

# Coursera Capstone Report

## Accommodation in Copenhagen, Denmark - Potential Areas for new Hotel

### Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
- [Results](#)
- [Conclusion](#)

## Introduction

As part of the Applied Data Science Capstone by IBM Coursera course, this data science project will focus on using data in order to **identify the optimal location for a new tourist accommodation service (hotel) in Copenhagen, Denmark**. To find the optimal location, the project will look into the following parameters:

- **placement** (i.e. low competitive proximity and short distance to city center)
- **strategy** (i.e. pricing, unique selling point etc)

As such, this report is heavily relevant for aspiring new start-ups in the accommodation service in the capital of Denmark.

The background for this project, is that I myself live in Copenhagen, where peer-to-peer accommodation services (primarily AirBnB) have become an increasingly popular way for tourists to visit the city. I use AirBnB myself for hosting tourists, but find that the market for accommodations are heavily skewed in the city, with areas with low or no accommodation offers. Therefore, it seems likely that there exists and unexploited potential here for new business to grow in a city, which is currently heavily renewing several neighbourhoods.

# Data

For this project, the following dataa sources will be used:

- **Foursquare** (API)
  - This project will use the search endpoint of the Foursquare data api to extract **data on the professional accomodation service providers in Copenhagen** (hotels, inns, hostels, bed & breakfasts etc.). With this data, it should be possible to identify the competitive situation in different areas of the city (proximity, distance to city center).
- **AirBnB** (*can be accessed through this [link \(http://insideairbnb.com/get-the-data.html\)](http://insideairbnb.com/get-the-data.html)*)
  - This project uses the public available AirBnB data (insideairbnb). Specifically, this project will use the **listings.csv.gz**, which is a dataset of listings in Copenhagen with data on the hosts' ratings, location and apartment/home characteristics. With this data, it is possible to identify areas with relative more accomodations (popular areas) as well as areas with competitive advantages in terms of the different ratings (placement, value for money etc.) the accomodations have received.
  - From the AirBnB page it is also possible to download a **Neighborhoods** geojson file, which contains info on the neighborhood data for Copenhagen. This is used in order to map the different data analysis plots on a geographical map to better visualize the differences between the different neighborhoods.

# Methodology

To reach the goal of this report, this analysis will apply two methodological approaches:

1. The analysis will start of with **an exploratory analysis** of the differences between the different neighborhoods in Copenhagen. With this explorartory analysis, the neighbourhod with the greatest potential for a new professional accomodation service will be selected:
2. Then a **unsupervised machine learning model (k-means)** will be used to identify clusters of listings in a specific neighbourhod in order to identify the optimal location for a new accomodation service (hotel) in Copehagen.

# Analysis

First, the data will be presented before the explorative analysis is applied. The two datasets are called `df_fsq` (Foursquare API Data, 138 professional accommodation services in Copenhagen) and `df_abnb` (AirBnB Listing Data, 28195 AirBnB listing in Copenhagen). The first five rows of the two dataframes are presented below:

```
In [10]: df_abnb.shape, df_fsq.shape
```

```
Out[10]: ((28195, 106), (138, 15))
```

```
In [25]: # Foursquare Dataframe
```

```
Out[25]:
```

	Unnamed: 0		id	location.address	location.cc	location.city	location.country	location.distance	location.lat	location.lng	lc
0	0	4adcdaf9f964a520ba5c21e3		Rådhuspladsen 57	DK	København	Danmark	85	55.676001	12.569647	
1	1	4c1ffeb78b3aa593af5b9e5f		Løngangstræde 27	DK	København	Danmark	297	55.675367	12.572863	
2	2	5bf19af6b9b37b00398d7d7b		H. C. Andersens Blvd. 12	DK	København	Danmark	134	55.676677	12.566421	
3	3	5b28df571ffe97002cb5a0ae		Helgolandsgade 3	DK	København	Danmark	591	55.672745	12.561002	
4	4	5bae844e73fe25002ca31c2e		NaN	DK	NaN	Danmark	2	55.676098	12.568337	

In [24]: `# AirBnB Listings Dataframe`

Out[24]:

	id	listing_url	scrape_id	last_scraped	name	summary	space	description	experier
0	6983	<a href="https://www.airbnb.com/rooms/6983">https://www.airbnb.com/rooms/6983</a>	20200322041115	2020-03-23	Copenhagen 'N Livin'	Lovely apartment located in the hip Nørrebro a...	Beautiful and cosy apartment conveniently loca...	Lovely apartment located in the hip Nørrebro a...	
1	26057	<a href="https://www.airbnb.com/rooms/26057">https://www.airbnb.com/rooms/26057</a>	20200322041115	2020-03-23	Lovely house - most attractive area	Our lovely house in the center of the city is ...	Totally charming old 150 m2 stone house from y...	Our lovely house in the center of the city is ...	
2	26473	<a href="https://www.airbnb.com/rooms/26473">https://www.airbnb.com/rooms/26473</a>	20200322041115	2020-03-23	City Centre Townhouse Sleeps 1-12 persons	TOWN HOUSE ON KNABROSTRÆDE - located in the ab...	The house is a total of 240sqm divided on 4 fl...	TOWN HOUSE ON KNABROSTRÆDE - located in the ab...	
3	29118	<a href="https://www.airbnb.com/rooms/29118">https://www.airbnb.com/rooms/29118</a>	20200322041115	2020-03-22	Best Location in Cool Istedgade	NaN	The apartment is situated in the middle of the...	The apartment is situated in the middle of the...	
4	29618	<a href="https://www.airbnb.com/rooms/29618">https://www.airbnb.com/rooms/29618</a>	20200322041115	2020-03-23	Artsy and familyfriendly home in lovely Copenh...	Artsy, bright and spacious flat, close to the ...	It's a three bedroom apartment with a spacious...	Artsy, bright and spacious flat, close to the ...	

## Results

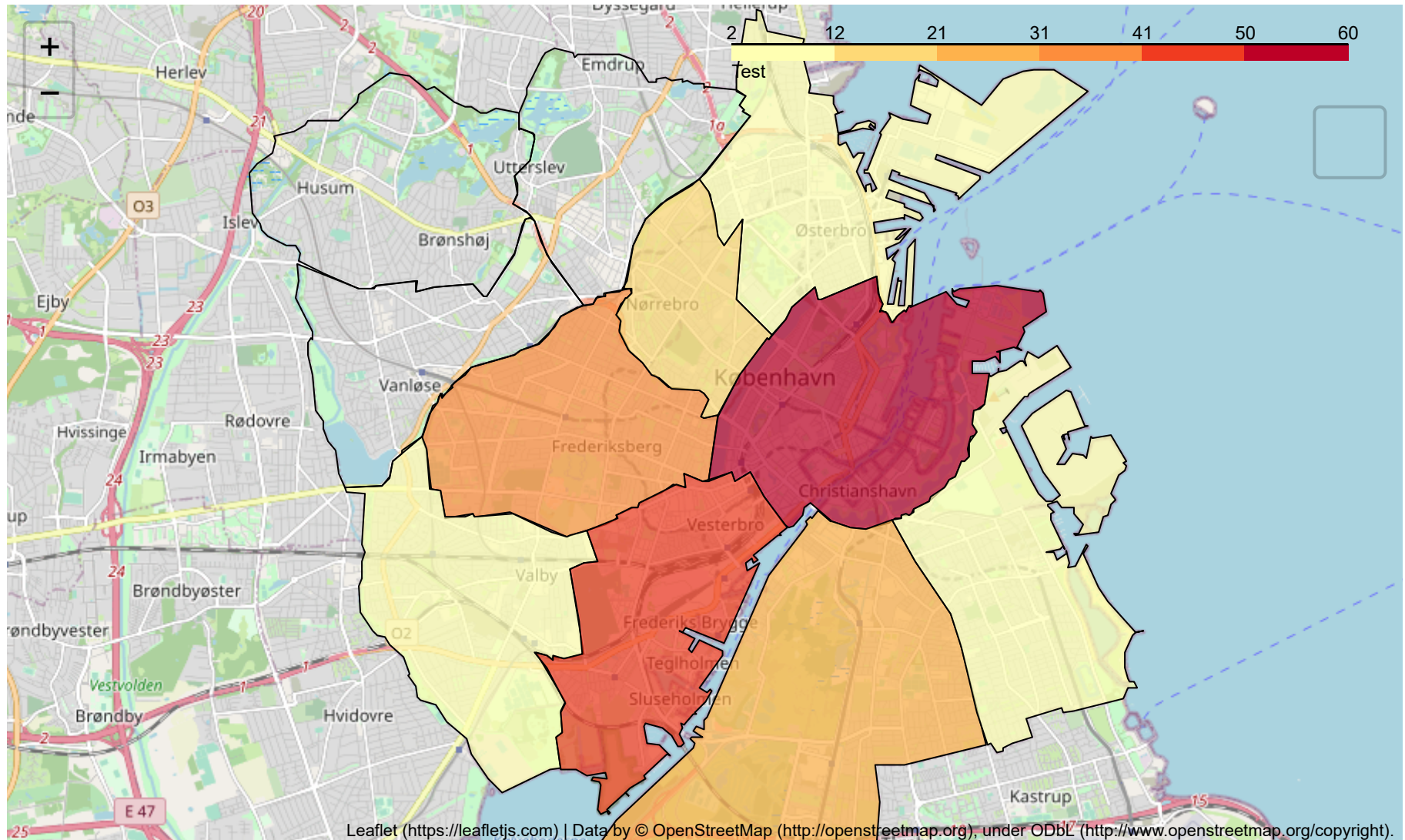
First, we have a look at the Foursquare data. In the map-chart below, the number of professional accommodation services (hotels, hostels, inns etc) per neighbourhood in Copenhagen are visualized. It is clear that the professional accommodations services are centered in the inner-city neighbourhood. Vesterbro and Frederiksberg follows along, while areas like Østerbro has a significantly lower competitive density - i.e. this neighbourhood definitely has potential for a new professional accommodation service when we look solely on the competitive situation

In [516]: *# Map visualization of commercial accomodate services that are located in each Copenhagen neighbourhood*

C:\Users\45521\AppData\Local\Continuum\anaconda3\lib\site-packages\folium\folium.py:415: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.

FutureWarning

Out[516]:

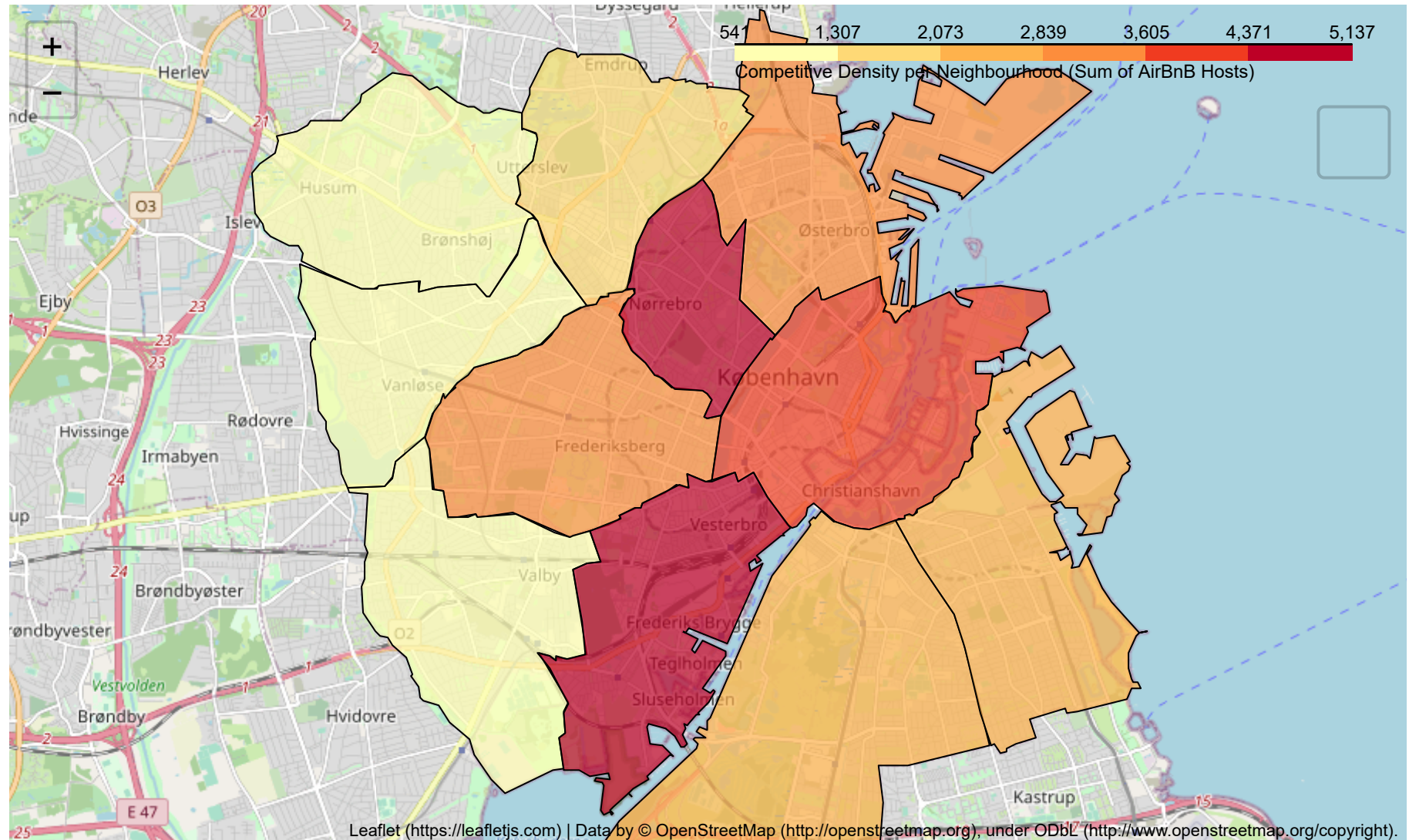


If do the same plot of the volume of AirBnB in the different Copenhagen neighbourhoods (see map below), there is a clear pattern here. The inner city, Vesterbro and Nørrebro neighbourhoods both have a high concentration of professional accommodation services (hotels, hostels, inns etc.) and of AirBnB listings. In contrast, Østerbro has a very low competitive density of professional accommodation services, but have relative more AirBnB-listings. Again this neighbourhood shows great potential, as it is well-used for AirBnB accommodations, but with a low density of professional competitions.



In [183]: # Map visualization of AirBnB listings in each Copenhagen neighbourhood

Out[183]:



Furthermore, Østerbro also have the lowest availability (within the past 365 days) amongst the Copenhagen neighbourhoods (see table below). This indicates that the even though Østerbro actually has an above-average volume of AirBnB listings, but with the lowest availability among these listings. As a result, one could suspect that increasing the availability of accommodation services in Østerbro will prove lucrative as the demand seems to exist.



In [511]: *# Table for the average days of AirBnB listing availability per Copenhagen neighbourhood*

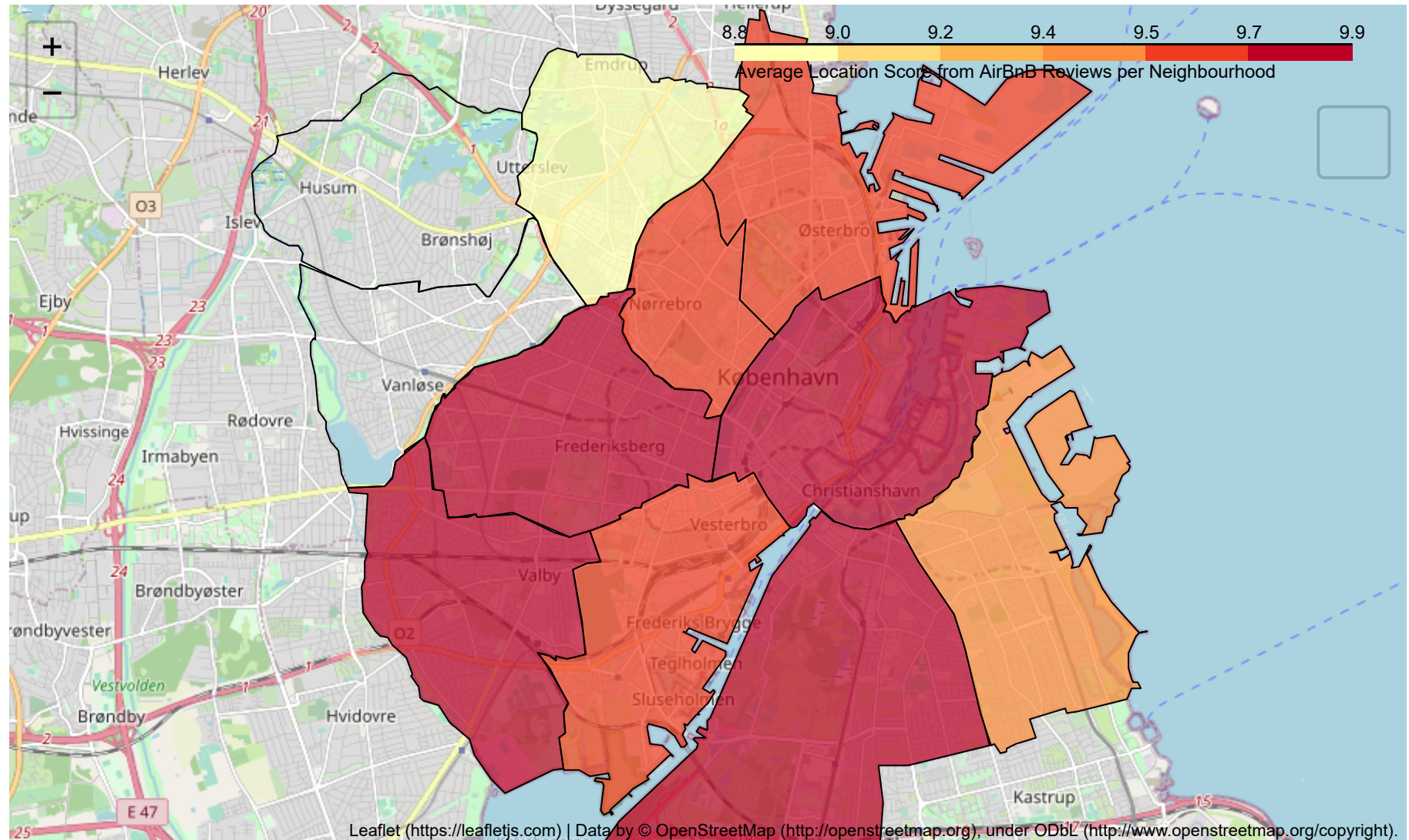
Out[511]:

	neighbourhood_cleansed	availability_365
6	Valby	132.00
2	Bispebjerg	124.50
4	Indre By	117.14
0	Amager Vest	116.87
3	Frederiksberg	109.54
1	Amager st	98.44
7	Vesterbro-Kongens Enghave	80.16
5	Nrrebro	67.26
8	sterbro	66.67

From the map below, we can see that Østerbro does not have the best score on the AirBnB-review score regarding location of the listing. This neighbourhood does not have a particular poor location, but it is outcompeted by the Inner City, Frederiksberg, Valby and Amager Vest.

In [184]: # A visualization map for the AirBnB-review score for Listing Location per Copenhagen Neighbourhood

Out[184]:



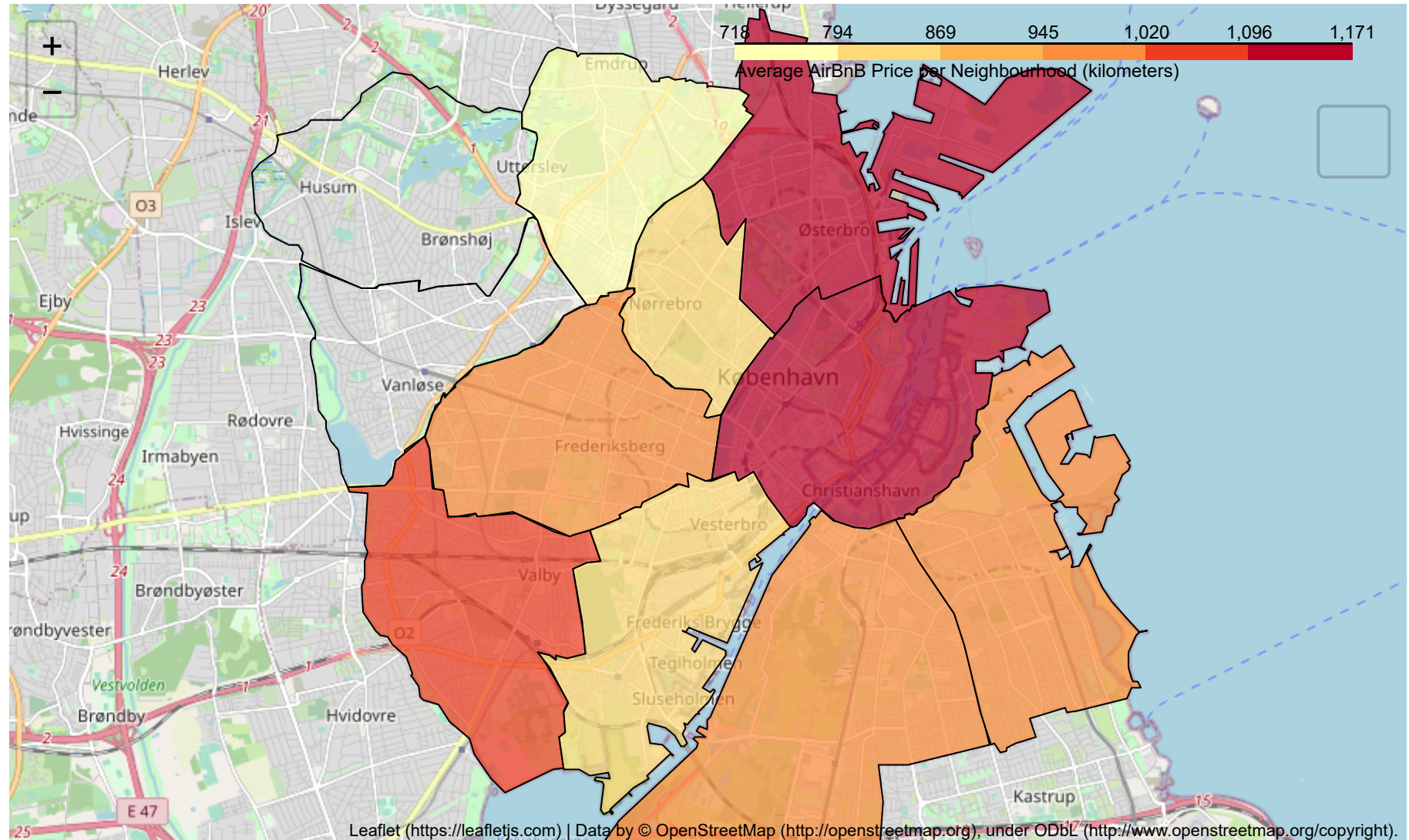
At last, we have a look at the average price per AirBnB-listing in the different Copenhagen neighbourhoods. Again, Østerbro proves lucrative as it has the second highest average price per listing - only behind the Inner City-neighbourhood. When we combine this with "value for money"-score (see table below the map) for the different neighbourhoods, Østerbro is further supported as a lucrative accommodation area within Copenhagen. The tourists are willing to pay a high price for accommodations in this area - and they find that the accommodations in Østerbro is worth the investment.

```
In [538]: # Map viualization of the average AirBnB listing price per Copenhagen neighbourhood
```

```
C:\Users\45521\AppData\Local\Continuum\anaconda3\lib\site-packages\folium\folium.py:415: FutureWarning: The choropleth
  method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example no
  tebook 'GeoJSON_and_choropleth' for how to do this.
```

FutureWarning

Out[538]:



In [510]: *# Table over the "value for money"-review score for the AirBnB Listings per Copenhagen neighbourhood*

Out[510]:

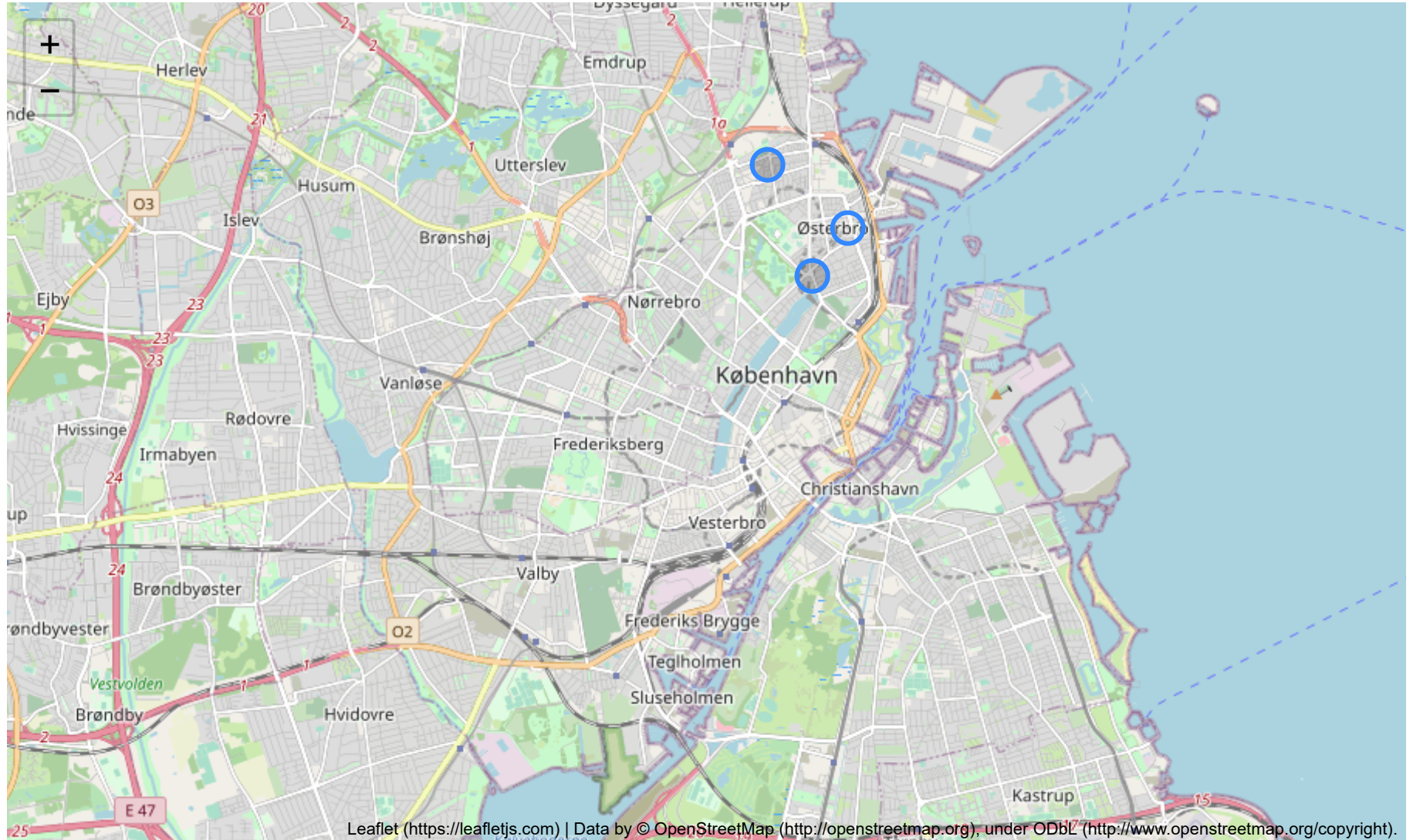
	neighbourhood_cleansed	review_scores_value
7	Vesterbro-Kongens Enghave	9.59
8	Østerbro	9.58
1	Amager st	9.56
3	Frederiksberg	9.54
0	Amager Vest	9.53
5	Nrrebro	9.40
6	Valby	9.29
4	Indre By	9.26
2	Bispebjerg	9.00

Based on the explorative analysis above, we have identified Østerbro as a lucrative neighbourhood for a potential new professional accommodation service in Copenhagen. In the following, we will use a unsupervised machine learning clustering analysis (K-means) to identify hidden groupings of the AirBnB-listings in this neighbourhood. Based on a K-means clustering, it is possible to identify three clusters based on the coordinates of the listings in Østerbro. The centers of these three clusters are visualized in the map below.



In [537]: # Map visualization of the centers of the three Østerbro clusters

Out[537]:



Of the three clusters, the one furthest to the south is the most lucrative in terms of the potential for a new professional accommodation service. This cluster has the shortest distance to the city center, the highest average listing price and review-scores (especially on the "good location"-reviewscore. Furthermore, this cluster has a relative lower average reviews per month and availability among the its AirBnB-listings. This indicates a potential for a new accommodation service here to meet the demand. Furthermore, the placement furthest to the south in Østerbro, is close to some of the biggest tourist attractions and locally-beloved areas such as the little mermaid statue, the lakes and Fælledparken. Thus, the Østerbro neighbourhood is highly recommended for new accommodation service providers in the Danish capital area. Focus should therefore be on the location attractions with a slightly below-average price-strategy to make the new accommodation service lucrative in this area. Furthermore, the value for money is definitely a unique selling point that will help the new service to thrive in this area

In [23]: *# Table over the 3 clusters in Østerbro (Labels).*  
*# Label 0 is the cluster with the center furthest to east*  
*# Label 1 is the cluster with the center furthest to north*  
*# Label 2 is the cluster with the center furthest to south*

Out[23]:

	Distance_to_Citycenter	price_num	reviews_per_month	availability_365	review_scores_rating	review_scores_accuracy	review_scores_clean
<b>Labels</b>							
<b>0</b>	3.38	828.84	0.53	49.66	95.48		9.74
<b>1</b>	4.07	840.22	0.63	47.24	94.81		9.70
<b>2</b>	2.58	963.99	0.53	47.83	95.93		9.82

## Conclusion



In this report, we have used data to investigate and identify the optimal location for a new professional tourist accommodation service (fx hotel, hostel, inn) in Copenhagen, Denmark. To find the optimal location, this project has leveraged data from Foursquare (professional accommodation services like hotels, hostels, inns etc.) and AirBnB listings data. With these data sources, we were able to identify Østerbro as a highly relevant and lucrative neighbourhood for aspiring new start-ups in the professional accommodation service in the capital of Denmark.

Østerbro has the a relative low competitive density of other professional accommodation services. However, the volume of AirBnB listings are relative higher indicating a demand for accommodation options in the area. Furthermore, Østerbro has the second highest average listing price on AirBnB as well as the second highest score on the "value for money"-review metric. This indicates a room in the market for a slightly below average price option which will provide the tourists with true value for money.

Within Østerbro, this project has identified a specific lucrative area in the south part of the neighbourhood. This area has the highest review-scores and average listing price, but with relative lower availability and accommodations volumes - i.e. this area is particularly in need of a new accommodation service that can meet this demand.