

Linear Regression Analysis for the Kings County's (Seattle, WA) House Market.

by Y. Kostrov

Outline

- 1 Overview
- 2 Business Problem
- 3 Data Description
- 4 My Python Package

Overview

The purpose of this project is to analyze a data set containing data about houses sold in Kings County (Seattle, WA).

Overview

During the analysis:

Overview

During the analysis:

- 1 I will perform necessary data wrangling first.

Overview

During the analysis:

- 1 I will perform necessary data wrangling first.
- 2 I will build a Linear Regression Model with one explanatory variable.:

Overview

During the analysis:

- ① I will perform necessary data wrangling first.
- ② I will build a Linear Regression Model with one explanatory variable.:
 - I will check statistical assumptions for the linear regression model
 - I will explain the model, including intercept, coefficient for the explanatory variable, R^2 , and ANOVA

Overview

During the analysis:

- ① I will perform necessary data wrangling first.
- ② I will build a Linear Regression Model with one explanatory variable.:
 - I will check statistical assumptions for the linear regression model
 - I will explain the model, including intercept, coefficient for the explanatory variable, R^2 , and ANOVA
- ③ I will build a Multiple Linear Regression Model with many explanatory variables.

Overview

During the analysis:

- ① I will perform necessary data wrangling first.
- ② I will build a Linear Regression Model with one explanatory variable.:
 - I will check statistical assumptions for the linear regression model
 - I will explain the model, including intercept, coefficient for the explanatory variable, R^2 , and ANOVA
- ③ I will build a Multiple Linear Regression Model with many explanatory variables.
 - I will check statistical assumptions for the multiple linear regression model
 - I will explain the model, including intercept, coefficients for the explanatory variables, R^2 , and ANOVA

Business Problem

- The fair price of the house is a hard quantity to assess.

Business Problem

- The fair price of the house is a hard quantity to assess.
- Both sellers and buyers would like to know the best price for the house.

Business Problem

- The fair price of the house is a hard quantity to assess.
- Both sellers and buyers would like to know the best price for the house.
- Which features of the property would be the best predictors of the value?

Business Problem

- The fair price of the house is a hard quantity to assess.
- Both sellers and buyers would like to know the best price for the house.
- Which features of the property would be the best predictors of the value?
- I will build a regression model that helps predict the value of the house.

Business Problem

- The fair price of the house is a hard quantity to assess.
- Both sellers and buyers would like to know the best price for the house.
- Which features of the property would be the best predictors of the value?
- I will build a regression model that helps predict the value of the house.
- I will, also, check the necessary statistical assumptions for the regression model and explain the model's parameters.

Data Description

- The file called “kc_house_data.csv” in the data folder of the project holds the data for this project.

Data Description

- The file called “kc_house_data.csv” in the data folder of the project holds the data for this project.
- This project will use this data about Kings County's(Seattle, WA) housing market to create Linear Regression Model.

Data Description

- The file called “kc_house_data.csv” in the data folder of the project holds the data for this project.
- This project will use this data about Kings County’s(Seattle, WA) housing market to create Linear Regression Model.
- The data file contains numerous columns with information about properties sold such as price, size of the living area, size of the basement, number of bedrooms, etc.

My Python Package

- While working on this project, I have created my own Python package with helping functions.

My Python Package

- While working on this project, I have created my own Python package with helping functions.
- The most important function in this package is “evaluate_model.py” (in the “src” folder). This function:

My Python Package

- While working on this project, I have created my own Python package with helping functions.
- The most important function in this package is “evaluate_model.py” (in the “src” folder). This function:
 - creates the model from the data frame.

My Python Package

- While working on this project, I have created my own Python package with helping functions.
- The most important function in this package is “evaluate_model.py” (in the “src” folder). This function:
 - creates the model from the data frame.
 - prints out the model summary of Linear Regression.

My Python Package

- While working on this project, I have created my own Python package with helping functions.
- The most important function in this package is “evaluate_model.py” (in the “src” folder). This function:
 - creates the model from the data frame.
 - prints out the model summary of Linear Regression.
 - performs the checks for the statistical assumptions of the Linear Regression.

My Python Package

- While working on this project, I have created my own Python package with helping functions.
- The most important function in this package is “evaluate_model.py” (in the “src” folder). This function:
 - creates the model from the data frame.
 - prints out the model summary of Linear Regression.
 - performs the checks for the statistical assumptions of the Linear Regression.
 - performs a lot of different visualizations.

Modeling

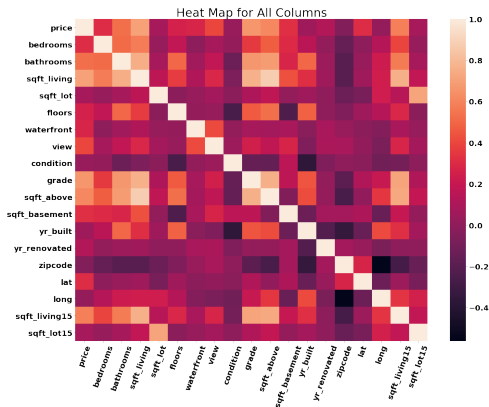
- My first goal was to create linear regression model with one independent variable.

Modeling

- My first goal was to create linear regression model with one independent variable.
- I created the correlation matrix and heat map for visualization purpose.

Modeling

- My first goal was to create linear regression model with one independent variable.
- I created the correlation matrix and heat map for visualization purpose.



Selection of the explanatory variable for the Linear Model

- “sqft_living” has the highest correlation of 0.71 with the “price”.

Selection of the explanatory variable for the Linear Model

- “sqft_living” has the highest correlation of 0.71 with the “price”.
- I build a regression model for the “price” to be predicted by “sqft_living”.