

Australian Cities clustering using population and venues

Lahiri Bellarykar

January 28 2021

1. Introduction:

1.1 Background

Australia is one of the most attractive tourist destinations, due to its sunny weather and endless options. Here are some official data if you may want to take a look

<https://www.tra.gov.au/Data-and-Research/publications>

Australia is a big country. Its land area is 7.692 million sq km.

Enormous landmass and low population make it a different country altogether. Please refer to the following link for more on Demographics

https://en.wikipedia.org/wiki/Demography_of_Australia

While traveling across the Australia, travelers do not know much about every Australian city. But in the age of internet, it is very common that, they do have an idea about certain major cities/towns/settlements. The knowledge can be augmented with a Data science based solution.

1.2 Problem

When an outsider traveler enters Australia, they have many questions about what to expect from their journey to a particular city/town. As a traveler, it is common to have a basic idea about a new country and few major/well-known cities in it.

The gap however is when a traveler wants to travel to a certain city/town it is difficult to guess what to expect at the new place. In this situation, a similar city or set of similar cities can be presented to the traveler. Thereby a traveler learns about a new place by comparing it with other cities that he already knows about.

e.g If a traveler is going to visit Victoria, Melton, then a comparison with Melbourne would be helpful to understand what to expect.

There can be different ways to help them set their expectations right. One of the most intuitive ways is to group similar cities. Which makes it easier for a traveler to learn comparatively about a city (A good read on the Human judgment by using comparison:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222713/>).

For a traveler, venues in each city are more important, because that is what a normal traveler is out to explore. Thus, using venues in a city as a feature to group similar cities can be a productive approach.

It would be a good help indeed to have a model to compare Australian cities/towns with each other, to plan a trip better. This model can group similar cities/towns making it easy for a traveler to know which cities are similar to each other and which are not.

Thus, we propose to use venues in each city, city rank, city population, city population growth, state, to group similar cities. Travelers can be presented with the groups of cities. So, they can know which group a given city belongs to. And in comparison, to other known cities travelers can reasonably guess what they can expect from a given city

1.3 Interest

A traveler is definitely interested in this model, as he can easily see a group of cities and can instantaneously know what the new city like. The model creates groups of the cities and it can be readily made available over a web API using a variety of technologies, user interfaces, and communication methods.

2. Data acquisition and cleaning

2.1 Data sources

To get the required data we use following sources and their purpose.

Data source	Purpose
https://en.wikipedia.org/wiki/List_of_cities_in_Australia_by_population	To get data about Australian cities list, city population, population growth, city rank
Python Geopy library	To get Latitude, Longitude of each Australian city
Foursquare API	To get the venues data for each Australian city

2.2 Data cleaning

Data about Australian cities need to be scraped from

https://en.wikipedia.org/wiki/List_of_cities_in_Australia_by_population

As the data comes from an HTML page, it needs a good cleanup to prepare it for EDA

As part of the cleanup, we do the following.

Data	Cleanup done
Growth and percent_of_national_population City fields Scraped from Wikipedia URL	+, - , % signs are removed from the numerical string so that it can be used for Data analysis
Population2019, Growth ,percent_of_national_population City fields	These fields are casted to Integer, float, float respectively
Venue_Category for each city taken from Foursquare API	Data from Foursquare API does return empty category data for some venues. Such 'Not assigned' venue category records for a city are of no use for Data analysis, hence these records are removed.

2.3 Feature selection

It is important to understand which features give distinct and important information about a feature for it to be valuable for building clustering model

Following table lists features and feature categories

Features used	Features dropped	Reason for dropping features
Rank		
Population2019		
Growth		
	percent_of_national_population	High correlation with Population2019 and does not represent any new information useful for clustering cities
State		
	Population2011	2011 population does not add any new information about a city for clustering the cities
Venue Categories (there are 477 venue categories). Please refer Appendix 1		

3. Exploratory Data Analysis

As most of the features used for clustering in this case are categorical, their mean values for each city describe that category distinctively without any repetition of information. Thus, no need to check correlation among categorical features.

We pick numerical values though and check their correlation

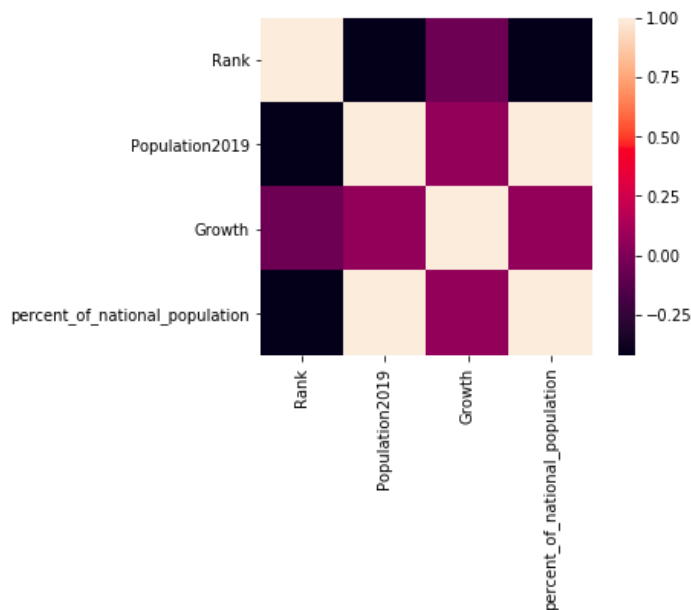


Figure 1. Correlation of numeric features.

From correlation we can see there is High correlation between **percent_of_national_population** and **Population2019**. And **percent_of_national_population** does not represent any new information useful for clustering cities, thus we drop it.

4. Clustering model

To cluster Australian cities, we use K-Means clustering. To find the best **k** value we use **distortion** and **inertias** and graphically find the elbow.

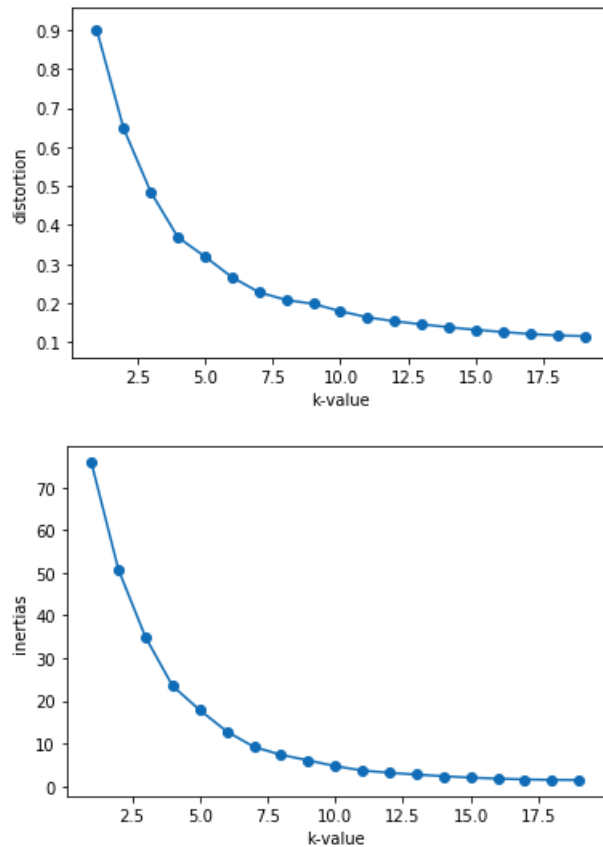


Figure 2. Scatter plot of distortion and Inertias against k-value

Elbow occurs at k = 5

Thus, we use $k = 5$ and then cluster the Australian cities. Let's take a look at the clustered Australian cities in the Map

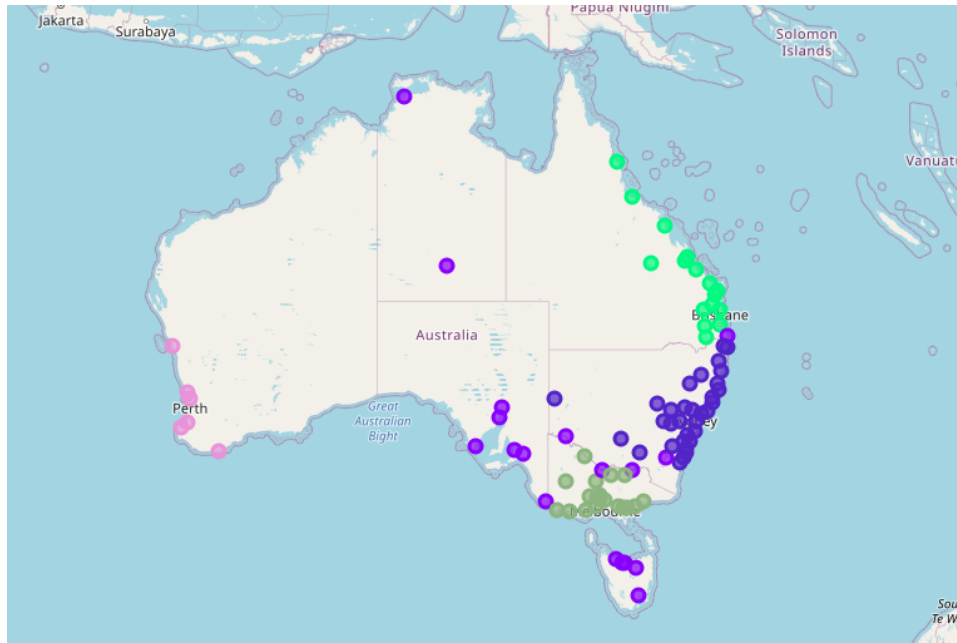


Figure 3. Clustered Australian cities displayed on the Folium Map (colors indicate clusters)

5. Conclusion

Australian cities within a particular state are largely similar to each other.

They form a group as they have common venues. As we can clearly see Sydney and Melbourne are major cities and many consider them as similar in terms of a major role in Australia's economy. But our intention was to cluster the cities for the purpose of travelers, that is why we see that Sydney and Melbourne belong to a different cluster. And that is justifiable. Thus, it is very important to decide the purpose of clustering cities before we do because depending on purpose one city can move to a different cluster.

6. Future direction

Current venues data for cities can be improved by using different venue information providers or by using paid Foursquare services. We can also study traveler's exploration of Australia in the past and add applicable features build a clustering system even better.

There can be many good ideas to cluster the cities depending on specific purposes. E.g city clustering by the kind of food available, clustering the city by type of adventure sports, clustering cities by natural attractions.

All these clustering use cases are possible if we have enough data to model it.

Appendix

1. List of all features considered for clustering

City,Australian Capital Territory/New South Wales,New South Wales,New South Wales/Victoria,Northern Territory,Queensland,Queensland/New South Wales,South Australia,Tasmania,Victoria,Victoria/New South Wales,Western Australia,ATM,Accessories Store,Acupuncturist,Adult Education Center,Advertising Agency,Afghan Restaurant,African Restaurant,Airport,Airport Gate,Airport Lounge,Airport Service,Airport Terminal,Alternative Healer,American Restaurant,Animal Shelter,Antique Shop,Aquarium,Arcade,Architecture Firm,Art Gallery,Arts & Crafts Store,Arts & Entertainment,Asian Restaurant,Assisted Living,Athletics & Sports,Auditorium,Australian Restaurant,Austrian Restaurant,Auto Dealership,Auto Garage,Auto Workshop,Automotive Shop,BBQ Joint,Bagel Shop,Baggage Claim,Bakery,Bank,Bar,Baseball Field,Basketball Court,Basketball Stadium,Bath House,Bay,Beach,Bed & Breakfast,Beer Bar,Beer Garden,Beer Store,Betting Shop,Big Box Store,Bike Rental / Bike Share,Bike Shop,Bike Trail,Bistro,Blood Donation Center,Board Shop,Boarding House,Boat Launch,Boat Rental,Boat or Ferry,Bookstore,Border Crossing,Boutique,Bowling Alley,Breakfast Spot,Brewery,Bridal Shop,Bridge,Bubble Tea Shop,Building,Burger Joint,Burrito Place,Bus Line,Bus Station,Bus Stop,Business Center,Business Service,Butcher,Cafeteria,Café,Camera Store,Campground,Candy Store,Capitol Building,Car Wash,Caribbean Restaurant,Casino,Cave,Cemetery,Cheese Shop,Child Care Service,Chinese Restaurant,Chiropractor,Chocolate Shop,Church,City Hall,Climbing Gym,Clothing Store,Club House,Cocktail Bar,Coffee Shop,College & University,College Academic Building,College Administrative Building,College Bookstore,College Classroom,College Lab,College Library,College Quad,College Science Building,College Technology Building,College Theater,Comedy Club,Comfort Food Restaurant,Comic Shop,Community Center,Community College,Concert Hall,Conference Room,Construction & Landscaping,Convenience Store,Convention Center,Cosmetics Shop,Courthouse,Coworking Space,Credit Union,Cricket Ground,Cruise Ship,Cuban Restaurant,Cupcake Shop,Dam,Dance Studio,Daycare,Deli / Bodega,Dentist's Office,Department Store,Design Studio,Dessert Shop,Dim Sum Restaurant,Diner,Discount Store,Dive Bar,Dive Shop,Dive Spot,Doctor's Office,Dog Run,Donut Shop,Drugstore,Dumpling Restaurant,Duty-free Shop,Eastern European Restaurant,Electronics Store,Elementary School,Embassy / Consulate,Emergency Room,Entertainment Service,Escape Room,Event Space,Eye Doctor,Factory,Fair,Falafel Restaurant,Farm,Farmers Market,Fast Food Restaurant,Field,Filipino Restaurant,Financial or Legal Service,Fire Station,Fish & Chips Shop,Fish Market,Fishing Spot,Fishing Store,Flea Market,Flower Shop,Food,Food & Drink Shop,Food Court,Food Service,Food Stand,Food Truck,Football Stadium,Fountain,Fraternity House,French Restaurant,Fried Chicken Joint,Frozen Yogurt Shop,Fruit & Vegetable Store,Funeral Home,Furniture / Home Store,Gaming Cafe,Garden,Garden Center,Gas Station,Gastropub,Gay Bar,General College & University,General Entertainment,General Travel,German Restaurant,Gift Shop,Gluten-free Restaurant,Go Kart Track,Golf Course,Gourmet Shop,Government Building,Greek Restaurant,Grocery Store,Gun Range,Gym,Gym / Fitness Center,Gym Pool,Halal Restaurant,Harbor / Marina,Hardware Store,Health & Beauty Service,Health Food Store,Herbs & Spices Store,High School,Historic Site,History Museum,Hobby Shop,Hockey Field,Hospital,Hostel,Hot Dog Joint,Hotel,Hotel Bar,Hotel Pool,Housing Development,Hunting Supply,IT Services,Ice Cream Shop,Indian Restaurant,Indie Movie Theater,Indie Theater,Indonesian Restaurant,Insurance Office,Internet Cafe,Irish Pub,Italian Restaurant,Japanese Restaurant,Jazz Club,Jewelry Store,Juice Bar,Karaoke Bar,Kebab Restaurant,Kids Store,Korean Restaurant,Lake,Language School,Laundromat,Laundry Service,Lawyer,Leather Goods Store,Lebanese Restaurant,Library,Light Rail Station,Lighthouse,Lingerie Store,Liquor Store,Locksmith,Lounge,Luggage

Store, Malay Restaurant, Market, Martial Arts School, Massage Studio, Medical Center, Medical Lab, Medical Supply Store, Mediterranean Restaurant, Meeting Room, Memorial Site, Men's Store, Mental Health Office, Mexican Restaurant, Middle Eastern Restaurant, Middle School, Military Base, Miscellaneous Shop, Mobile Phone Shop, Modern European Restaurant, Molecular Gastronomy Restaurant, Monument / Landmark, Mosque, Motel, Motorsports Shop, Mountain, Movie Theater, Moving Target, Multiplex, Museum, Music Store, Music Venue, Nail Salon, National Park, Nature Preserve, Newsstand, Nightclub, Nightlife Spot, Non-Profit, Noodle House, Nursery School, Office, Opera House, Optical Shop, Organic Grocery, Other Great Outdoors, Other Nightlife, Other Repair Shop, Outdoor Event Space, Outdoor Supply Store, Outdoors & Recreation, Outlet Store, Paper / Office Supplies Store, Park, Parking, Pedestrian Plaza, Performing Arts Venue, Perfume Shop, Pet Service, Pet Store, Pharmacy, Photography Lab, Physical Therapist, Piano Bar, Pie Shop, Pier, Piercing Parlor, Pilates Studio, Pizza Place, Platform, Playground, Plaza, Police Station, Pool, Post Office, Power Plant, Preschool, Print Shop, Professional & Other Places, Pub, Public Art, Public Bathroom, RV Park, Racetrack, Radio Station, Real Estate Office, Record Shop, Recording Studio, Recreation Center, Recruiting Agency, Recycling Facility, Religious School, Rental Car Location, Rental Service, Residential Building (Apartment / Condo), Resort, Rest Area, Restaurant, River, Rock Climbing Spot, Rock Club, Sake Bar, Salad Place, Salon / Barbershop, Sandwich Place, Sausage Shop, Scenic Lookout, School, Science Museum, Sculpture Garden, Seafood Restaurant, Shoe Repair, Shoe Store, Shop & Service, Shopping Mall, Shopping Plaza, Shrine, Singaporean Restaurant, Skate Park, Skating Rink, Smoke Shop, Snack Place, Soccer Field, Soccer Stadium, Social Club, Sorority House, South American Restaurant, Southern / Soul Food Restaurant, Souvenir Shop, Souvlaki Shop, Spa, Spanish Restaurant, Speakeasy, Spiritual Center, Sporting Goods Shop, Sports Bar, Sports Club, Stables, Stadium, State / Provincial Park, Stationery Store, Steakhouse, Storage Facility, Street Food Gathering, Strip Club, Student Center, Summer Camp, Supermarket, Supplement Shop, Surf Spot, Sushi Restaurant, Swim School, Synagogue, TV Station, Tailor Shop, Tanning Salon, Tapas Restaurant, Tattoo Parlor, Taxi Stand, Tea Room, Tech Startup, Temple, Tennis Court, Thai Restaurant, Theater, Theme Park, Thrift / Vintage Store, Tour Provider, Tourist Information Center, Town Hall, Toy / Game Store, Track, Track Stadium, Trade School, Trail, Trailer Park, Train Station, Tram Station, Transportation Service, Travel & Transport, Travel Agency, Tree, Turkish Restaurant, University, Used Bookstore, Vacation Rental, Vegetarian / Vegan Restaurant, Veterinarian, Video Game Store, Video Store, Vietnamese Restaurant, Vineyard, Voting Booth, Warehouse Store, Water Park, Waterfall, Wedding Hall, Whisky Bar, Wine Bar, Wine Shop, Winery, Wings Joint, Women's Store, Yoga Studio, Rank, Population2019, Growth