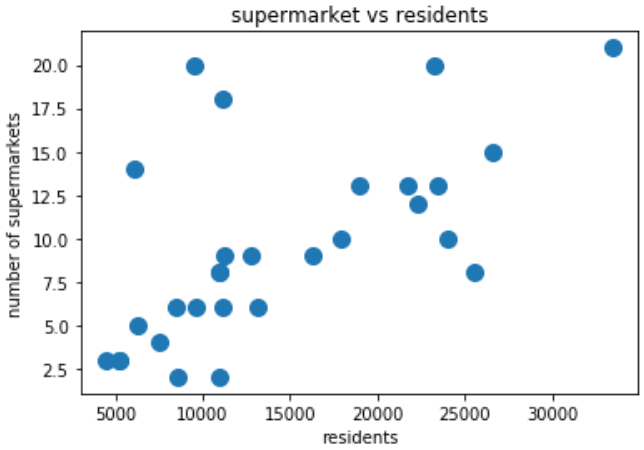
We study the relation between number of supermarkets and residents in the city of Zurich. We quantify the relation between these two variables and make predictions for the location of new supermarkets in the city. We extend the analysis studying the relation between number of supermarkets in a certain area and the number of people passing by that area.

All necessary files are available in this link:

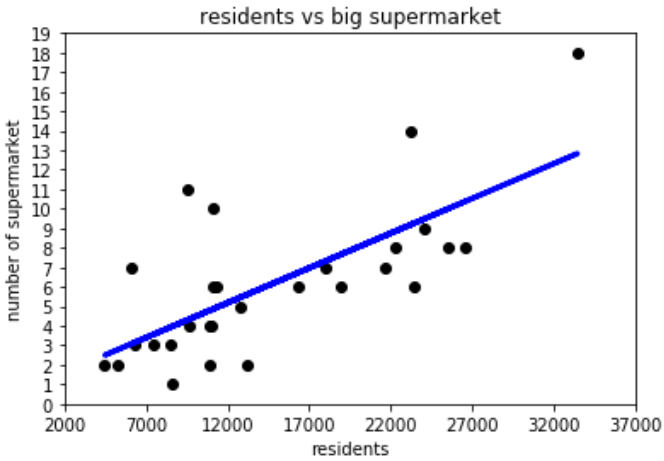
https://drive.google.com/open?id=1SCOKzfdO6kJF6sKHYqUMNFATcXdzq-S8

Results:

Each dot in the graph is a Quartier in the city of Zurich. On the x-axis we count the residents in that

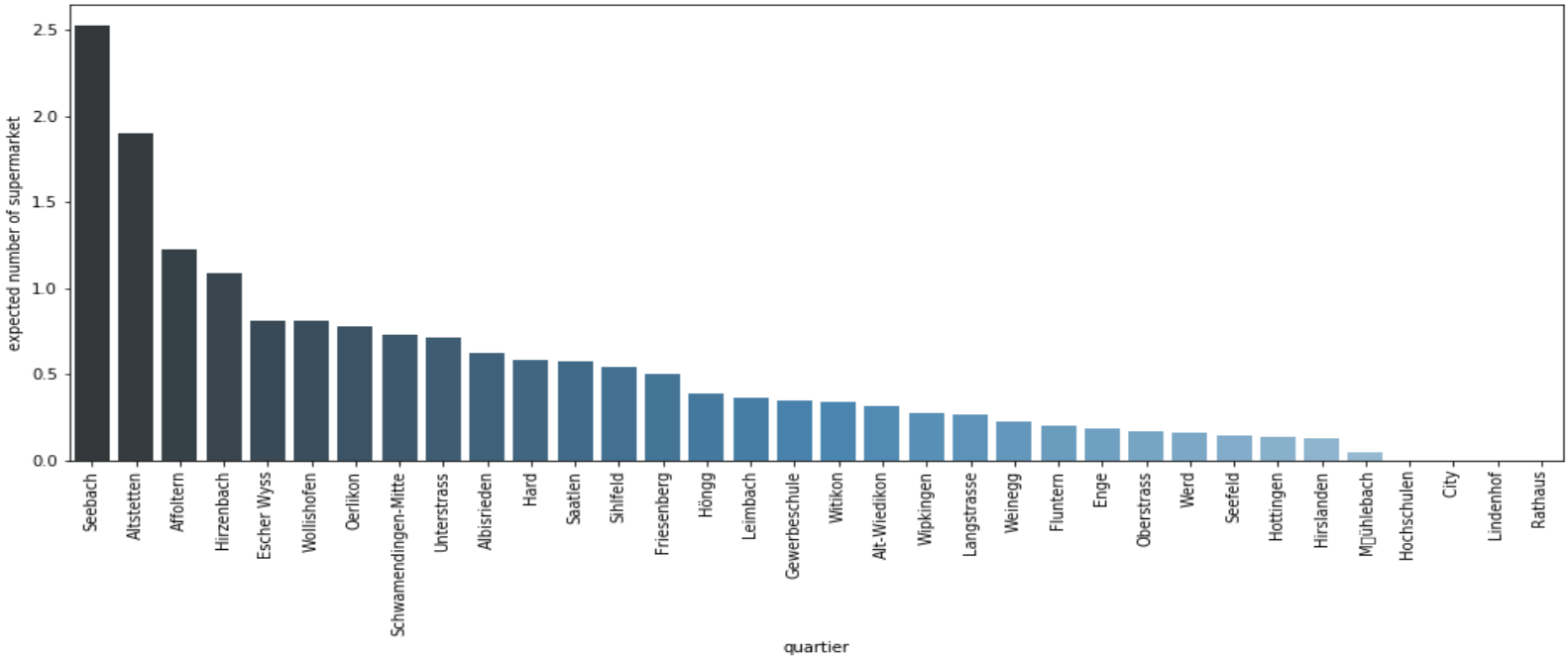


Quartier, on the y-axis the number of supermarkets. When we analyzed what the points in the north-west of the graph we realized that they are representing quartier that are not very populated but were a lot of people spend their days for job-related reasons. We decided to analyze further following two directions. First we distinguished between large supermarket and small grocery shop. Second, we try to include in the study data about traffic.
The result for the first analysis gives us the following:

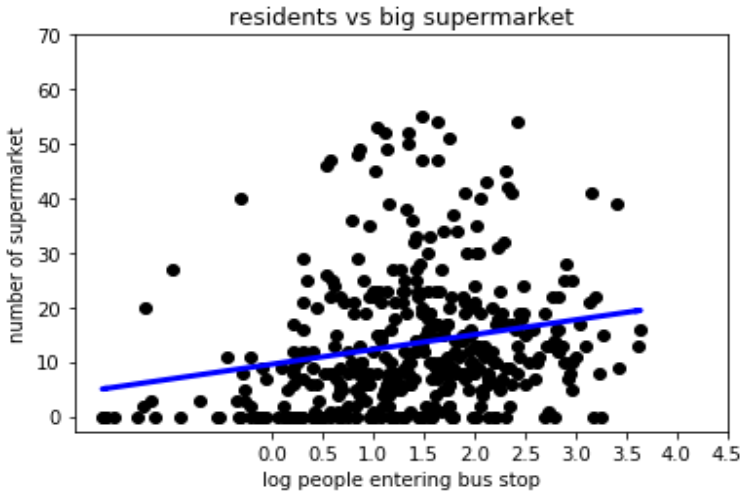


The graph now includes only large supermarkets. We see that the relation between this number and the number of residents in a certain quartier is stronger. Residents of a certain quartier are likely to shop (big shopping in a big supermarket) close to their own house.

We compute that one big supermarket serves about 2800 residents. Using the forecast on population growth in Zurich we can compute the number of new big supermarket we expect:



In the second part of the analysis, we follow a similar pattern but we focus on data about traffic. We know how many people pass through a certain station (bus, train …) and we want to relate this data to the number of supermarkets close to that point:



Each point in the graph is one station in the city of Zurich. On the x-axis the log of the number of people entering in that station, on the y-axis the number of supermarkets in a 1-km (result robust to other specifications) distance from that station.