

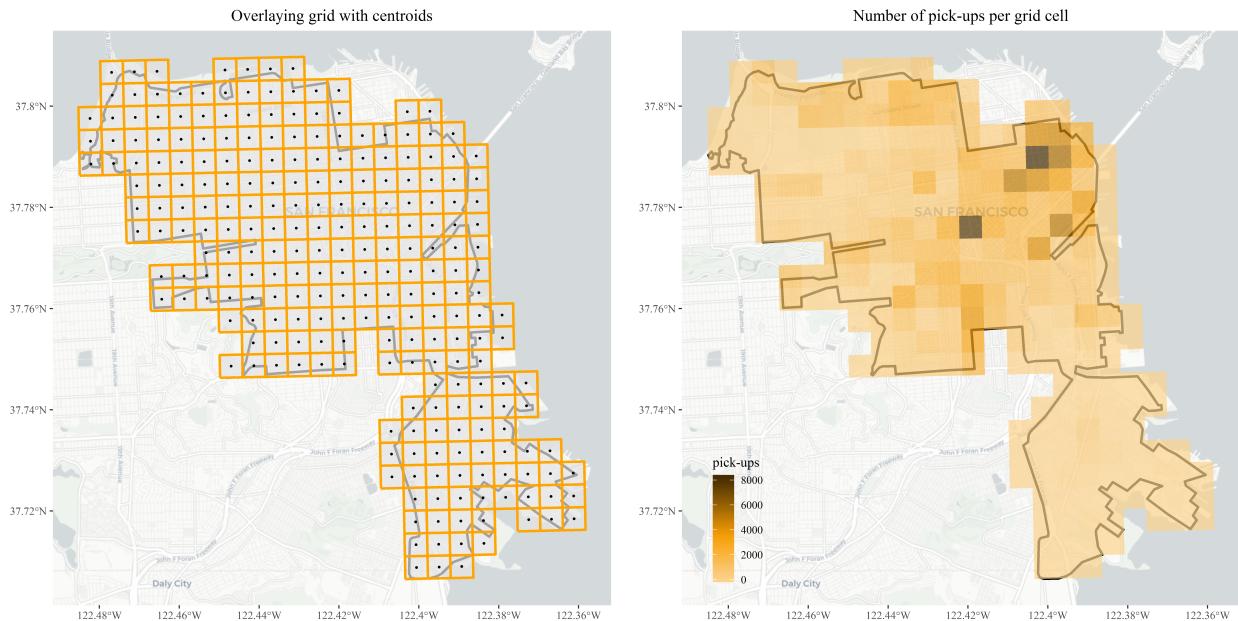
# Chapter 5: Results & Discussion

Draft version

This chapter presents and discusses the results of the experiment described in Chapter 4. It is structured as follows. The first section shows the clusters that resulted from the cluster loop, along with their main characteristics, and the chosen locations of the model points. Section two provides the structure of the models that were build in the model loop, and the residual diagnostics for each them. Then, the third section covers the accuracy of the forecasts, while in the fourth and last section, the limitations of DBAFS are discussed, and recommendations for possible improvements are given.

## 5.1 Clustering

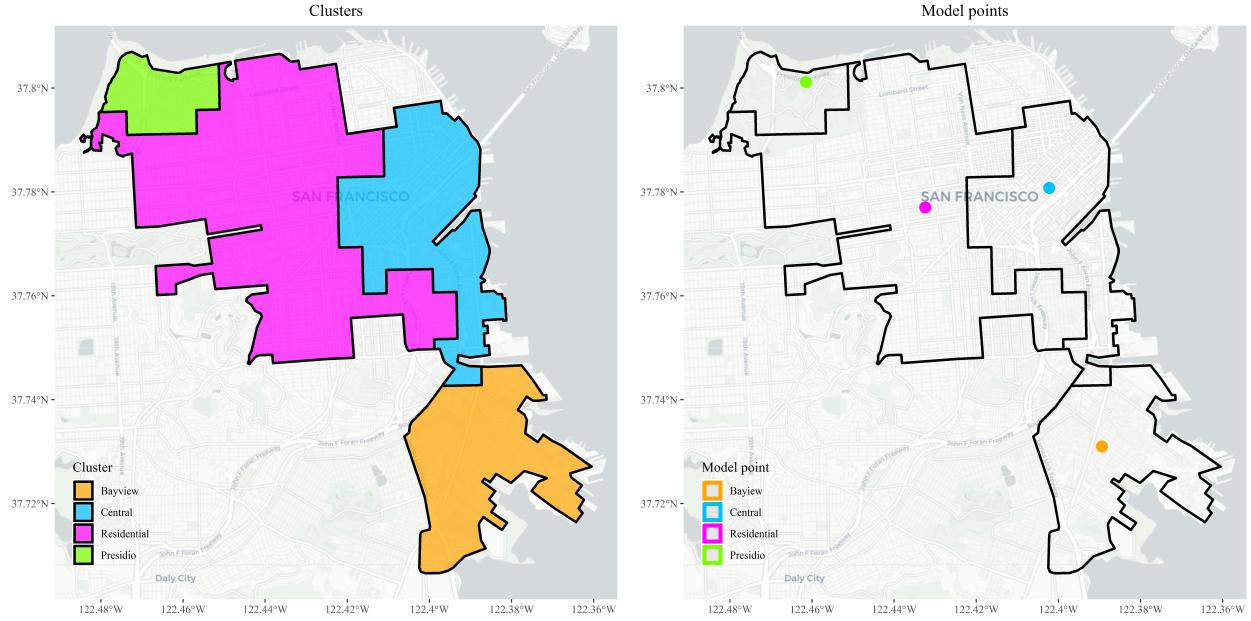
Figure 5.1a shows the grid overlaying the JUMP Bikes system area in San Francisco, including the centroids of all the grid cells. In total, the grid contains 249 grid cells, each 500 meter high and 500 meter wide. Figure 5.1b shows the calculated number of pick-ups per grid cell, during the training period. On average, there were 938 pick-ups per grid cell. The maximum number of pick-ups in a grid cell was 8189, while in 14 of the 249 grid cells, there were no pick-ups at all. It can be seen that high counts of pick-ups occur in the grid cells along the diagonal axis from south-west to north-east. Mainly in the south-eastern corner of the system area, the usage intensity is very low.



The automatic procedure of defining the number of clusters  $k$  and the mixing parameter  $\alpha$  in the spatially constrained hierarchical clustering, lead to a definition of  $k = 4$  and  $\alpha = 0.6$ . This resulted in a partition containing four fully spatial contiguous clusters. The geographical outlines of these clusters are shown in Figure 5.2a. The centroid of all grid cell centroids in each cluster, weighted by the number of pick-ups in the corresponding grid cells, are shown in Figure 5.2b. These weighted centroids serve as the model points in DBAFS.

Roughly speaking, and based on a large study of neighborhood indicators in San Francisco (San Francisco Department of Public Health 2014), the four clusters can be characterized as follows. The blue cluster mainly contains neighborhoods where the residential density is high compared to the job density, and the area zoned for commercial usage is relatively small. The blue cluster forms the city center of San Francisco,

containing the neighborhoods with the highest population densities, but also with a relatively high job density compared to the residential density, and large areas zoned for commercial usage. The orange cluster covers the Bayview/Hunters Point neighborhood, which is a rather isolated area, with a high percentage of low-income households and relatively high crime rates. Finally, the green cluster covers the Presidio Park, a recreational area with few inhabitants, and a relatively high number of bike lanes. For the sake of clarity, the purple, blue, orange and green clusters are from now on referred to as the *Residential*, *Central*, *Bayview* and *Presidio* clusters, respectively.



## 5.2 Model building

## 5.3 Forecasting

## 5.4 Limitations & Recommendations

## **References**

San Francisco Department of Public Health. 2014. “The San Francisco Indicator Project.” <https://www.sfindicatorproject.org/>.