FINDING AREAS IN A CITY SIMILAR TO AREAS OF INTEREST

CAPSTONE PROJECT - IBM APPLIED DATA-SCIENCE COURSE ON COURSERA

AUTHOR: DAN-STEFAN FLORESCU

PROBLEM CONTEXT

- International travel is more affordable and more accessible than ever before
- People are changing their jobs more often than ever before
- People are living in more places during their lives
- People are working abroad more than ever

PROBLEM

Using data – science, finding areas in a city, similar to areas of interest.

In this project, the problem was particularized to:

City:

- Zurich, CH
- 48 areas

Areas of interest:

- South Kensington – London UK
- Greenwich – New York, NY USA
- West Loop – Chicago, IL USA.

BUSINESS CASE & STAKEHOLDERS

- In the relocation business, a firm such as a real estate agency, is hired by an employer to help a relocating employee find a place in a new city
- The firm works with the employee to find a suitable place for them to move to
- Finding areas in the new city that are similar to areas that the employee already finds favorable (areas of interest) can help the company speed up the process and give them a great place to start the search for a new home

Stakeholder

- Real estate agency working in the relocation business

DATA

Location data of the areas

- Names of the areas
- Latitude and longitude of the center of the areas
- Radius of the areas

Characterizing data of the areas

- Top 100 local venues in each area via Foursquare API

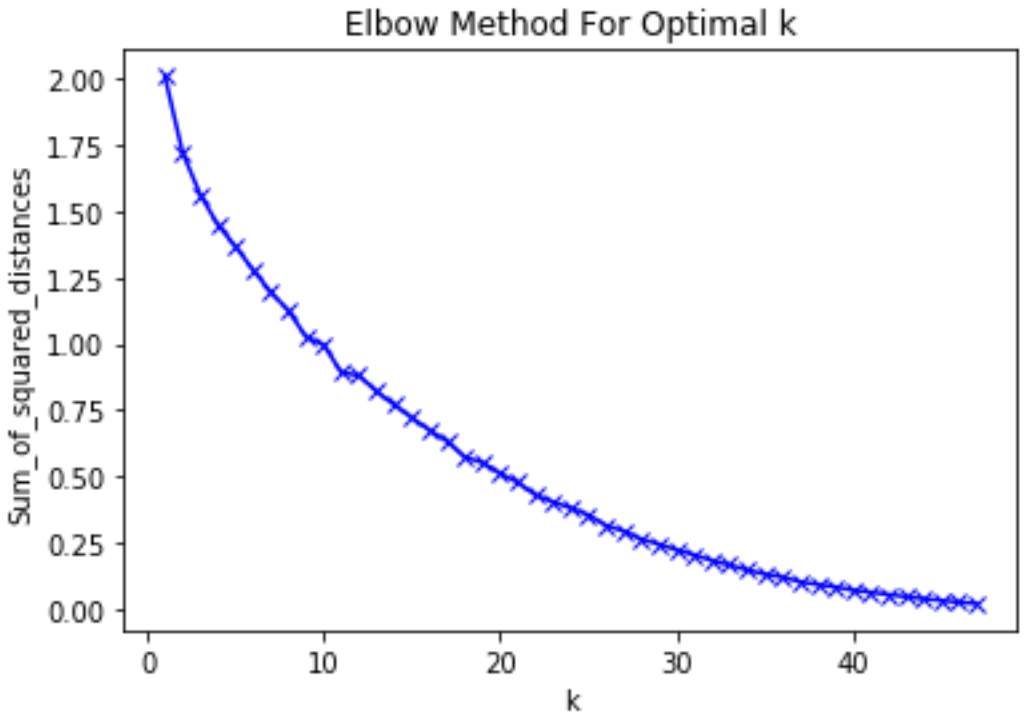
METHODOLOGY

- Data regarding venues in each were used to characterize each area
- Top 100 data regarding local venues were gathered for all areas
- Unsupervised machine learning clustering algorithm, K – means, was used to cluster similar areas
- Visualizing color coded (by cluster) areas on a folium map highlights what areas in the city are similar to areas of interest

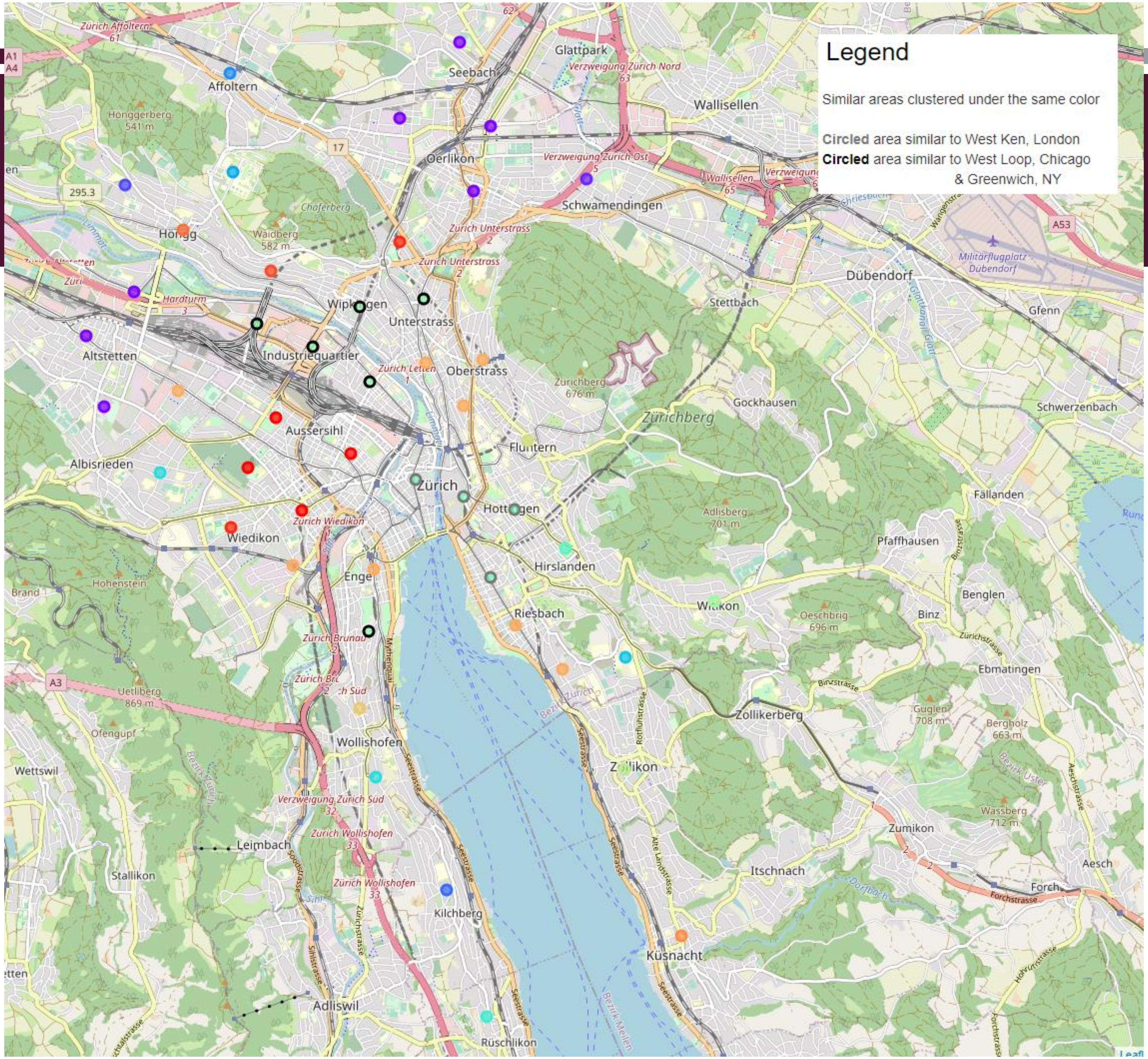
K-MEANS CLUSTERING

- sklearn KMeans function in Python was used to cluster the areas
- Data regarding the areas was imported to a pandas dataframe, sorted and one hot encoded
- Elbow method was used to find an optimum K (22), number of clusters for the model

ELBOW METHOD GRAPH



RESULTS


Legend
Similar areas clustered under the same color
Circled area similar to West Ken, London
Circled area similar to West Loop, Chicago
& Greenwich, NY
Glattpark
Seebach
Affoltern
Wallisellen
Oerlikon
Schwamendingen
Dübendorf
Stettbach
Gfenn
Wipkingen
Unterstrass
Oberstrass
Industriequartier
Altstetten
Aussersihl
Albisrieden
Fluntern
Zürich
Hottingen
Wiedikon
Gockhausen
Schwerzenbach
Fällanden
Pfaffhausen
Hirslanden
Enge
Riesbach
Benglen
Ebmatingen
Zollikerberg
Wollishofen
Leimbach
Stallikon
Kilchberg
Zumikon
Itschnach
Küsnacht
Adliswil
Rüschlikon
Wettswil
Aesch
Forch

RESULTS

Areas similar to South Kensington, London:

- Districts 1, 7, & 8; areas also known as:
  - Altstadt
  - Hochschulen
  - Seefeld
  - Muhlebach

Areas similar to West Loop, Chicago and Greenwich New York:

- Districts 5 & 10; areas also known as:
  - Escher Wyss
  - Industriequartier
  - Wipkingen