

projet_2pages_detail

Group:9 CHEN , XU

My Dataset and the problem

This Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran.

It contains 8 project physical and financial variables(which is form v1 to v8 in data set), 19 economic variables and indices in 5 time lag numbers (from v11 to v29 in data set), and two output variables that are construction costs and sale prices.

Obviously, this dataset should be treated as an Linear Regression Problem, because the attributes and the goals are continuous, and we want to predict the construction costs and the the sale prices with the different physical & financial variables and the economic variables.

The dimension of the data: “observation” “variable”

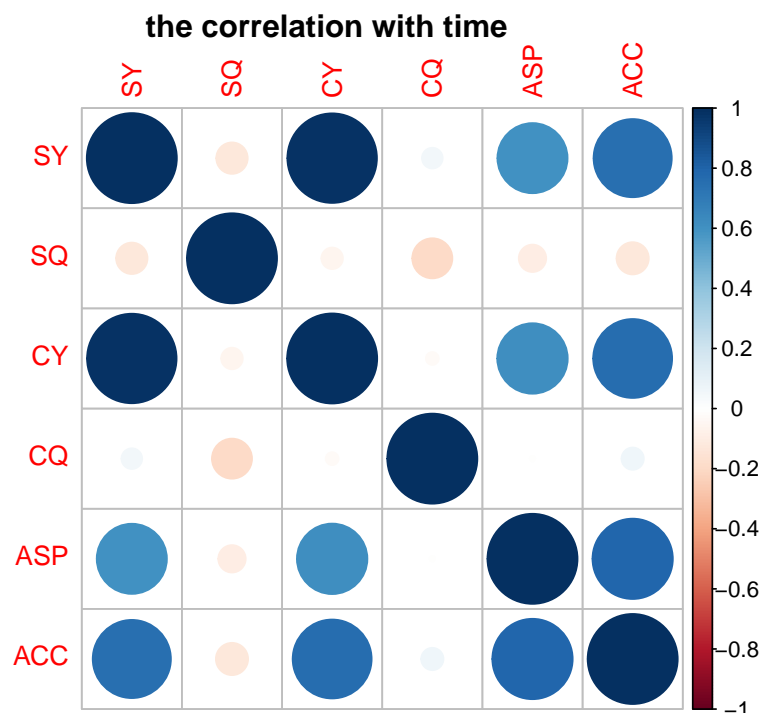
[1] 372 109

The variables

the target variables

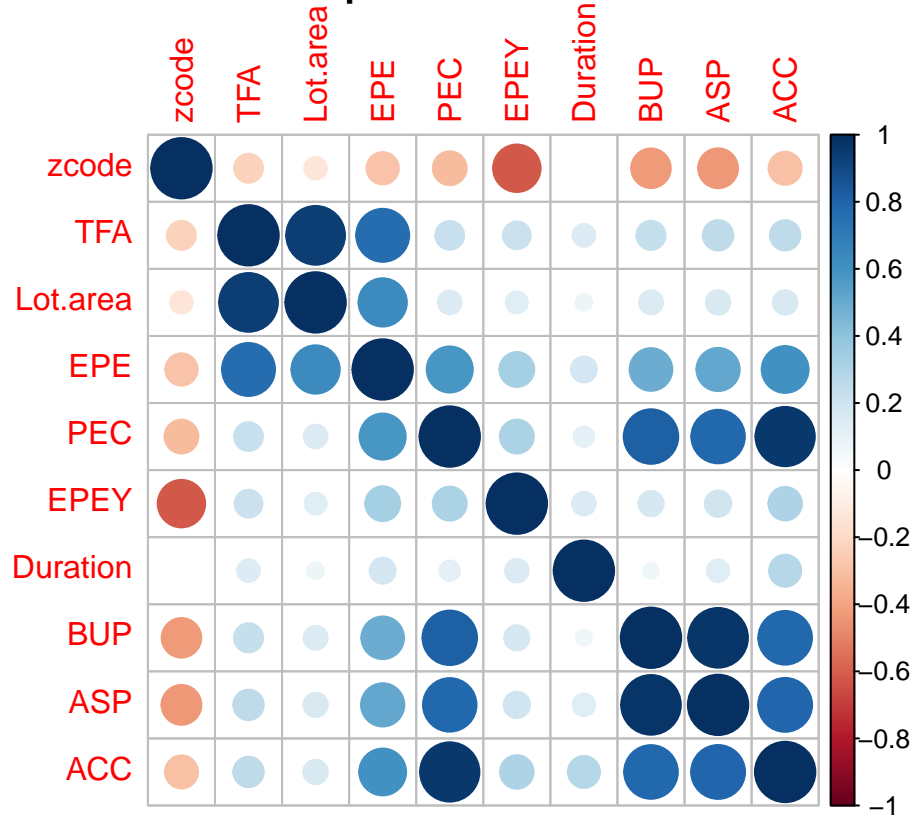
The variables ASP and ACC are the target variables who is actual construction costs and sale prices. At first, we check the correlation between the target variables with others:

For the others variables, there are three groups:



So for this dataset, the first 4 variables are “start year”, “start quarter”, “completion year” and “completion quarter”. These four variables describe when they build this building. And there is a relation between the variables about year and the targets. What’s more, the interval between the start year and the end year maybe an important variable.

the correlation with physical & financial variables



There are so many variables so I’d like to introduce the variables which has more correlation. We find that the variables zipcode, EPE, PEC and BUP are more important, which are “Project locality defined in terms of zip codes”, EPE: “Total preliminary estimated construction cost based on the prices at the beginning of the project”, PEC: “Preliminary estimated construction cost based on the prices at the beginning of the project” and BUP: “Price of the unit at the beginning of the project per m2”.

ECONOMIC VARIABLES AND INDICES

These variables are in 5 time lag numbers, so there are almost 95 variables. Each 19 variables describe the economic situation of the society for one instance. So perhaps they will influence the target variables, or, which is more possible, there are many variables who are not interesting for the target variables.