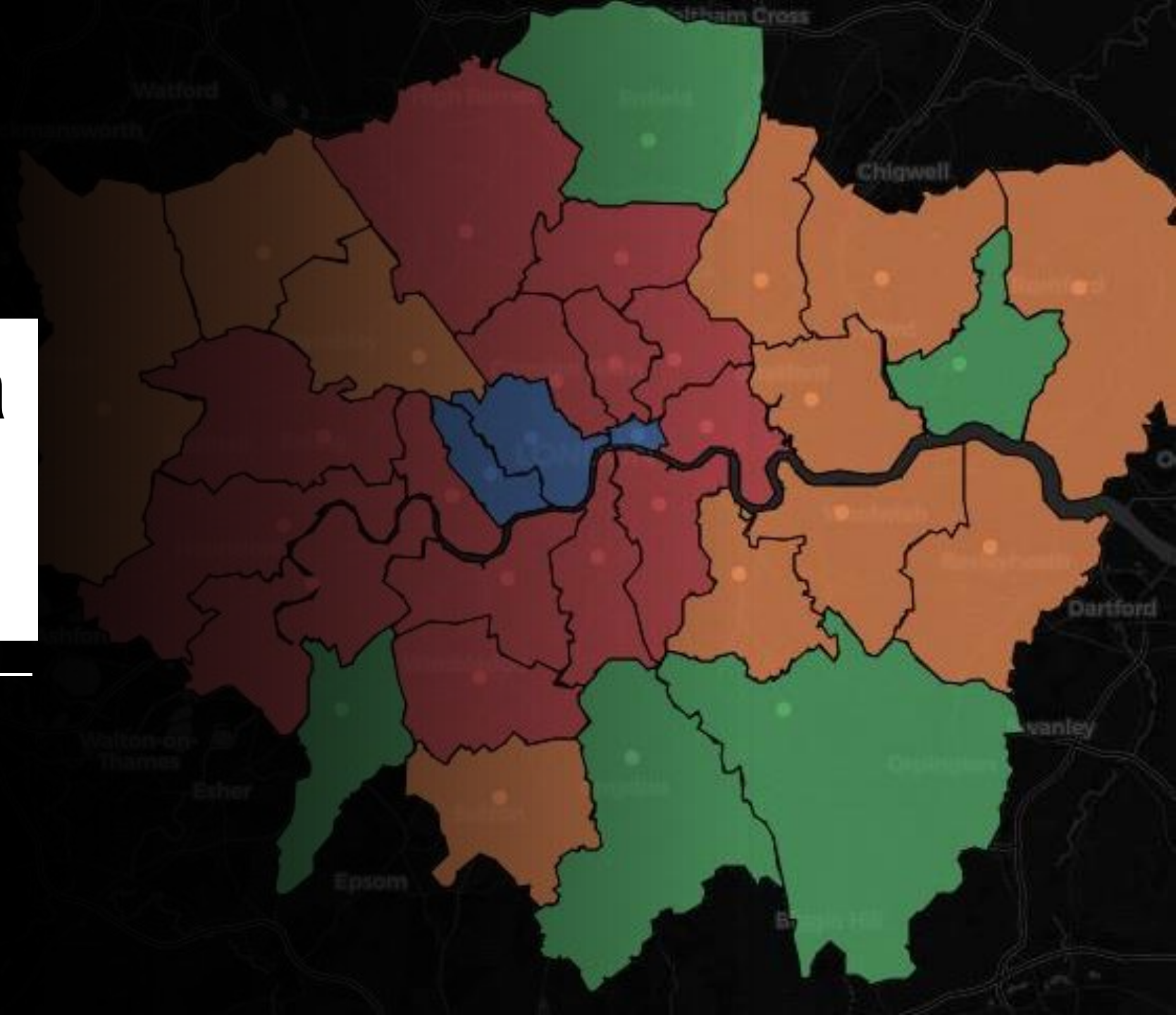
IBM Coursera Capstone Project

Clustering Neighbourhoods



Introduction

- One of the most common type of traveling is a business or non-business **city trip**,
- Travellers are in need of **quick orientation** in an unknown metropolis with its overwhelming impressions, opportunities, venues and wide variety of unique districts with their individual characters.
- Solution: a **clustering system** for neighbourhoods of each individual city based on several dimensions of up-to-date information

Data

Data used in the project to build a a meaningful and characteristic feature set for each neighbourhood of a specific city:

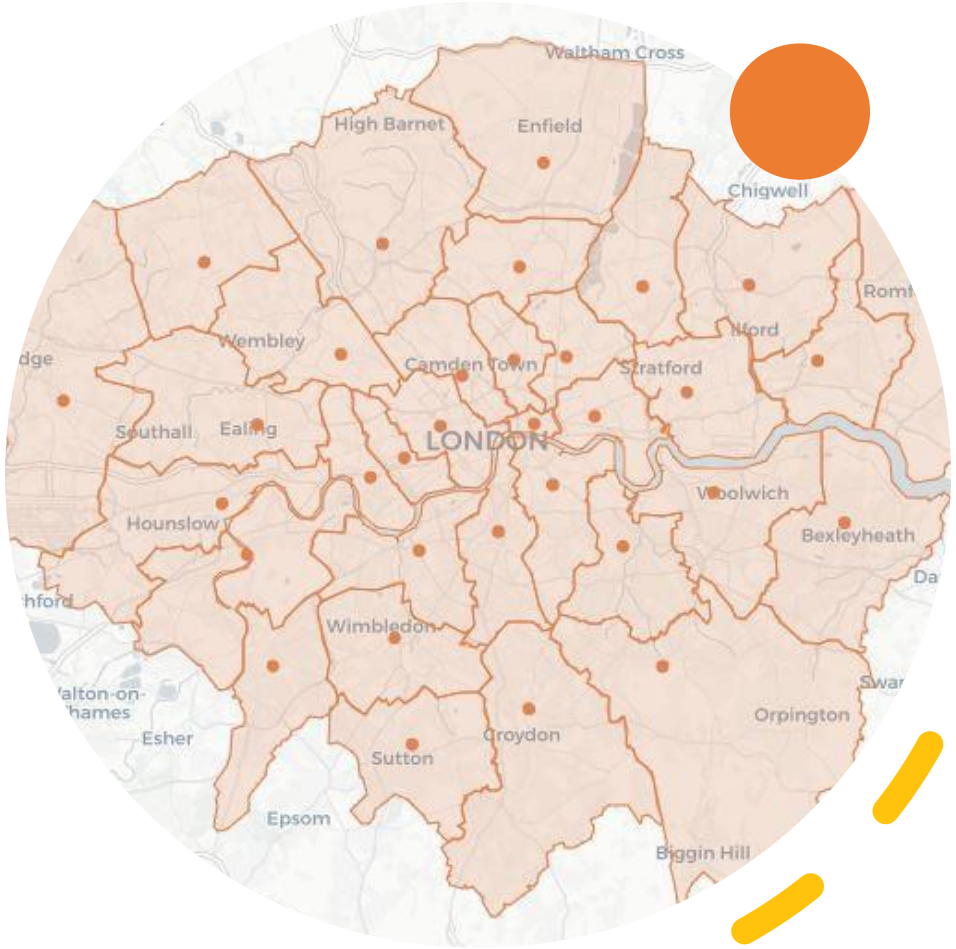
1. **Airbnb** listings data (from http://insideairbnb.com/index.html)

   - Airbnb price level as a price indicator per neighbourhood
   - Airbnb room type data

2. **Foursquare API**
   for analyzing **venue data** for the neighbourhoods of a specific city.

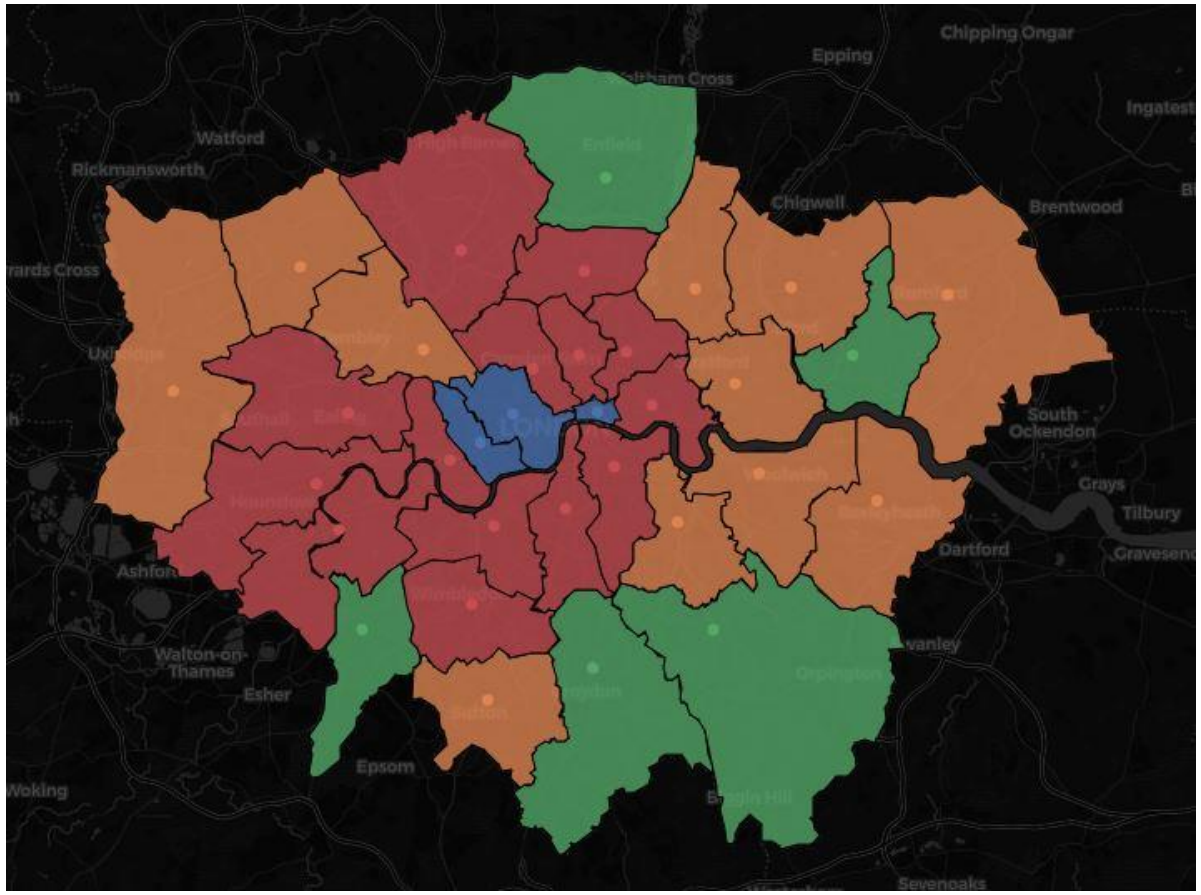
London

We will use data for the city of **London** for an example workflow



Clustering

After building our **feature set** we apply K-Means and gain 5 clusters for London's neighbourhoods:



Cluster Profiles

Cluster Profiles

- **Cluster 0** (e.g. Westminster, City of London): high price level, few private rooms, many restaurants
- **Cluster 1** (e.g. Brent, Newham): relatively low price level, many private rooms, many shops
- **Cluster 2** (e.g. Bromley, Croydon): low price level, many private rooms, many outdoors & recreation venues
- **Cluster 3** (e. g. Camden, Islington): moderate price level, many private rooms, many nightlife venues