

# **Travel to Toronto, Sydney, London and New York: The Data Scientist Way!**

**David See**  
IBM – Applied Data Science Capstone

# Table of Contents

1. Introduction

2. Data

3. Methodology

4. Results

5. Discussion

6. Conclusion

# 1. Introduction

## The Problem Statement

- “What places should I visit in a given city such that I could maximise my travel experience and minimise visiting places that are similar?”
- “Given my travel interests, which countries should I visit and what are the differences/similarities between these countries?”
- **More specifically, I am a traveller who is keen to explore and review cafés, following which I would like to determine which neighbourhood and country I should travel to in this report.**
- Travellers who wish to visit the following four cities/countries would be interested in this report:
  - Toronto, Canada
  - Sydney, Australia
  - London, United Kingdom
  - New York, United States.
- Note that the relevant map is visualised as the **background** on each slide (we have the world map with the four cities labelled on this slide).

## 2. Data

# City Information and Foursquare

- The postal code, borough, neighbourhood, latitude and longitude for each city were obtained from Wikipedia, doogal, CostlessQuotes and Distancesto.
- Using the latitude and longitude for each of the neighbourhoods above, we applied the ‘venues/explore’ endpoint via Foursquare to discover nearby venues that are situated within 500 meters.

### 3. Methodology

## Machine Learning: k-Means Clustering

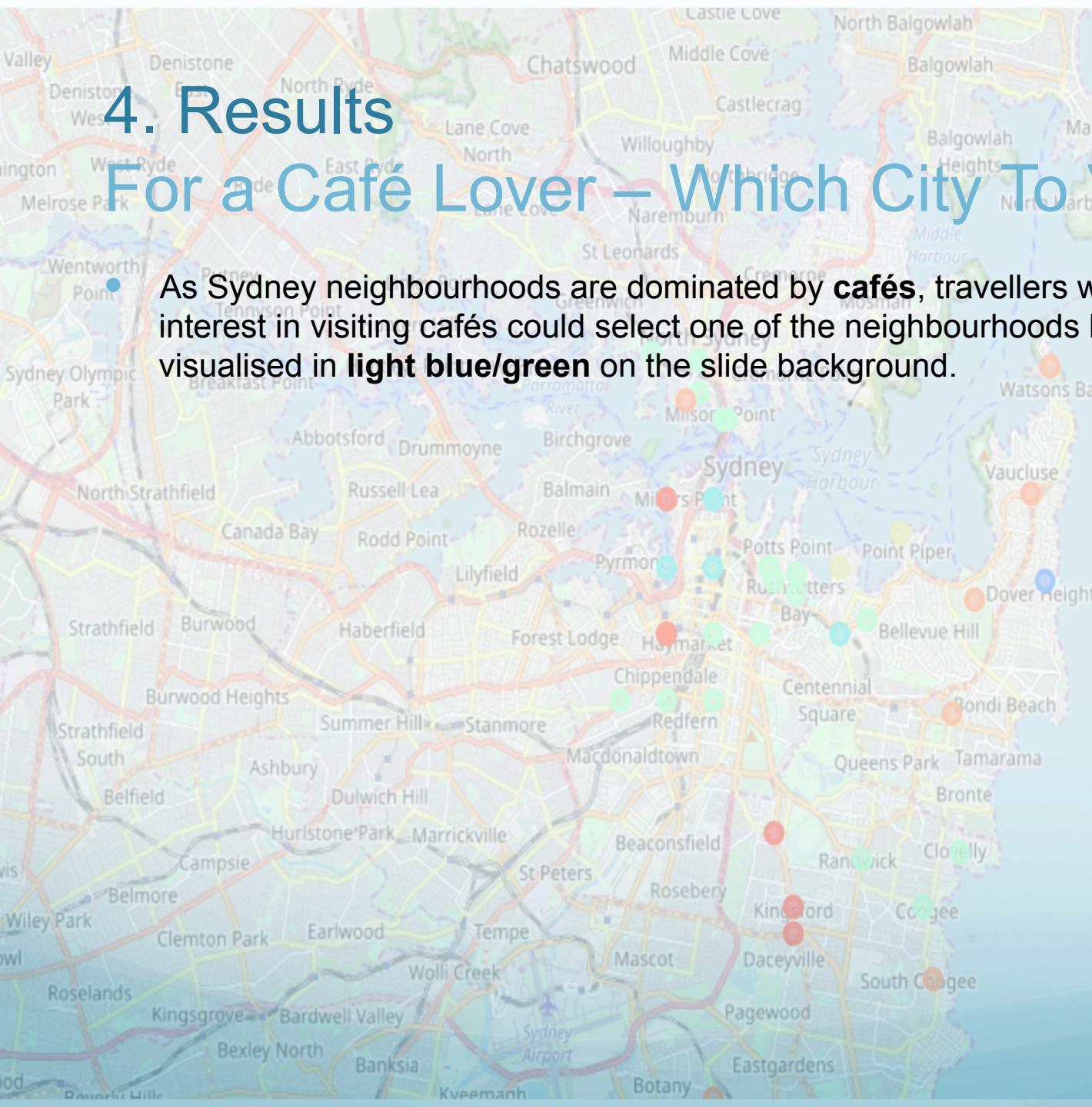
- Using the information from the previous section, we labelled the neighbourhoods in each city according to their latitude and longitude (The map for Sydney is shown as the background of this slide as an example).
- We applied machine learning (specifically k-means clustering) to the data. We assume that travellers will spend 7 days in the city of their choice and as such we have selected 7 clusters using the k-means clustering algorithm. The travellers are expected to spend 1 day in each of the 7 clusters.
- **The places *within* each cluster are relatively similar, however the places are relatively different *between* each cluster.**

### 3. Methodology Machine Learning: Data Visualisation

- The idea is that neighbourhoods that are similar will be grouped within the same clusters using machine learning.
- We use a map to visualise the results from the k-means clustering, where similar clusters share the same colours (the background of this slide) – this provides the basis to answer the questions posed at the start of the presentation.
  - For example, travellers who have an interest in visiting **cafés** could select one of the neighbourhoods (e.g. The Rocks) listed under Cluster 0, which are visualised in **red** on the slide background.
  - Conversely, Cluster 4 is suitable for travellers looking to explore both **Chinese restaurants and Indonesian restaurants** as visualised in **green** on the slide background.

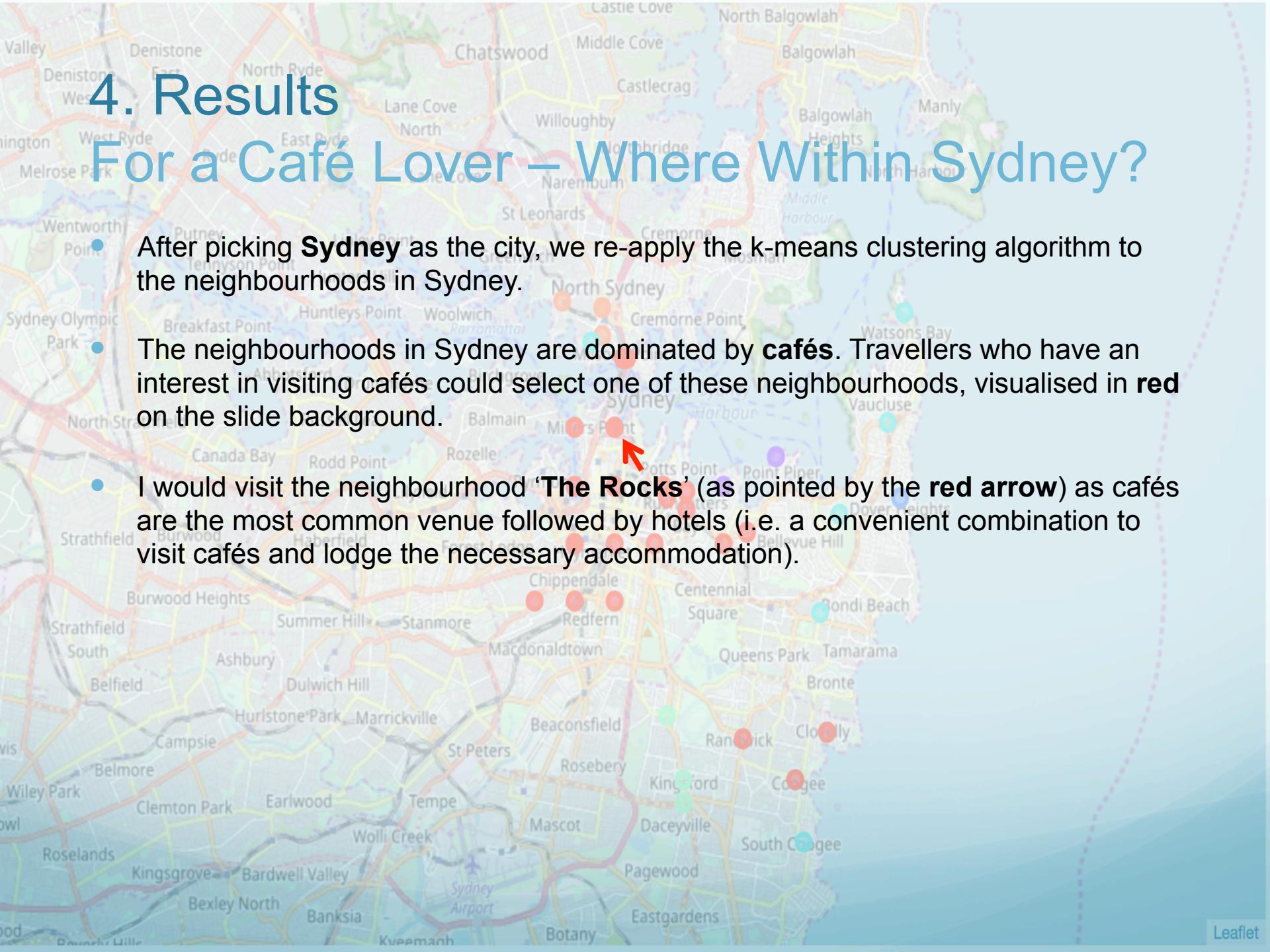
# 4. Results For a Café Lover – Which City To Visit?

- As Sydney neighbourhoods are dominated by **cafés**, travellers who have an interest in visiting cafés could select one of the neighbourhoods located in **Sydney**, visualised in **light blue/green** on the slide background.



# 4. Results For a Café Lover – Where Within Sydney?

- After picking **Sydney** as the city, we re-apply the k-means clustering algorithm to the neighbourhoods in Sydney.
- The neighbourhoods in Sydney are dominated by **cafés**. Travellers who have an interest in visiting cafés could select one of these neighbourhoods, visualised in **red** on the slide background.
- I would visit the neighbourhood '**The Rocks**' (as pointed by the **red arrow**) as cafés are the most common venue followed by hotels (i.e. a convenient combination to visit cafés and lodge the necessary accommodation).



# 5. Discussion

## London as a Hub for Coffee Shops

- The City of London is dominated by coffee shops, making it a popular destination for coffee lovers.
- Restaurants (in particular Italian restaurants) are the second most common venue after coffee shops – a point that is highly relevant to pasta and pizza enthusiasts.

# 5. Discussion

## New York – Ideal for Families

- The popular venues in New York are parks, gyms, clothing stores, American restaurants, Japanese restaurants, theater, spa, exhibitions, supermarket, hotel and ice cream shops, making it an ideal travel destination for families.



## 5. Discussion

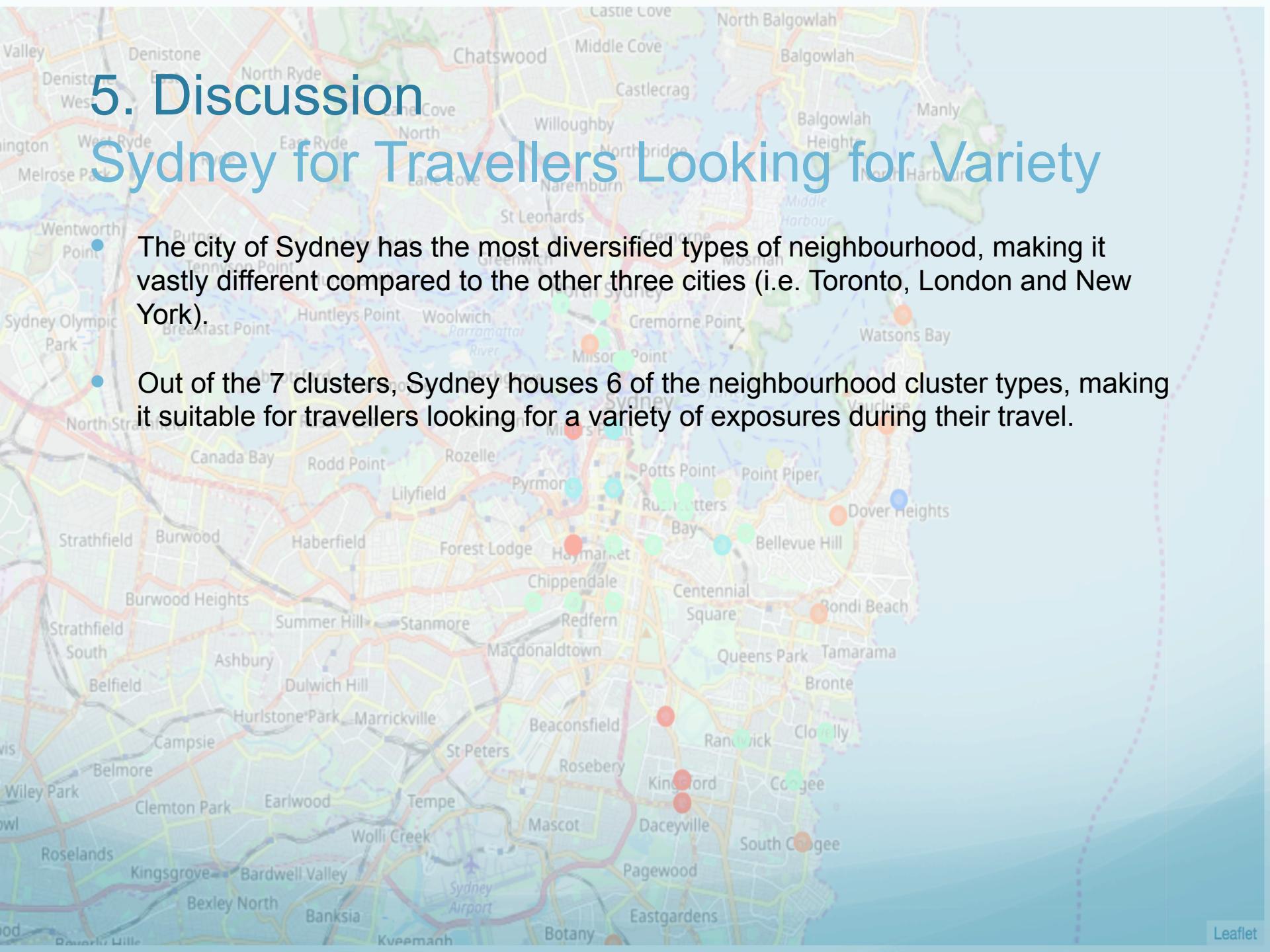
### Toronto – Mixture of London and New York

- Even though Toronto is geographically closer to New York, the city of Toronto houses more neighbourhoods that are more similar to London than New York.
- In essence, Toronto has a mixture of neighbourhood styles from both London and New York. Therefore, travellers are able to experience the neighbourhood cultures from both London and New York just by visiting Toronto alone. This proposition is relevant for travellers that have travel budget constraints but would like to make full use of their travel expenses.

# 5. Discussion

## Sydney for Travellers Looking for Variety

- The city of Sydney has the most diversified types of neighbourhood, making it vastly different compared to the other three cities (i.e. Toronto, London and New York).
- Out of the 7 clusters, Sydney houses 6 of the neighbourhood cluster types, making it suitable for travellers looking for a variety of exposures during their travel.



## 6. Conclusion

### The Problem Statement Answered

- Being a café lover myself, where is the best place to visit? The answer is **The Rocks, Sydney, Australia.**
- From the analysis, we draw the following conclusions for each of the four different cities/countries:-
  - Toronto as a Mixture of London and New York
  - Sydney for Travellers Looking for Variety
  - London as a Hub for Coffee Shops
  - New York – Ideal for Families.