Clustering stations in London based on Arts & Entertainment venues

Clustering stations this way helps tourists

- ▶ Filter stations based on their preferred venue types
- ▶ Better plan their trips
- ▶ Choose an optimal hotel location
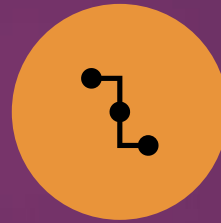- ▶ Gain a better understanding of the city
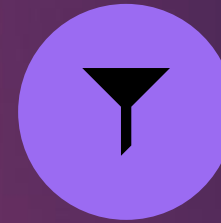
Data Acquisition & Cleaning

STATION DATA WAS COLLECTED FROM DOOGAL.CO.UK
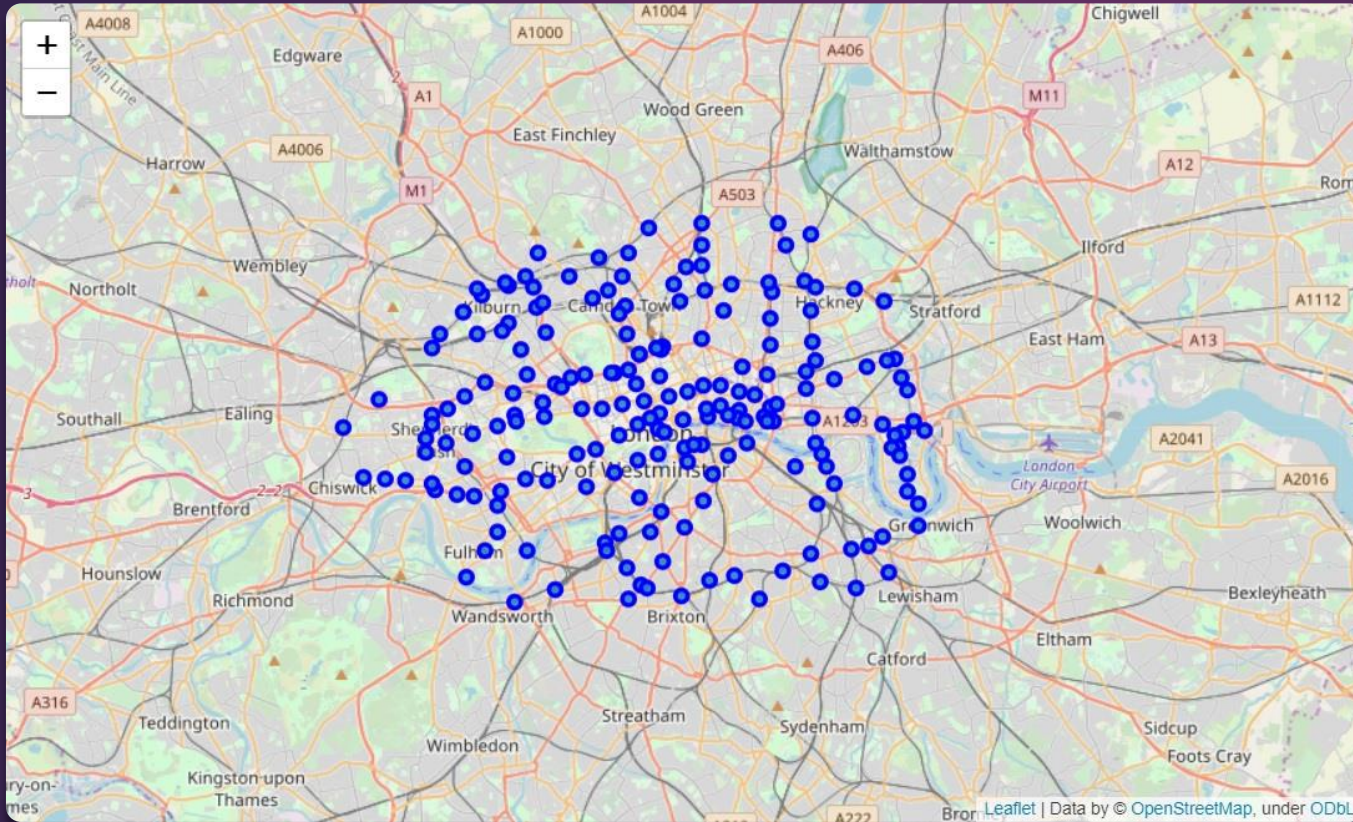
VENUE INFORMATION WAS COLLECTED USING THE FOURSQUARE API

ONLY STATIONS IN ZONES 1 AND 2 WERE KEPT

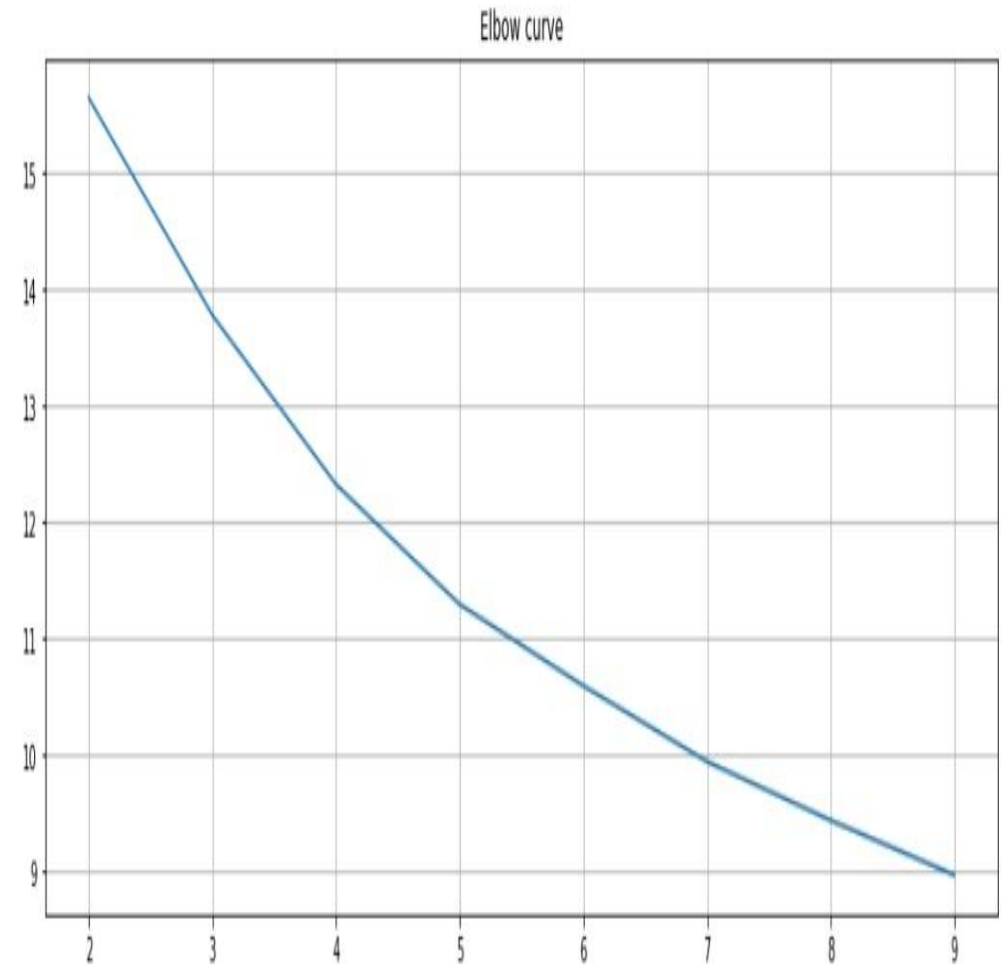**STATIONS WITH LESS THAT FOUR UNIQUE VENUE CATEGORIES WERE REMOVED**

TOTAL OF 117 STATIONS ACROSS LONDON

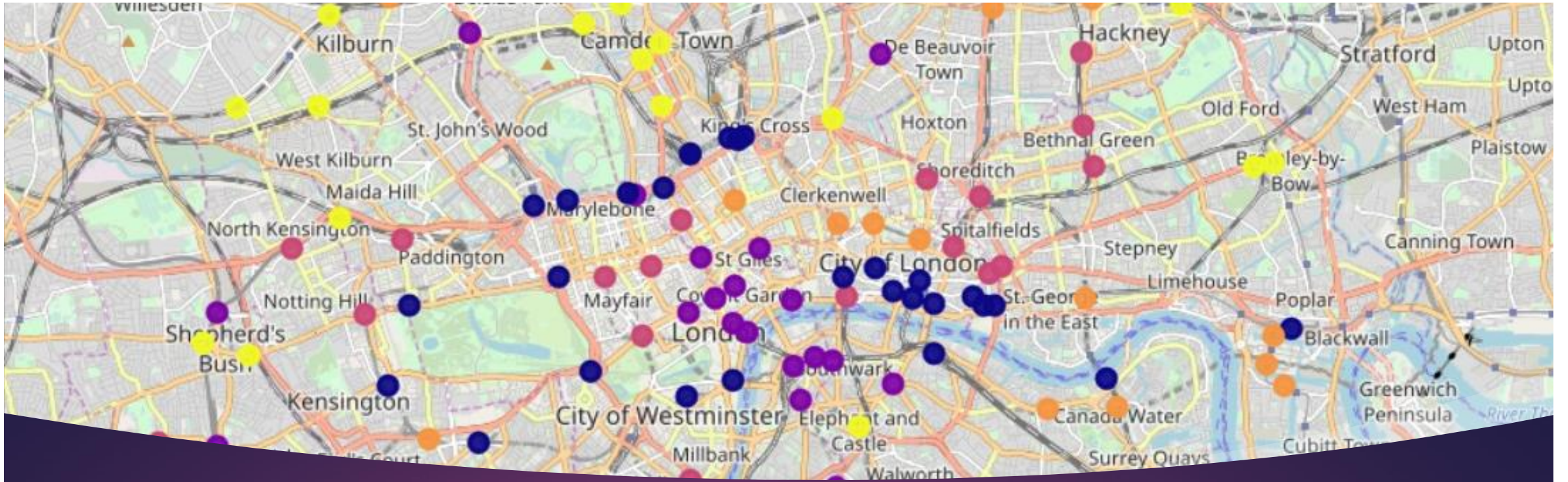Original stations before data cleaning

K-Means was run

- Performed one hot encoding

- Got the frequency for the top 4 categories

- Implemented k-means for 5 clusters since elbow curve was inconclusive

- Run for a total of 12 times with different centroid seeds

Elbow curve

Clustering had interesting results

▶ Stations were spread-out almost evenly across clusters

▶ The five main categories were: 'Music Venues', 'Museums', 'Art Galleries', 'Movie Theatres' and 'Theatres'

Conclusions

Arts & Entertainment venues are an effective way of clustering the stations in London

Tourists can gain valuable insight and better plan their trips

Further improvements can be made (e.g. selecting station and visualising its venues)