Problem Formulation

The question I try to resolve using machine learning is: Given the basic information about an apartment (i.e., date of buying, postal code, number of rooms), what is the approximate price of that apartment?

The datapoints in the problem are the apartments. The apartments are characterized by several traits: the time when the apartment price was estimated, the postal code, and the number of rooms. Since all these three traits can be obtained easily, they will be used as features for one datapoint. The label is the price of the apartments at the corresponding time.

For the data of this project, there is a record of the average prices of old dwellings in housing companies and numbers of transactions by postal code area from the first quarter of 2010 until the last quarter of 2021. Since my interest is solely in the price of the apartments in capital region (i.e., Helsinki, Espoo, Vantaa, Kauniainen), the number of transactions and information about other postal code areas are not included in my dataset. The original data can be downloaded from the database of Statistics Finland [1].

[1] https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__asu__ashi__nj/statfin_ashi_pxt_112p.px/