Week 7 Report

A Clustering Model Based on Japan’s housing transaction data

37237256 薛丁銘

1. **Background**

Similar to last week, I decide to do a cluster model by myself. This time I have got a dataset of Japan’s housing transaction. Then I clustered the transaction into 10 clusters, showing the 10 most frequent housing transaction areas in Japan.

1. **Dataset**

Two datasets are involved in this task. The first one is the detailed housing transaction data in Japan for the last four month with the machi code of the houses as shown in Figure 2.1. In order to cluster the locations, another dataset with latitudes/longitudes of Japan machi code was prepared as shown in Figure 2.2.



Figure 2.1 The Housing transaction data

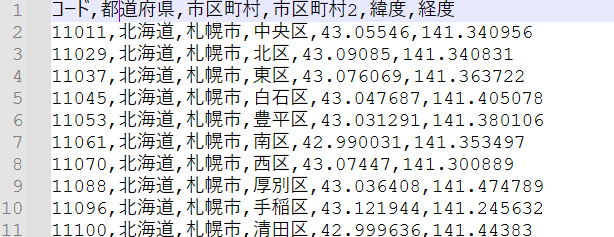


Figure 2.2 The latitude/longitude data

1. **Clustering Model**

In the first part I conducted the data preprocess including loading the data and merge the transaction data with the location data. There was something wrong with the length of the machi code so I had to convert the formation. The preprocess code is as shown in Figure 3.1.

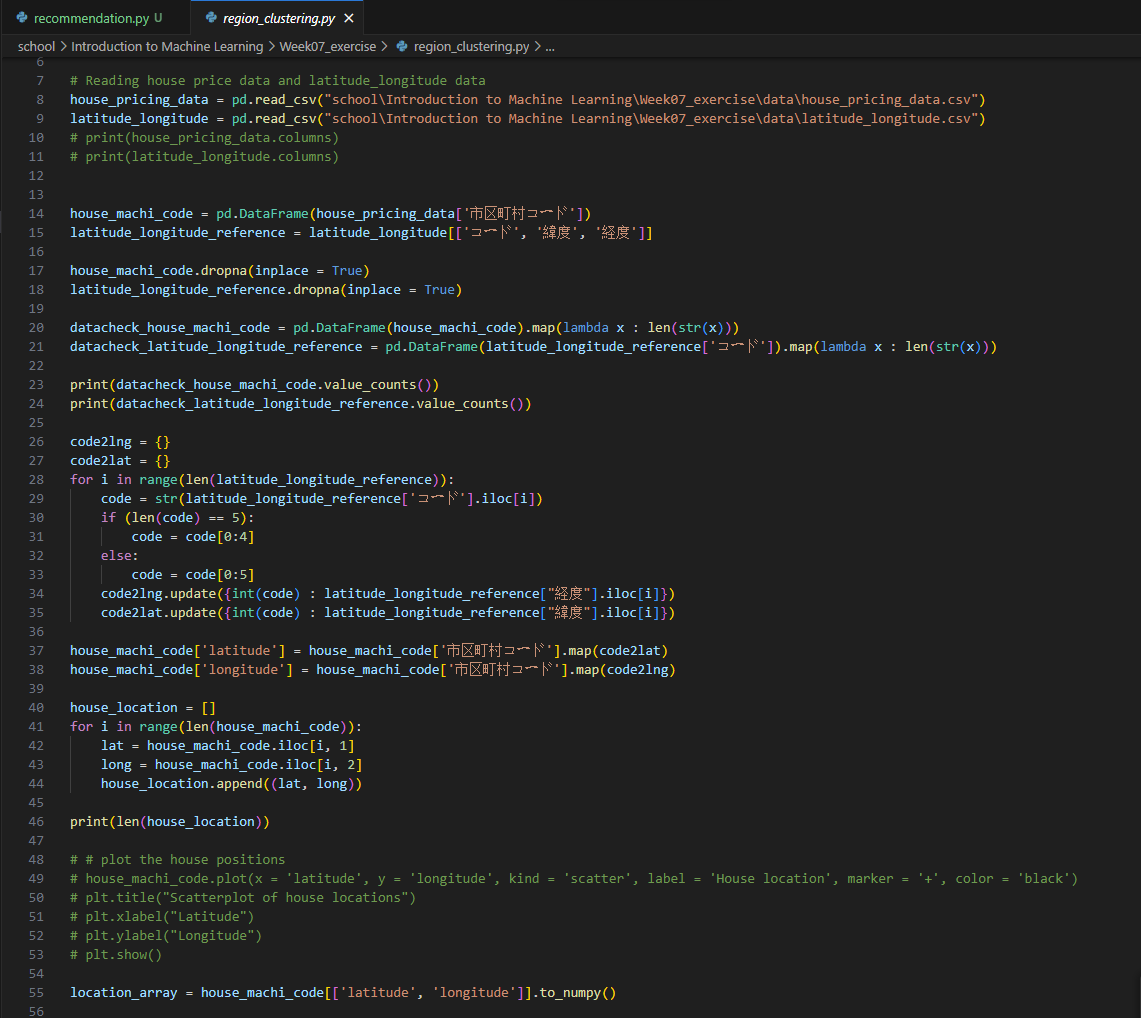


Figure 3.1 Preprocess of the data

After the preprocessing, I drew I map of the transaction data. The distribution of the transaction data looks like the map of Japan as shown in Figure 3.2. Then I used kmeans to cluster the housing transaction data into 10 clusters. I also drew another map of the clustering result with different colors. The result map is shown in Figure 3.3 and the clustering code is shown in Figure 3.4.

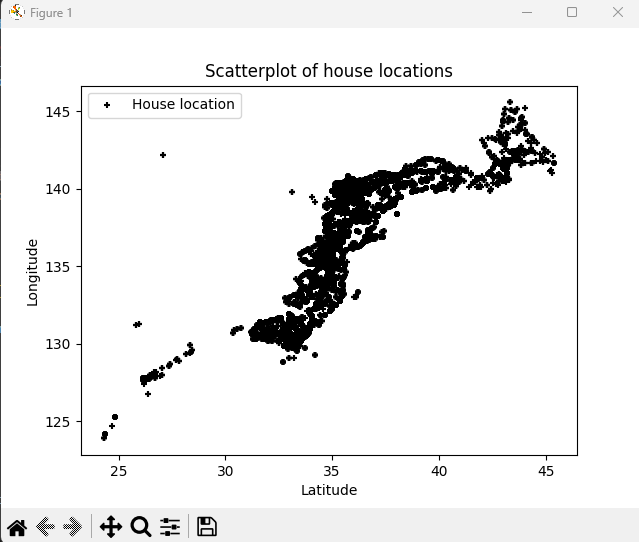


Figure 3.2 Shape of the transaction data

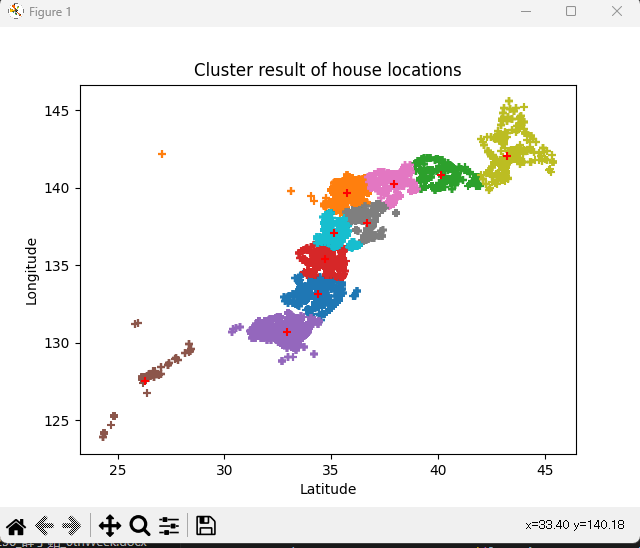


Figure 3.3 Shape of the transaction data

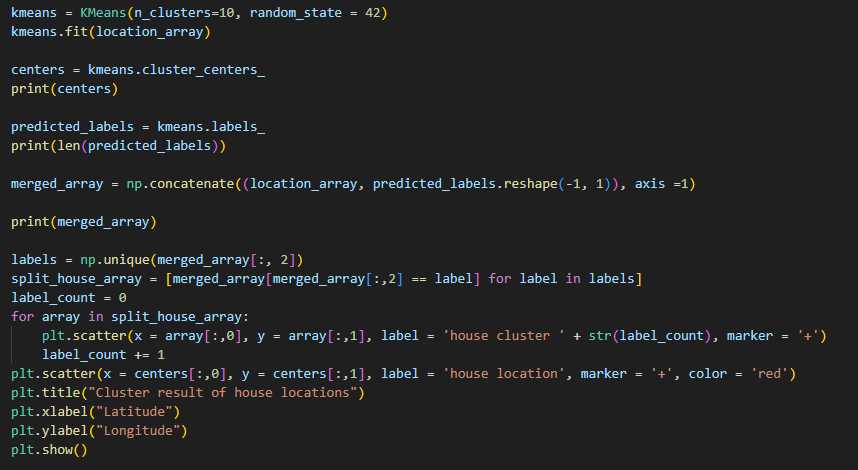


Figure 3.4 The clustering code

Finaly I ran the evaluation code as shown in Figure 3.5 and the results is shown in Figure 3.6. There are no other cluster methods for comparing so the clustering evaluation is just for reference. The evaluation took longer time than I expected. I guess it was because the dataset was too large.

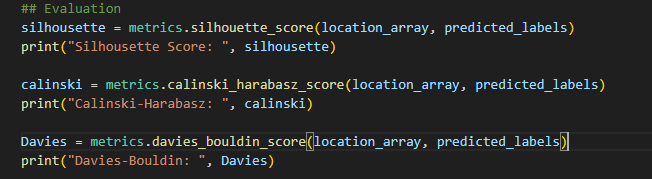


Figure 3.5 The evaluation code

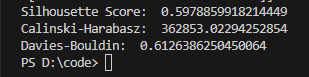


Figure 3.6 The evaluation result