Building the Next successful Lidl Butik/Supermarket in Norddjylland, Denmark.

1. Introduction

LIDL is one of the affordable or at best the cheapest supermarkets in Europe, but, it is not often widely represented in many European cities, especially in Scandinavian countries such as Denmark.

Surprisingly, the last five years, **LIDL Group** have started to build several **LIDL Supermarket** in different regions of Denmark, more particularly in the north of the country called **Nordjylland** in Danish language.

The purpose for this project therefore, is to investigate and find out where the next **LIDL Supermarket** will be built in **Aalborg,** the capital city of the region**.**The target audience is Aalborg Municipality or borough and the stakeholders are of course **LIDL Group** , and the main competitors among others are **Rema1000**, **Føtex**, **Fakta**, **ALDI and Netto.**

2. Methodology

This section of the project will be devoted to describing the type of dataset that will be used, to find a shared meaning and understanding of the project under investigation. Below are the thematic steps of the section.

- Acquisition of the data by using **Foursquare Api ;**
- Present an overview of data;

- Perform some exploratory Data Analysis (EDA) to find some patterns in dataset and to have an idea of what kind of useful Machine learning can be suitable to solve problem
- Build a model
- Test and explain the results.

2.1 Data Acquisition

To Tackle or solve this problem, the study will require a Dataset, which contains locations of the existing **Lidl** Supermarket and main competitors which already exist in Nordjylland, the focus of the project.The writer do not have an existing dataset for this study, hence, the **Foursquare API** will be used in order to create a proper dataset that will be used to investigate and propose solution to the problem. The dataset will contain:

- All main competitors such as **Rema1000**, **Føtex**, **Fakta**, **ALDI, Netto** and their location
- All the existing **LIDl** supermarket, their geographic location and the distance from the main city.
- Aalborg borough or neighborhood, the population and the number of person per km2.

The final Dataset to be used can be seeing the in the figure below

	name	categories	address	city	distance	lat	lng	postalCode
0	REMA 1000	Discount Store	Vestergade 22	Nørresundby	1255	57.059502	9.919370	9400
1	REMA 1000	Discount Store	Godsbanen 2	Aalborg	723	57.042319	9.924412	9000
2	REMA 1000	Discount Store	Østerbro 29 A	Aalborg	1054	57.046191	9.936409	9000
3	REMA 1000	Discount Store	Danalien 1 E	Aalborg	2562	57.029152	9.943098	9000
4	REMA 1000	Discount Store	Stenbjergvej 1	Aalborg Øst	5300	57.040576	10.005793	9220

2.2 Data Analysis and Visualization

After Data wrangling or Data munging, the study will perform some exploratory Data analysis (EDA) and visualization for a better understanding of the dataset and tries to figure out which Machine Learning algorithm can be suitable or useful in solving or sharing more light into the investigation.

Surprisingly, After a simple grouping by postalCol, I realized that there are more **REMA 1000 supermarket** in **Nordjylland** than any other supermarket mentioned above, the results are presented in the following table.

	name	postalCode
0	Fakta	26
1	Føtex	14
2	Lidl	9
3	REMA 1000	37

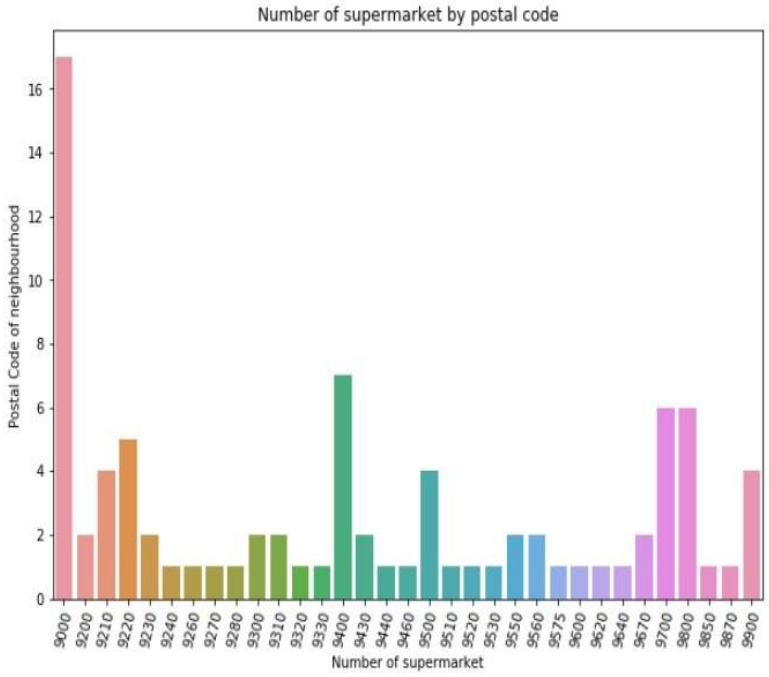
It can also be seen that there are **9 Lidl** supermarkets, located in various parts of the region, which are represented in the table below:

	name	categories	address	city	distance	lat	lng	postalCode
41	Lidl	Discount Store	Thistedvej 1d	Nørresundby	1753	57.063187	9.910343	9400
42	Lidl	Discount Store	Diskovej 3	Aalborg	3062	57.023636	9.942100	9210
43	Lidl	Discount Store	K Christensensvej 1 - 3	Aalborg	4116	57.013425	9.896384	9000
44	Lidl	Discount Store	K. Christensens Vej 1	Aalborg SV	4138	57.013435	9.895280	9200
45	Lidl	Discount Store	Østergade 52	Brønderslev	24062	57.263749	9.949852	9700
46	Lidl	Discount Store	Hostrupvej 68	Hobro	45469	56.645578	9.793760	9500
48	Lidl	Discount Store	Sæbygårdvej 29	Sæby	47406	57.333284	10.503293	9300
49	Lidl	Discount Store	Åstrupvej 2	Hjørring	45567	57.456340	9.977785	9800
51	Lidl	Discount Store	Hjørringvej 125	Frederikshavn	57039	57.447173	10.513747	9900

As can be inferred in the table above, there is only one out of the 9 **Lidl** supermarkets located in the city center. The remaining 8 are located very far from Aalborg center. There must be an array of reasons behind this choice and the study will enumerate some of them in the next sections.

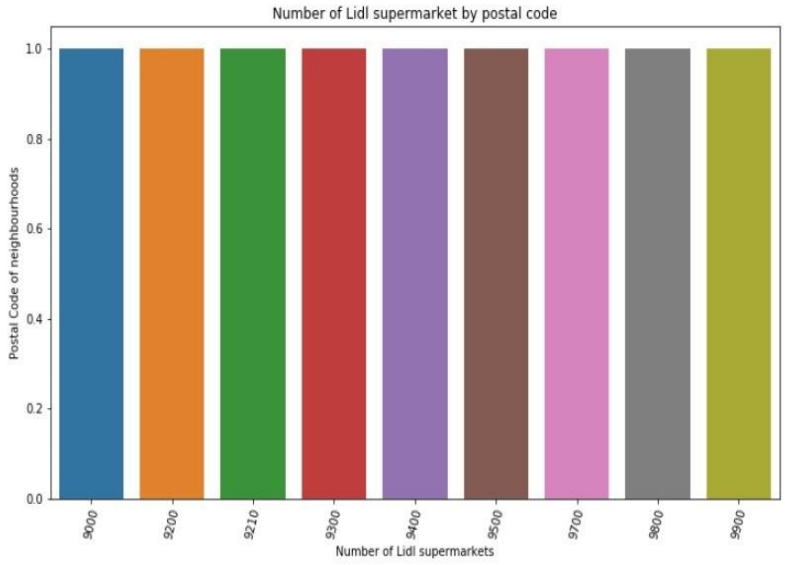
One can also Visualize all supermarkets in **Nordjylland** by postal code in the following figures:

```
df_bar = df.groupby('postalCode')['name'].count().reset_index()
plt.figure(figsize=(10,7))
plt.xticks(rotation=75)
ax = sns.barplot(x = 'postalCode', y = 'name', data=df_bar)
ax.set(xlabel='Number of supermarket', ylabel='Postal Code of neighbourhood', title='Number of supermarket by postal code')
plt.show()
```



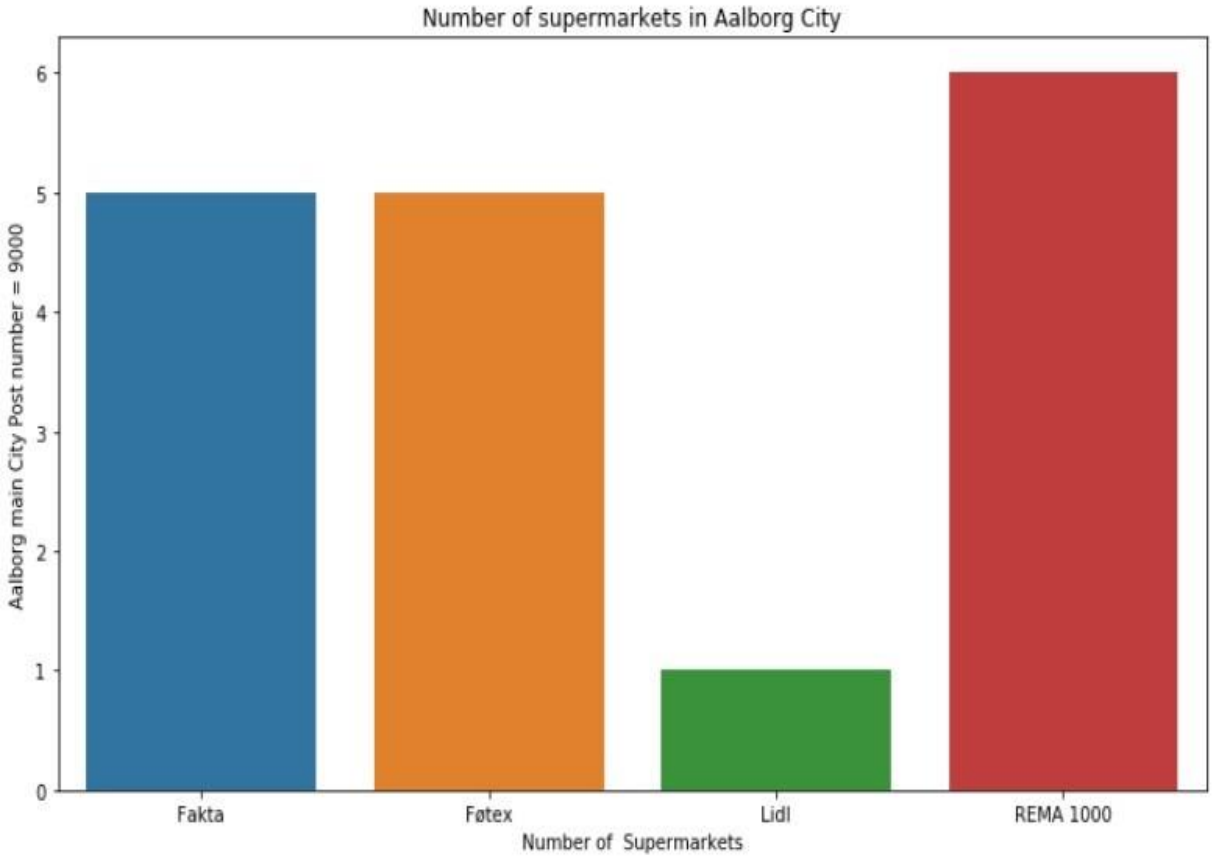
Number of **Lidl** supermarket in Aalbrog are presented in the following figure:

```
lidl_filtered_df = df[df['name'] == 'Lidl']
plt.figure(figsize=(12,7))
plt.xticks(rotation=75)
ax = sns.countplot(x = 'postalCode', data=lidl_filtered_df)
ax.set(xlabel='Number of Lidl supermarkets', ylabel='Postal Code of neighbourhoods', title='Number of Lidl supermarket by postal code')
plt.show(ax)
```

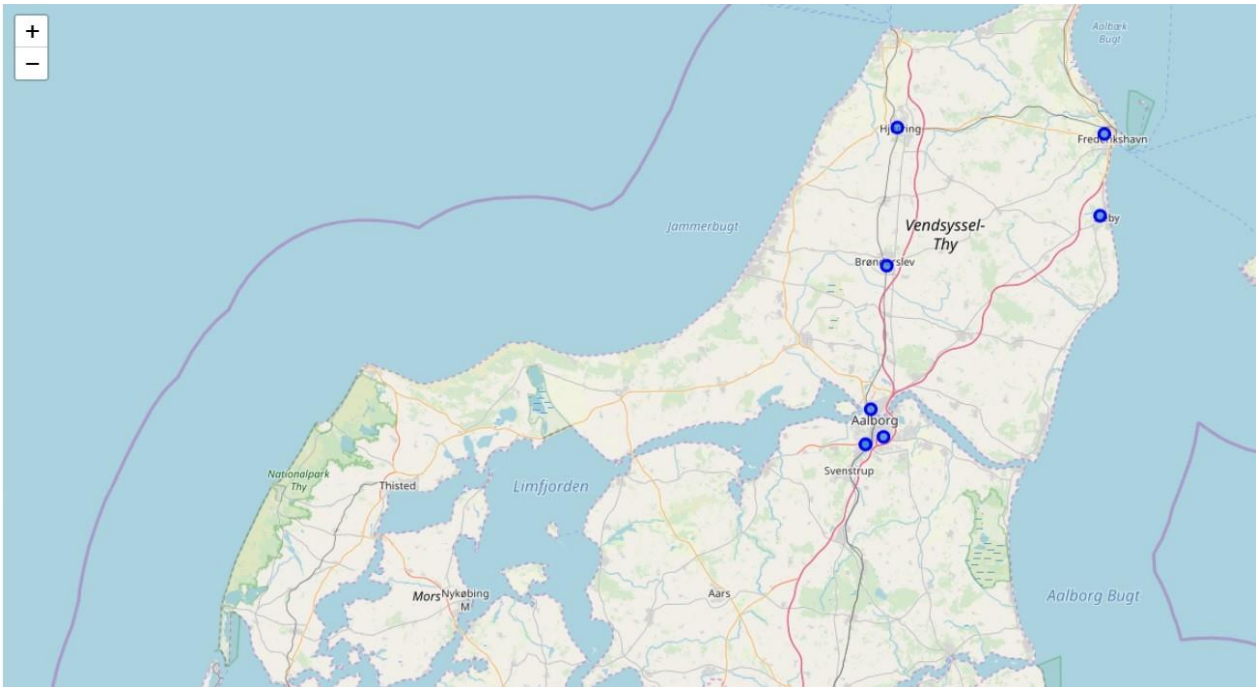


From the above graph we can see that **LIDL** has decided to build only one **Supermarket** per neighborhood (by Postal code). The reason of this choice can be very difficult to understand. But on the surface, one could see that, they established the supermarkets based on the Districts/Kommune, however, all the Kommunes do not have the same population size, hence, the reason and decision must be reconsidered

The number of the **supermarkets** in Aalborg city centre are represented as below.



Now one can Create a map of Aalborg with **Lidl** Supermarkets their locations and postalCode of neighbourhood using **folium**



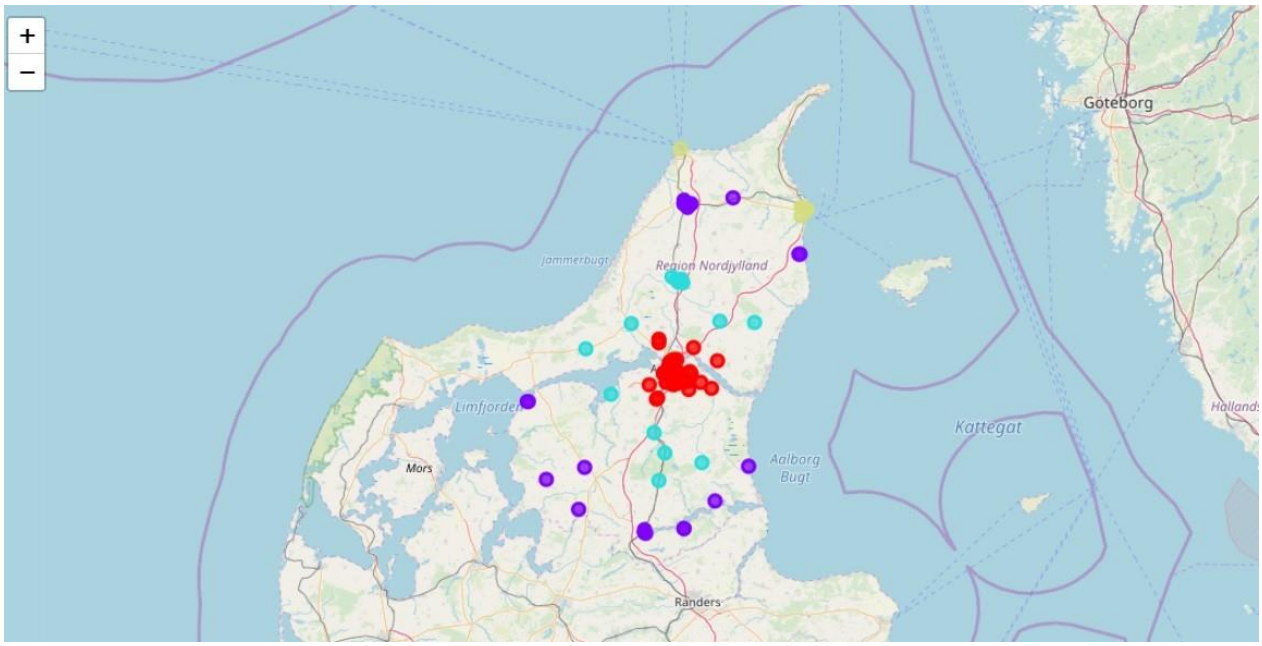
2.3 Modelling

With a better understanding of the dataset, one can use unsupervised Machine learning algorithm as a suitable method to enhance a better analysis of the issue at hand. Therefore, in the next step, the investigation will be using a **k-means** algorithm thus using the location data and the distance to give readers a good meaning of the study.

Below is a representation of the clustering results taking into consideration a reasonable number of clusters.

	name	categories	address	city	distance	lat	lng	postalCode	cluster_labels
41	Lidl	Discount Store	Thistedvej 1d	Nørresundby	1753	57.063187	9.910343	9400	0
42	Lidl	Discount Store	Diskovej 3	Aalborg	3062	57.023636	9.942100	9210	0
43	Lidl	Discount Store	K Christensensvej 1 - 3	Aalborg	4116	57.013425	9.896384	9000	0
44	Lidl	Discount Store	K. Christensens Vej 1	Aalborg SV	4138	57.013435	9.895280	9200	0
45	Lidl	Discount Store	Østergade 52	Brønderslev	24062	57.263749	9.949852	9700	2
46	Lidl	Discount Store	Hostrupvej 68	Hobro	45469	56.645578	9.793760	9500	1
48	Lidl	Discount Store	Sæbygårdvej 29	Sæby	47406	57.333284	10.503293	9300	1
49	Lidl	Discount Store	Åstrupvej 2	Hjørring	45567	57.456340	9.977785	9800	1
51	Lidl	Discount Store	Hjørringvej 125	Frederikshavn	57039	57.447173	10.513747	9900	3

Using **folium** we can visualize the different cluster and their locations.



3. Results and Prediction

The cluster analysis show that many of **Supermarket** are located in the main City **Aalborg** but this is not the case when we consider **Lidl**, which is our target for this project. When we look at the first **Cluster** which represent the group of clusters gethered in main City with Postal code between **9000 to 9430**. There are **4 Lidl** assigned to this cluster and only one of them is located in main center with **PostalCode 9000**. It has also been discovered that, each **Lidl supermarket** has different location very far from each other.

According to the results from the different data combined with cluster analysis, the study found the need and suitability to build the next **Lidl** supermarket in the main city area (postalCode 9000), where the other supermarkets are widely represented. There might be strong competition, but with an existing good will and brand the market could still play in their (LIDL) favor.

Another factor that can play in favor of **Lidl** is that, it is more represented in many European countries and will therefore, have a visibility advantage as compare to the other competitors as stated earlier.

Besides, the population is growing in the main city and equally diverse in nature, and with simple demand and supply knowledge, the diverse population growth means increase in demand for basic goods and services. Accepting this as given, the establishment of new Lidl supermarket in the main city means it will not be left out with the advantages associated with the multicultural growth of the city.

Conclusion and Discussions

The main purpose of this project was to investigate where **Lidl Goup** will build the next **Lidl** supermarket in **Nordjylland.** With the objective to help stakeholders in their decision to find the optimal location for the next successful **Lidl** supermarket in the region.

Having an EDA and performing clustering algorithm according to the different locations and their respective distances to city centre, coupled with the growing population as established, the study came to the conclusion that, **Lidl** should make efforts to establish the next supermarket in Aalborg City Centre. However, the final decision of this, is not with the writer or investigator, but, the stakeholders (Lidl Group) based on the specific characteristics that defines the setting up of their new supermarkets decisions which is not privy to the study.

Further Development

The following are suggestions about how this project could be further implemented:

1. Integrate communal plan with the project.
2. Using more automated Artificial Intelligence (AI) tools for better performance of the algorithm.