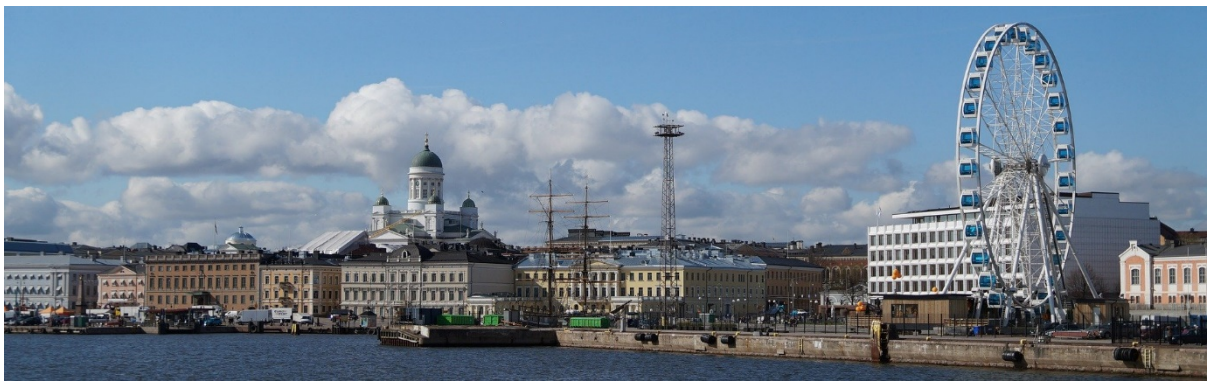
REPORT

IBM Applied Data Science Capstone

Most liked restaurant types in Helsinki, Finland

By: Mikael Björkqvist

January 2020



Introduction

Helsinki, capital of Finland, is full with restaurants. Restaurants are almost always categorized by their cuisine, prize range and style. This helps customers to choose restaurant more easily when searching them through applications like yelp, google, tripadvisor etc.

In example, user want's to eat "Italian" food, and then searches restaurants belonging to "Italian restaurant" category, with some search application.

To understand what types of restaurants people of Helsinki likes the most, and where these restaurants are in located, is valuable information for entrepreneurs, investors, critics etc.

Business problem

What is the most liked restaurant type in Helsinki, and where are such restaurants located.
Using already existing location and user/customer input data as variables. This report and methods build for it, will also act as an solid base for further processing of similar ideas.

Target audience

Investors, loan givers, entrepreneurs, city management etc.

Data

To solve the problem, we will need the following data:

• List of all restaurants in Helsinki area, which will be the scope of this project.

• Data from those restaurants. I.E. Type, user score and coordinates. This data can be used to make data frames to rank restaurant types, and plot clusters of restaurant types

Sources of data and methods to extract them

We use Helsinki map, wiki and Python Geocoder for coordinates and area selection. Restaurant location data we get from Foursquare service using their API queries. Foursquare has one of the largest databases of 105+ million places and is used by over 150,000 developers. These venues are updated with data by over 13 billion venue check ins starting 2009. Foursquare API will provide many categories of the venue data which is good when we want to expand our research.

All data scraping, cleaning, wrangling, machine learning (K-means clustering) and map visualization (Folium) is done in Python 3 using Jupyter notebook, that is also shared for review.