**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 (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:

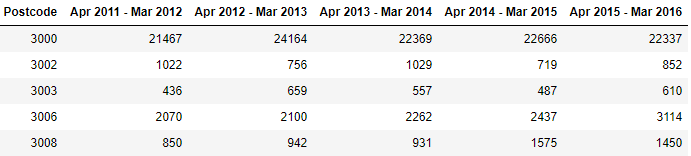
* 5 suburbs of Melbourne for settling down (based on crime rate and frequency of nearby venues)
  + All suburbs will be within 150 KM of distance from the city center
  + I will also report the best suburb
* Worse 5 suburbs of Melbourne (based on crime rate)

|  |
| --- |
| **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”** information, which reports the total number of crimes for five years, ranging from 2011 to 2016. The crimes are reported against the post codes. The crimes data looks like



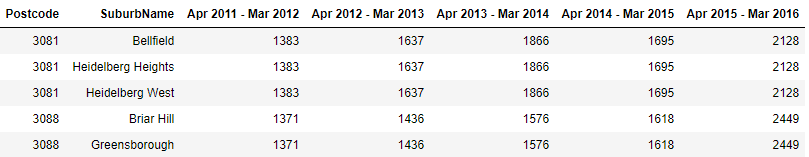
The above table lacks the names of suburbs.

**2.1 Scrapping Melbourne Suburb Table from Wikipedia**

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

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

After extracting the suburb names, the table looks like

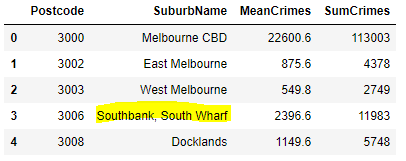


* 1. **Data Cleansing**

I compute the mean and sum of the crimes for each suburb. And, I keep only this information in the data for simplicity.

While merging the suburbs, I realised that some suburbs have similar post codes. Due to this issue, we get duplicates of crime statistics in the table. We need to record crime statistic for a post code once only.

* So, I grouped the suburbs with similar post codes to keep one crime statistics



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

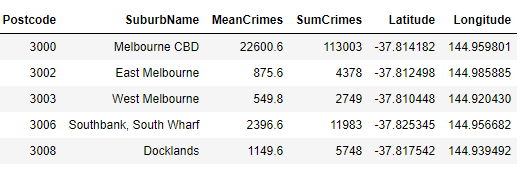
I need the location information for each suburb, so that we can extract the information about nearby venues, such as schools or parks for each suburb.

**3.1 Geolocator to extract location**

I have used 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
* I have grouped several suburb names with similar post codes. For such cases, I compute the mean location.

The table looks like

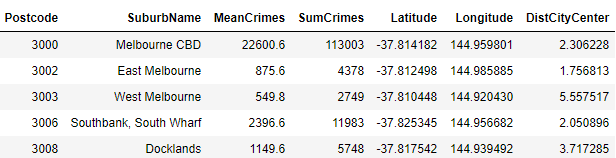


**3.2 Distance from City Center**

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

To avoid picking the remote suburbs, I consider suburbs, which are within 150 KM distance from the Melbourne City Center

The table looks like



**3.3. Feature Selection using Foursquare API**

I have extracted the information about the following venues (features) for each suburb using “Foursquare API”

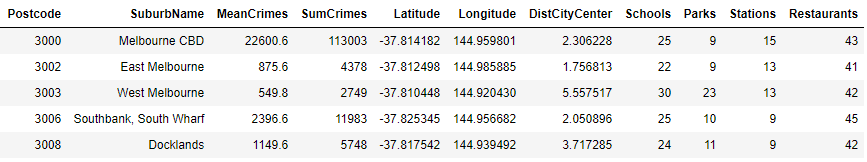
* The frequency of nearby schools
* The frequency of nearby parks
* The frequency of nearby train stations
* The frequency of nearby restaurants

The search radius is kept 3 km for each suburb.

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

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

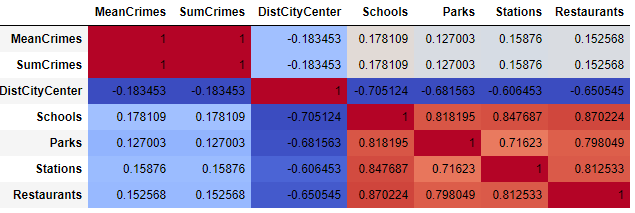
I have performed exploratory analysis to understand the relationship between various features in my data in this section. The data looks like



One suburb has a very high crime rate i.e. over 20000. It seems not correct as it very high compared to other crime rates. So, I consider it as an outlier and remove it from my analysis.

**3.1 Correlation**

I analysed the correlation between various features.



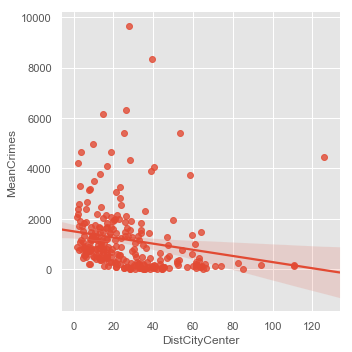
From the above correlation, we cannot find a feature which is highly correlated to crime rates.

**3.2 Regression Plots**

**3.2.1 Crimes vs Distance from City Center**

I plotted mean crimes against the distance from the city center. The plots indicate an exciting fact i.e. crime rates tend to decrease as we move away from the city center.

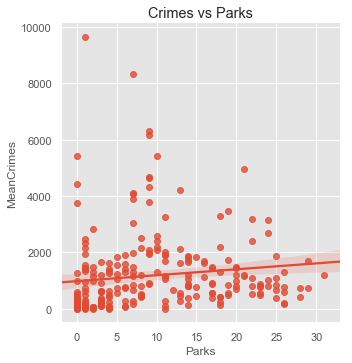
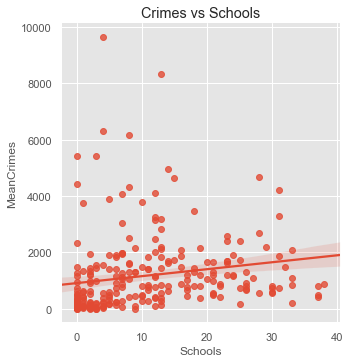
It makes sense because a lot of people try to live near the city. That may result in more crimes compared to farthest suburbs.

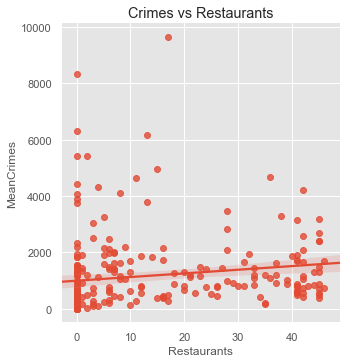
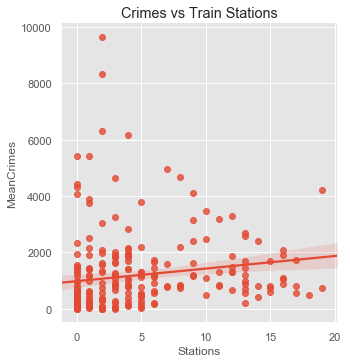


**3.2.2 Crimes vs Other Features**

I plotted the remaining features against the crime rates.

There is a weak correlation between crime rates and the remaining features



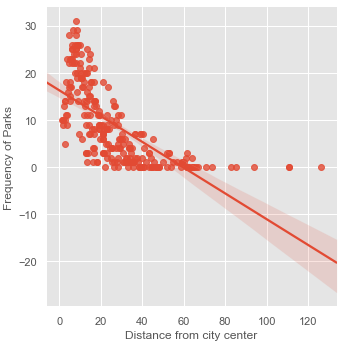
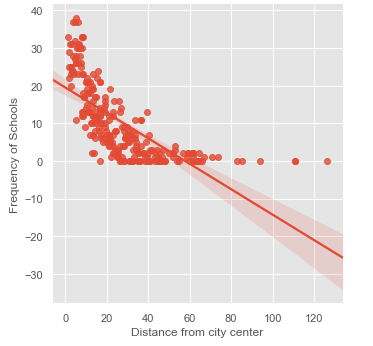


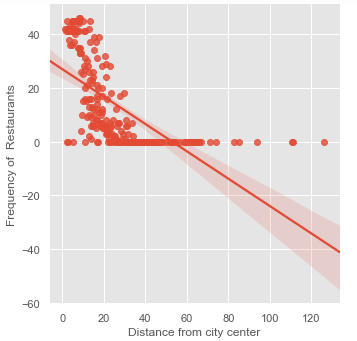
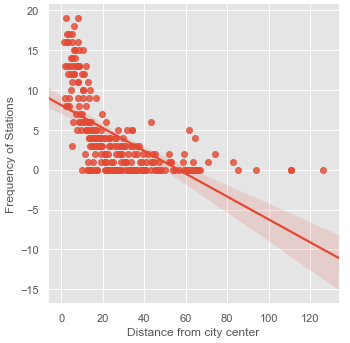
**3.2.3 Distance from City Center vs Other Features**

I plotted the frequencies of remaining features against the distance from the city center.

The regression plots reveal an interesting pattern. We get more schools, parks, stations and restaurants for suburbs, which are near to the city center (within 3 km). As we move away from the city center, the frequency of these venues starts decreasing.

So, for out of the city suburbs, the venues are bit far away. In other words, someone will have to drive more to reach to those venues.





**3.3 Box Plots**

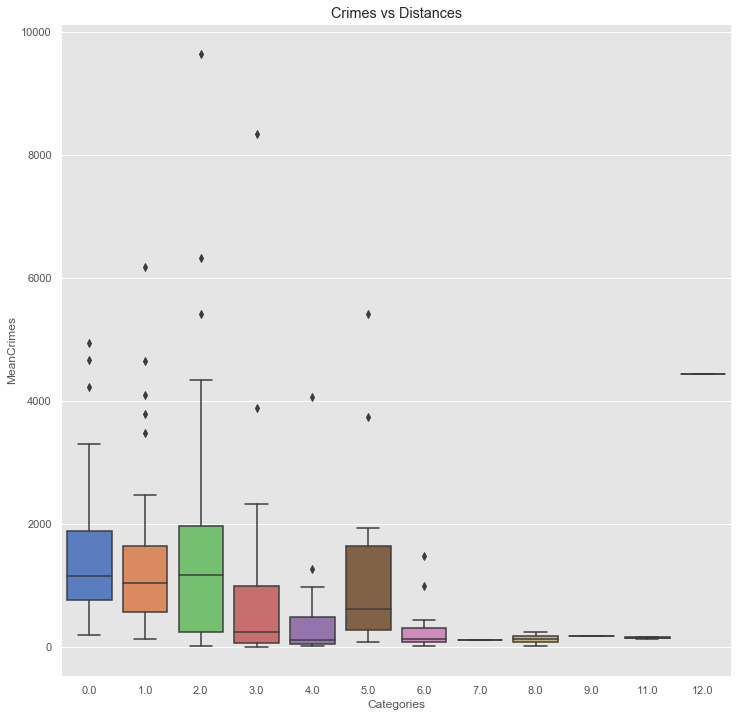
For better visualization, I decided to apply the concept of **binning** using the box plot.

In binning, I grouped the distances from city centers. Such as, the first bin covers 0 -10 kms, the second bin covers 10-20 kms and so on. I refer to bins as categories in this report.

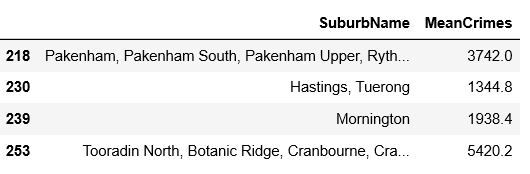
Then, I plotted the features against these categories for analysis.

**3.3.1 Crimes vs Distance from City Center**

The plot supports our hypothesis that crime rates reduce as we move away from the city center.



Interestingly, we find a very high crime rate for Category 5. The reason is that some of the suburbs have very high mean crime rates compared to others as shown below (suburbs with mean crimes less than 1000 are not shown).

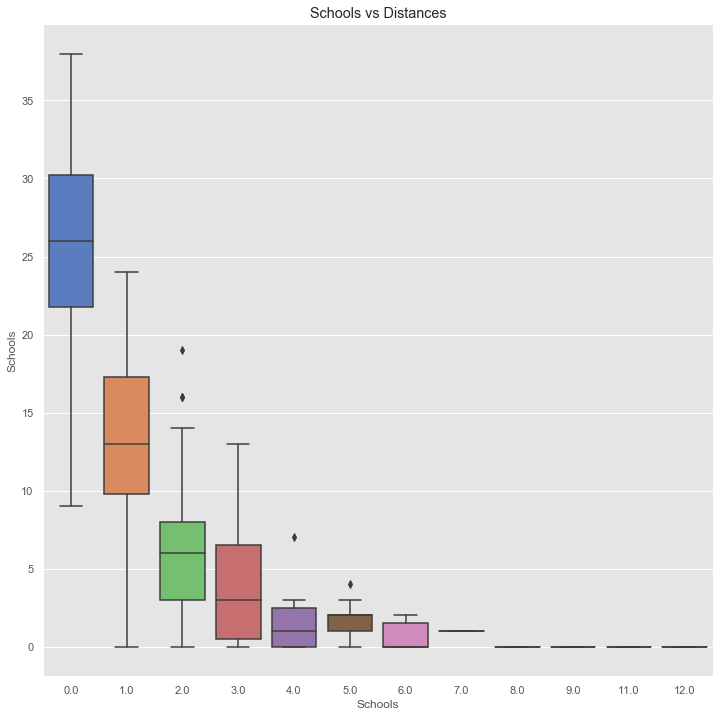


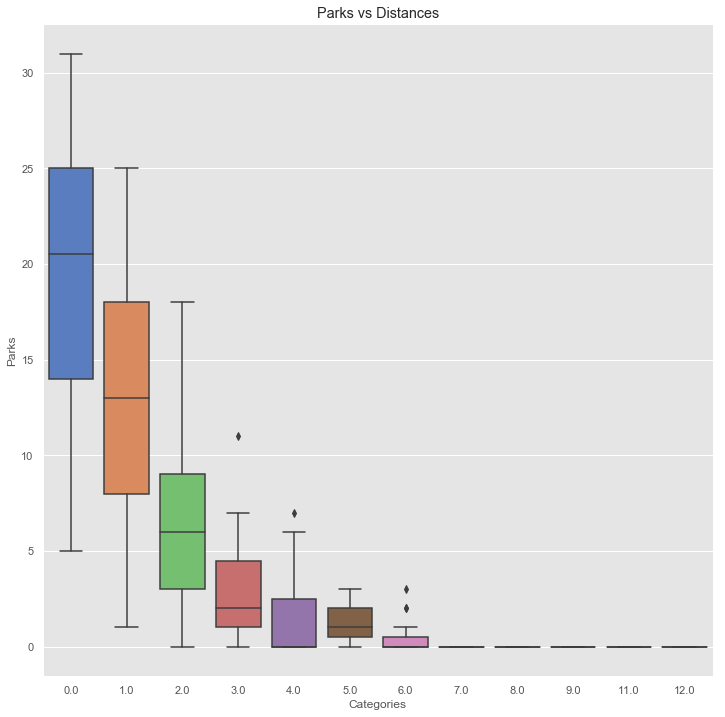
**3.3.2 Distance from City Center vs Other Features**

I used the concept of binning to plot frequencies of all features against the distance from the city center i.e. categories.

The results support the previous hypothesis i.e., the frequencies of features decrease with an increase in the distance from the city center.

I have reported box plots for schools and parks in this report. But, I found the similar pattern with the stations and restaurants.

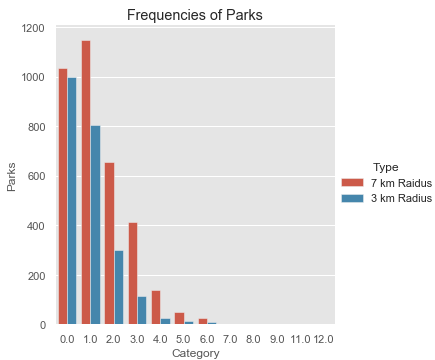
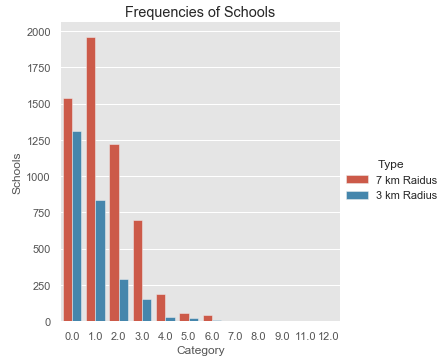




**3.3.3 Effect on increasing 3 Km Radius**

I hypothesised that for remote suburbs, the venues are far away and we need to drive more.

To verify this hypothesis, I decided to increase the search radius to 7 km from 3 km to find the nearby venues using the ‘Forusquare Api’. With an increase in the radius of the search, the frequencies of all features (schools, parks or restaurants) also increase compared to 3 km radius. I have just included two plots in the report for illustration.



|  |
| --- |
| **4. Results** |

In the previous section, I have hypothesized that crimes rate decrease as we move away from the city center. The frequencies of nearby venues, such as school, or restaurants also decrease as we move away from the city center.

* Though, one can travel a bit more to reach to those venues
* The experiment of changing the radius from 3 km to 7 km demonstrated that

In this section, I will report the

* Top 5 and Worse 5 suburbs based on crime rates
  + Top 5 venues (within 10 km) for the top 5 suburbs for more insights
  1. **Top and Worse Suburbs**

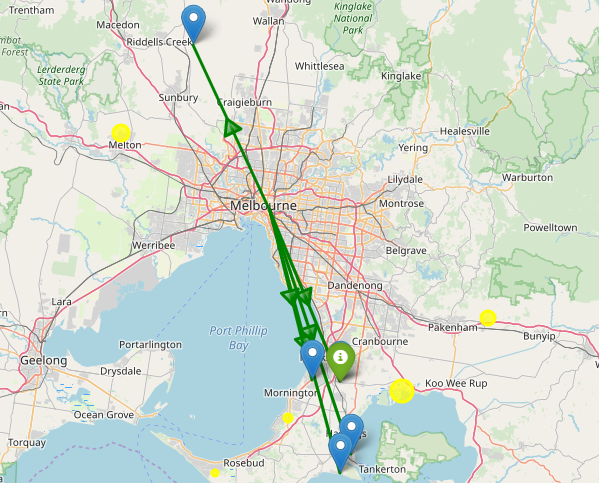
Overall, the crime rate is low from 40 Km onwards from the city center with the exception of some of the suburbs in 50 Km range.

We do not want to find a suburb very far away from the city center.

So, I limit my range from 40-70 Km to determine the top 5 suburbs.

I have used the folium map to highlight the suburbs as shown below

* The yellow markers indicate the 5 suburbs with the highest crime rates. The size of the marker is proportional to mean crimes for that suburb
* The top 5 suburbs are displayed with location marker symbols
  + The best suburb is indicated by green icon (location type marker)



The details of the top 5 suburbs including top 5 venues can be found in the following table. The best suburb for living is Koo Wee Rup North, which is

* within 50 Km from the city center and has the least crime rate

The top 3 suburbs have grocery stores as the one of the common venues. So, we can hypothesize that these suburbs have nearby grocery stores, which results in frequent visits for the people.



|  |
| --- |
| 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>