

Exploring Suburbs of Melbourne City (Australia)

Nabeel Khan

September 06, 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 the above information. But foreigners and immigrants will have no idea about these issues.

The property agents also not provide complete information. Therefore, it becomes essential to find such information (recommended suburbs) before relocating to Melbourne.

The safe suburbs should have a low or zero 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
2. Park
3. Restaurants
4. Train Stations

1.3 Target Audience

This project will suit people with young families, who are

- Relocating to Melbourne especially immigrants

1.4 Project Goal

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

In this project, I will report:

- Top 5 suburbs of Melbourne (based on crime rate) using a map
- Worse 5 suburbs of Melbourne (based on crime rate) using a map
- Geospatial distribution of crimes using a map

2. Data Acquisition

The crime statistics of Melbourne suburbs are required for this project. 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 postcodes as shown below.

Postcode	Apr 2011 - Mar 2012	Apr 2012 - Mar 2013	Apr 2013 - Mar 2014	Apr 2014 - Mar 2015	Apr 2015 - Mar 2016
3000	21467	24164	22369	22666	22337
3002	1022	756	1029	719	852
3003	436	659	557	487	610
3006	2070	2100	2262	2437	3114
3008	850	942	931	1575	1450

2.1 Scrapping Melbourne Suburb Table from Wikipedia

The following Wikipedia link has suburb names listed against the postcodes

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

I extracted the suburb names and merged this with the crime data. Now, the table looks like

Postcode	SuburbName	Apr 2011 - Mar 2012	Apr 2012 - Mar 2013	Apr 2013 - Mar 2014	Apr 2014 - Mar 2015	Apr 2015 - Mar 2016
3081	Bellfield	1383	1637	1866	1695	2128
3081	Heidelberg Heights	1383	1637	1866	1695	2128
3081	Heidelberg West	1383	1637	1866	1695	2128
3088	Briar Hill	1371	1436	1576	1618	2449
3088	Greensborough	1371	1436	1576	1618	2449

2.2 Data Cleaning

Some suburbs have similar postcodes. So, we get duplicate rows of crime statistics after merging it with the crime data. Though, we need to remove duplicates and store this information once only.

- Therefore, I grouped the suburbs with similar postcodes to record duplicates. This makes sure to have one crime statistics per post code.
- I also computed the mean and sum of the crimes for all years and kept it in the table.

- One suburb has a very high mean crime rate i.e. over 20000. It seems not correct compared to other crime rates. So, I treated it as an outlier and removed it.
- I removed the unnecessary columns

After cleaning, the new data looks like this

	Postcode	SuburbName	MeanCrimes	SumCrimes
0	3000	Melbourne CBD	22600.6	113003
1	3002	East Melbourne	875.6	4378
2	3003	West Melbourne	549.8	2749
3	3006	Southbank, South Wharf	2396.6	11983
4	3008	Docklands	1149.6	5748

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.

3.1 Geolocator to retrieve location

I have used 'Geo locator api' to obtain the location (latitude and longitude) information for each suburb.

- I removed any rows, where we cannot find the location information for a suburb
- For multiple suburbs against a postcode, I computed the mean location.

3.2 Distance from City Center

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

In this project, I have considered suburbs, which are within 120 KM distance from the Melbourne city center. Otherwise, we may end up finding very remote suburbs.

The table now looks like this.

Postcode	SuburbName	MeanCrimes	SumCrimes	Latitude	Longitude	DistCityCenter
3000	Melbourne CBD	22600.6	113003	-37.814182	144.959801	2.306228
3002	East Melbourne	875.6	4378	-37.812498	144.985885	1.756813
3003	West Melbourne	549.8	2749	-37.810448	144.920430	5.557517
3006	Southbank, South Wharf	2396.6	11983	-37.825345	144.956682	2.050896
3008	Docklands	1149.6	5748	-37.817542	144.939492	3.717285

3.3. Feature Selection using Foursquare API

I have extracted the information about the following venues for each suburb using the “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 to provide more insights about suburbs.

3. Data Analysis

In this section, I have performed an exploratory analysis of data to understand the relationship between various features.

- Now, the data has frequencies of schools, parks, stations and restaurants as shown below.

Postcode	SuburbName	MeanCrimes	SumCrimes	Latitude	Longitude	DistCityCenter	Schools	Parks	Stations	Restaurants
3000	Melbourne CBD	22600.6	113003	-37.814182	144.959801	2.306228	25	9	15	43
3002	East Melbourne	875.6	4378	-37.812498	144.985885	1.756813	22	9	13	41
3003	West Melbourne	549.8	2749	-37.810448	144.920430	5.557517	30	23	13	42
3006	Southbank, South Wharf	2396.6	11983	-37.825345	144.956682	2.050896	25	10	9	45
3008	Docklands	1149.6	5748	-37.817542	144.939492	3.717285	24	11	9	42

3.1 Correlation

We are mainly interested in correlating some features against the crime rates for insights.

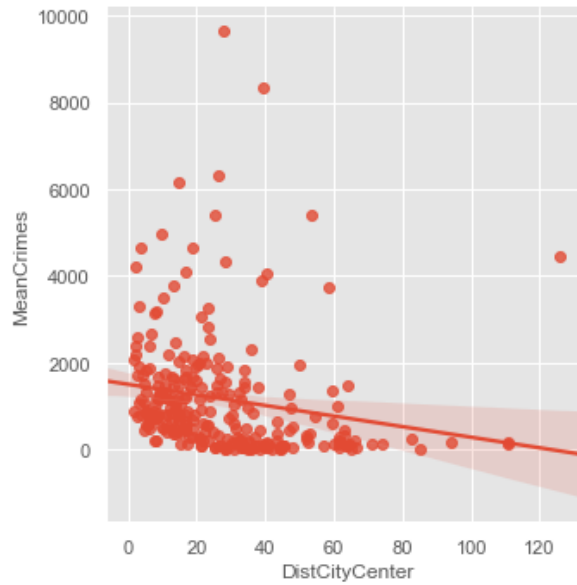
I analysed the correlation between various features. From the correlation, we cannot identify a feature that is highly correlated to crimes. So, I decided to experiment with regression plots for further analysis.

	MeanCrimes	SumCrimes	DistCityCenter	Schools	Parks	Stations	Restaurants
MeanCrimes	1	1	-0.183453	0.178109	0.127003	0.15876	0.152568
SumCrimes	1	1	-0.183453	0.178109	0.127003	0.15876	0.152568
DistCityCenter	-0.183453	-0.183453	1	-0.705124	-0.681563	-0.606453	-0.650545
Schools	0.178109	0.178109	-0.705124	1	0.818195	0.847687	0.870224
Parks	0.127003	0.127003	-0.681563	0.818195	1	0.71623	0.798049
Stations	0.15876	0.15876	-0.606453	0.847687	0.71623	1	0.812533
Restaurants	0.152568	0.152568	-0.650545	0.870224	0.798049	0.812533	1

3.2 Regression Plots

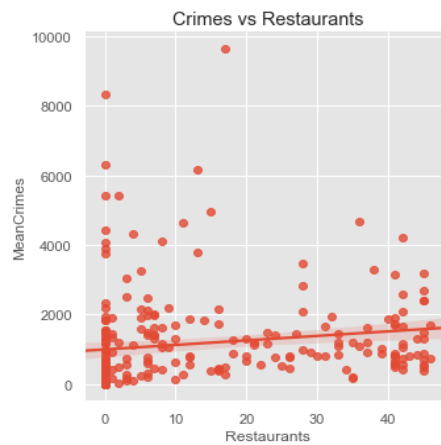
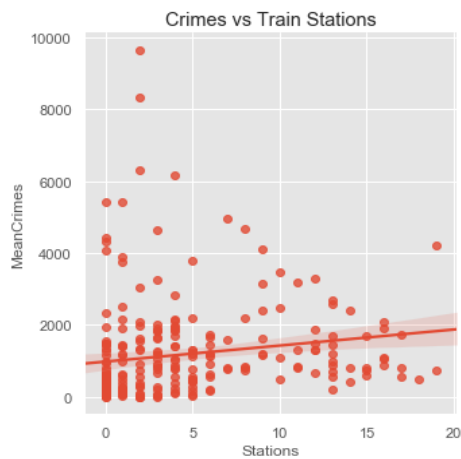
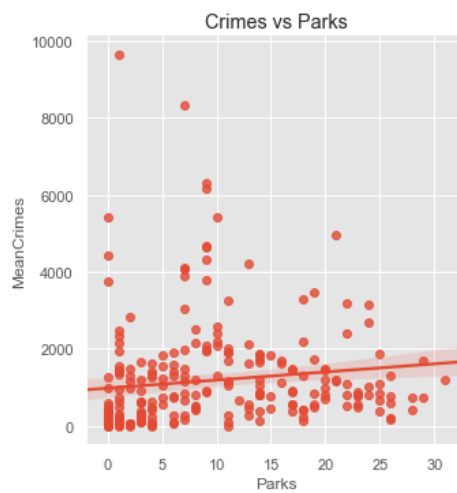
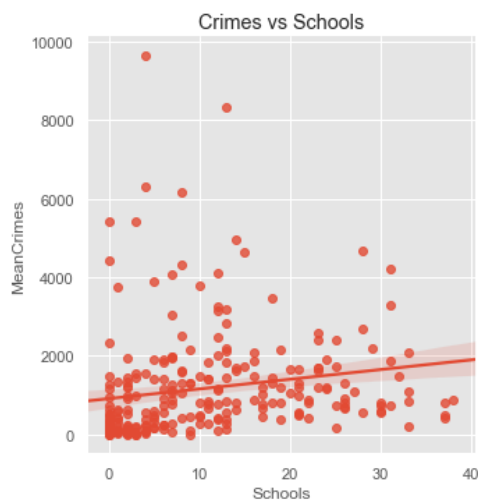
3.2.1 Crimes vs Distance from City Center

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



3.2.2 Crimes vs Other Features

I plotted the remaining features against the crime rates. But it is hard to make some hypothesis based on these plots.

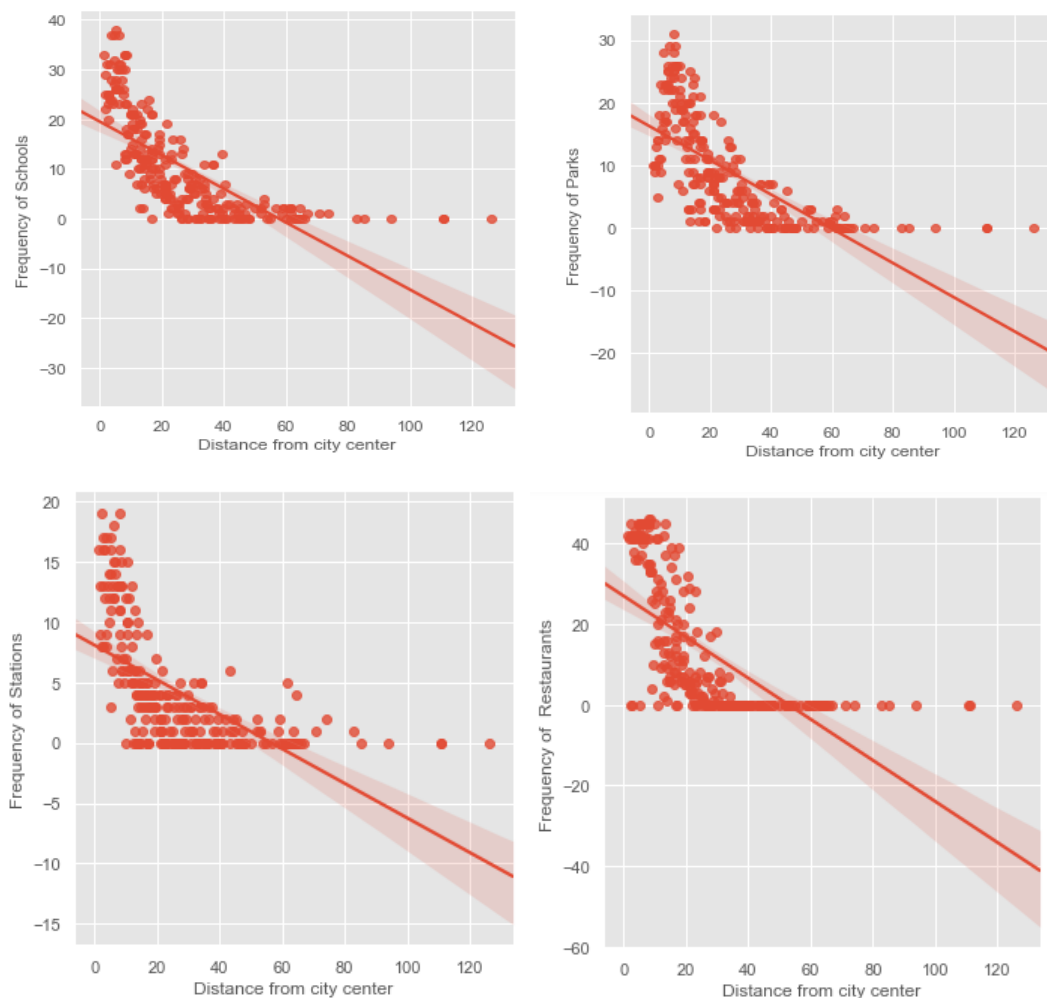


3.2.3 Distance from City Center vs Other Features

I used distance from the city center as a reference and plotted remaining features against it.

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

Please note that we are using a 3 km radius search for our suburbs. So, maybe, people need to drive more for faraway suburbs to get to these venues. This hypothesis is verified later in section 3.3.3.



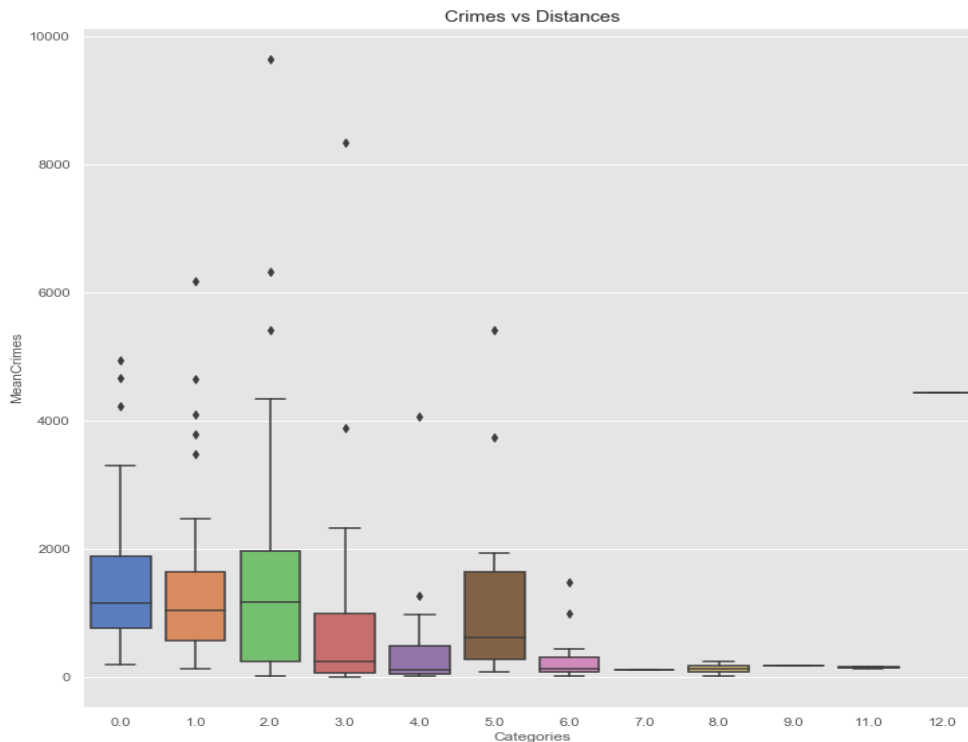
3.3 Box Plots

I decided to use a box plot because it handles large data easily and can provide better visualisation.

I applied the concept of binning or categorization with my data for better visualisation. In **binning**, I grouped the distances from city centers into categories. Such as, the 1st bin covers 0 -10 km, the 2nd bin covers 10-20 km and so on. And, then I plotted the features against these categories for analysis.

3.3.1 Crimes vs Categories [Distance from City Center]

The plot verifies our hypothesis that crime rates reduce with an increase in distance from the city center.



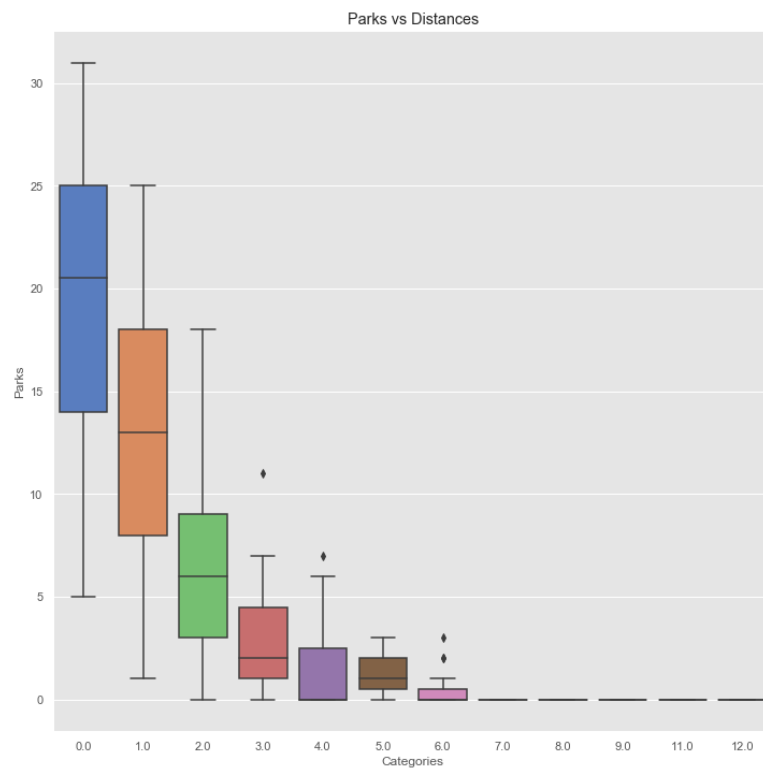
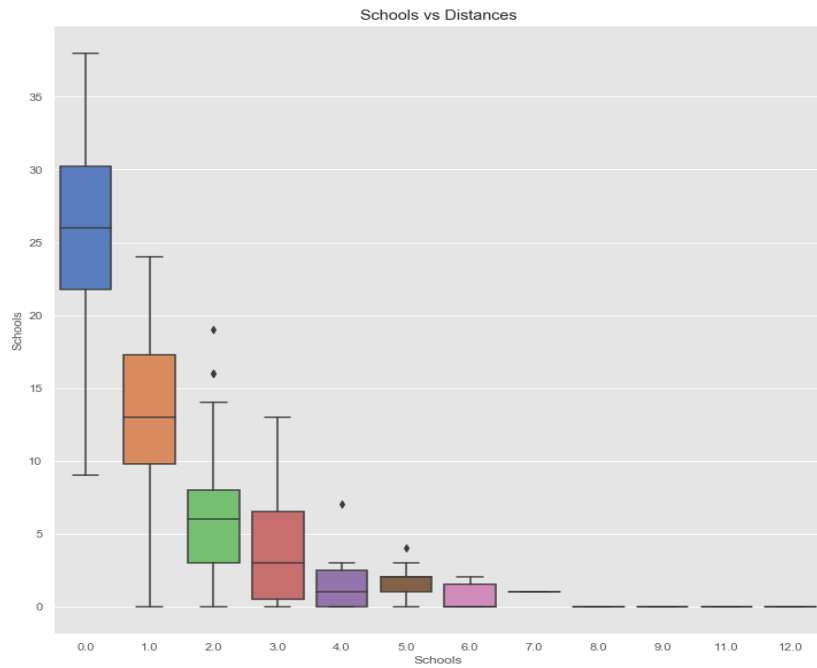
Interestingly, we find a very high crime rate for Category 5. The reason is that some of the suburbs in Category 5 have very high mean crime rates as shown below (suburbs with mean crimes less than 1000 are not shown).

	SuburbName	MeanCrimes
218	Pakenham, Pakenham South, Pakenham Upper, Ryth...	3742.0
230	Hastings, Tuerong	1344.8
239	Mornington	1938.4
253	Tooradin North, Botanic Ridge, Cranbourne, Cra...	5420.2

3.3.2 Features vs Categories [Distance from City Center]

I plotted frequencies of all features against the distance from the city center using the concept of binning.

The results verify the previous hypothesis i.e., the frequencies of features decrease with an increase in the distance from the city center. I have only reported box plots for schools and parks in this report.

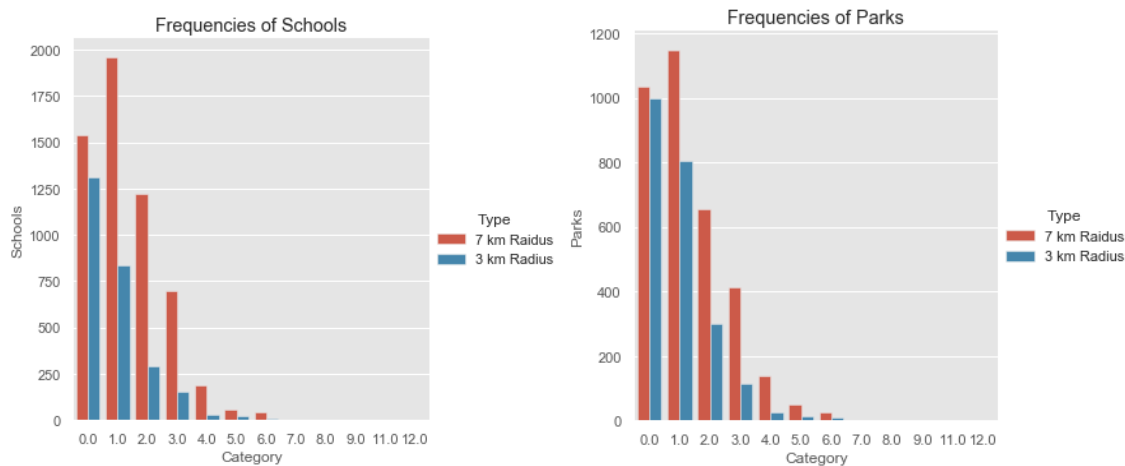


3.3.3 Effect of increasing the Search Radius

I hypothesised that the frequency of nearby venues decreases as we move away from the city center.

To verify this hypothesis, I increased the search radius from 3 km to 7 km. Then I used the 'Foursquare Api' to find the nearby venues (schools, or parks) for all suburbs. The bar plots indicate that the frequencies of all features (schools, parks or restaurants) increase compared to ones obtained using 3 km radius. These results validates the hypothesis that people living in faraway suburbs may have to drive more for basic things, such as groceries, or parks etc.

I have just included two plots in the report for illustration.



4. Results

In the previous section, I have performed an analysis to get some insights from the data. In this section, I will report the

- Top 5 suburbs based on crime rates
 - Top 5 venues (within 10 km) for the top 5 suburbs for more insights
- Worse 5 suburbs based on crime rates
- Clustered suburbs (based on distance from the city center) on a map
 - The size of the marker is proportional to mean percent crime
 - The map provides geospatial distribution of crimes in Melbourne

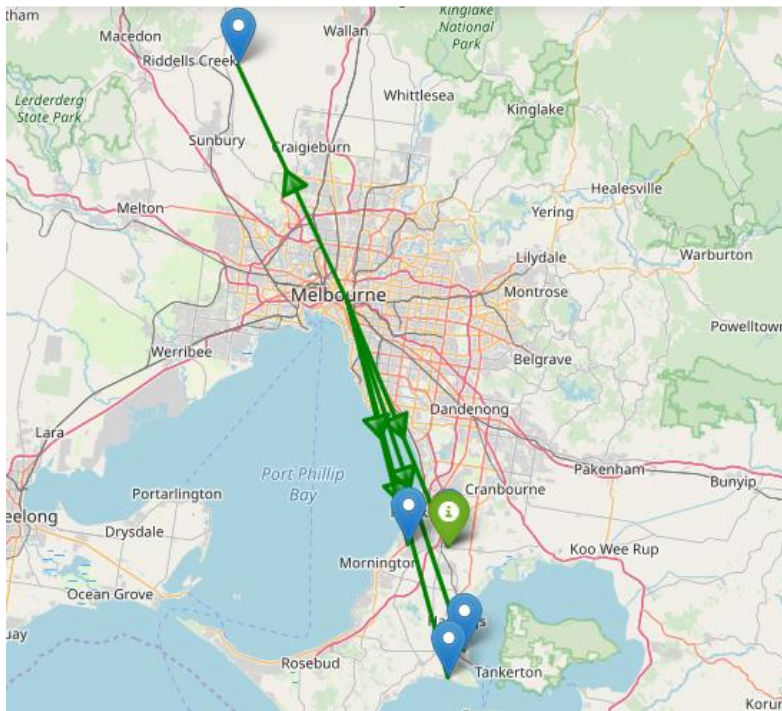
4.1 Best Suburbs

Overall, the crime rate gets lower as we move away from the city center. The suburbs from 40 km onwards (except for suburbs in 50 km range) have comparatively low crime rates.

As discussed before, we do not want to find a very remote suburb. So, I look for reasonable suburbs, which are 40 -70 km away from the city center and have the least crime rates.

I have used the folium map to highlight the best five suburbs (location icons) as shown below.

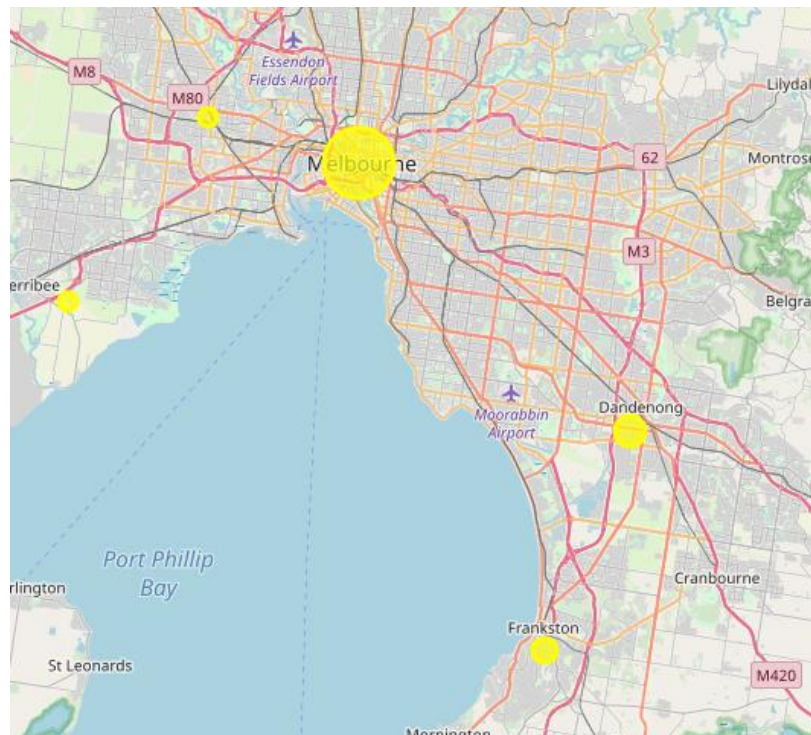
- The best suburb (least crime rate) is indicated by a green icon (location type marker)
- The suburb name is 'Kee Wee Rup North' and it is around 44 km away from the city center



4.2 Worse Suburbs

I have identified the worse five suburbs and plotted them on a map

- The yellow markers indicate the worse suburbs with the highest crime rates. The size of the marker is proportional to mean crimes for that suburb
- Interestingly, the city center has the highest crime rate.



4.3 Top Venues for best suburbs

We already know the top five suburbs. I identified the top 5 venues for each suburbs as shown below.

SuburbName	MeanCrimes	DistCityCenter	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Koo Wee Rup North	6.8	44.273578	Fast Food Restaurant	Grocery Store	Café	Gas Station	Supermarket
HMAS Cerberus	10.4	65.045797	Beach	Café	Grocery Store	Australian Restaurant	Hotel
Main Ridge	11.2	41.983425	Café	Supermarket	Beach	Gastropub	Convenience Store
Clarkefield	20.0	44.332088	Train Station	Thai Restaurant	Construction & Landscaping	Café	Bar
Merricks, Point Leo, Shoreham	32.6	61.551920	Grocery Store	Bakery	Fast Food Restaurant	Airport	Pizza Place

The top 3 suburbs have grocery stores as popular venues. This is important as everyone often visits them to buy grocery items.

However, if someone prefers a beach, then, there are two suburbs with nearby beaches.

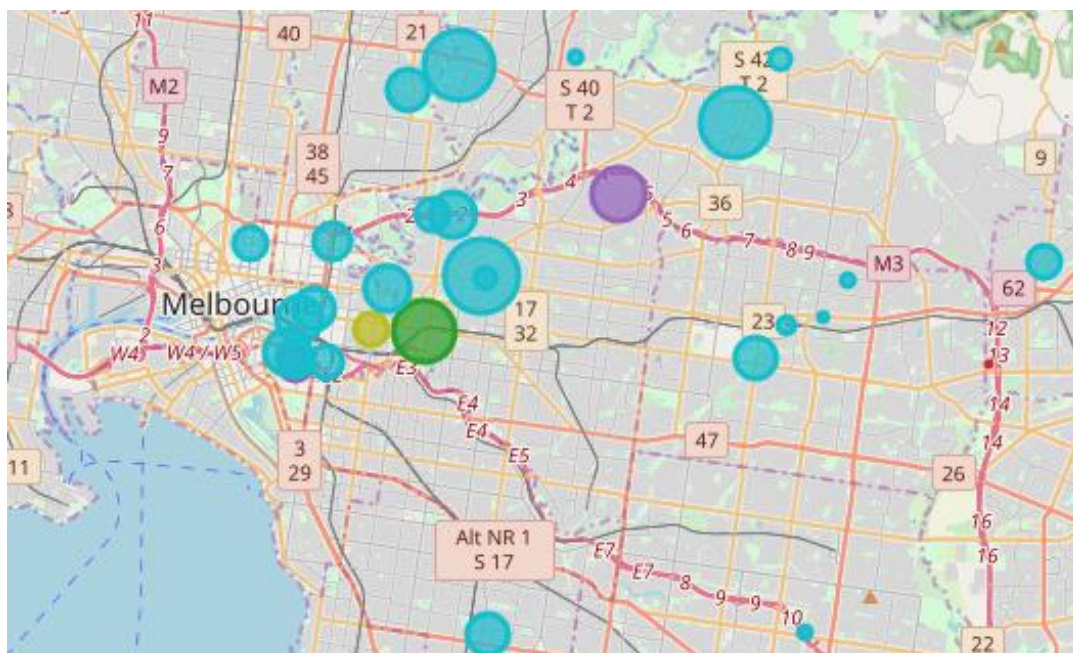
This information provides comprehensive insight, and one can choose the suburb based on personal preferences.

4.4 Clustered Suburbs with mean percent crimes

In this map, I clustered the suburbs of Melbourne into 50 clusters using the distance from the city center.

Then, I computed the crime percentage for each cluster by dividing the sum of crimes (mean) for each cluster by the sum of crimes (mean) for the Melbourne city.

I plotted the clusters on a map, which provides a rough geospatial distribution of crimes in Melbourne. The size of the marker is proportional to the percent crime.



5. Conclusion

The project indicates that it is not safe to live in the city center due to high crime rates. The crime rates decrease as we move away from the city center. But the frequency of venues also starts decreasing and requires more commute to reach the famous places. I believe that this hypothesis will be valid for any big city in the world. Moreover, rental or house prices also reduce if we choose to live a bit away from the city center.

I have reported the top five suburbs to live based on the least crime rates.

- I have also shown the geospatial crime distribution using a map, which will be useful for immigrants to avoid areas with high crime rates.

6. Future Work

In the future, I will like to link this project with the housing market.

I will like to see the impact of crimes in a suburb on the sale or rent of houses. Such analysis will provide extra help to immigrants. Such as, if some home is available for an extra-low price, then that can be associated with crime rates in that suburb for more investigation.

7. References

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