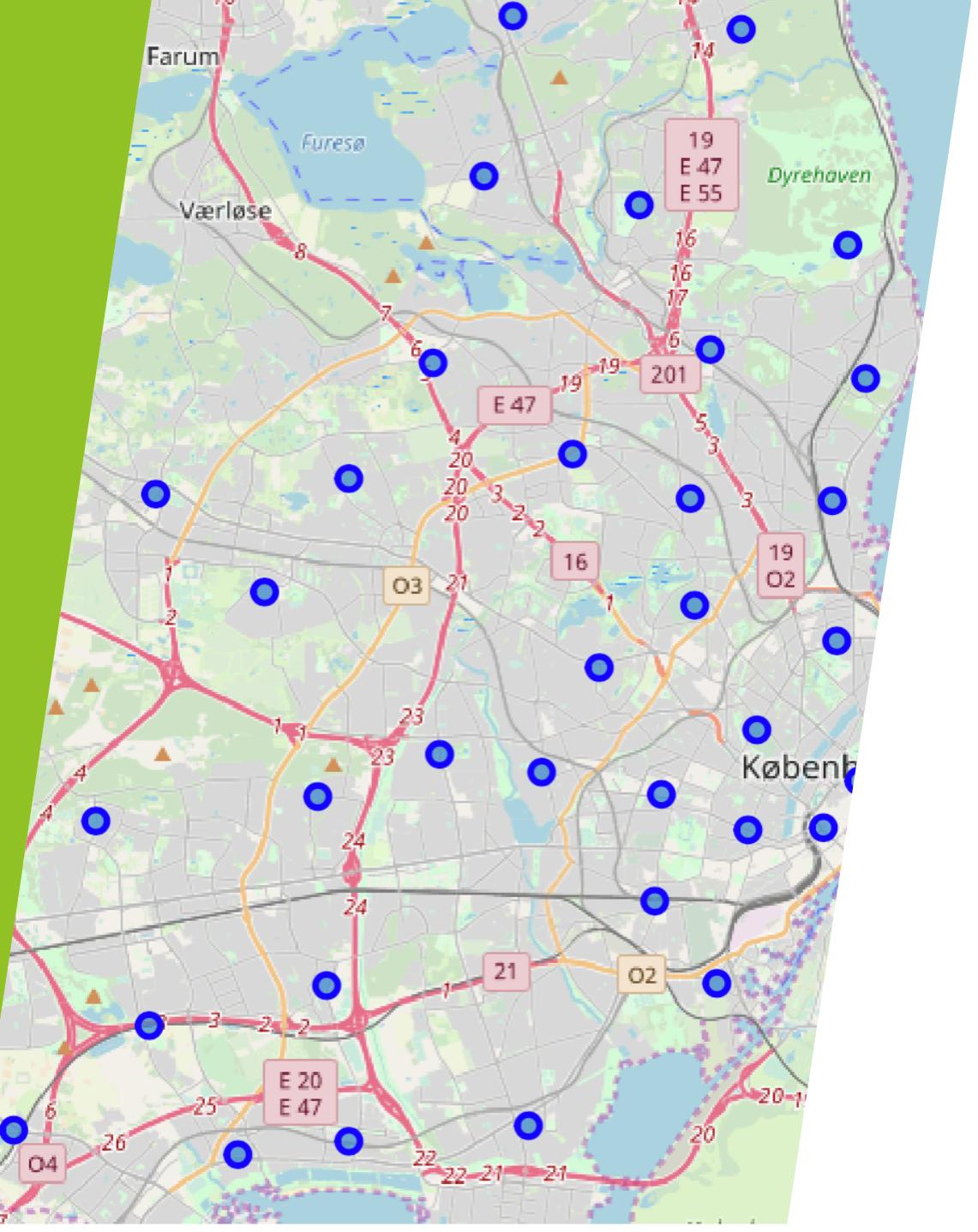


# A new restaurant in Copenhagen

In which region would a new restaurant most likely survive?

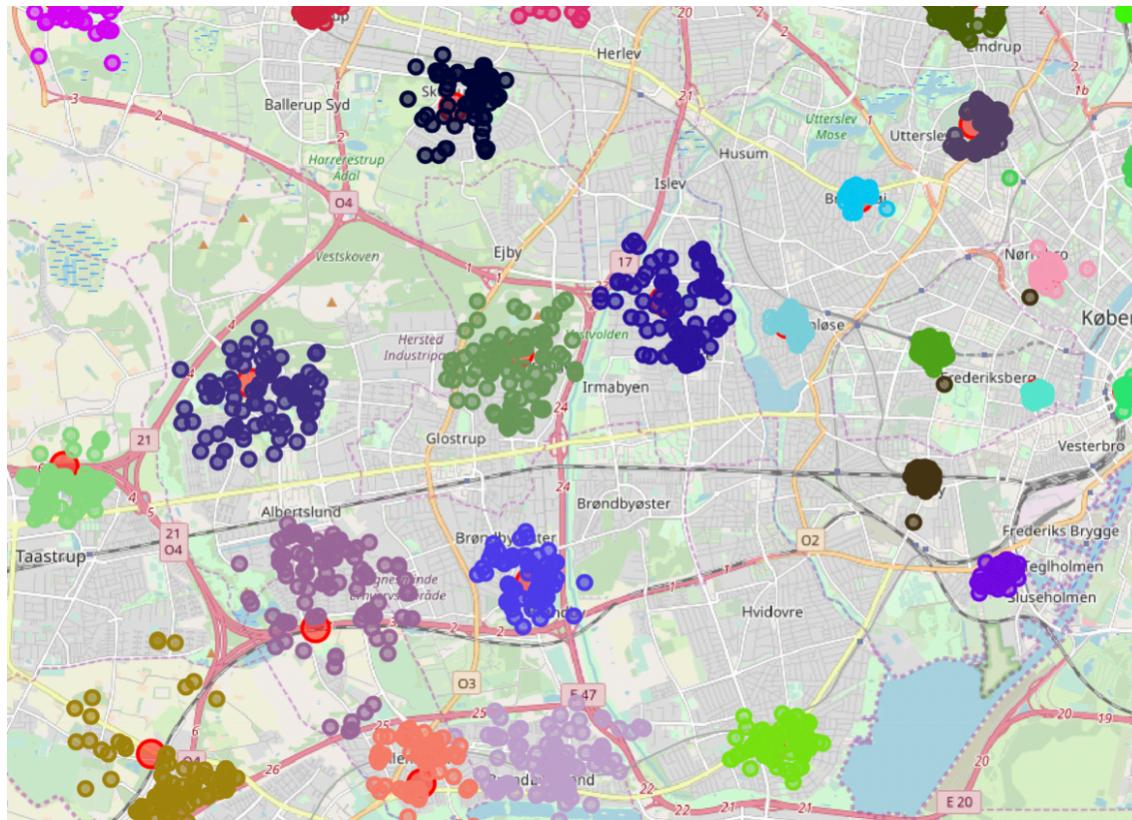


The region centers on map.  
How to find the best ones to  
open a new restaurant?

- ▶ Pre-process Data
- ▶ Determine Cluster number
- ▶ Build Clustering Model
- ▶ Refine Model
- ▶ Calculate Competition Level
- ▶ Calculate Determine Feature
- ▶ Analyze Result

# Venues Distribution

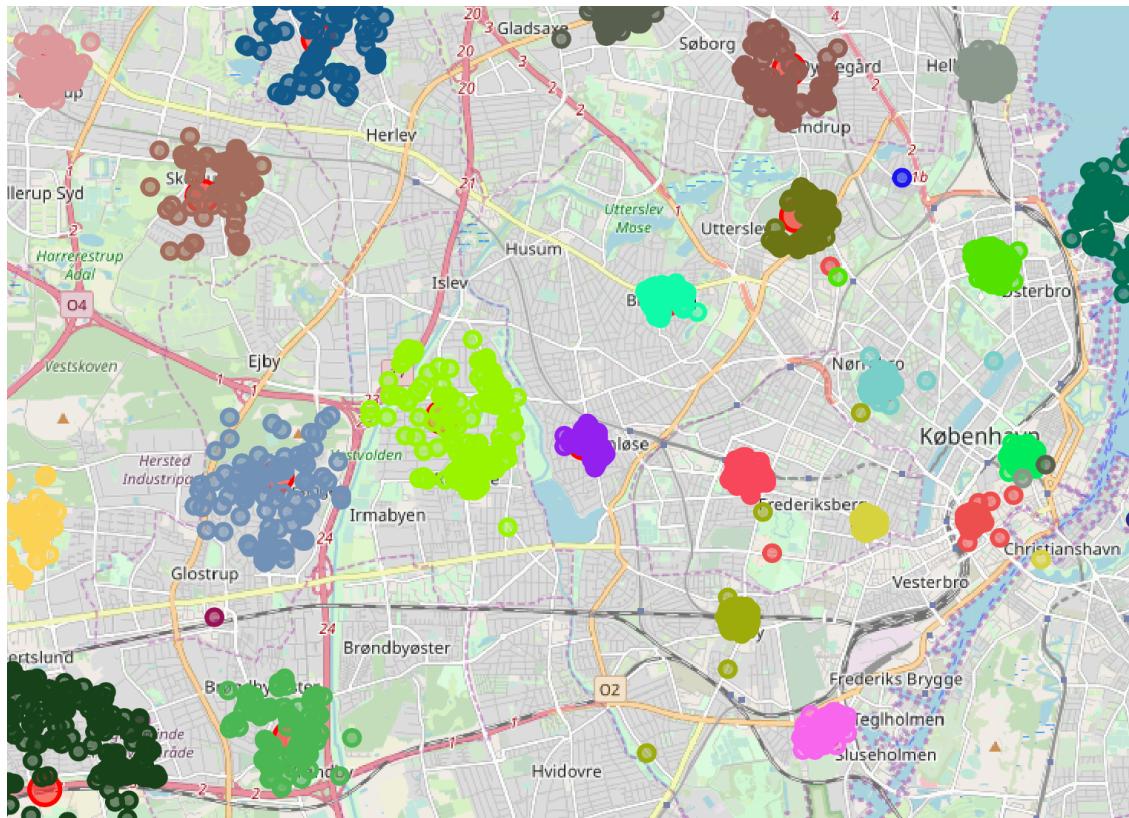
- limit 100
- Radius 1000m



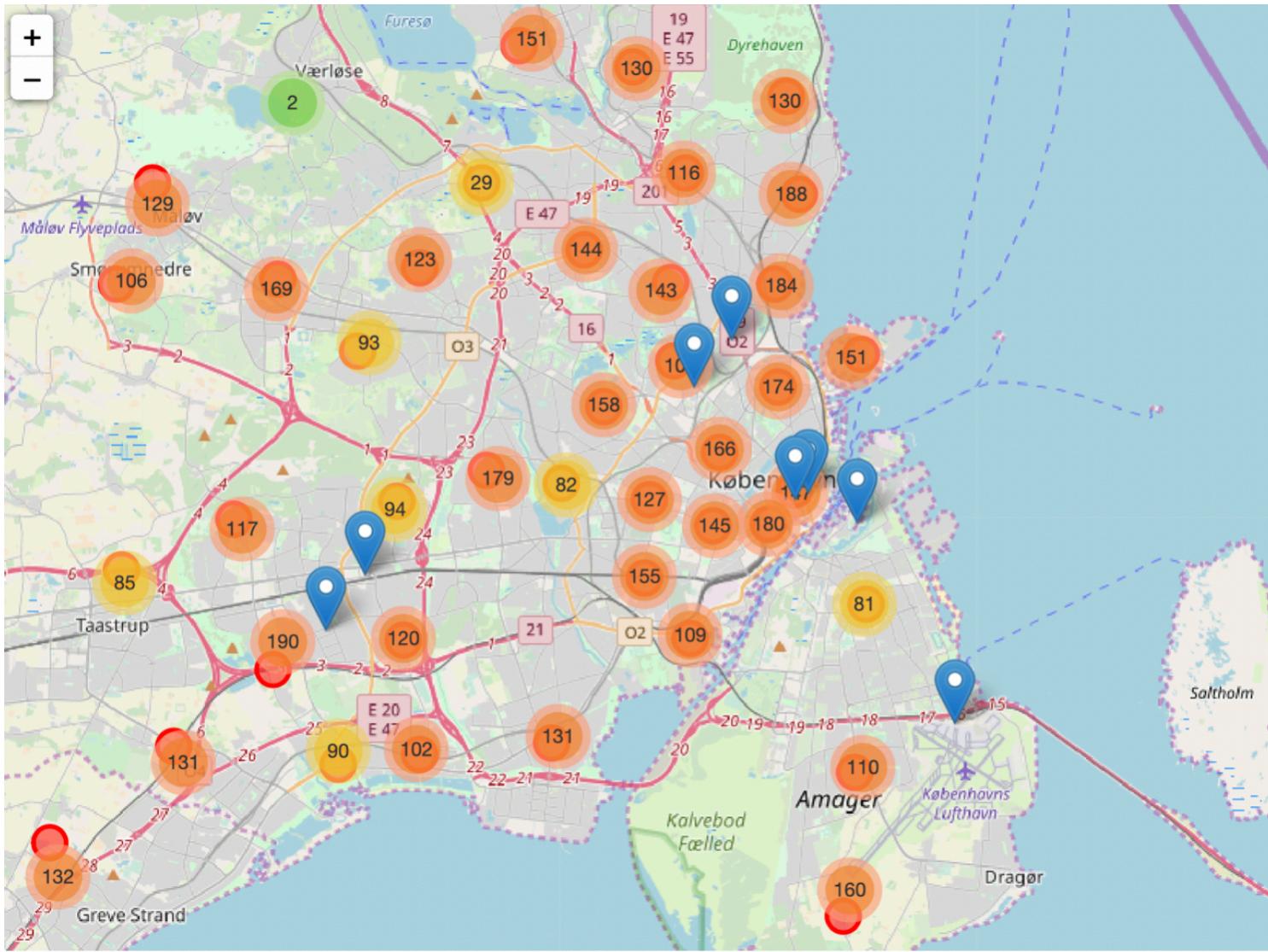
- ▶ Remote regions are more spread out, some region even overlapped to each other, while region close to city center are more centralized as the venues locate closer to each other .
- ▶ Need refine the data quality by increasing data point limit and decreasing the data point radius.

# Venues Distribution

- limit 200
- Radius 500m



- ▶ Compared to dataset with limit 100 and radius 1000m, this dataset distribution looks more centralized and less overlapped.
- ▶ Can be used for further clustering analysis

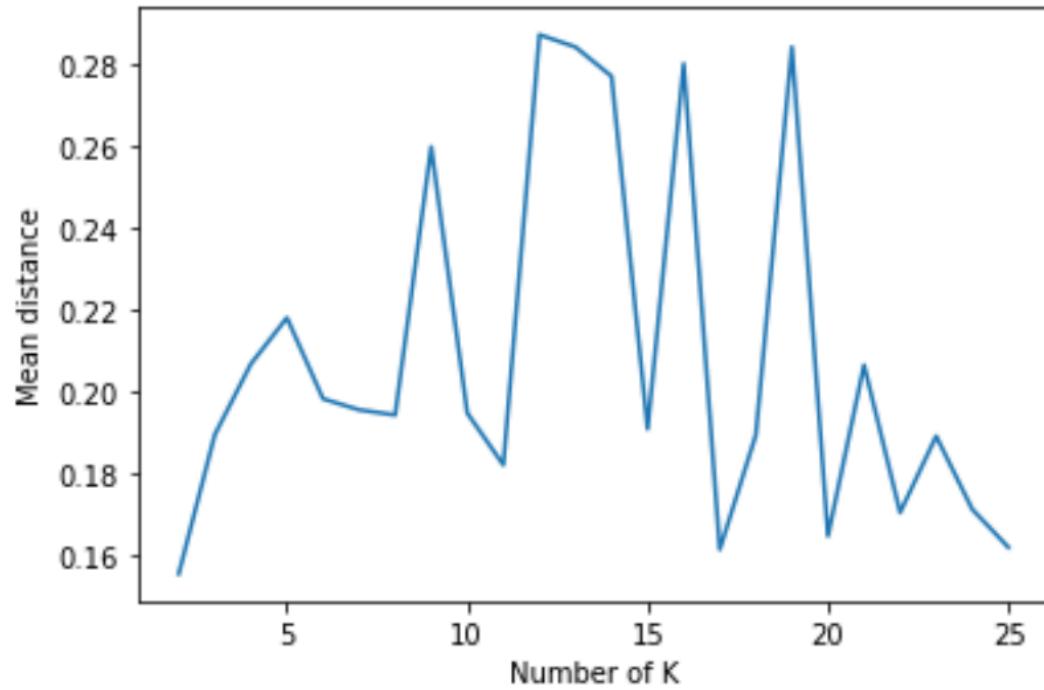


An intuitive presentation of venue numbers in each region.

# Top venue categories in each region

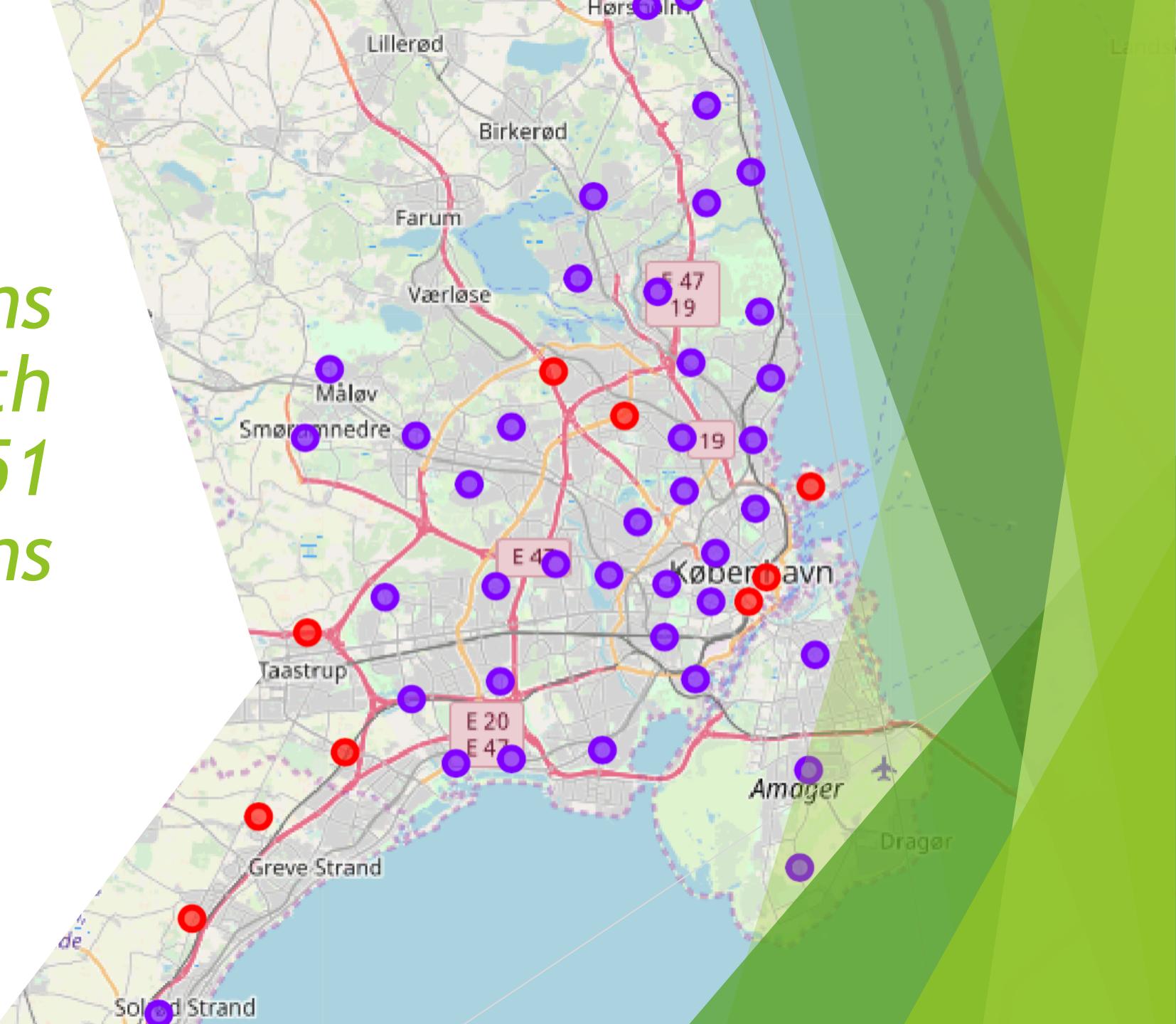
Region	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Copenhagen N	Residential Building (Apartment / Condo)	Food Truck	Doctor's Office	Bar	Bike Shop	Office	Thrift / Vintage Store	Bus Line	Salon / Barbershop	Garden
1 Copenhagen NV	Residential Building (Apartment / Condo)	Medical Center	Cemetery	Building	Café	Bus Line	Doctor's Office	Sushi Restaurant	Pizza Place	Grocery Store
2 Copenhagen S	Pizza Place	Salon / Barbershop	Coffee Shop	General Entertainment	Arts & Crafts Store	Sushi Restaurant	Flea Market	Pet Store	Café	Rock Club
3 Copenhagen SV	Office	Residential Building (Apartment / Condo)	Convenience Store	Plaza	Other Great Outdoors	Café	Professional & Other Places	Building	Bank	Pizza Place
4 Copenhagen V	Office	Bar	Music Venue	Building	Convenience Store	Art Gallery	City Hall	Coffee Shop	Fast Food Restaurant	Courthouse

# *mean distance with different K value*



- ▶ The mean distance has the smallest value when the K value is 2, selected from a range between 2 and 25 .
- ▶ Making 2 clusters for the 51 regions would lead to a better clustering result

*K-means  
clustering with  
2 clusters for 51  
regions*



*Top 5 regions best for opening a new restaurant :  
Vallensbæk, Vedbæk, Rødovre, Dragør and  
Copenhagen Ø.*

Area	Postal Code	Venue_count	restaurant_count	restaurant_rate	max_rest_rate	New_Rest
Vallensbæk	2625	190	6.0	0.031579	0.184615	29
Vedbæk	2950	190	7.0	0.036842	0.184615	28
Rødovre	2610	179	6.0	0.033520	0.184615	27
Dragør	2791	161	4.0	0.024845	0.184615	25
Copenhagen Ø	2100	174	8.0	0.045977	0.184615	24
Holte	2840	159	6.0	0.037736	0.184615	23

►  $restaurant\_rate =$   
 $restaurant\_count / Venue\_count$

►  $max\_rest\_rate =$   
 $\max(restaurant\_rate)$

►  $New\_Rest = Venue\_count *$   
 $max\_rest\_rate -$   
 $restaurant\_count$

# Resources and Tools

- ▶ Wikipedia: Post codes
- ▶ Google map geocoding API: Coordinate info
- ▶ FourSquare API: Venues information
- ▶ Folium: map creation
- ▶ SKlearn: machine learning python package
- ▶ Pandas: Data frame handle package