Classifying Urbanizing Sectors of Sydney Greater Capital Area

By Vrinda Maru Kansal, December 4, 2019

Introduction to the opportunity

Sydney GMA is a growing city with many opportunities for entrepreneurs and working professionals. 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 growing urban footprint in underdeveloped neighborhoods.

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

1. Classifying neighborhoods as highly developed, downtown and less/under developed

2. Understanding how urban footprint of Sydney will expand

3. Exploring underdeveloped neighborhoods looking at remarkable population growth

4. Identifying business opportunities in urbanizing neighborhoods

Data

We need data from reliable sources for analysis. To understand our problem and quantify results we will use following data:

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 2 (SA 3)** – There are 45 SA3s.
- **Statistical Areas 4 (SA 2)** – 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 neighborhood is retrieved using Geocoder from Geopy Library of Python.

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

Foursquare API is used to explore types of venues and their frequencies in each neighborhood. This data is used to classify neighborhoods based on their urban development.