**Introduction**

House Prices Differ upon many variables , maybe it’s possible to guess the range of the prices

Without asking by yourself in a particular town or area.  
we will dive through House prices Dataset and get some insights that can be useful through our analysis .

We will perform exploratory analysis on the whole dataset that will include answering our research questions and finally we will end up training a linear regression model will help us Predicting House prices.

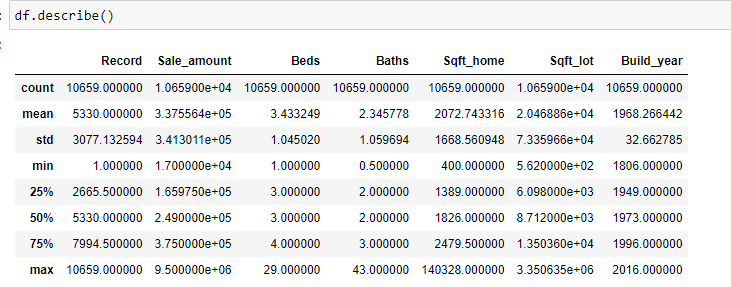
Our Main steps:

·      Explanatory data analysis (EDA) and Descriptive analysis

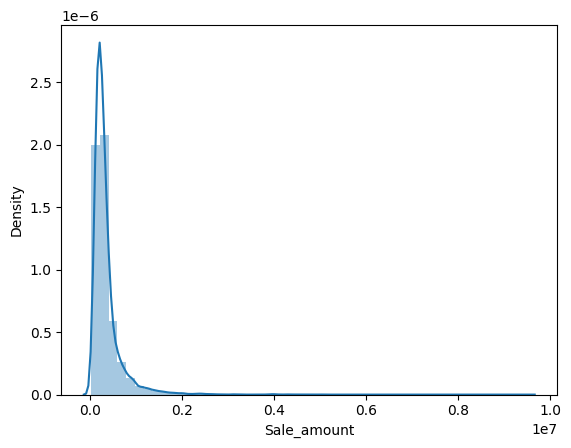
·      Feature Selection and Preprocessing

·      Modelling

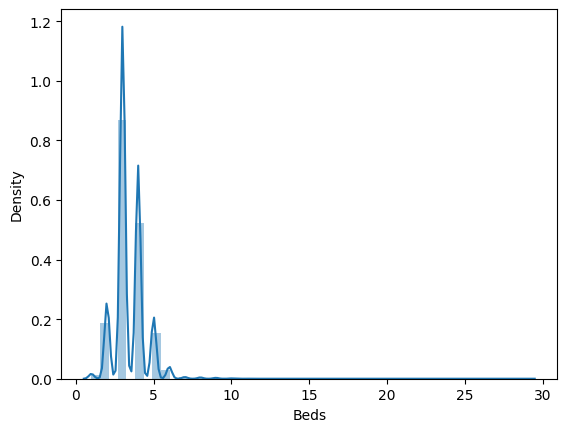
**Descriptive analysis table**

****

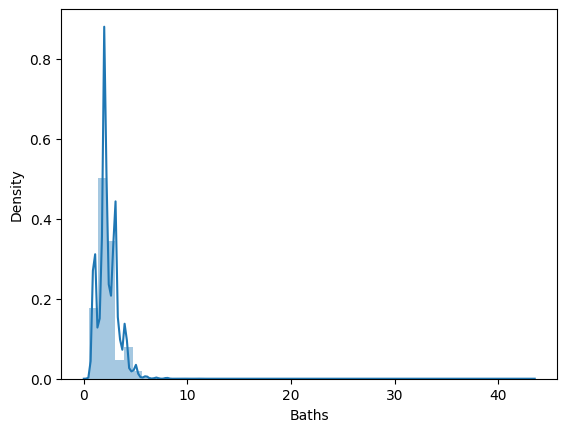
Distribution of Sale Amount Variable which is our target Variable

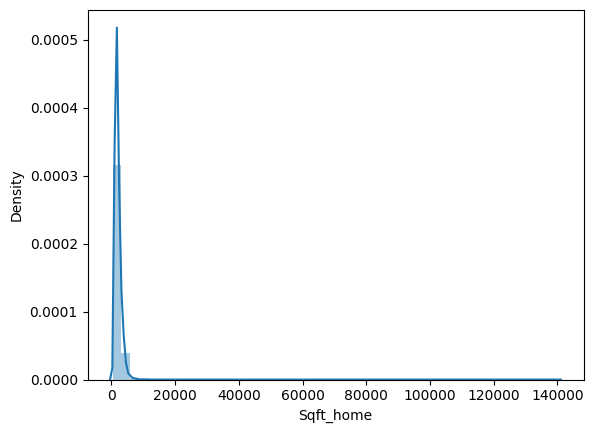


Distribution of number of beds

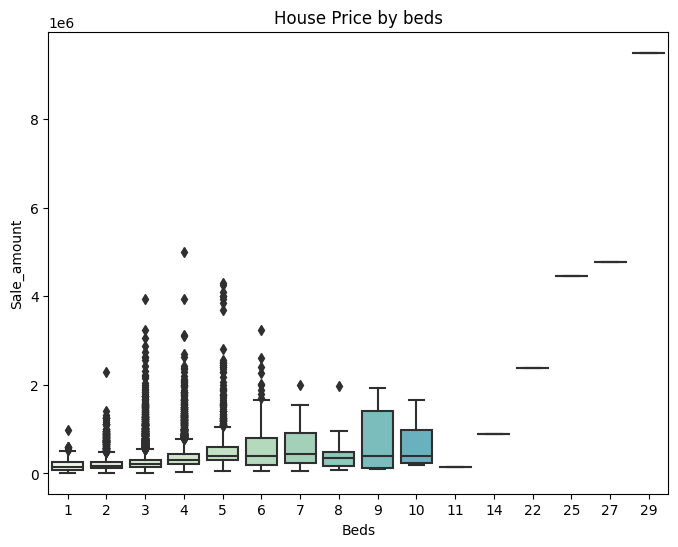


Distribution of number of beds

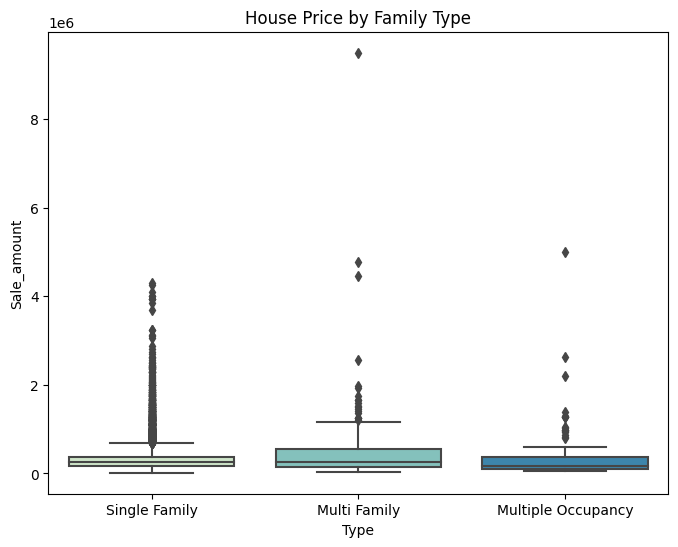


Distribution of sqft\_home

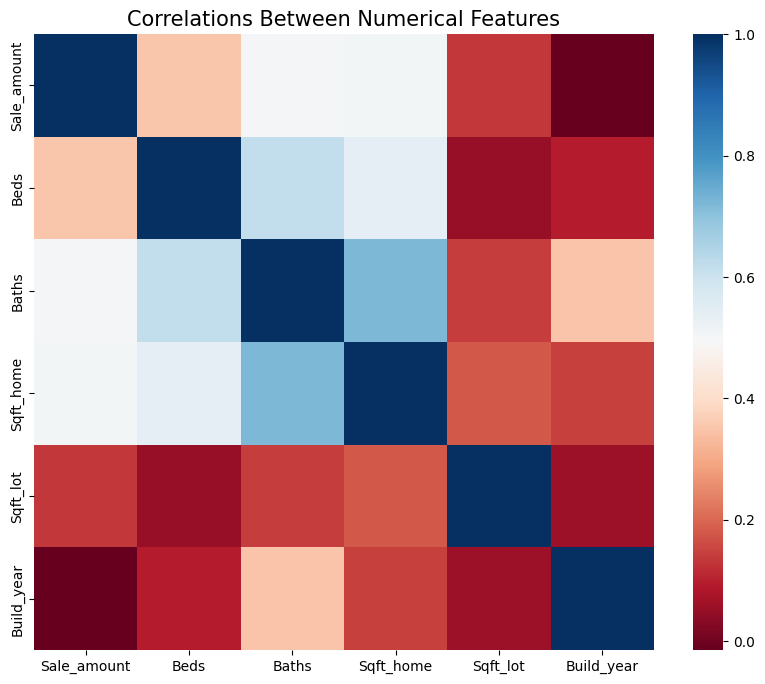
Box Plot of number of beds and Sale amount

****

Box Plot of Family Type against Sale\_amount

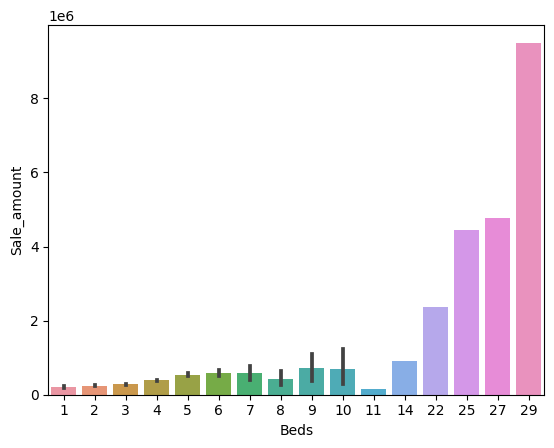
****

Correlation between all the numerical features

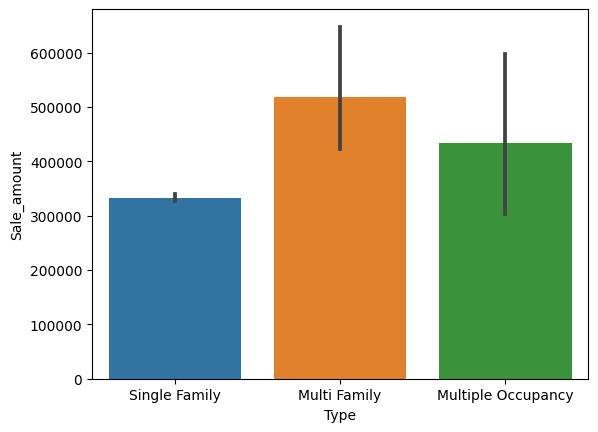
****

Bar Plot of number of beds against Sale\_amount.

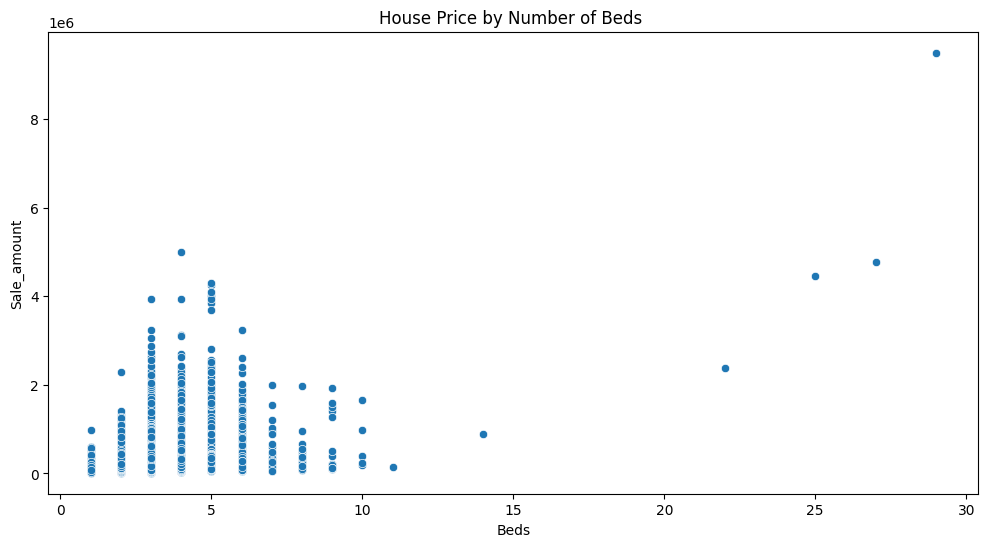
as number of beds increase sale\_amount increases

****

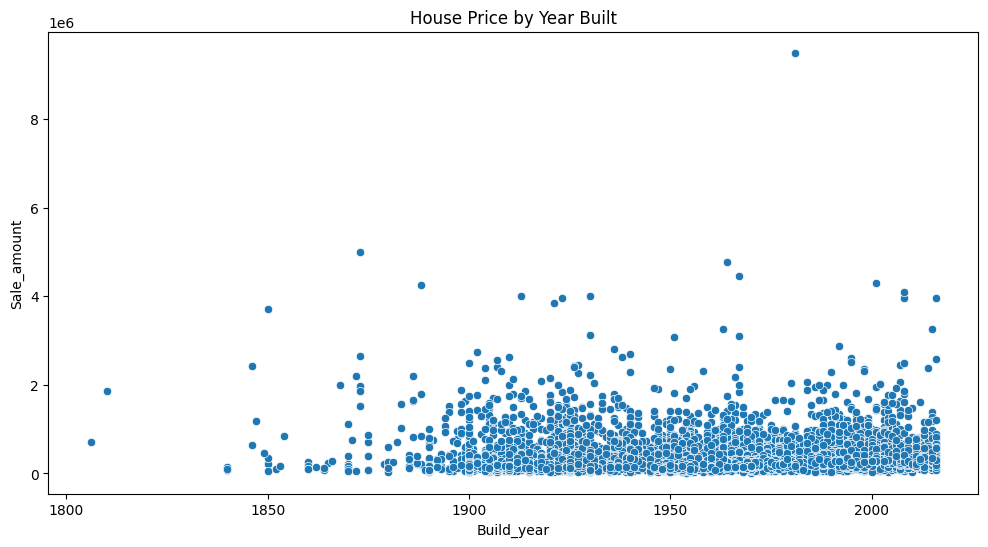
Bar Plot of number of Type of Families against Sale\_amount.

****

Scatter Plot of house Prices and Number of Beds

****

Scatter Plot of house Prices and Number of Beds



**Research Questions**

In order to help People whose willing to investigate the prices of the houses to make better decisions buying their dream house.

We will focus on important variables that has great influence on our research and prediction process.

**Anova Part**

1. Is there a significant difference in the House Sale Price for types of the families category?
2. Is there a significant difference in the House Sale Price for number of bedrooms category?

From the Above research questions we concluded those hypothesis

First question:

H0 = The mean Sale\_amount is equal for types of the families category.

H1 = At least one of the families category has a mean price that is not the same as the other condition categories.

Result :



Our Null hypothesis is rejected as we got P-value less than 0.05 , and we conclude that At least one of the families category has a mean price that is not the same as the other condition categories

Second question:

H0 = The mean Sale\_amount is equal for number of bedrooms category.

H1 = At least one of the bedrooms category has a mean price that is not the same as the other condition categories

Result:



Our Null hypothesis is rejected as we got P-value less than 0.05, and we conclude that Number of bedroom number changes the mean price in at least one category of the bedrooms

**Regression Part**

1. Can we Predict the House Sale Price using Our features of each record of houses ?

Answering this question we will go through few modeling Steps:

**Feature Selection :**

We will drop Sale\_date , Record columns as they have no significant to our target variable.

**Data Preprocessing :**

1-Encoding the categorical Variable using One Hot Encoding in order to feed the data to our model.

2-Normalization for the numerical Data

**Data Split :**

Splitting our data to Train and test with 80% for train, 20% for test.

**Model:**

We used Linear regression model as it’s simple and good choice for continuous numerical targets.

**Result:**

We evaluated our model using few metrics like MAE , MSE , RMSE , R2 score  
and this our Results :

