Ironhack Bootcamp Final Project, Regression

Equipo 4

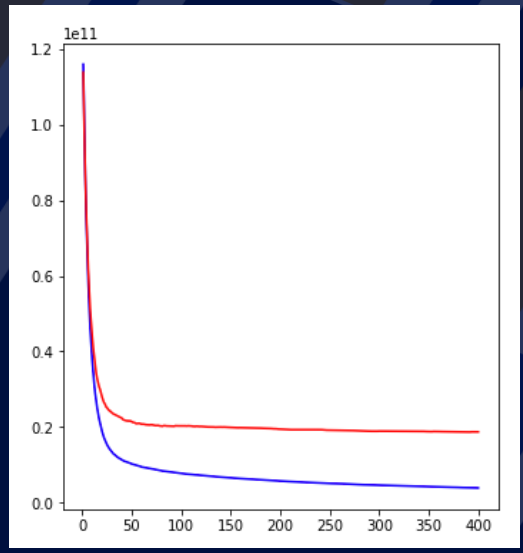
Model

- Analyse all the provided data
- Predict an estimated Price
- Plot different graphics with all the characteristics of the properties

Linear Regression

- Train 90%, test 10% (around 26000 rows of data is more than enough)
- Initially we score 70%, not valid
- Gradient Boosting Regressor, we obtain a valid score of 86.3%

```
In [30]: reg.fit(x_train,y_train)

Out[30]: LinearRegression()

In [31]: reg.score(x_test,y_test)

Out[31]: 0.7053915417435492

In [32]: clf = ensemble.GradientBoostingRegressor(n_estimators = 400, max_depth = 5, min_samples_split = 2,
                learning_rate = 0.1, loss = 'ls')

In [33]: clf.fit(x_train, y_train)

Out[33]: GradientBoostingRegressor(max_depth=5, n_estimators=400)

In [34]: clf.score(x_test,y_test) #our score is right above 85% so it should be valid

Out[34]: 0.8627271940148936
```



Model analysis

- Properties over 650k, mostly depends on:
  - Latitude
  - Longitude
  - Zip codes
  - Grades

Analysis

- Data_análisis.ipynb document where most of the análisis is done
- Answered all the asked questions in a second notebook