Locating where to start a business in Copenhagen, Denmark

Nick Lagrilliere

February 27, 2021

1. Introduction

1.1 Background

Copenhagen, Denmark, would appear to be an interesting location to start a business given the increased worldwide focus on sustainability and the city's aspiration to become the first carbon neutral capital in the world by 2025. Also it's consistently high rankings in quality of living surveys (Figure 1) and Denmark's reputation as the happiest country in the world are attractive features [(1)]. Finally, it also has a high GDP per capita of $56.300 and is ranked as the 7th best country in the world for doing business [(2)].

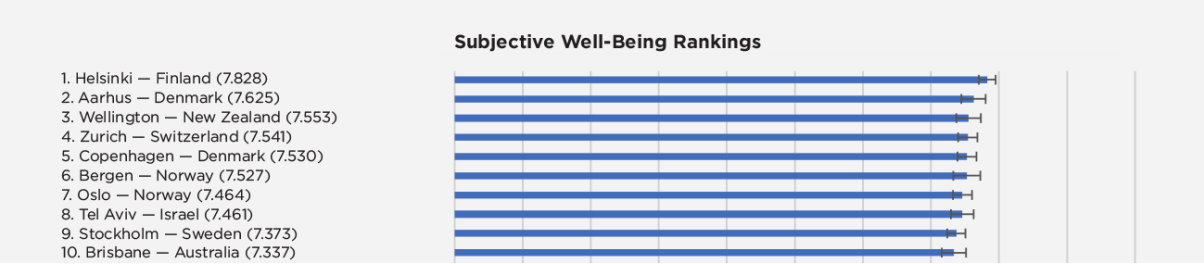

Figure 1. Global Ranking of Cities — World Happiness Report

Investigating these issues quantitatively and pulling in additional data are important steps to confirm whether Copenhagen is indeed an attractive place to start a business and, more specifically, where in the city would be a good location to start a particular kind of business.

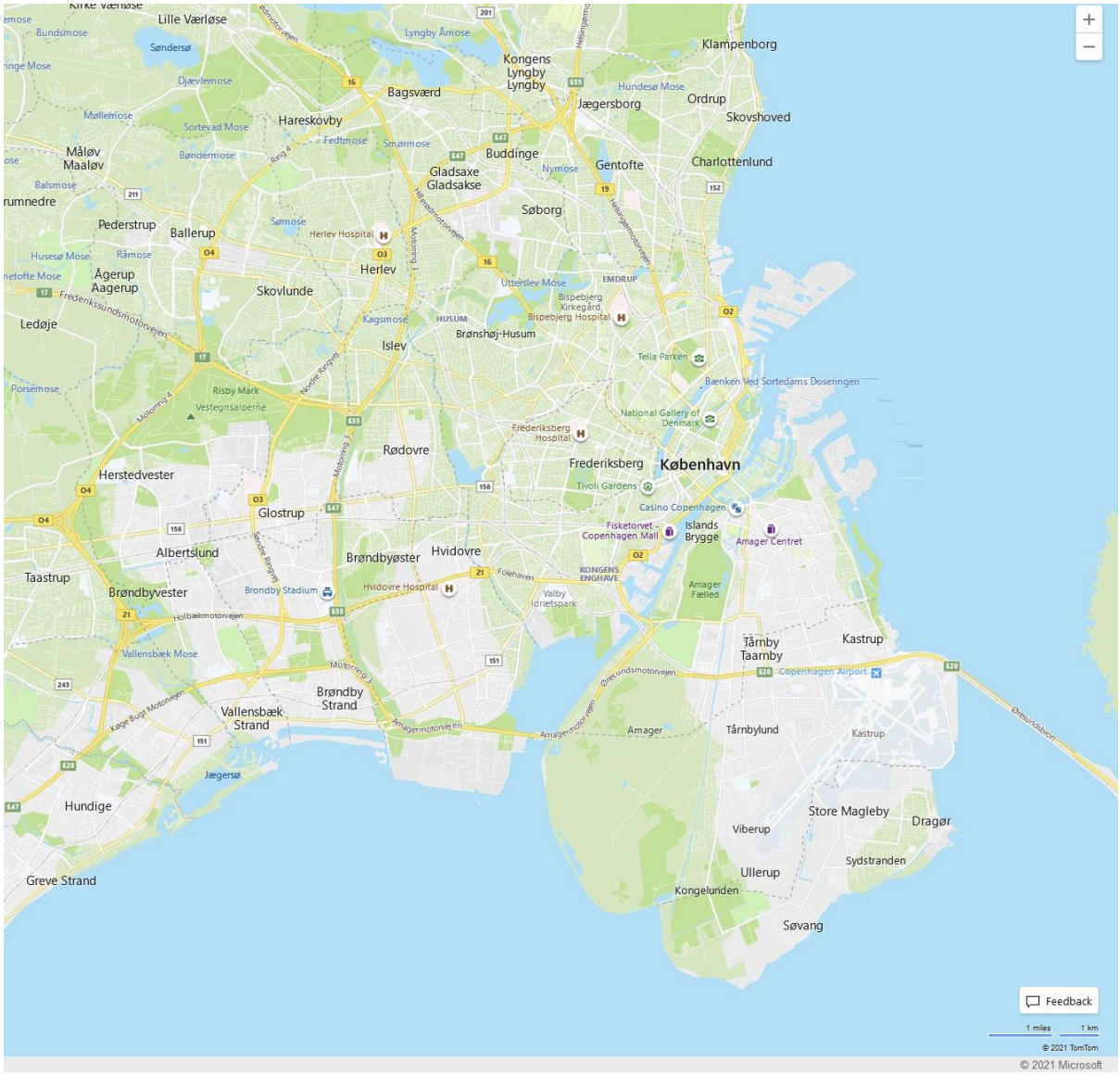

Figure 2. Map of the Copenhagen metro area.

1.2 Problem

A number of key questions need to be answered for a future business owner to be feel comfortable investing, related to future profitability of the enterprise:

- What are the demographic trends in Copenhagen (e.g. is the population expected to grow)?
- What are the economic trends in Copenhagen (e.g. is economic growth likely)?

Also investigating what particular geographical area would be of interest and what type of business to invest in would be important questions to answer.

- What areas of the city would be attractive for investing in a business (e.g. proximity to major transport hubs)?
- What types of business are interesting investment opportunities (e.g. there are few coffee shops in area)?

1.3 Interest

The target audience of the study would to clients who are interested in starting a new business venture in Copenhagen, Denmark and the methodology would help these stakeholders answer whether they would like to invest in the city, and more specifically in what geographical area they would best do so.

2. Data acquisition and wrangling

2.1 Data

A number of data sources were used to investigate the attractiveness of Copenhagen, Denmark as a place to invest in a new business venture.

The website of **Statistics Denmark** (3) has a data portal that features a wide range of data, including :

- Population by municipality (4)
- Gross Domestic Product (GDP) by region (5)
  - GDP per capita by region(6)
  - Population projections by municipality (7)

To investigate what particular geographical area would be of interest for starting a new business the **Open Data DK** [8] data portal features extensive datasets covering Copenhagen:

- Station location maps as .geojson files

An overview of the largest and smallest stations in Denmark by number of daily passengers was available on the Jyllands-Posten newspaper website and locally saved as a .csv file spreadheets [9]

The **Foursquare API** [10] location data were leverage to explore and compare station areas:

- search for specific type of venues around a location
- retrieve location data

2.2 Data wrangling

The data from Statistics Denmark were available as .cvs files for download and data cleaning was limited to remove blank columns, modifying the headers and changing indices to make the pandas dataframes more user friendly.

Importing data from Open Data DK required flattening of the nested geojson files in order to load them into the dataframe. Columns with irrelevant features were dropped and the additional characters of the Danish alphabet corrected. The geographical coordinates required further wrangling in order to be converted to latitudes and longitudes. They were first cast to string formats in order to split the original column into two separate columns, after which the original column was removed. Unwanted characters were removed and the data type converted to float in order to make the coordinate data useable for the geographical applications.

The coordinates to generate the map of Copenhagen were derived from the GeoPy Nomatim library and the station location data were then plotted using the Folium library.

Station passenger data were available as a .csv download and were only modified by adding the relevant headers.

Finally, the Foursquare venue data were generated by calling the Foursquare API and subsequently manipulated to identify the top venue types per station and the most common venues per station.

3. Exploratory Data Analysis

Explore basic insights regarding the macroeconomic and population context in Copenhagen.

3.1 Gross Domestic Product (GDP)

Gross Domestic Product in Copenhagen has been steadily rising since the start of available data, almost tripling from around 150 billion danish kroner in 1993 to 450 billion danish kroner in 2019. Even te 2008 financial crisis left only a minor imprint. This indicates that the local economy is very robust and steadily growing.

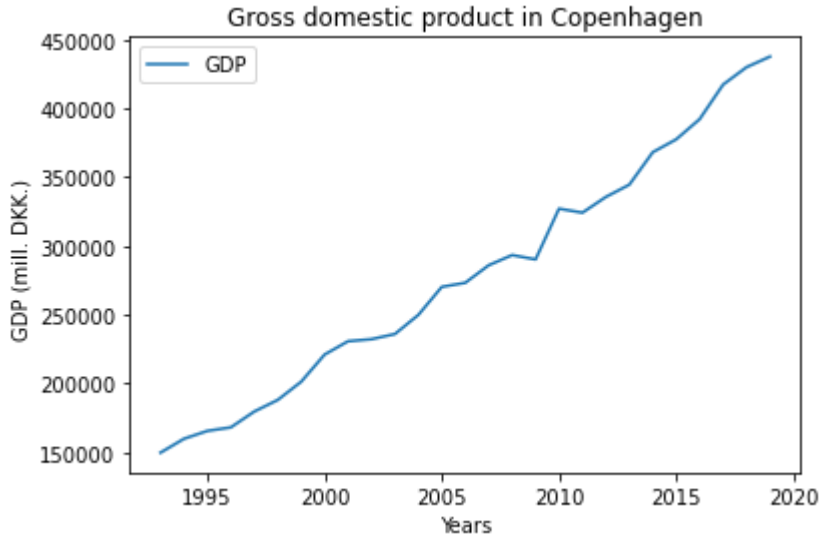

Figure 3. Line plot of Gross Domestic Product (GDP) in Copenhagen

3.2 Population

Population in Copenhagen has been steadily increasing over the period for which data is available from 509.861 in 2008 to 638.117 in 2021. Growth seems to deccelarate towards 2020, likely driven by increasing property prices which are pushing lower income households out to the suburbs. Also the COVID-19 pandemic could have had impact on immigration to to city.

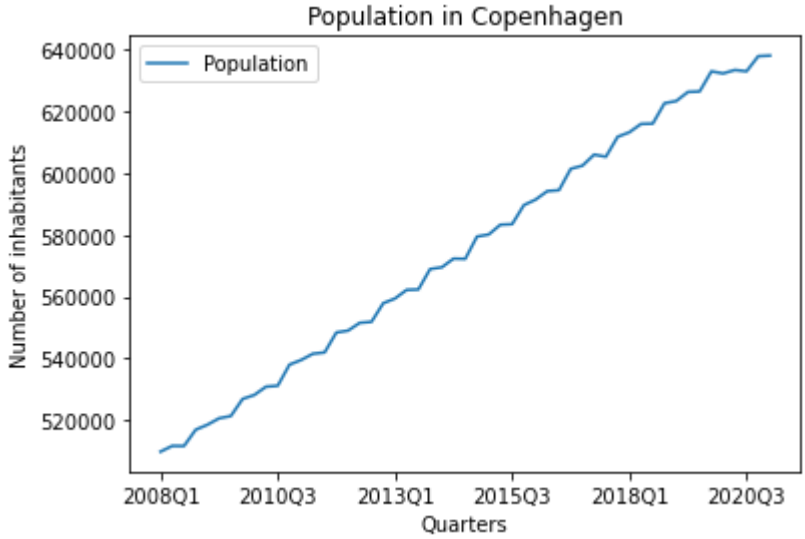

Figure 4. Line plot of population in Copenhagen

3.3 GDP per capita

GDP per capita has been increasing steadily, from 247.000 danish kroner per capita in 1993 to 554.000 danish kroner in 2019. This near doubling indicates that economic growth has been outpacing population growth in the city which ought to be indicative of a robustly growing economy and increasing wealth for the population.

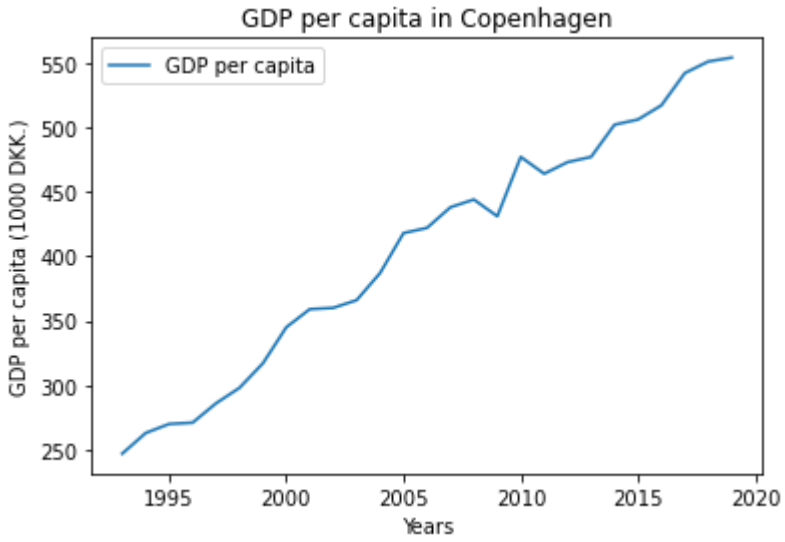

Figure 5. Line plot of GDP per capita in Copenhagen

3.5 Population projections

Population projections for the city of Copenhagen show the current growth trend continuing over the next decade with the population reaching around 880.000, before then slowing down somewhat towards 2045 when the population is forecast to reach 920.236.

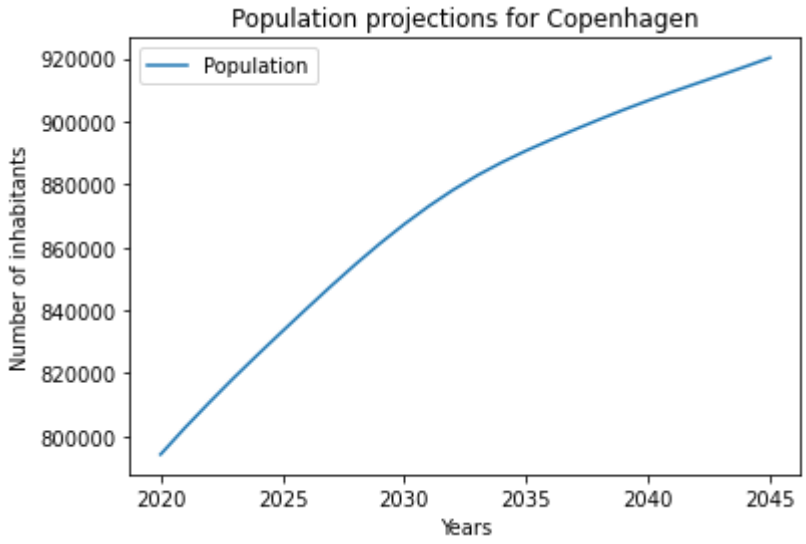

Figure 6. Line plot of population projections in Copenhagen

3.6 Geographical data of train stations

The geojson data from the Open Data DK data portal contain the coordinates, station type, station name and the name of the town in which the station is located.

There are two station types: the metro stations and the stations serving the suburban rail line called the S-train. Regional and intercity trains from outside the Copenhagen area would arrive at the same stations and therefore these were not included as separate stations.

The Copenhagen metro area is made up of several towns surrounding the city of Copenhagen and these are listed in the Town column for each station. The geographical coordinates are displayed in latitude and longitude in the EPSG 4326 coordinate reference system.

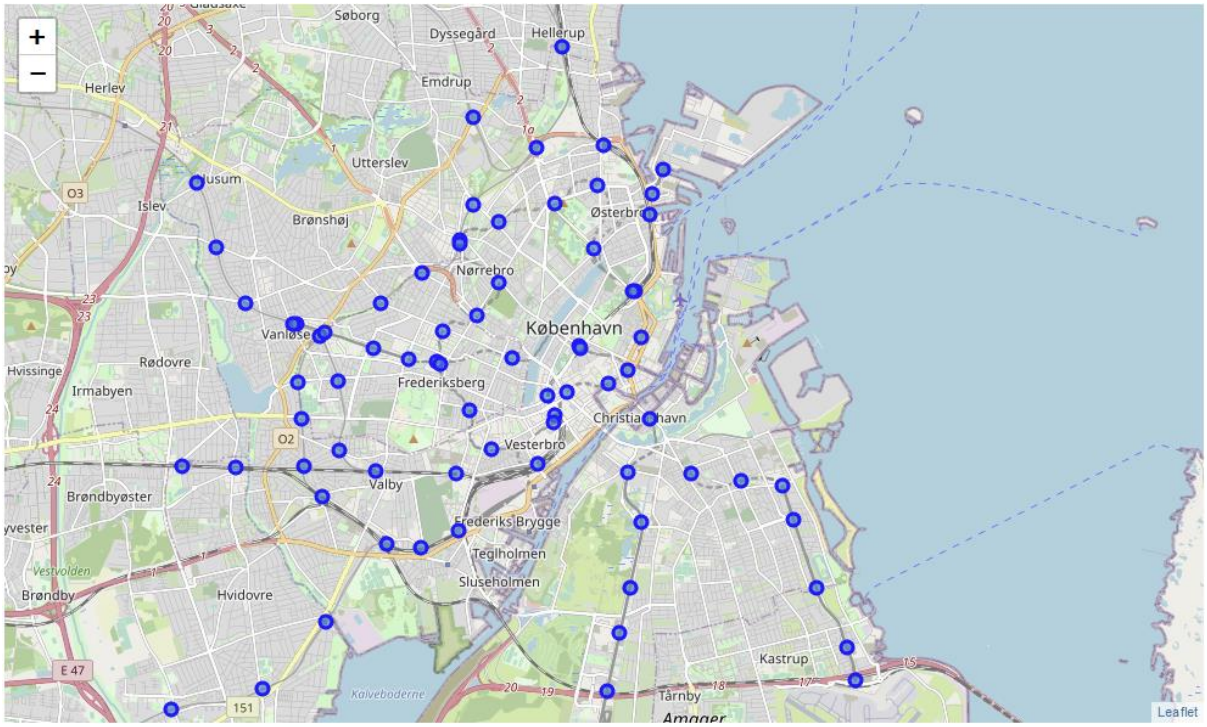

Figure 7. Map of the geographical station location in Copenhagen

3.7 Station passenger data

The passenger data were filtered down to the top 15 train stations by daily passenger numbers in the Copenhagen metro area. The clear outliers here are Nørreport station, Denmark's busiest with 107.800 daily passengers, and København H, Copenhagen's central station with 103.600 daily passengers. Both stations are regional transport hubs for metro, suburban S-train, regional and intercity trains and buses.

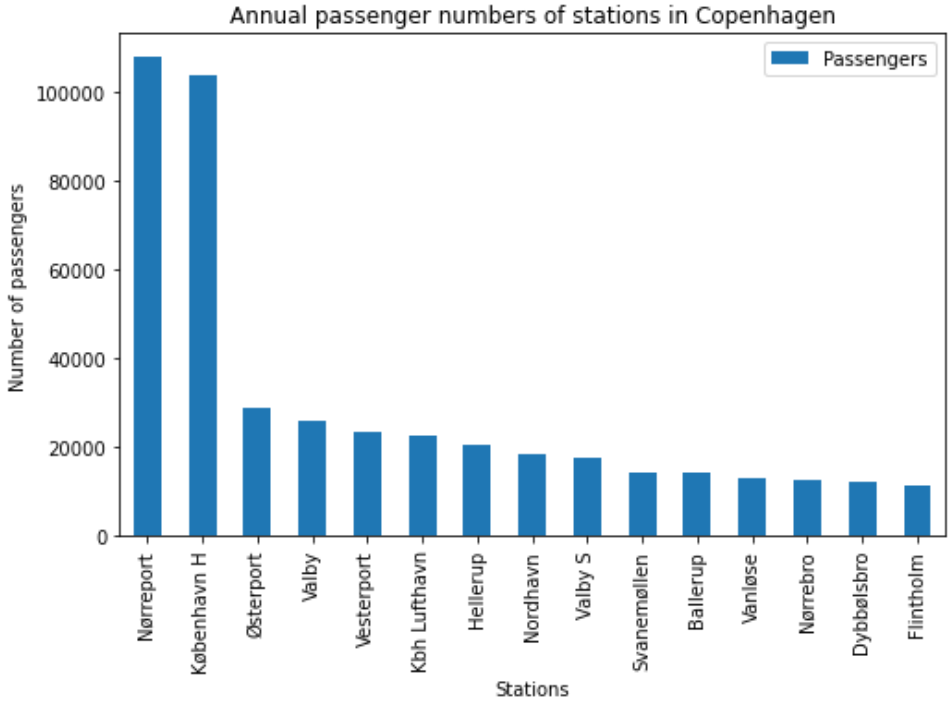

Figure 8. Bar plot of passenger numbers per train station.

3.7 Foursquare data

The Foursquare API was used to get data on venues in the proximity of the stations in Copenhagen and leveraged to explore and compare station areas. The number of venues in a 500 m radius around the train stations were identified and counted. These were then grouped by venue type and subsequently used to identify

the top 10 most common venues around a station. This allows us to identify whether a particular type of business is already common in an area or whether there could be an opportunity to start a business type that is currently not well represented.

Name	Cluster Labels	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
Kastrup	1	Pizza Place	Supermarket	Racetrack	Mediterranean Restaurant	Liquor Store	Scandinavian Restaurant	Department Store	Fish & Chips Shop	Fast Food Restaurant
Langgade	1	Pizza Place	Convenience Store	Coffee Shop	Burger Joint	Supermarket	Bar	Record Shop	Train	Train Station
Peter Bangs Vej	1	Soccer Field	Cheese Shop	Pizza Place	Grocery Store	Scandinavian Restaurant	Sporting Goods Shop	Supermarket	Gaming Cafe	Thai Restaurant
KB Hallen	0	Soccer Field	Train Station	Sporting Goods Shop	Bus Station	Restaurant	Lake	Athletics & Sports	Gym	Park
Vigerslev Allé	8	Indian Restaurant	Moving Target	Park	Karaoke Bar	Gym	Department Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant
...	...	...	...	...	...	...	...	...	...	...
Gammel Strand	0	Plaza	Cocktail Bar	French Restaurant	Coffee Shop	Jazz Club	Bar	Furniture / Home Store	Hotel	Restaurant
Marmorkirken	0	Scandinavian Restaurant	Café	Restaurant	Hotel	Bakery	Art Museum	Plaza	Breakfast Spot	Coffee Shop
Trianglen	1	Bakery	Coffee Shop	Café	Italian Restaurant	Ice Cream Shop	Soccer Stadium	Bar	Sushi Restaurant	Vietnamese Restaurant
Kongens Nytorv	0	Scandinavian Restaurant	Restaurant	Coffee Shop	French Restaurant	Clothing Store	Steakhouse	Bar	Furniture / Home Store	Wine Bar
Orientkaj	3	Grocery Store	Scandinavian Restaurant	Furniture / Home Store	Harbor / Marina	Wine Bar	Italian Restaurant	Office	Department Store	Dessert Shop

Table 1: Top 10 most common venues per station

4. Machine learning

In order to help analyse the Foursquare API data on the different stations, the data were segmented and grouped in order to classify the stations into clusters to find similar stations in Copenhagen. Clustering is a form of unsupervised machine learning. In this instance a very popular clustering algorithm, called k-means, was used.

4.1 K-Means Clustering

One of the algorithms that can be used for data segmentation is K-Means clustering. It can group data based on similarity by using partitioning clustering which divides the data into K non-overlapping subsets or clusters. As such, it is an unsupervised algorithm.

Since we must must initialize K which represents number of clusters, a number of different values from 5 to 10 were used and the optimal value was considered to be XX since this clusters the stations in a meaningful way.

4.2 Clustering results

The red “cluster 0” stations (Figure 8) in the city center of Copenhagen is characterised by a large number of surrounding venues but due to high passenger foot traffic would also be an attractive area to invest in a business. The key hear would be to leverage the above Foursquare top venues data to identify a potential market gap.

The purple "cluster 1” stations (Figure 8) are all located in the densely populated central neigbourhoods outside the historical city centre. These would also represent attractive investment locations outside the expensive city centre.

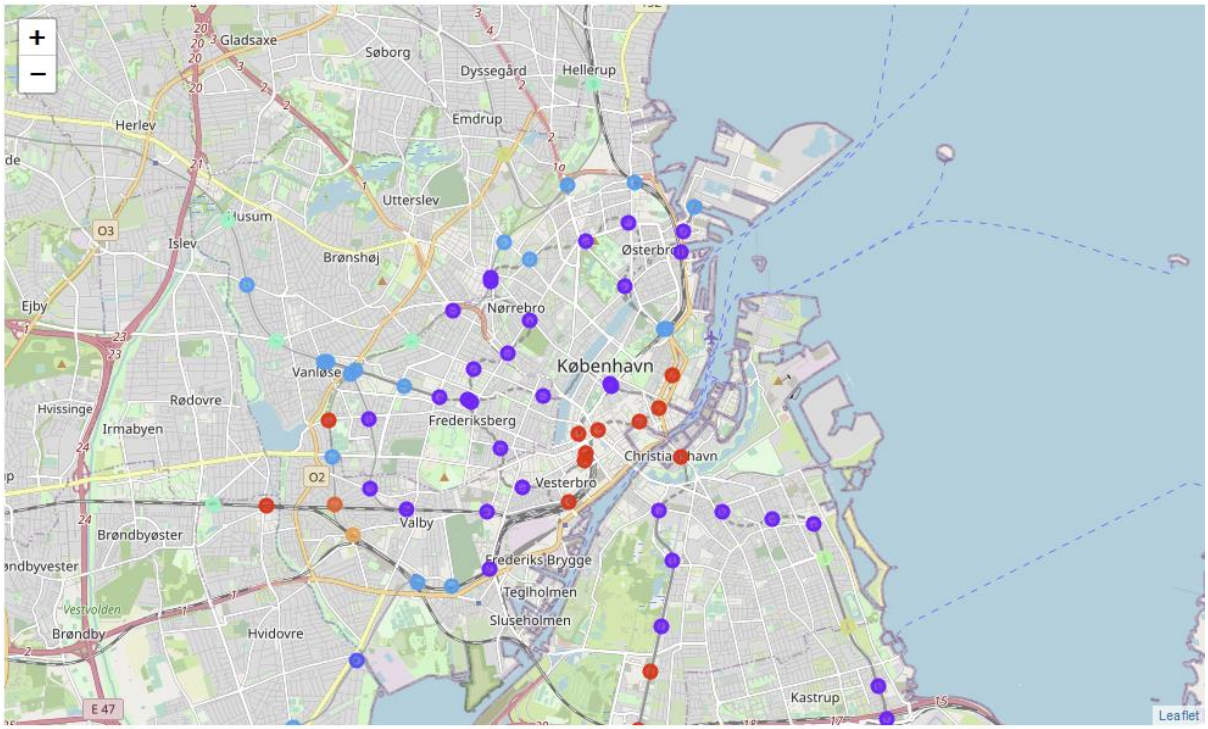

Figure 8. Map of the station clusters in Copenhagen

The other clusters represent a number of different features:

- Dark blue “cluster 2”is located near an industrial facility and sports and recreational facilities.
- Blue "cluster 3” stations are located in areas with a more suburban character.
- Cyan "cluster 4” is located near a housing estate and major motorway intersections.
- Turquoise "cluster 5” represent lower density residential areas in the suburbs.
- Green “cluster 6” and yellow "cluster 7” respresent unique locations due to their proximity to the city beach and outdoor recreational facilities, respectively.
- Orange "cluster 8” and dark orange "cluster 9” are both located in suburban transport hubs.

5. Conclusions

In addition to Denmark's reputation for sustainability, the high rankings in quality of living and happiness, and ease of doing business we were able to quantify the macroeconomic situation.

Gross Domestic Product (GDP), population, and GDP per capita in Copenhagen have been increasing steadily and population projections for the city show the current trend continuing for the foreseeable future. This substantiates quantitatively that the city is an attractive investment location.

Since our clients would want to invest in locations with high foot traffic, we analysed data on daily passenger numbers in the busiest train stations and identified in particular Nørreport station and København H as locations of interest.

The data on venues in the proximity of the stations allowed us to identify the top 10 most common venues around a station. This allows the client to identify whether their business idea would fill a niche in a particular area.

The clustering of train stations also showed two clear clusters in the historical city center and central neighbourhoods of Copenhagen

and could be of particular interest as locations for potential investors.

6. Future directions

Another dataset to integrate into te decision on where to locate a new business would be real estate costs, since the monthly rent and utilities make up a significant part of the monthly expenses for a physical business venue.

Another aspect that could be considered regarding the potential locations for the busy would be future infrastucture investments. Investing a business that is close to a metro or train station that is planned to open up in the near future could imply lower costs and future growth in the customer base.

7. List of Figures

8. List of Tables

9. References

1. World Happiness Report(https://worldhappiness.report/ed/2020/)

2. Forbes Best Countries for Business

(https://www.forbes.com/best-countries-for-business/list/)

3. Statistics Denmark (https://www.dst.dk/da)

4. StatBank Denmark (https://www.statbank.dk/) Tables FOLK1A

5. StatBank Denmark (https://www.statbank.dk/) Tables NRHP 1-2.1.1

6. StatBank Denmark (https://www.statbank.dk/) Tables NRHP 1-2.1.1

7. StatBank Denmark (https://www.statbank.dk/) Tables FRKM120

8. Open Data DK (https://www.opendata.dk/city-of-copenhagen/oversigtskort) station_oversigtskort.geojson

9. Jyllands-Posten (https://jyllands-posten.dk/indland/trafik/article6570259.ece/)

10. Foursquare Developers (https://developer.foursquare.com/)