# Activity report: Analysis of demographics and voting behavior in The Hague.

The following report represents the results work done for Work package 1 of Micro-targeting & Local elections project.

The goal of this study is as follows - to analyze how demographic characteristics of the citizens can “predict” voting behavior. By demographics, we mean age, ethnicity, income, employment, etc. In case of voting behavior, it is a number of votes given for particular party, for example, PVV, PvdA, D66, and so on. The analysis was conducted for The Hague on the district level. That is, 44 districts were examined. As voting datasets municipal elections of 2014 and 2018 were chosen. As main instruments 3 machine learning algorithms were used - NMF, multiple linear regression (MLR), and support vector regression (SVR). All manipulations were made in Python programming language under support of Scikit-learn library.

In short, the job was done in the following manner. First, the dataset for the 2014 year was analyzed. 33 demographics characteristics were carefully chosen to “cluster” citizens using NMF. As a result, 7 clusters-profiles were identified (Figure 1). The interpretation here is as follows. Let us consider column number 3 ("profile" number 3). Here we can observe grouping of old Dutch citizens with low-income received from pensions. Secondly, these profiles were matched to 44 districts (Figure 2). This figure can be read in the same manner. That is, previously mentioned profile number 3 can be mostly found in “Zuiderpark” district. These results were roughly validated by the citizens of The Hague. Consequently, people responsible for interviews are provided with a "map" that shows where they can find which people. In addition, results were plotted on the map (see Jupyter notebook).

Further, the same procedure was applied to voting data (Figure 3 and 4), 7 voting profiles were identified and aligned to the districts. The voting profiles here represent unique combinations of votes for various parties. Finally, there is a possibility to combine obtained results. Using voting profiles as predictive variables and a set of corresponding values for demographics profiles as predictors. That is, MLR and SVR were applied to get “predictions”. As a metrics to evaluate results R-Squared was chosen. In case of MLR, the results can hardly be named as “satisfactory”, since only 3 out of 7 voting profiles can be predicted with “accuracy” (R-Squared value) from 0.53 up to 0.75. Oppositely, SVR provides R-squared values from 0.75 to 0.86 for 6 out of 7 profiles.

Important to mention some limitations of the conducted analysis. First, assumptions of the methods used should be studied more carefully. Although, using NMF as a clustering algorithm allows to partly avoid multicollinearity (import for the adequate usage of MLR) or SVR, in its turn, is less strict then MLR and has non-linear kernels. Secondly, to see more clearly the connection between demographics and voting behavior, recent datasets in demographics should be studied. Currently some data, income values, for example, available only for 2014.

To conclude, it seems that microtargeting might work in The Hague. Using latest knowledge in machine learning and publicly available datasets it is possible to get satisfactory estimates. Therefore, decision-makers should be aware of such technologies.

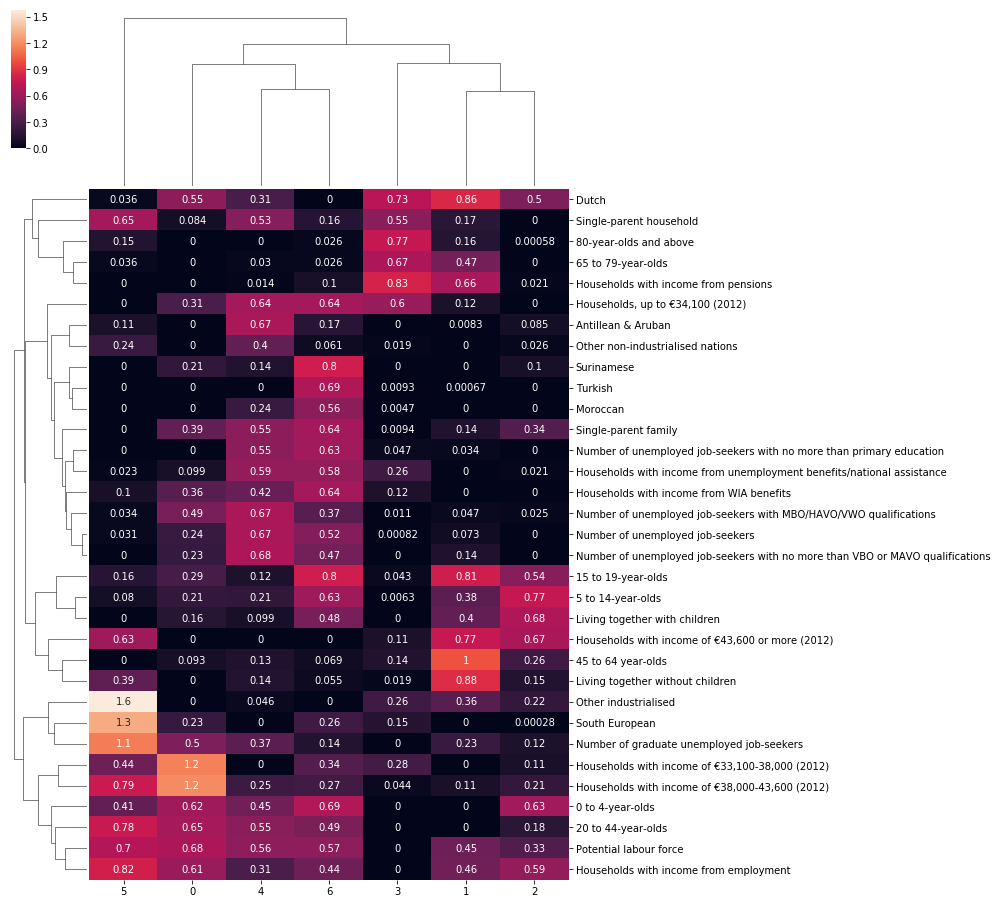
****

Figure 1. “Clustering” voting preferences into profiles

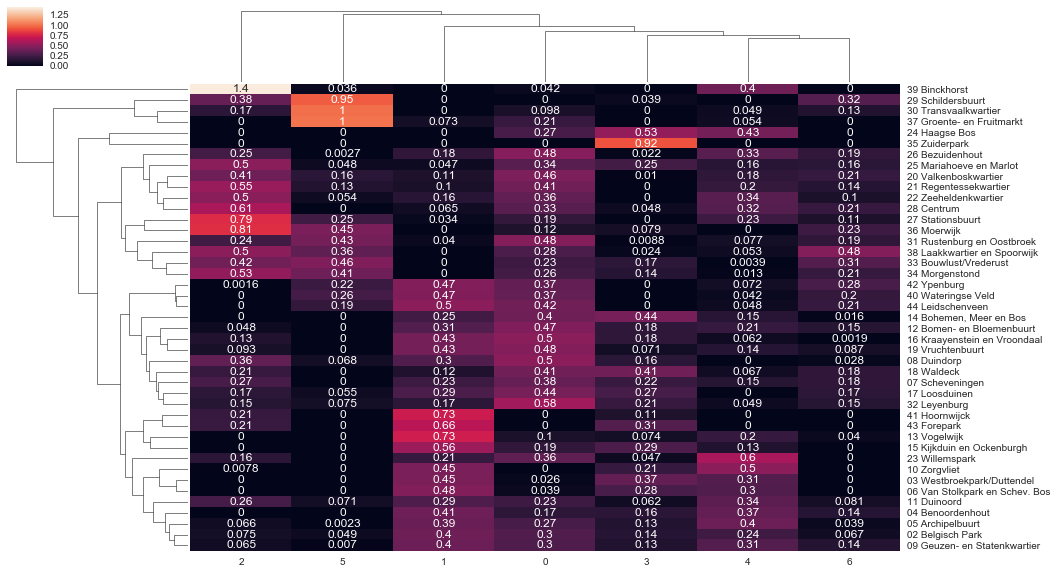


Figure 2. Distribution of demographics profiles across the districts



Figure 3. “Clustering” voting preferences into profiles

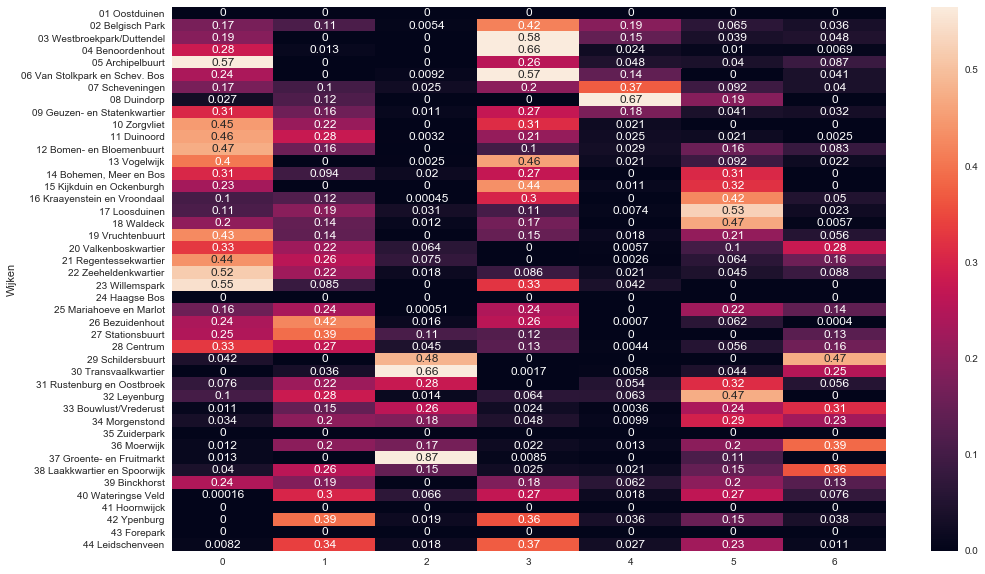


Figure 4. Distribution of voting profiles across the districts

# Mapping The Hague local election with folium

The results of our analysis can be mapped into interactive visualization using folium geomapping library in python as can be seen in Figure 5. The idea behind this visualization is to show in interactive way the results with real map of The Hague district.

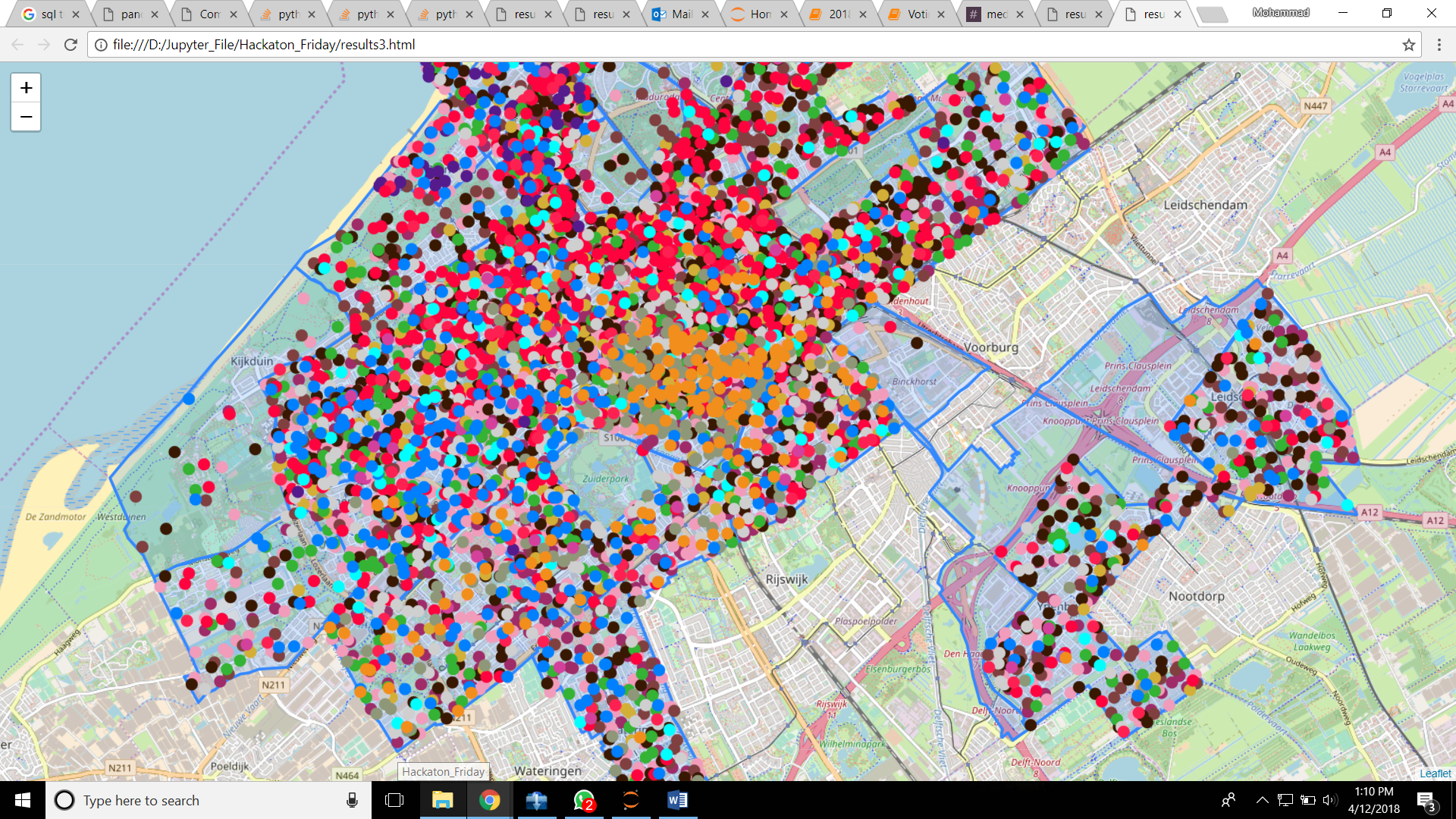


Figure 5. Distribution of voters in The Hague districts

The resulted map was in HTML format so it is easy to open and can just be opened in any computer that have internet browser. The dot (points) represent the people and the color show which political parties that particular people choose. Since the total number of voters in The Hague contained almost 200,000, this resulted map were scaled down such that **each dots (points) represents 30 people**.

User can easily zoom in and zoom out to see detailed spread of voters in their own district as seen in Figure 6. User can also click the dots (points) to show the detailed information of that particular citizen. Each citizen will have unique number and show which political parties that those citizens choose. Although right now it is not connected to our **profile** results yet and **assigned** **randomly**, the information will also try to show which profile of that particular citizen belong to correspond to its voting behavior.

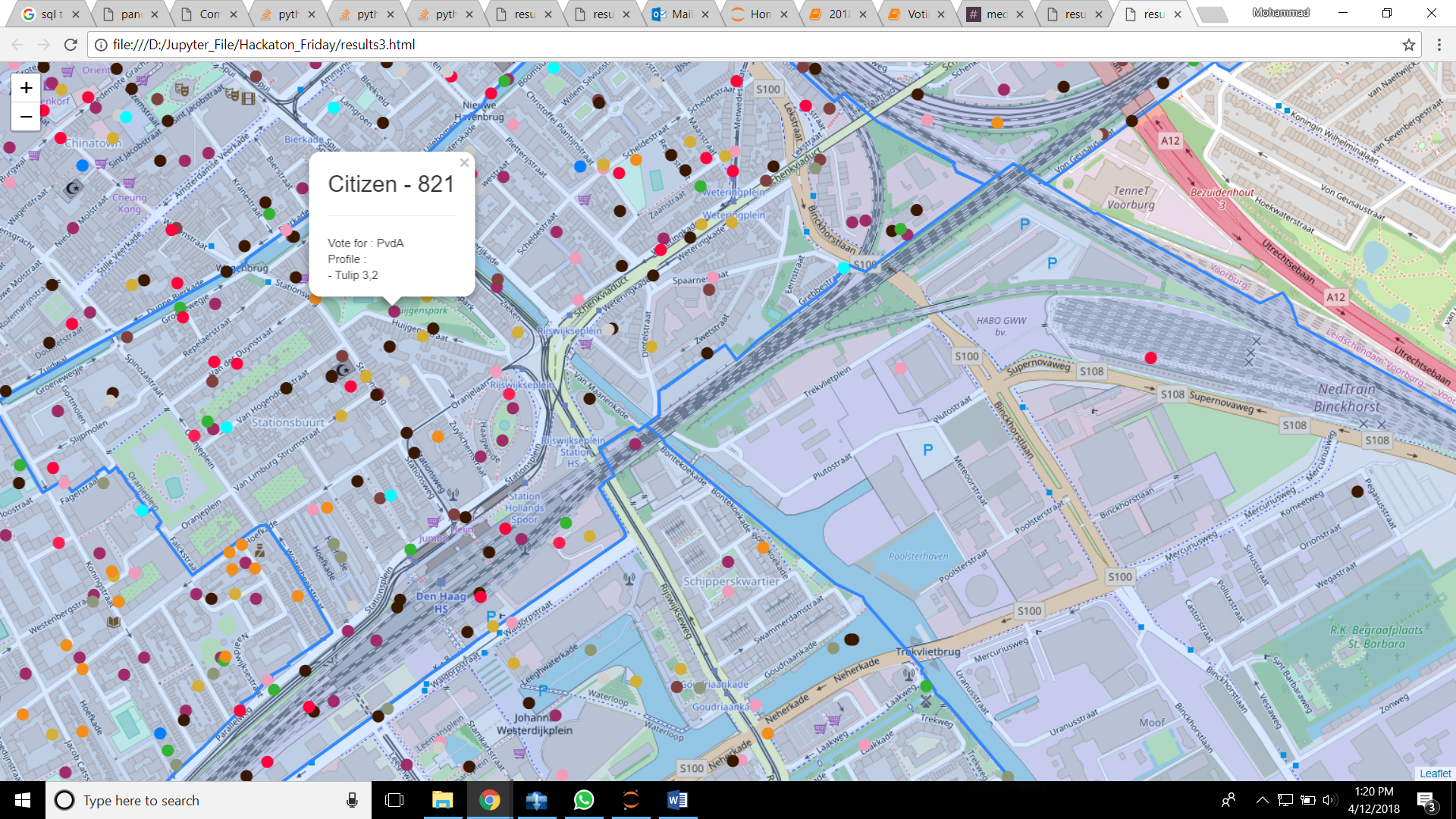


Figure 6. Distribution of voters in The Hague districts with detailed information