

Classification of Greater Sydney's Suburbs under Urbanisation

IBM Coursera Data Science Final Capstone Project Week 1 Submission

James Berry, 12/04/2020

Introduction and Project Overview

Sydney is a major commercial city in Australia with consistent annual population growth. The Transport Authority “Transport for New South Wales (TfNSW)” posted population projection data for the city from 2016-2056 which highlights spatial distribution of projected population and its implications.

As the city grows and develops, it becomes increasingly important to examine and understand it quantitatively.

Entrepreneurs, Investors, City Planners and Developers have an interest in identifying early opportunities and signs of urbanisation in underdeveloped neighbourhoods. In order to make such projections, data such as the following used throughout this project can play a vital role in guiding the decision making process identifying opportunities.

This project will use population projection data and foursquare API for following analysis:

1. Classifying neighbourhoods as highly developed, downtown and less/under developed
2. Understanding how urban footprint of Sydney will expand
3. Exploring underdeveloped neighbourhoods looking at remarkable population growth
4. Identifying business opportunities in urbanizing neighborhoods

Data

Data will be obtained from reliable official data sources only for analysis. In order to best illustrate the topic of this project and present the results and findings, the following data will be used:

1. Population Projection from <https://opendata.transport.nsw.gov.au/dataset/population-projections>

The population dataset, an excel file, is provided by TfNSW Open Data Hub under Creative Commons Attribution 4.0 International (CC BY 4.0) Licence. It aggregates population projections for different geographical divisions from 2016 to 2056.

Spatial geographies used in this dataset are:

- Sydney Greater Capital City Statistical Area (SGCCSA) – This excludes areas of Newcastle and Wollongong from Sydney Greater Metropolitan Area. We will analyze only this area and henceforth refer to it as just Sydney.
- Travel Zones (TZ) – There are 2345 TZ's.
- Statistical Areas 2 (SA 2) – This geography is approximately the same size of a suburb and can be useful for reporting and reviewing of results at a local neighborhood level. This is the area we will use for our analysis and refer to it as neighborhoods. There are 249 neighborhoods in Sydney.
- Statistical Areas 3 (SA 3) – There are 45 SA3s.
- Statistical Areas 4 (SA 4) – Sub-regional geography used for collecting demographic data. There are 14 SA4's. However, it is too broad for our analysis to mean anything.
- GSC District- There are 6 districts.
- Local Government Areas (LGA) - These are political boundaries which may not always align with functional land use areas.

Data of Estimated Population Projection (ERP) from 2016- 2036 is retrieved in a Pandas DataFrame and grouped by SA2 (renamed as Areas). From this data a new column is created containing % of population growth from 2016 to 2036.

2. Latitude and Longitude of each neighbourhood is retrieved using Geocoder from Geopy Library of Python.

3. Foursquare Developers Access to venue data: <https://foursquare.com/>

Foursquare API is typically used to explore types of venues and their frequency of occurrence in each of the dataset's neighbourhoods in the Greater Sydney area. This data will form the basis in order to classify neighbourhoods in Sydney relative to the rate of urbanisation and general development within them.