Feature Engineering Real Estate Analytics

Problem Statement

While searching for the dream house, the buyer looks at various factors, not just the height of the basement ceiling or the proximity to an east-west railroad. Using the dataset, find the factors that influence price negotiations while buying a house.
There are 79 explanatory variables describing every aspect of residential homes in Ames, Iowa.

Objectives

- To Assess the data and prepare a fresh dataset for training and prediction

- To create a box plot to identify the variables with outliers

Tasks To Perform:
1) Import the necessary libraries
1.1 Pandas is a Python library for data manipulation and analysis.
1.2 NumPy is a package that contains a multidimensional array object and several derivative ones.
1.3 Matplotlib is a Python visualization package for 2D array plots.
1.4 Seaborn is built on top of Matplotlib. It's used for exploratory data analysis and data visualization.

2) Read the dataset
2.1 Understand the dataset
2.2 Print the name of the columns
2.3 Print the shape of the dataframe
2.4 Check for null values
2.5 Print the unique values
2.6 Select the numerical and categorical variables

3) Descriptive stats and EDA
3.1 EDA of numerical variables
3.2 Missing value treatment
3.3 Identify the skewness and distribution
3.4 Identify significant variables using a correlation matrix 3.5 Pair plot for distribution and density

Project Outcome

• The aim of the project is to help understand working with the dataset and performing analysis.

• This project will assess the data and prepares a fresh dataset for training and prediction
• To create a box plot to identify the variables with outliers