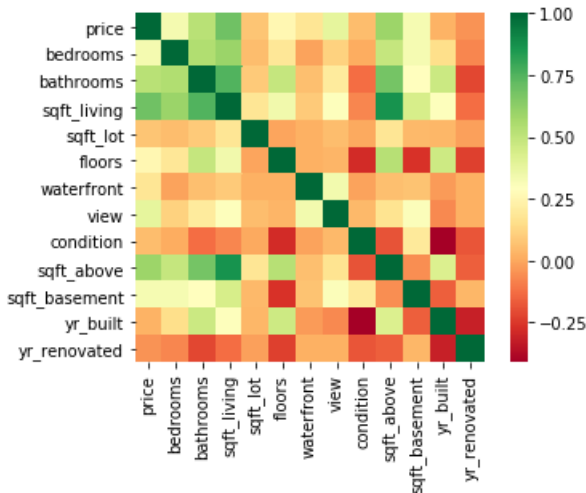# House price in the USA

28 July, 2019

Motivations

- The real estate markets, like those in the USA, present an interesting opportunity for data scientists to analyze and predict where property prices are moving towards.

- Prediction of property prices is becoming increasingly important and beneficial.

- Property prices are a good indicator of both the overall market condition and the economic health of a country.

The problem

In this project we want to estimate the **price** of an house that have to been sold based on information regarding houses that have been previously sold. These information include the **location** of the house, its **surface**, its **condition**, the **number of rooms** and the **number of floors** and the **year** of construction.

I use the freely available data set present on Kaggle at
`https://www.kaggle.com/shree1992/housedata`

Correlation between the data

$\Rightarrow$ the most correlated with price are **sqft living** and **sqft above** $\Rightarrow$ I will focus on them !!!

sqft living vs price

Price vs sqft_living

there is a **linear correlation** between these two variables !!

sqft above vs price

Price vs sqft_above

there is a **linear correlation** between these two variables !!

Conclusions

- We found a linear correlation between the price of the house and its surface (living and ground floor)

- It would be interesting to extend such analysis taking into account also different independent variables.

- It would be interesting to extend such analysis taking into account a bigger data set