**Exploring Suburbs of Melbourne City (Australia)**

**Nabeel Khan**

**September 03, 2019**

|  |
| --- |
| 1. **Introduction** |

**1.1 Problem Statement**

Find a reasonable suburb of Melbourne for living.

**1.2 Discussion**

Melbourne has been named the world’s most liveable city for seven years in a row [1]. In 2018, Melbourne performed best in healthcare, education and infrastructure. Melbourne not only maintained its score in stability but also gained points in culture and environment.

All these factors make Melbourne a natural choice for people to live, work, or study. But there are two issues:

* Firstly, all neighborhoods are not safe to live due to high crime rates
* Secondly, a neighborhood may be safe, but it can be very far away from the city center

The locals of Melbourne will know about it. But people living in other Australian states, foreigners, or immigrants will have no idea about a better suburb to relocate. Property agents do not provide a complete information. Therefore, it becomes essential to find such information (recommended suburbs) before relocating to Melbourne.

The safe suburbs should have a very low crime rate. At the same time, it will be good for a suburb to have some of the following venues in nearby premises

1. School (important for young families)
2. Park (good for kids)
3. Restaurants (good outing)
4. Train Stations (necessary to commute if not driving)
   1. **Target Audience**

This project will suit people with young families, who are

* Relocating to Melbourne especially immigrants
  1. **Project Goal**

The project will help to determine a suitable neighborhood of Melbourne for living.

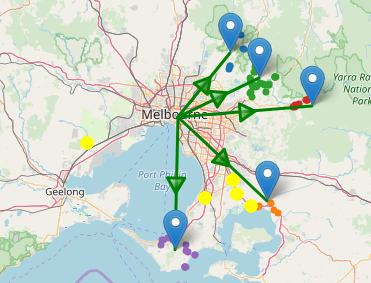
In this project, I will report:

* 5 suburbs of Melbourne for settling down (based on crime rate and frequency of nearby venues)
  + All suburbs will be within 100 KM of distance from the city center
  + I will also report the best suburb
* Worse 5 suburbs of Melbourne (based on crime rate)

The main outputs of the project will comprise of two graphs:

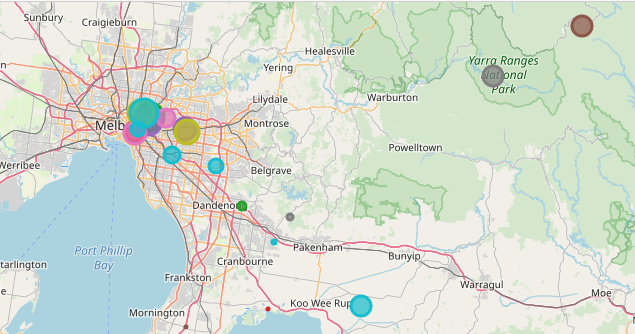
**Graph-1**

1. The yellow markers indicate the worse suburbs
2. Top 5 suburbs are represented by “location” symbol



**Graph-2**

1. I will cluster suburbs based on distance from the city center.
2. Then, I will compute the percent crime for each cluster. The size of the cluster is proportional to the percent crime. This graph will provide an overall picture of crime stats in Melbourne

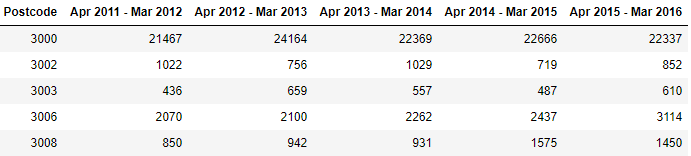


|  |
| --- |
| **2. Data Acquisition** |

The crime statistics of Melbourne suburbs are required to identify safe neighborhoods. The crime stats can be download (excel sheet) from the following link

<https://www.crimestatistics.vic.gov.au/sites/default/files/embridge_cache/emshare/original/public/2016/06/dc/1f02f731b/offencesdatatable-yearending31march2016.xlsx>

The excel sheet has lots of information. I extracted the **“Offenses”** recorded information, which reports the total number of crimes against Post Codes for five years, ranging from 2011 to 2016. The crimes data looks like



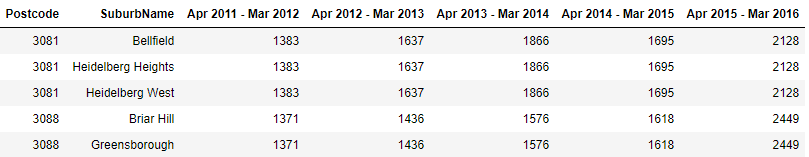
The problem is that above table has only Post codes. There is no information about the **names of suburbs**

**2.1 Scrapping Melbourne Suburb Table from Wikipedia**

I used the following Wikipedia link to find suburb names against the post codes

<https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs>

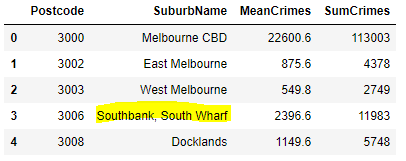
We can place suburb names in crimes data. The table will look like



**2.2 Data Cleansing**

I compute the mean and sum of the crimes for each suburb and just keep that in the table. Quite a few suburbs have similar post codes. That results in duplication of crimes, which will affect our analysis.

* So, I group the suburbs with similar post codes and record a single entry for crimes
* I compute the mean and sum of crimes for 5 years to keep data simple. The table looks like



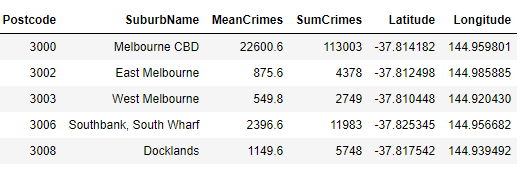
|  |
| --- |
| **3. Data Preparation** |

**3.1 Geolocator to extract location**

I use Geo locator Api to extract the location (latitude and longitude) information for each suburb.

* I remove any rows, where we cannot find the location information for a suburb
* We have grouped several suburb names (based on similar post code). For such cases, I compute the mean location.

The table looks like

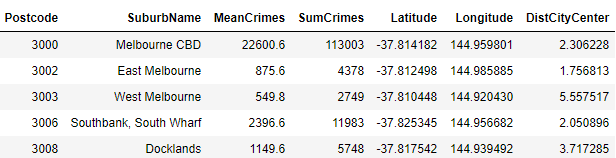


**3.2 Distance from City Center**

I extract the location information for Melbourne city center. Then, I compute the distance of each suburb from the city center using Haversine formula [2].

We only consider suburbs, which are within 150 KM distance from the Melbourne City Center to avoid very including remote suburbs in our analysis.

The table looks like



**3.3. Feature Selection using Foursquare API**

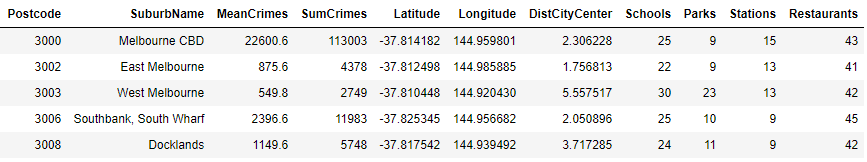
I have decided to use extract the following features for each suburb using “Foursquare API” within a radius of 3 km

* The frequency of nearby schools
* The frequency of nearby parks
* The frequency of nearby train stations
* The frequency of nearby restaurants

The above factors will help us to determine which suburb is better than the others despite having very low or very similar crime rates

|  |
| --- |
| **3. Analysis** |

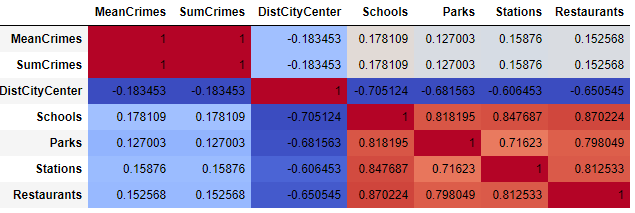
In this section, I perform analysis to understand the relationship between various features. The data looks like



One suburb seems to have a very high crime rate i.e. over 20000. It looks like an anomaly. So, I removed that suburb from our table for robust analysis

**3.1 Correlation**

I analysed the correlation between various features. The idea is to see any dependency or pattern between features. The correlation matrix is



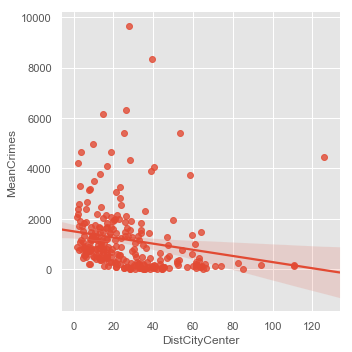
From the above correlation, we cannot find a feature which is highly correlated to crime rates. So, this statistics is not enough to make any hypothesis

**3.2 Regression Plots**

I decided to use regression plots to analyse the correlations between crimes and various features.

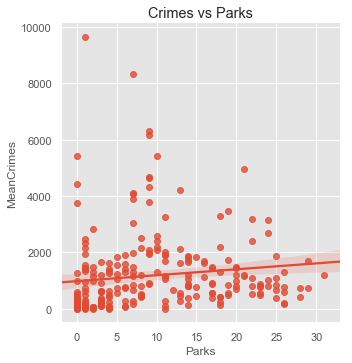
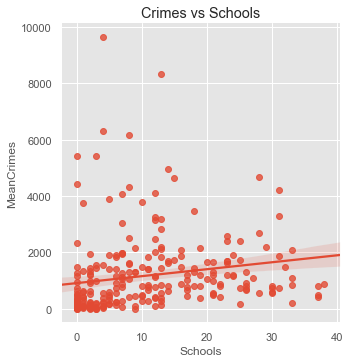
**3.2.1 Crimes vs Distance from City Center**

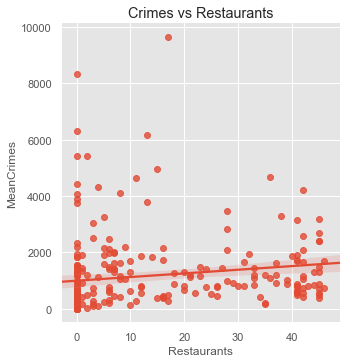
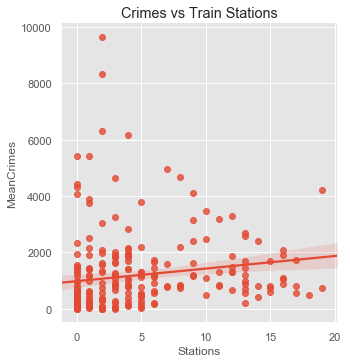
I plotted crimes (mean) against the distance from the city center. The plots indicate a very interesting thing i.e., crime rates tend to decrease as we move away from the city center. It also makes sense because lot of people try to live near the city. That may results in more crimes compared to farthest suburbs



**3.2.2 Crimes vs Other Features**

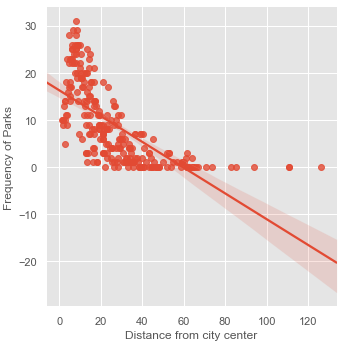
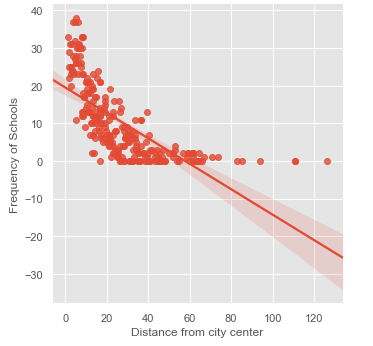
I plotted the remaining features against the crime rates. We can observe a weak correlation that an increase in feature leads to an increase in crime rates. But, I think the correlation is not very strong.

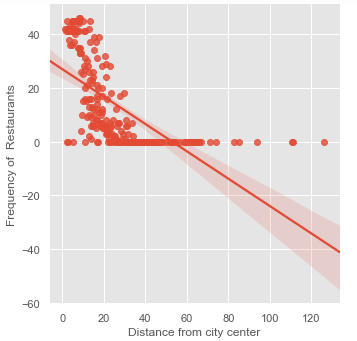
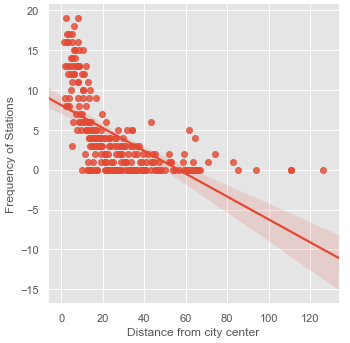




**3.2.3 Distance from City Center vs Other Features**

The regression plots reveal a very interesting pattern. We get more schools, parks, stations and restaurants for suburbs, which are near to the city center. As, we move away from the city center, the frequency of these venues start decreasing. The possible reason is that we are only looking within 3 km radius. For remote suburbs, we may find more venues but we need to go beyond the 3 km radius limit.



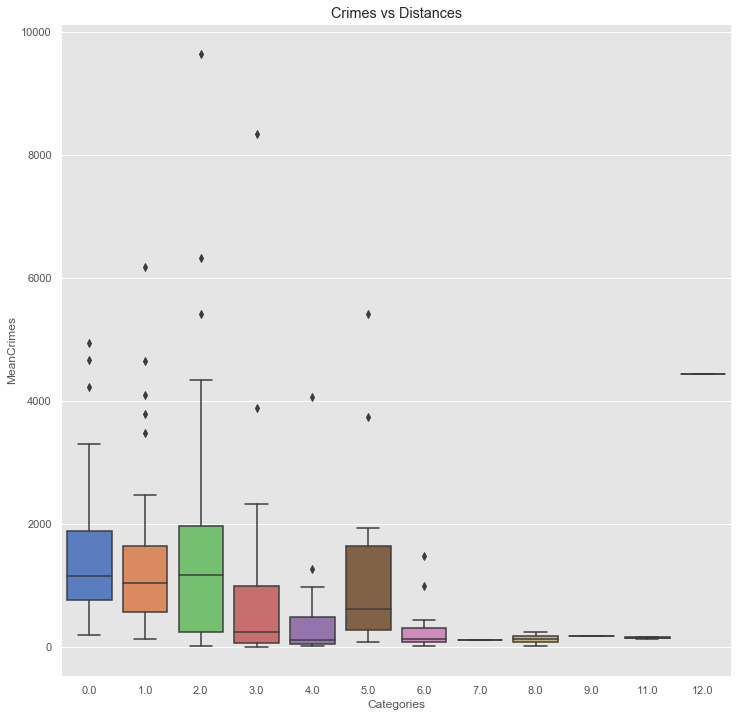


**3.3 Box Plots**

I decided to verify the above hypothesis using box plots. For better visualization, I used the concept of binning or categories. Such as, the first category cover 10 km distance from the city center, the second category cover 10-20 km distance from the city center and so on.

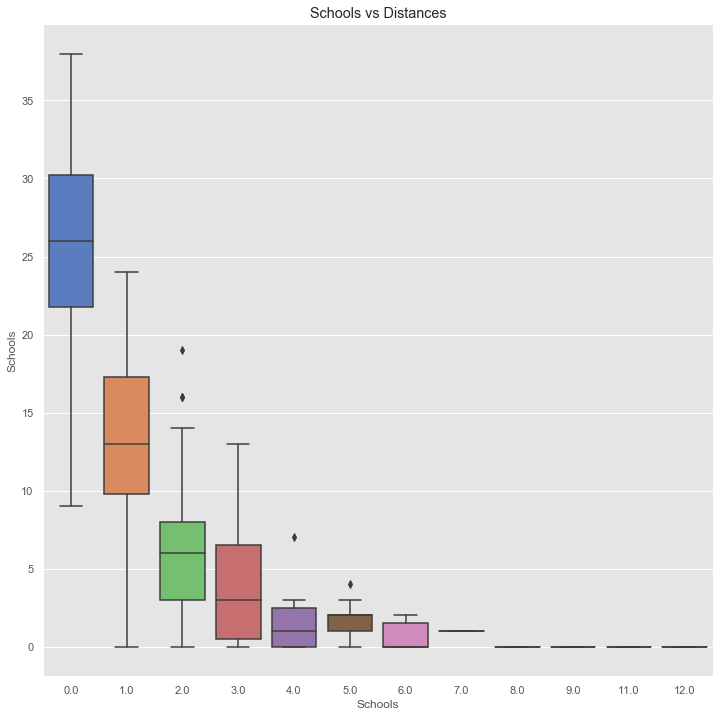
**3.2.1 Crimes vs Distance from City Center**

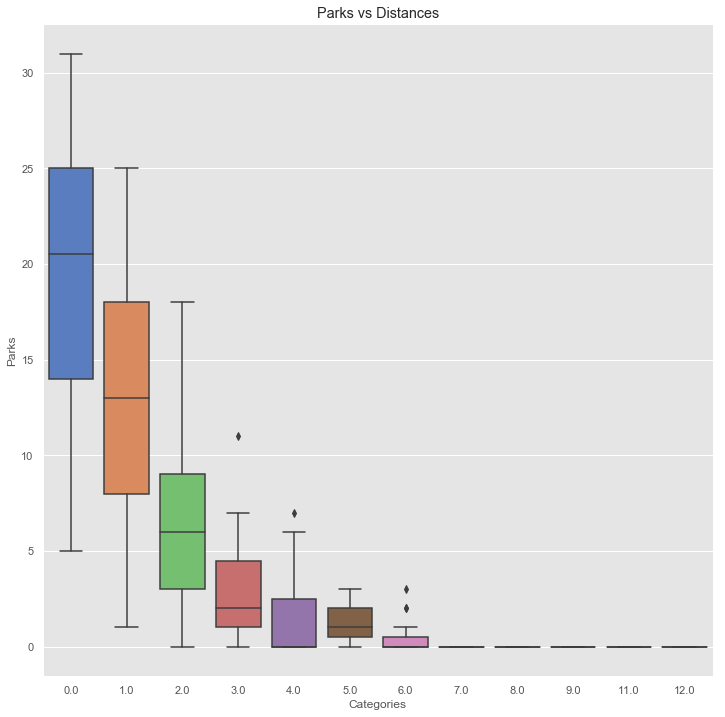
The plot support our hypothesis that crime rates start to reduce as we move away from the city center



**3.2.1 Features vs Distance from City Center**

To verify my hypothesis, that frequency of schools, or parks decrease with an increase in the distance from the city center. I used a similar box plot

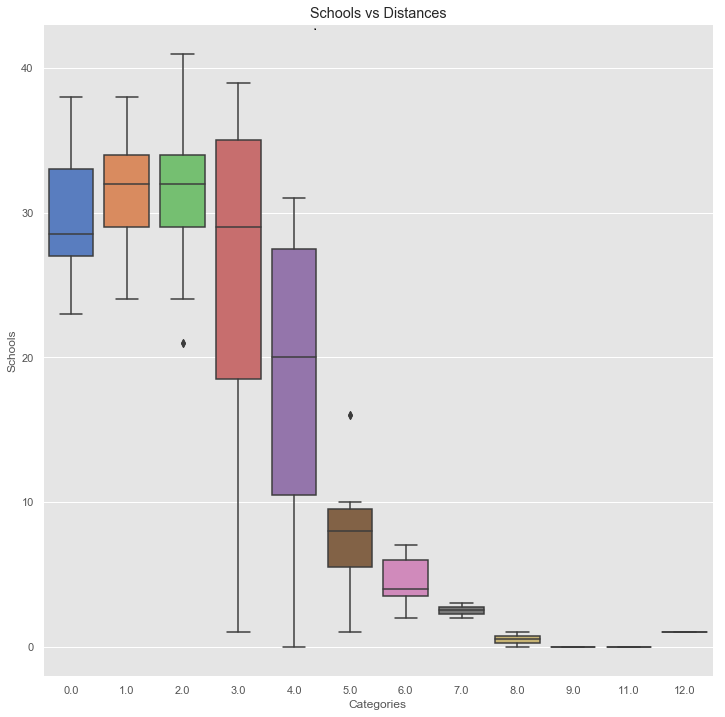


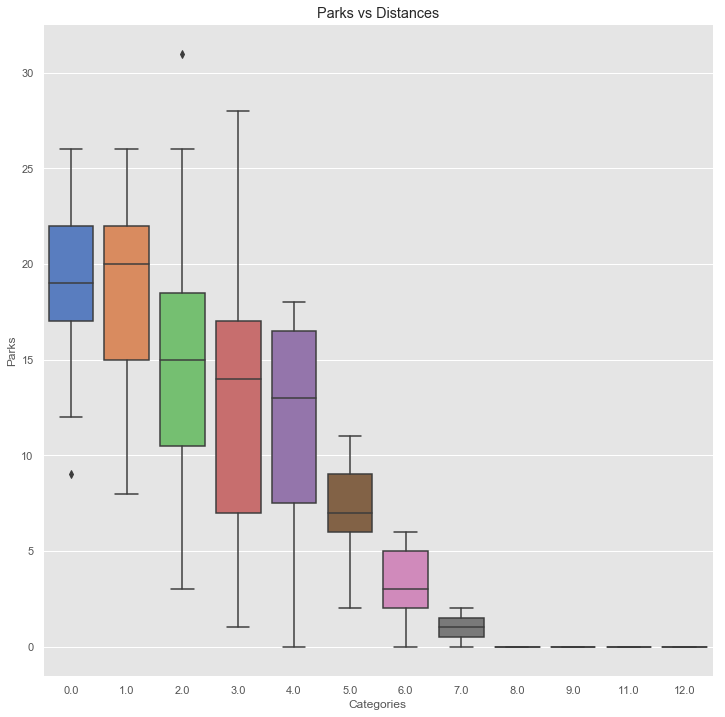


I found the similar patterns with the remaining two features.

**3.2.1 Increasing 3 Km Radius range**

As soon as we move away from the city center, the number of nearby venues start decreasing. It makes sense and this pattern applicable to all big cities of the world. To verify my hypothesis, I increased the range from 3 Km radius to 10 km and used foursquare api again to determine the frequency of schools and parks





The above box plots indicate that as we move away from the city, we can still find the venues but on a large travel time

|  |
| --- |
| **4. Suburbs Classification** |

I have performed the exploratory data analysis. In this section, I will discuss my approach to report the best and worse suburbs on a map for the users

|  |
| --- |
| 1. **Conclusion** |

|  |
| --- |
| 1. **Future Work** |

Refereces

1. <https://www.businessinsider.com.au/melbourne-liveable-city-vienna-2018-8?r=US&IR=T>
2. <https://en.wikipedia.org/wiki/Haversine_formula>