

Capstone Project - The Battle of the Neighborhoods

1. Introduction

The aim of this data-analysis is to create a city-map visualisation of public data to generate insights helpful in determining a variety of questions. The deliverable of the project is an interactive city-map of a city of choice in the Netherlands stating the demographical and venue characteristics values per postcode. For this report the city-map will generate helpful insights in determining potential locations to open a new high-end restaurant in the Rotterdam. The restaurant customer segment will target customers aged 25-44. The criteria to determine the ideal area based on postcode to open a high-end restaurant are:

- Postcode area with high population density
- Postcode area with high concentration of potential customers aged 24-44
- Postcode area with high average housing prizes
- Postcode area with high income (determined by low governmental financial aid)
- Postcode area with low competition by other restaurants (determined by # restaurants)

The interactive city-map is created using public data from the Dutch Central Bureau of Statistics to determine the demographics. The venues characteristics to be placed on the map are derived from FourSquare as is required by the Capstone Project assignment. Another outcome of the data analysis is a list of the postcodes with the corresponding area and neighbourhood.

1. Data

Borough data

The data for the demographic characteristics are derived from the CBS website. To create the list of the postcodes with the corresponding area and neighbourhood the following datasets are used and combined.

Datasets from CBS to create list of postcode information

- Postcode - pc6hnr20190801_gwb.csv
- Municipality - gem2019.csv
- Neighbourhood - buurt2019.csv
- Area - wijk2019.csv

Demographic data

As mentioned the aim of the project is to create a city-map visualisation of public data to generate insights helpful in determining a variety of questions. Therefore, multiple demographical parameters are inserted into the city-map to allow for multifunctional use. The demographical datasets are also derived from the CBS website. These datasets are combined with the dataset of the postcodes to create one dataset with all the demographical characteristics of the postcode areas.

Datasets from CBS to derive the demographic and characteristic data per postcode

- Demographics - 2014_PC6_demografie_en_wonen.csv
- Characteristics - 161010-Kenmerken-postcode-mw.csv

Venue data

The venue data is obtained by means of the FourSquare API to explore the postalcode neighborhood postalcode venues. To further define the categories a dataset from <https://gist.github.com/Zoltrix/d69a52e9f91f4211be93eea1f1398424> is saved and used.

Map latitude and longitude data

In order to create the map with the correct postalcode zones latitude and longitude data is required. The geojson with the polygon type coordinates of all postalcodes is found here: https://raw.githubusercontent.com/openstate/hackdetoekomst/master/datablog/nlmaps-stages/cbs_pc4_2017.geo.json. The postcode_coordinates.csv file from <https://github.com/bobdenotter/4pp> is used to derive the right lat/long for the postalcodes. Furthermore, to identify the coordinates for the folium map the data is fetched by importing the Geolocator package.

1. Methodology

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
- Results section where you discuss the results.
- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
- Conclusion section where you conclude the report.

The dataset described above will be used to create a city-map with different data visualisation techniques to derive the required insight for answering the question in where to best open a high-end restaurant in Rotterdam. The code consists of 4 parts, next the parts will be described on more detail.

PART A - PREPARE THE DATASET FOR [CITY]

The first part is created to prepare the dataset for the city. First the datasets pc6hn20190801_gwb.csv, gem2019.csv, buurt2019.csv, wijk2019.csv are read, formatted and merged to create one 4-figure postalcode dataset for all the borough data. Secondly, the demographics and characteristics (2014_PC6_demografie_en_wonen.csv and 161010-Kenmerken-postcode-mw.csv) datasets are read, formatted and merged with all relevant columns from the postalcode 4 dataset. Lastly, a dataset solely for the city of choise (in this case Rotterdam) is created and saved an output excel file.

PART B - CREATE CHOROPLETH MAP TO VISUALISE THE DATASET FOR [CITY]

In the second part of the code the map is created and the first visualisation technique 'choropleth mapping' is coded to visualise the demographics of each postalcode of the city. First the basic map is created by means of the Geolocator and Folium map created. After the basic map is created the postalcode hover-over layer is added with GeoJson based on the polygon type data of the postalcodes of the Netherlands. Once the basics of the maps are set, the choropleth layers (1) Population Density, (2) Household Size, (3) Age 25-44, (4) House value, (5) Financial aid, (6) Cultural Diversity are added. Below is a

description of the definitions of the demographics. All label indicators of the demographics on the top of the map are set to quantiles to show the division of the data in a clear context.

Demographic	Definition
Population Density	Number of addresses in postalcode
Household Size	Average household size in postalcode
Age 25-44	% of the postalcode population aged 25-44
House value	Average house value in postalcode (value * 1000)
Financial aid	% of the postalcode population receiving governmental financial support
Cultural Diversity	Dutch governmental cultural diversity score based on the number of immigrants in postalcode

PART C - ADD MARKERS TO [CITY] MAP

In the third part of the code the venue data is derived from FourSquare and based on this dataset the markers are added to the city-map. All sets of markers are added as a layer for the functionality that they can be shown and hidden. The first set of markers added to the map are the corresponding town to the postalcode area's based on the postcode_coordinates.csv file. Secondly, the venue markers are added. The data is collected with the FourSquare API and saved in an excel file to not overuse the API every time the code is run for the same city. As the FourSquare data categories are very defined, the dataset is merged with the broader category dataset. Lastly, the markers of the venue categories (1) Athletics & Sports, (2) Bar, (3) Clothing Store, (4) Food & Drinks Shop, (5) Restaurant, (6) Train Stations are added to the map.

PART D - SHOW RESULTS

In the last part of the code the interactive city-map is saved as .html and opened in the webbrowser.

1. Results

The objective of the business problem was to help entrepreneurs identify the best area in Rotterdam to open a high-end restaurant. The issue is addressed by creating an interactive city-map to explore the area's conforming to the set criteria. The results will be presented and analysed per criteria with a screenshot of the map. The output excel is used as reference for the potential postalcode areas. The initial selection of potential area's is based on the population density.

- Postalcode area with high population density
- Postalcode area with high concentration of potential customers aged 24-44
- Postalcode area with high average housing prizes
- Postalcode area with high income (determined by low governmental financial aid)
- Postalcode area with low competition by other restaurants (determined by # restaurants)

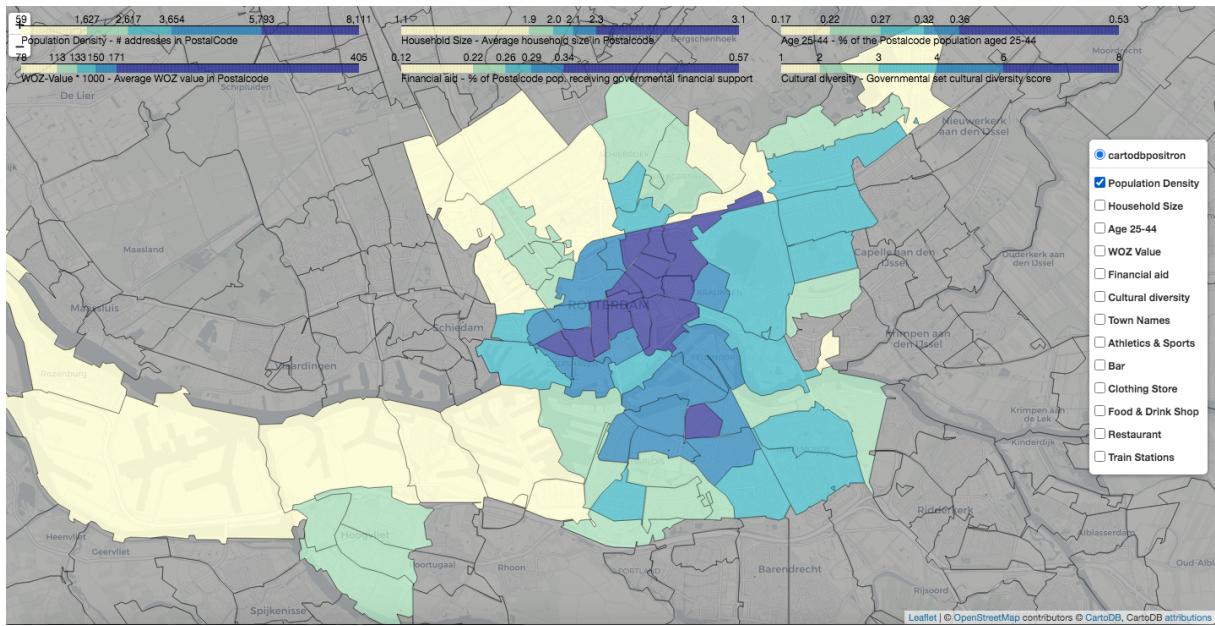


Figure 1. Choropleth – Rotterdam - Population Denisty

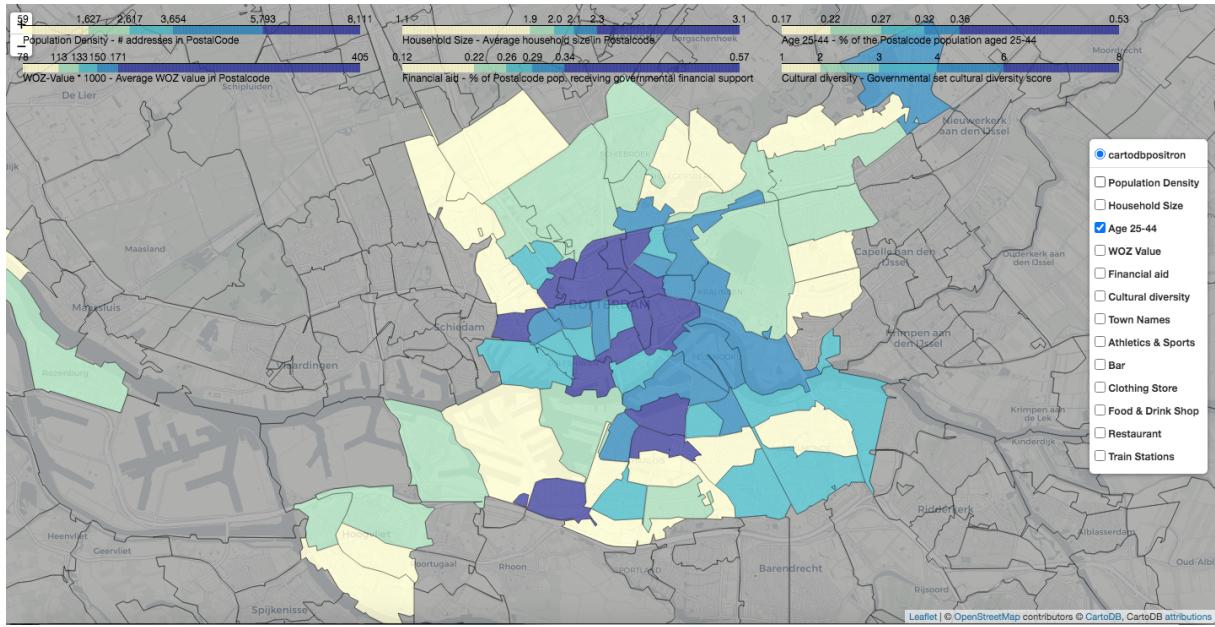


Figure 2. Choropleth – Rotterdam - Age 25-44

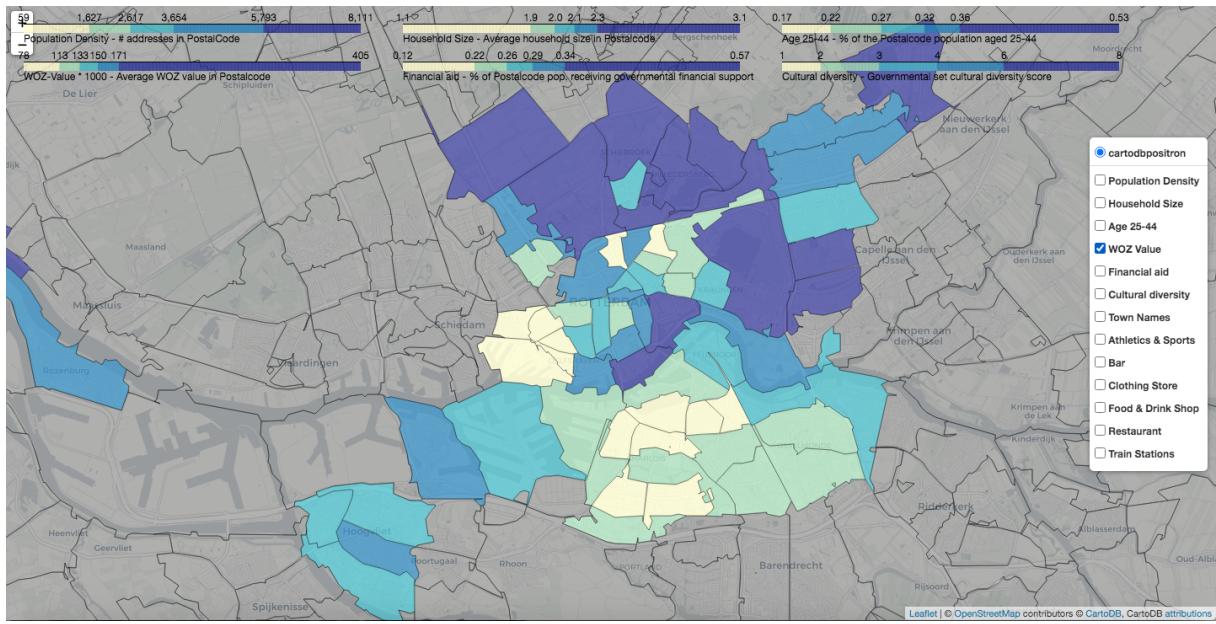


Figure 3. Choropleth – Rotterdam - House prize

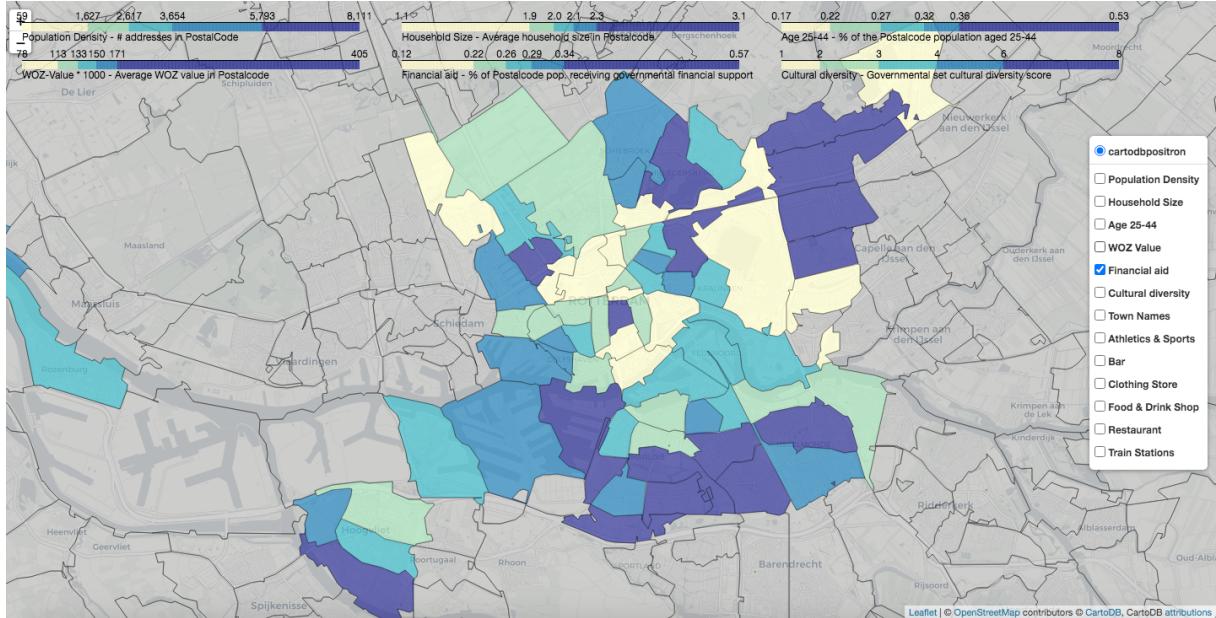


Figure 3. Choropleth – Rotterdam – Financial aid

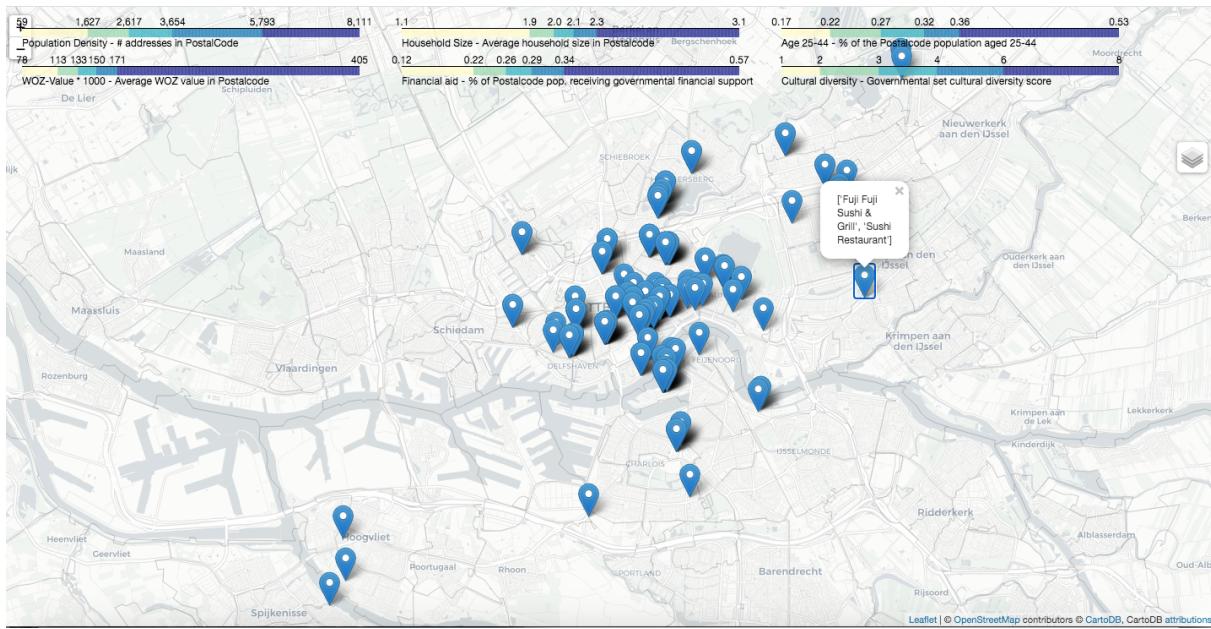


Figure 3. Choropleth – Rotterdam – Restaurants

1. Conclusion

According to the analysis it is stated to best open the high-end restaurant is postalcode area 3016. This is derived from comparing the outcomes on the map with the desired demographics and competition. Note that the postalcode is visible when hovering over the area's. All the postalcodes are inspected and not one of the postalcodes actually did satisfy all criteria. Having that set, the criteria were than prioritized as below with priority from high to low.

1. Postalcode area with high average housing prizes
2. Postalcode area with high income (determined by low governmental financial aid)
3. Postalcode area with low competition by other restaurants (determined by # restaurants)
4. Postalcode area with high population density
5. Postalcode area with high concentration of potential customers aged 24-44

As the allure of the area is of vital importance in the location of a high-end restaurant the housing price and income of the area is placed highest priority. Furthermore, the fourth and fifth criteria are still met as the surrounding areas do satisfy the criteria's. Also in this area there is little competition as only two other restaurants are placed here.