

Restaurants in Copenhagen

Christian S. Jacobsen

November 27, 2020

1 Introduction

1.1 Background

Restaurant placement is a difficult and important business problem. Deciding where to locate ones place of business can impact competition, profit and even employee recruitment. In a free market economy one would expect to eventually approach some sort of equilibrium in restaurant placement in a given area, simply because a gap in the market would attract prospective business owners, and an overabundance of restaurants in a given area would lead to strong competition, which in extreme cases could lead to business closure.

1.2 Problem

We wish to investigate whether such patterns can be identified from the location data of existing restaurants in an area, which could lead to more informed decisions about whether or not to open a particular type of restaurant in a given area. Specifically, we are interested in the central Copenhagen area, where disposable income and population density is generally higher than the rest of Denmark.

1.3 Customers

Potential customers for this sort of investigation is obviously prospective restaurant owners, but it could also be relevant for local city councils who wish to attract investment and new business to their part of town.

2 Data

2.1 Data sources

The municipality of Copenhagen is divided into districts, which we can use to focus our investigation to different sections of the city. Basic information about these districts is available on Wikipedia.

For information about already existing restaurant venues in Copenhagen we will use the Foursquare API.

Denmark is a very digitized country, and as such it has a lot of data about citizens and location readily available to the public, such as disposable income. The level of disposable income is very useful in deciding on restaurant placement, based on the hypothesis that people with a larger disposable income tend to eat out more often. This data is provided by Statistics Denmark, a government body that provides statistics to the public. Unfortunately, disposable income is only available at the municipal level, and since the central Copenhagen area consists of only two municipalities, the resolution is simply too low for this to be useful.

2.2 Data cleaning

The Foursquare API returns all sorts of venues, so for data cleaning we selected only the venue categories that contained the word 'restaurant'. Restaurants are then categorized by the nationality of the cuisine, which makes it easy to identify a potential gap in the market.

3 Methodology

From the 11 Copenhagen districts we identify 778 venues using the Foursquare API. 169 of these turn out to be labelled as restaurants. To reduce the complexity of the dataset we decide to apply a simple clustering algorithm, K Means. We identify the optimal amount of clusters to be 2, which effectively divides the city into two parts, one which is relatively dense with restaurants, and one which is relatively less dense.

4 Results

Although, as mentioned, data for disposable income is not available on the district level, this clustering fits well with an expectation of which city districts are relatively more affluent, although this could of course be a coincidence. Without more data we are not in a position to determine the level of correlation.

We shall label the two clusters the inner city, and outer city, since the density of restaurants also correlates with distance to the city center.

5 Discussion

While we are not in a position to directly recommend one district over another based on the data available, we can certainly identify some main results. Despite the density of restaurants in the inner city, there are gaps in the market that could potentially be exploited. For example, restaurants featuring african and south american cuisine are completely absent in all inner city districts except one. On the other hand, sushi restaurants are common in both clusters, which indicates strong competition, and as such, opening yet another sushi restaurant is probably ill-advised.

6 Conclusion

We have investigated the density of restaurants in the greater Copenhagen area using the Foursquare API. We applied a K Means algorithm to divide the city into clusters, and identified two clusters. The inner city cluster was relatively dense in restaurants of various types, but had some noticeable gaps such as african and south american cuisine. The outer city cluster was generally less dense in restaurants, which means less competition, but possibly also less customers.

For future work, the model could be enriched by including ease of access, such as availability of public transport, level of disposable income, crime rate, etc. Such data exists, but it is not readily available to the public for data protection regulation reasons.