# DATA

## Data Sources

Two main sources of data are applied for this project:

• Foursquare location data: Foursquare is a local search-and-discovery service. it features a developer API that lets third-party applications make use of Foursquare’s location data. I use this API to search and find out about different venues and their categories of each neighbourhood.

• [Melbourne Housing Snapshot](https://www.kaggle.com/dansbecker/melbourne-housing-snapshot): I download this dataset from Kaggle.com to understand the price of houses in Melbourne. This dataset contains a lot of features but I only use the ‘price’ to normalize housing price and location of the house.

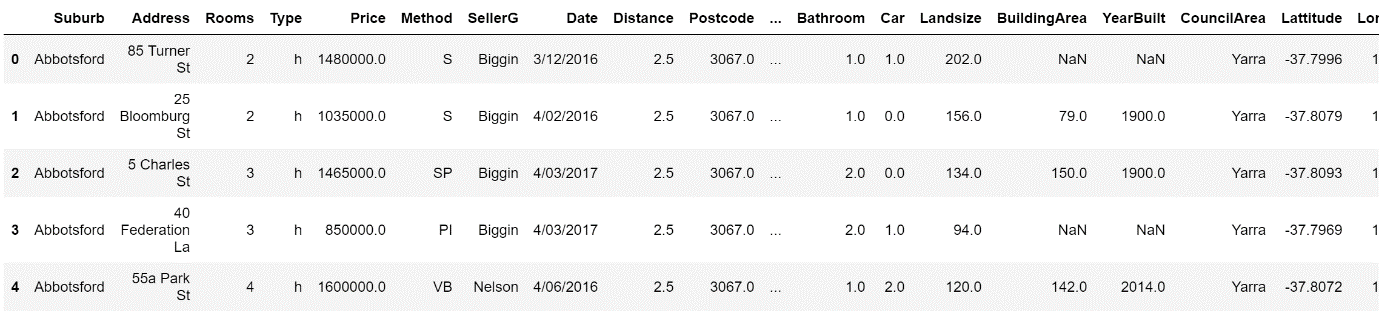


Table 1: Melbourne Housing Dataset

## Data Cleaning & Feature Selection

To simplify this problem, we do not consider rural areas, only metropolitan areas are investigated. In order to find out what is the location of each neighbourhood, the dataset is grouped by region and council area.

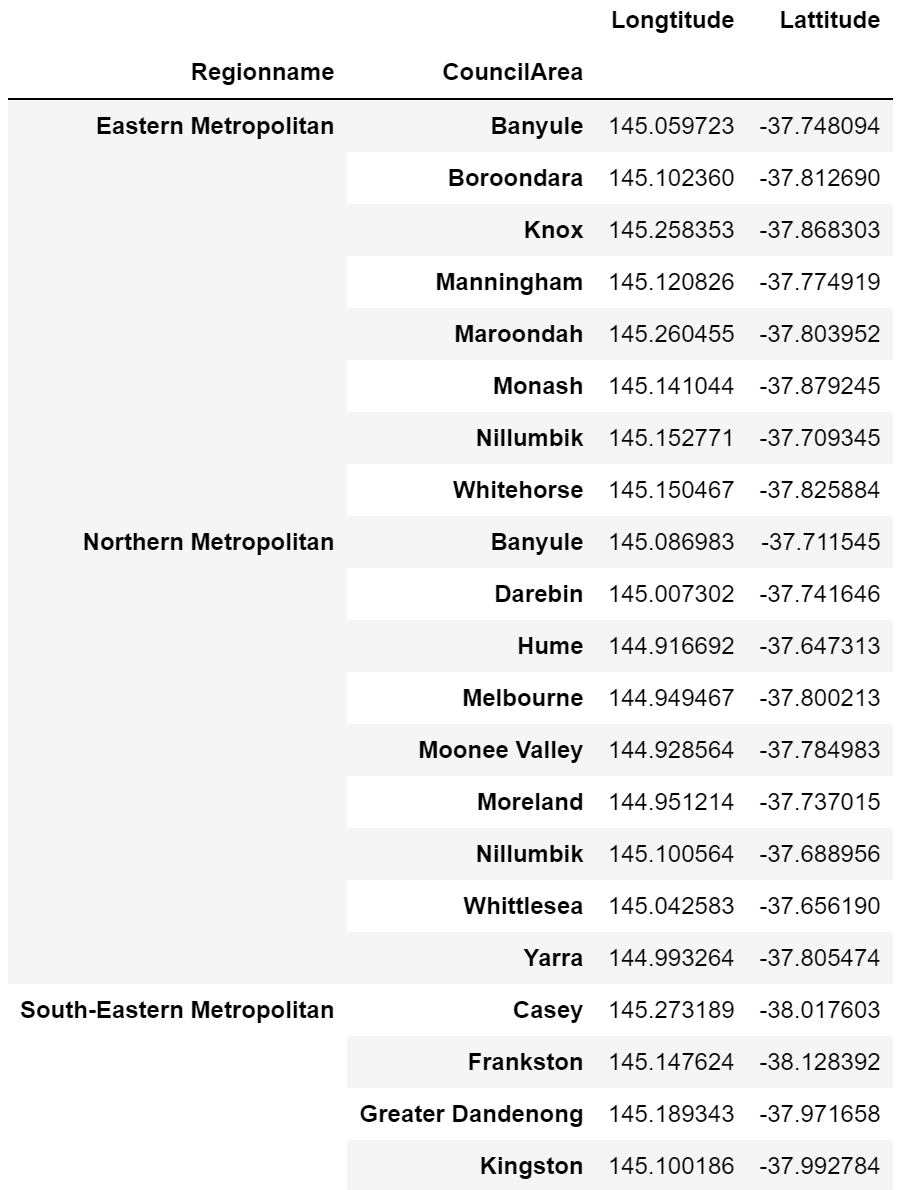


Table 2: Grouped Dataset