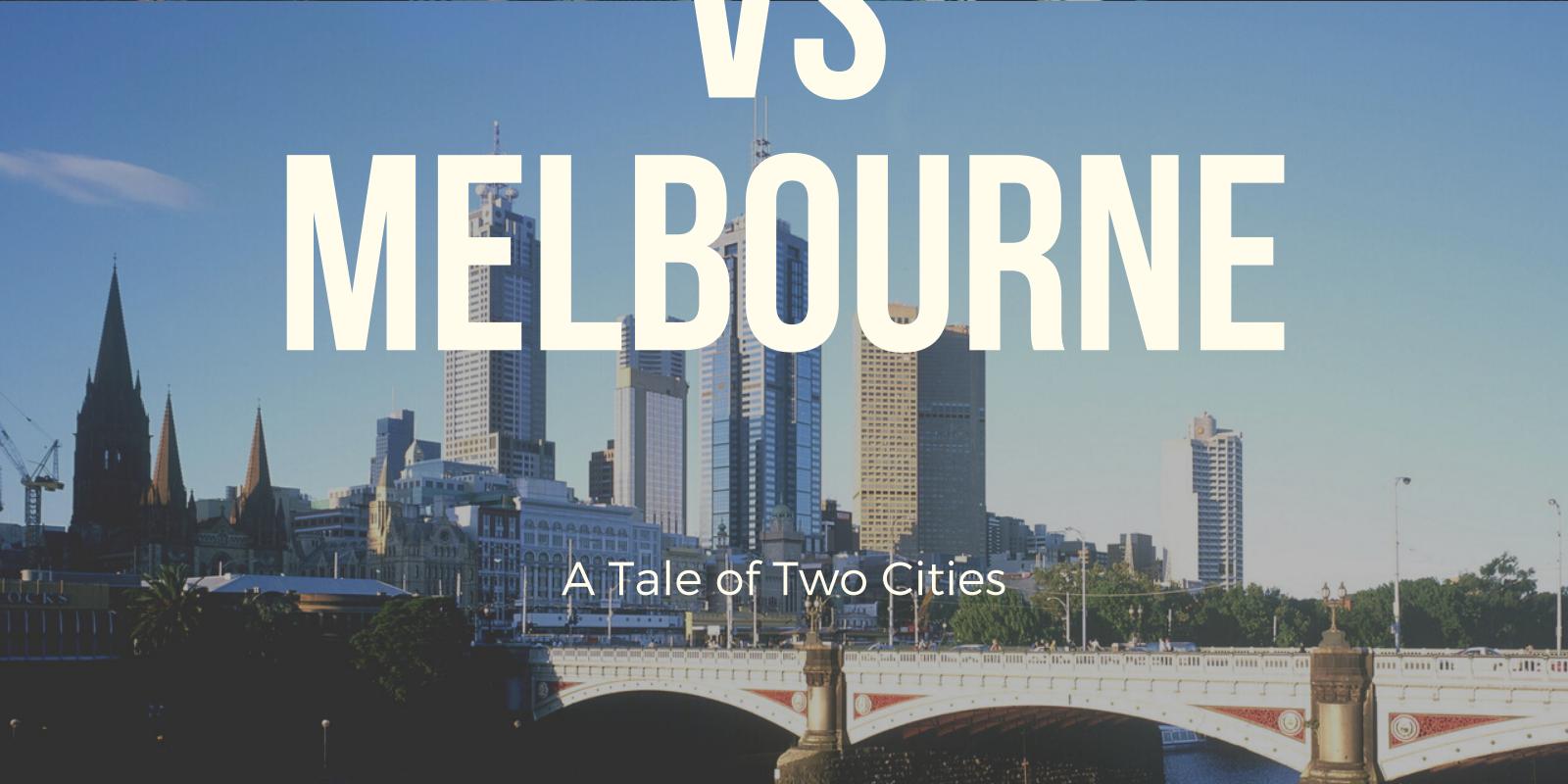


APPLIED DATA SCIENCE
CAPSTONE

NEW YORK VS MELBOURNE



A Tale of Two Cities

CONTENTS

1. Introduction
2. Data
3. Methodology
4. Results
5. Recommendations
6. Conclusion

INTRODUCTION

1.1 Business Case Scenario

I'm a post-graduate applicant considering offers from two universities - **Columbia University** and the **University of Melbourne**. Both academic programs have equal value to me and the tie-breaker is city livability in the university surrounds, since I am looking to rent in a neighbourhood close to the university. I am measuring this by evaluating both cities based on two questions:

- What **amenities** are available in each university area?
- Which university area has the **lowest crime rate**?

1.2 Key Considerations

- Walking distance to each university <= 5 km (or 3.11 miles)
- Local attractions and venues available
- Crime rate

1.2 Other Target Audience

- University students considering relocation
- University academics and staff considering relocation
- Property investors targeting the university body
- Business owners targeting the university body

DATA

2.1 New York Dataset

- FourSquare API Explore call
- **[Dataset A]** City of New York (NYPD Complaint Data):
<https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>
- **[Dataset B]** NYC Open Data: <https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz>
- **[Dataset C]** City of New York (Boundary Data):
https://cocl.us/new_york_dataset

2.2 Melbourne Dataset

- FourSquare API Explore call
- **[Dataset D]** Data VIC - Criminal incidents by LGA - Dec 2019:
<https://discover.data.vic.gov.au/dataset/crime-by-location-data-table>
- **[Dataset E]** Australia Post data:
https://www.matthewproctor.com/australian_postcodes
- **[Dataset F]** City of Melbourne Open Data - Postcodes:
<https://data.melbourne.vic.gov.au/Property/Postcodes/m7yp-p495>

DATA

2.3 Data Tools

- FourSquare API will be used to collect venues and neighbourhoods
- The visual maps will be created using Folium
- Pandas will be used to transform and load data onto dataframes



METHODOLOGY

3.1 FourSquare API + Kmeans Clustering

Using the the FourSquare API Explore call and Kmeans Clustering, top venues (within a 5km radius) around each university would be crawled and then grouped into clusters

3.2 Choropleth map

Using Folium, the neighbourhood boundary data of Melbourne and New York would be plotted onto a choropleth map, graduated by crime rate statistics

3.3 Combine elements

FourSquare cluster markers would then be plotted on its corresponding Choropleth map for a more comprehensive picture of the university surrounds and its relative safety in terms of crime rate.

RESULTS

4.1 Dataframe Load

We first loaded the neighbourhood boundary coordinates of both cities onto a dataframe

Melbourne

```
print(df_mel.shape)
df_mel.head()

(12, 3)

      Neighborhood        lat        long
0    North Melbourne -37.800600  144.943559
1        Melbourne   -37.817403  144.956776
2   East Melbourne   -37.818517  144.982207
3         Carlton   -37.803569  144.966112
4   Carlton North   -37.786971  144.967240
```

New York City

```
4
5     neighborhood_latlon = data['geometry']['coordinates']
6     neighborhood_lat = neighborhood_latlon[1]
7     neighborhood_lon = neighborhood_latlon[0]
8
9     neighborhoods = neighborhoods.append({'Borough': borough,
10                                         'Neighborhood': neighborhood_name,
11                                         'Latitude': neighborhood_lat,
12                                         'Longitude': neighborhood_lon}, ignore_index=True)
13 neighborhoods.head()
```

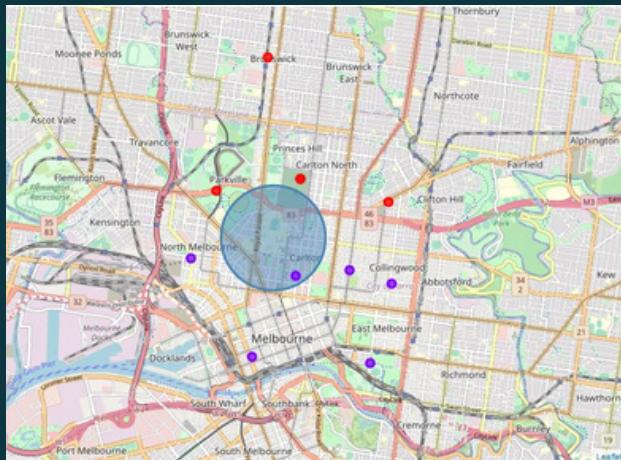
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

RESULTS

4.2 FourSquare API and Kmeans Clustering

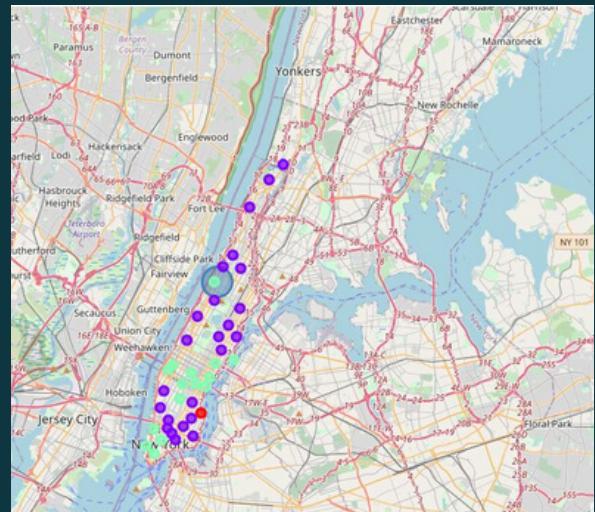
After using the FourSquare API to make an Explore call to gather venues for each city, Kmeans clustering was then applied to both to group the venues.

A preliminary Folium plot was then created for both cities



University of Melbourne and K-Clusters

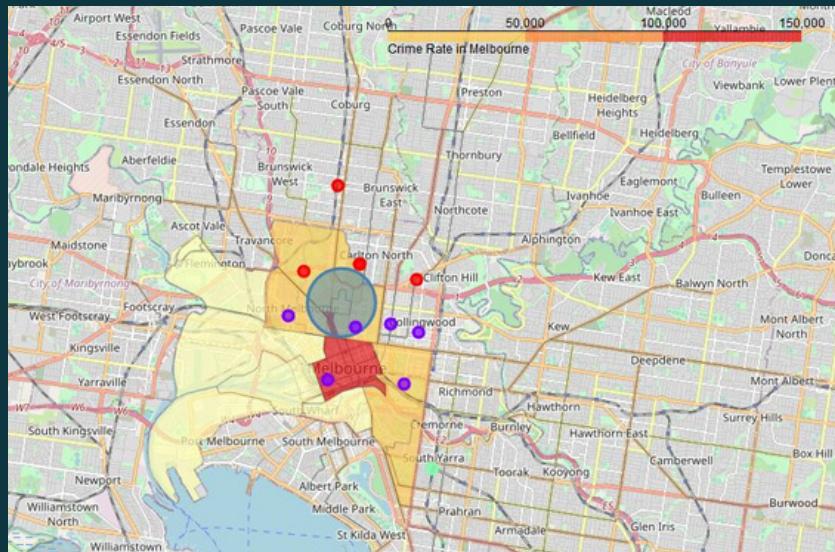
Columbia University and K-Clusters



RESULTS

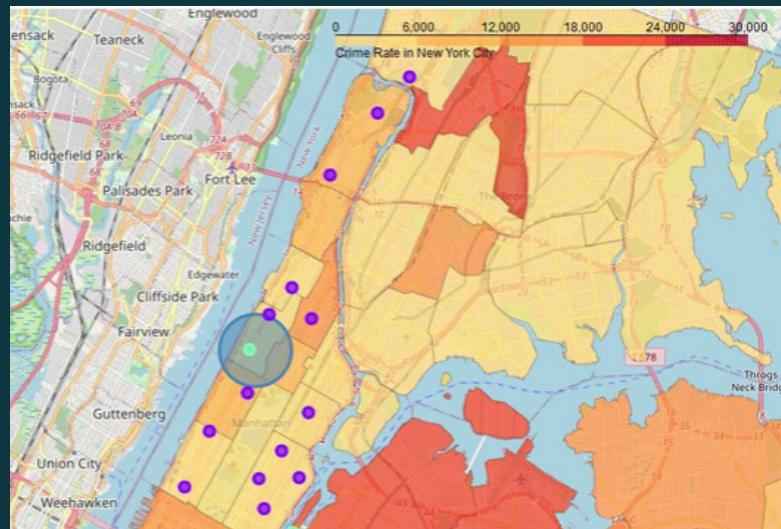
4.3 Plotting the Final Map

Taking the above elements we combined them with a Folium choropleth map which was generated using boundary JSON files based on **postcodes** (for Melbourne) and **police precincts** (New York City) and Crime Statistics for both cities.



University of Melbourne Surrounds

Columbia University Surrounds



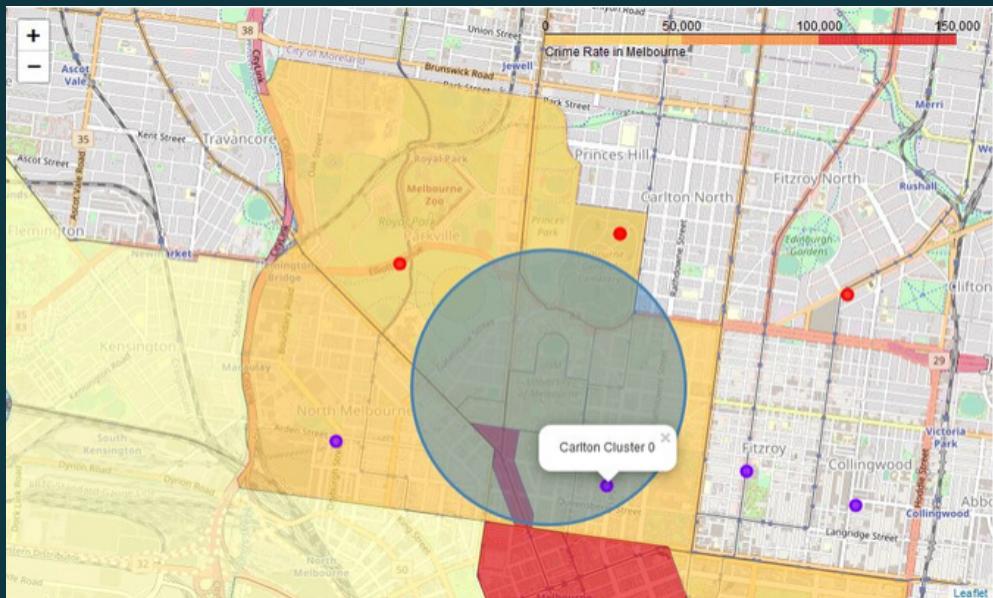
RECOMMENDATIONS

5.1 University of Melbourne

Although we started with shortlisting venues within walking distance of the university, we want to further refine our criteria so that we shortlist just one or two prime clusters for easy decision-making. To this end, we reduced the radius to 1 km (0.62 mile) walking distance to the university.

The **Carlton** cluster comes up as a clear choice and which also happens to be a low crime rate area. This would be a neighbourhood I can consider renting in Melbourne.

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
1	Carlton	0	Café	Coffee Shop	Wine Bar	Bakery	Cocktail Bar	Ice Cream Shop	Park	Whisky Bar	Vegetarian / Vegan Restaurant	Bar

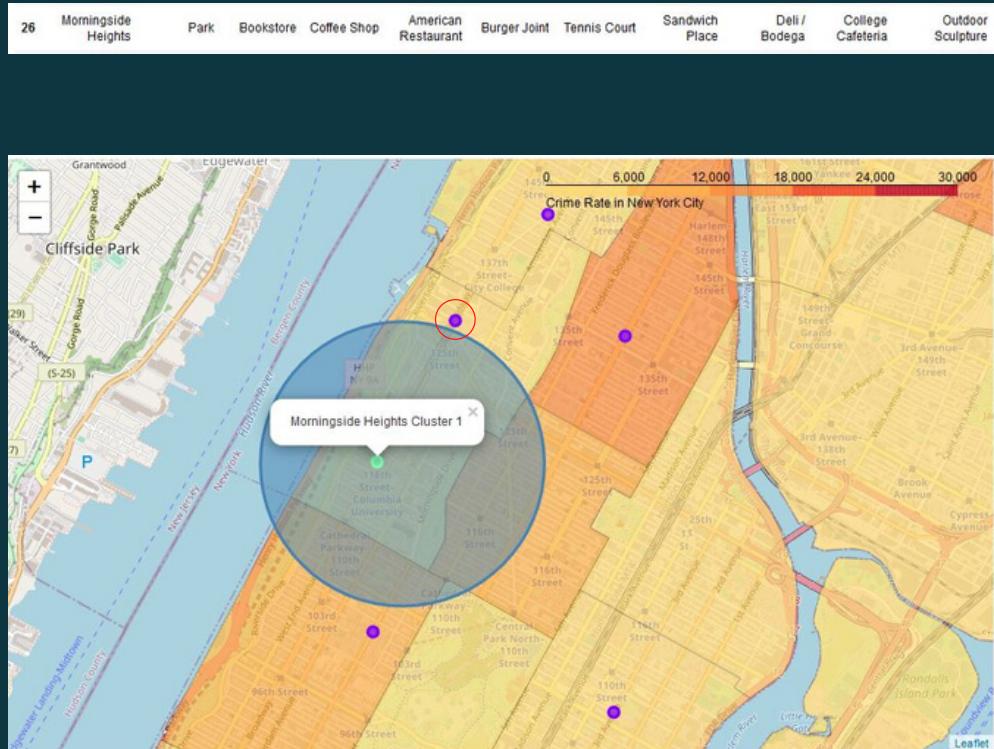


RECOMMENDATIONS

5.2 Columbia University

Again, we reduced the radius to 1 km (0.62 mile) walking distance to the university.

The **Morningside Heights** cluster comes up as a clear choice and which also happens to be a low crime rate area. A second cluster that comes close is the **Manhattanville** Cluster which sits close to our radius limit (circled in red)



CONCLUSION

6.1 Prime Neighbourhood Clusters

Two neighbourhood clusters, one for each city, have emerged for our consideration:

- **Carlton** in Melbourne
- **Morningside Heights** in NYC

6.2 Info Gap and Further Analysis

However, before we can make a final decision, additional information could be gathered in these other areas:

- Listings and average prices of rental apartments in each area
- Proximity to public transport (e.g. Melbourne has a free tram zone within the central city limits)
- Availability of late night amenities when working late
- The Melbourne City postcode json boundaries is limited, a more expansive dataset should be obtained if possible