; Large shopping mall	St Bernard's Rd	4,788	26,200
Paradise	Yes	O-Barn Interchange, High Traffic Volume	Darley Rd	6,776	37,600
Redwood Park	No	Residential	Golden Grove Rd	5,417	23,000
Ridgehaven	Yes	Residential; Large shopping mall (Newton Central)	Golden Grove Rd	4,008	26,400
Saint Agnes	Yes	Residential; shopping mall	North East Rd	4,136	17,100
Surrey Downs	No	Residential; Shopping; Too close to Golden Grove Tavern	Golden Grove Rd	3,326	16,700
Tea Tree Gully	No	Residential; Shopping; Too close to Tea Tree Gully Hotel	North East Rd	3,244	7,000
Tranmere	Yes	Residential; shopping mall	Glynburn Rd	3,757	28,800
Vista	No	Recreation Park	Lower North East Rd	971	2,500
Wynn Vale	Yes	Residential; shopping mall	The Golden Way	7,756	28,200
Yatala Vale	No	Rural	n/a	252	0

Conclusion

The methodology section describes the data exploration that were conducted and how they relate to the original business problem of providing two or more locations in Metropolitan Adelaide for a new hotel based on criteria:

a. High gaming revenue area
b. Competitor location, avoid locations where the competitor hotel has a presence
c. Located on a main arterial road with medium to high traffic volume
d. Located in proximity to shopping complex or transport hub
e. High population density
f. Provide Geodemographic classification of city amenities by applying a clustering algorithm
g. Customer demographics for customer insights and segmentation

The discussion section provided explanation and interpretation of results. It also explained determinants of Hotel Location and proposed a list of Cities

This project explored data insights that contribute to the identification of hotel locations in Metropolitan Adelaide