Australian Population Characteristics & Venues Data Analysis for young professionals

1. Description of the Capstone Project / Business Scope /

Australia is one of the largest countries in Asia Pacific region with population over 25 million people and shy population density of 3 persons per square kilometer. The specific census organisation /aka stakeholder/ approached us to conduct a deeper investigation of how population of Australia is spread across the continent and which cities accommodate majority of its citizens. Once the biggest urban area is selected, the stakeholder wants to know which suburbs are the most popular with young professionals and what venues/categories are associated with those suburbs. This will allow local state departments to build smarter infrastructure planning strategies and create more venues if necessary.

2. Data Sources

To provide the analysis that meet stakeholder's requirement, the following data sources have been taken into consideration:

2a. Live population statistics from the 'Worldometers' website:
https://www.worldometers.info/world-population/australia-population/

2b. Geo location information (latitude/longitude) from Google map

2c. Real estate statistics from 'OpenAgent' website that indicate urban suburbs popular with young professionals:
https://www.openagent.com.au/blog/best-suburbs-to-live-in-sydney

2d. **Foursquare API** to get the most common venues and its categories of given suburbs.

3. Data Preparation

All data have been carefully analysed and assessed against the scope of the project.

3a. Live population statistics from 'Worldometers' will be scraped from the website and injected into multiple Pandas DataFrame within Jupyter notebook.

3b. Geo location information will be recovered from Goggle Map's 'Search' engine and injected into csv files accordingly. Those files then will be uploaded to Jupyter notebooks and modified into Pandas Data Frame.

3c. Real estate statistics from ‘OpenAgent’ site will be scraped from website and injected into manually created csv file with names of suburbs popular for young professionals and its geo locations. Then the csv file will be uploaded into Jupyte notebook and translated into Pandas Data Frame.

4c. Venus information will be extracted from the **Foursquare API** and injected into Pandas Data Frame.

4. Data Description and Analysis

All collected data will be modified into Pandas-related Data Frames for further investigation. Once injected into Juyter notebook, all data frames will be correctly indexed and displayed in a form of tables to emphasise the business scope.

The bar graph, scatter plot and folium maps with markers will visualise the corresponding data and its validity against the business objectives.