Applied Data Science Capstone

Melbourne East vs Melbourne West

Business problem:
Melbourne, Australia enjoys its super rich diversity both culturally and socially. It attracts tourists from all around the world to experience its unique character and food culture. Due to the fact that Melbourne is identified as the most diverse city across Australia, it welcomes various kinds of food and beverages (F&B) services from various countries for its residents.

The effect of historic events and the direction of urban development/growth has favoured more towards the eastern areas from the Central Business District (CBD) whilst the western and northern areas are less prioritised. This has caused a clear segregated demographic profile of communities. The map below highlights areas based on an Index of Relative Socio-economic Disadvantage (IRSD) SEIFA scores such that higher intensity of red are areas that are disadvantaged in terms of income, occupation and education. It is evident that higher socio-economic communities are located at the east whilst more disadvantaged communities are located at the west and north.

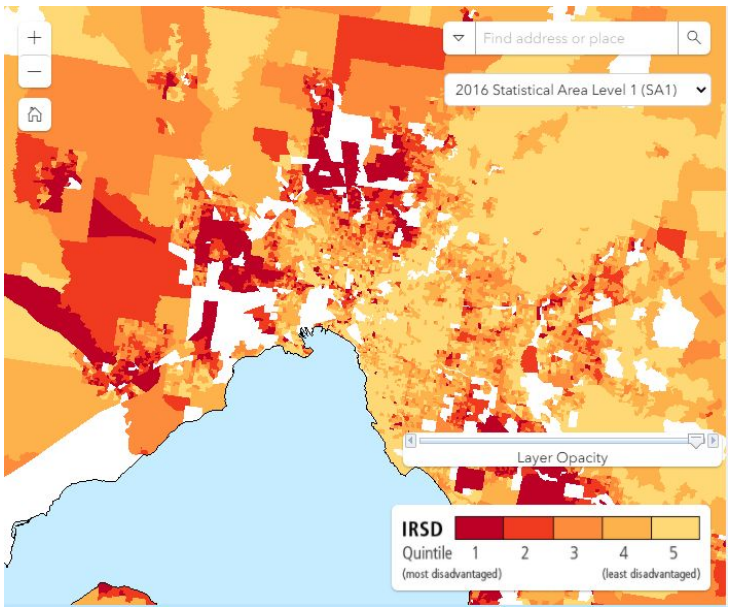

Image sourced from
https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~IRSD%20Interactive%20Map~15

The goal of this study is to understand what are the top food and drink premises found in Melbourne inner east suburbs and Melbourne's West. This is so that informed decisions can be made on what sort of F&B services should be pursued that the community likes.

Research questions:

1. Do Melbourne's east and west have similar clusters of F&B services?
2. Is there a frequent occurrence of the same type of businesses in both areas?

3. What would be the best F&B services to be pursued in both west and east of Melbourne?

Data extraction:
Data of suburbs boundaries will be acquired from the Australian Bureau of Statistics (ABS). Moreover, metadata such as suburbs within melbourne western and eastern boundaries will be taken from the ABS Statistical Area Level 2 (SA2) areas. SA2 areas are used for the dissemination of population estimates and to provide a human perspective by using a scale that can be readily understood as a locality.

A geographic information software known as QGIS will be used to visualise these spatial data and undertake a geoprocessing process to acquire the centroids (points) of each suburb as well as its latitude and longitude.

After that is done, the data will be exported to a CSV file for it to be read and further analyse it with python coding and jupyter notebook.

Methodology:
Foursquare location data will be used here to disseminate the top 10 most frequent or occurrence of venue categories that are located within Melbourne east and west. Scikit python library will be used to undertake a K means clustering to carry out the clustering of similar venue categories within the neighbourhood. Venue categories other than food and drinks will be excluded from the study if it is identified within the top 10 and will be replaced with the next one below the line.