A COURSERA CAPSTONE PROJECT

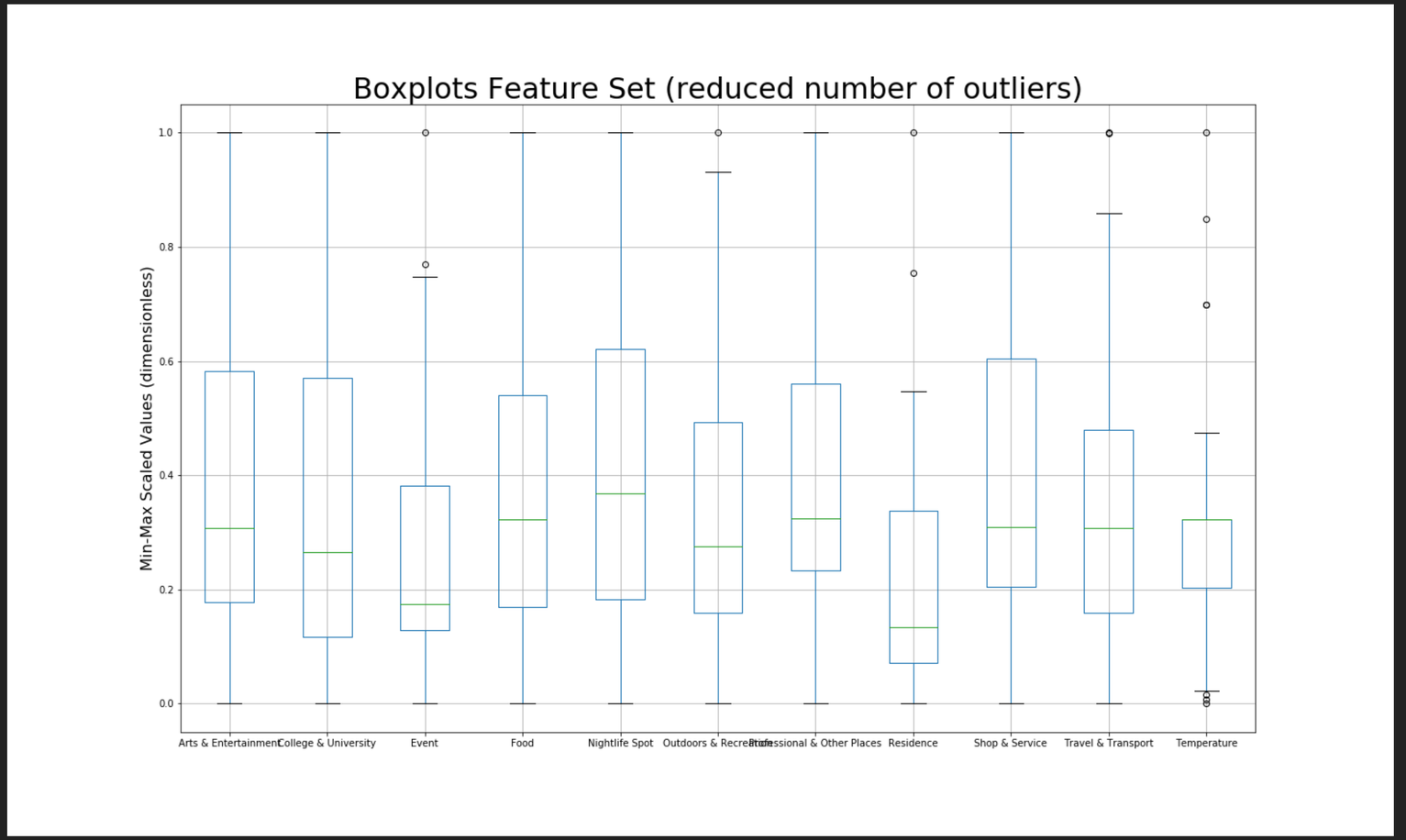
CLUSTERING EUROPEAN CITIES

INTRODUCTION

- People tend to travel to same, well-known cities
- Use Foursquare data (and more) to find similarities between European capitals
- Enable travel recommendations

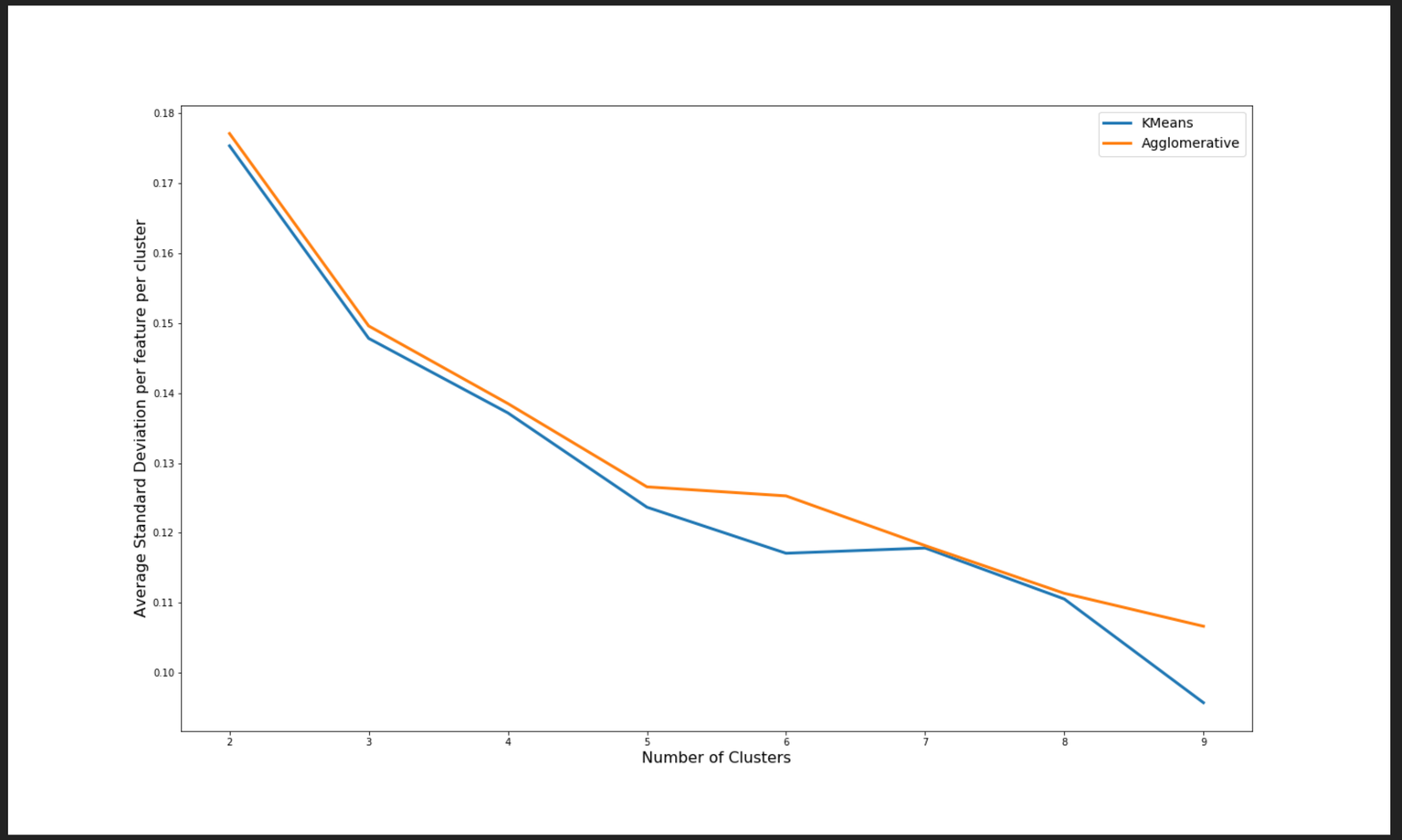
DATA

- Foursquare API
- Population numbers (indirectly)
- Average temperature
- Preprocessing: Drop cities with <= 500000 inhabitants



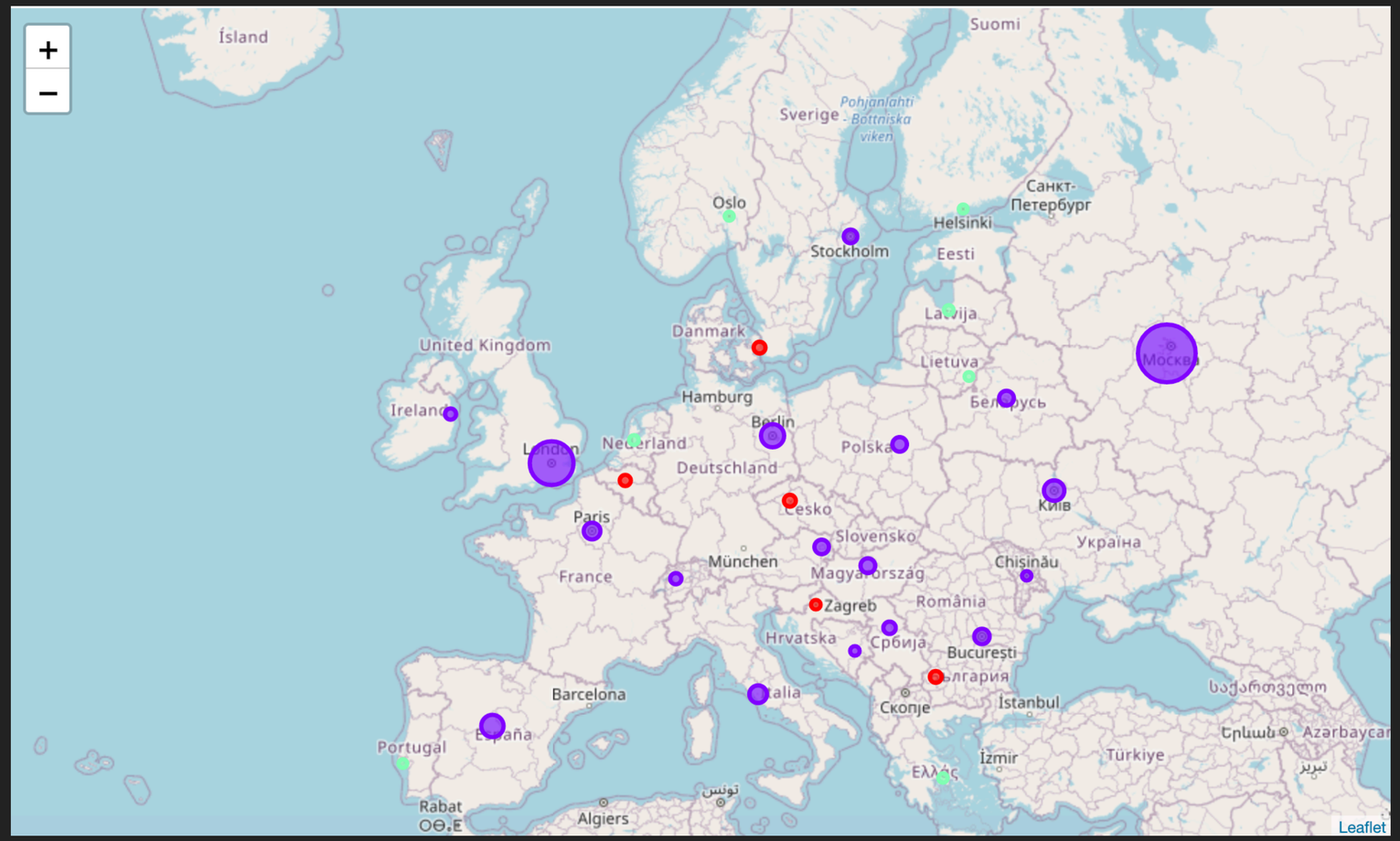
OPTIMAL CLUSTERING

- KMeans or Agglomerative?
- Optimal k?
- Elbow method:
  3- or 5-Means clustering



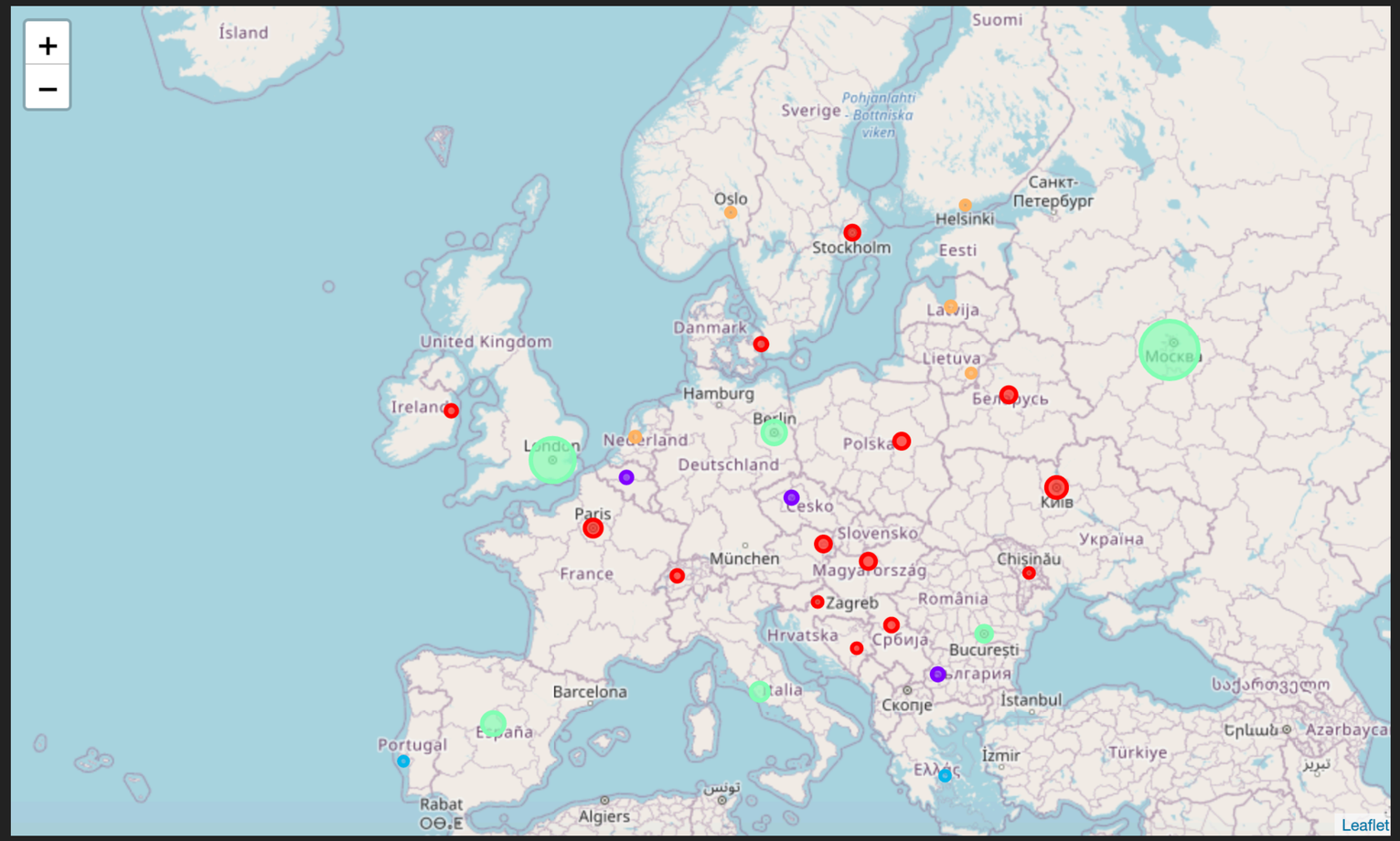
3-MEANS

- Purple: Large cluster with bigger cities
- Red: High venue density cluster
- Green: maller cities



5-MEANS

- More diversity between clusters
- Red: Largest cluster (almost 50%)
- Green: Bigger cities
- Purple: High venue density
- Blue: Warmest cities
- Orange: Coldest cities



RESULTS / LEARNINGS

- 5-Means clustering seems more appropriate
- Possible improvement: average age of population or city
- Need parameters that better describe flair