Analyzing what might influence the housing value

From INFSCI2415, I learnt much knowledge about data visualization. To use it in practical fields and combine data visualization with real life, I looked through various information and data source. Finally, I determine to study the field in Real estate. According to the data source: <https://www.kaggle.com/datasets/camnugent/california-housing-prices/discussion?resource=download> , we found there are several attributes that may influence the medium price of the houses in one block. In this report, we began to study in two different parts:

1. Study the relation between one single attribute and median house value in one block of California.
2. Analyze the relation between two or more combined relation and median house value.

Especially, median of house price could be more appropriate to represent the general situation rather than average value of houses.

In this report, I choose four different attributes, which are population, median\_income, housing\_median age and total bedrooms. Before implementing these plots, based on my experience, I assume median\_income, housing\_median age and population may have strong relation to house prices. While total bedrooms may have nothing to do with housing prices. By using python, I will generate four scatter plots separately. These figures are shown below:

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图片包含 图表

描述已自动生成

图表, 散点图

描述已自动生成

These four plots have completely overthrown previous assumption. It is obviously that median\_income, total bedrooms, and population will influence the house price in one block of California. The correlation between median\_income, total bedrooms, and population and housing price presents a certain degree of positive correlation, because there are more pots stayed on higher place as x increased. On the other hand, we couldn’t find any correlation between housing age and housing price.

However, when it comes to housing\_median\_age, we cannot find any correlation between housing\_median\_age and house price.

After we study the single element, then we began to combine two attributes which can influence the house price together to see the trend. Firstly, we do the normalization to two attributes named median\_income and population. The function of normalization is (x – min) / (max – min). Then We defined the processed median\_income weight as 0.5 and the processed population weight as 0.5. After that we should generate the plot which shows the correlation between the combined attribute and housing price.

图表, 散点图

描述已自动生成

According to this data visualization, people can find that median\_income, total bedrooms, and population are key factors which will influence the house prices in a block of California. Also, even though we look through two elements together like median\_income and population, we can also get their relation between the combined attribute and housing price.

Due to the universality of real estate, if the data remained constantly, we could know the housing price by attributes. In one location, if we have the data of local median\_income, total bedrooms, and population, we can get the relative price in one place, which will be more precisely than personal experience.

Github Link: <https://github.com/StuRinDQB/INFSCI2415>