**Capstone Project – Battle of Neighborhoods**

Table of Contents

[Chapter 1 : Introduction 1-2](#_Toc33811589)

[1.1 Background 1-2](#_Toc33811590)

[1.2 Problem Description 1-2](#_Toc33811591)

[1.3 Project Objectives 1-3](#_Toc33811592)

[1.4 Project Scope 1-3](#_Toc33811593)

[Chapter 2 : Data 2-4](#_Toc33811594)

[2.1 Introduction 2-4](#_Toc33811595)

[2.2 Data Requirements 2-4](#_Toc33811596)

[2.3 Source of Data 2-4](#_Toc33811597)

[2.4 Validity / Reliability of Data 2-5](#_Toc33811598)

# 

# Introduction

## Background

Our client, Mrs. Grace Smith is a retired school teacher who lives in a suburb named Dockland in Melbourne. In Australian English the term ‘suburb’ is largely synonymous with what is called a ‘neighborhood’ in other countries. Now, due to some personal reasons, Mrs. Smith needs to move to another suburb in Melbourne. She finds Dockland very comfortable, so she wishes to move to a neighborhood that is very similar to Dockland.

## Problem Description

In order to proceed further, we needed to understand the problem more specifically. Hence we needed precise answers to the following questions.

* Is Mrs. Smith okay with moving out of Melbourne, i.e. to a neighborhood in any other nearby city in Australia?
* What did Mrs. Smith mean by ‘very similar’ to Dockland?

So, we conducted interviews with Mrs. Smith to understand her lifestyle and preferences. We understood that she is 61 years and lives with her husband, she leads an active life, has friends mostly living in Melbourne, goes for a walk in the park every morning and evening, buys groceries herself , uses public transport, prefers to eat home cooked food and so on….

From these interviews, we gathered that

* Mrs. Smith does not wish to move outside Melbourne city.
* By ‘very similar’ to Dockland, she meant – the presence of venues like parks, grocery stores, supermarkets, good transportation facilities like car rentals, light rail stations, etc. She was not particular about venues like cafes, restaurants, schools or colleges.

## 

## Project Objectives

Make a recommendation to our client, Mrs. Smith on the following :

* Choice of a suburb (neighborhood) in Melbourne that is very similar to Dockland.
* The similarity should be based on the presence of particular venues that the client mentioned/needs to ensure that she can live comfortably.

## Project Scope

Neighborhoods could be compared based on several different parameters including cost of living, demographics, culture, presence of schools, emergency services, law and order situation, noise, etc. But these parameters are outside the scope of this project.

This project aims at recommending a neighborhood that is very similar to Dockland as per the needs of Mrs. Smith only.

# Data

## Introduction

Here, we shall look into the data that will be required to solve our problem and the source of the data. We shall also ensure the validity and reliability of this data.

## Data Requirements

Our problem is about comparing neighborhoods in a city. So we will need to leverage data from a location data provider to know about which neighborhoods are present in Melbourne city and about the venues in each of these neighborhoods. Location data platforms mostly work based on geographic coordinate data, which we would need to provide.

So, we could list our data requirements as follows:

* Names of all neighborhoods in Melbourne city.

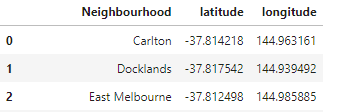
Example data include : Carlton, Parkville, East Melbourne

* List of venue categories that are preferred by Mrs. Smith

Example venue categories include: Supermarket, Park, Light Rail Station, Hospital

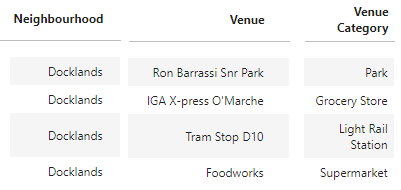
* Geographic coordinates (Latitude and Longitude) of Melbourne city and all the neighborhoods in Melbourne city.

Example:



* Venues present in each of the neighborhoods that match Mrs.Smith’s preferences

Example:



## Source of Data

* **Names of all neighborhoods in Melbourne city.**

The website - <https://www.melbourne.vic.gov.au/about-melbourne/melbourne-profile/suburbs/Pages/suburbs.aspx> is a government website which provides information about all the suburbs in Melbourne city. The required data needs to be extracted using web-scrapping.

* **Venues present in each of these neighborhoods.**

Location data provider called Foursquare provides up-to-date information about venues.

* **List of venue categories that are preferred by Mrs. Smith**

Firstly, we would collect the venue categories that Mrs. Smith is looking for in the neighborhood. These would then be matched with Foursquare’s Venue Categories. The ‘Developer’ section of Foursquare provides a list of all Venue Categories used in Foursquare at <https://developer.foursquare.com/docs/resources/categories>.

* **Geographic coordinates (Latitude and Longitude) information of Melbourne city and all the neighborhoods in Melbourne city.**

At the moment the plan is to use Geocoder to generate this data.

## Validity / Reliability of Data

The data in Foursquare is believed to be reliable based on geospatial coverage, accuracy and update frequency.

The initial plan was to extract the names of neighborhoods in Melbourne city from the Corra website-<http://www.corra.com.au/australian-postcode-location-data/>. This was an attractive option because it provided neighborhood data along with the latitude and longitude information in a csv file format. This meant that extracting data would be a simple straight-forward process. But since the data was last updated on March 19th, 2013 and the site itself claimed, “ PLEASE NOTE: THIS DATA IS NOW OUT OF DATE BY QUITE A BIT”, the data didn’t seem fit for our project. So the Australian government website was chosen over this one, even though extracting data would need web-scrapping.

We will be using Geocoder to generate data on geographic coordinates. A problem with Geocoder is that sometimes it fails to generate data for some neighborhoods. Incase this happens, we would have to look for other options. But that shouldn’t be a problem.