BootCamp Project. House prices.

Made by Jabbo Ramirez
Jennifer Mejia and Hanna Bustamante

What are we accomplishing?

Develop and evaluate a model that predicts house prices in Ames, IOWA by using tools learned throughout the course,

Use all the information given by the Datasets we predict a house price depending on the different criteria inserted into the model.

Tools used

- Python
- PGadmin/SQL
- Jupyter Notebook
- Visual Studio Code
- Tableau
- SQL Alquemy
- Google Slides

Our code can estimate the price of a house as long as we submit **all of the available criteria.** From 5 rooms, to having a pool to even having a fireplace!

What data set are we using?

The dataset used in this project comes from a KAGGLE prediction competition compiled by Dean De Cook for educational purposes.

The data is from a period between 2006 and 2010.

Features 80 different criteria for houses and 1459 residential properties in Ames, IOWA that are taken into consideration when buying a house.

Tools used

Python
Developed a code to clean the data of null and irrelevant values

Jupyter Notebook
Tested the code to make sure all the irrelevant values and columns were erased

PgAdmin
Imported the cleaned CSV files to create a database

SQLAlchemy
Import the CSV file to python and use machine learning on the dataset.

Tableau
Imported the database and created an easy to follow dashboard

Overview of the preprocessing steps

- Dropped features with over 50% missing values
- Filled categorical columns with null values with 'None'
- Filled numerical columns with missing values with mean of entire column
- Dropped columns with low variance
- Applied Label Encoding (all text data becomes numerical) to categorical columns

Machine learning!?

Machine Learning algorithms used
- Linear Regression
- Random Forest Regressor

Which worked best
- Random Forest Regression worked best.

Random Forest Regression:
Score: 0.8980621096596039
MAE: 16616.091589041094
MSE: 714106425.9779499
RMSE: 26722.769803632815

Linear Regression:
Score: 0.8458913960375708
MAE: 21253.85812555094
MSE: 1079578398.3813806
RMSE: 32856.938359825625