Data Analytics Program : DataScience Track

Capstone Project : Real estate price prediction

Student Name : Dana Hammad

Date : 17th Jan 2025

* Introduction

Real estate price prediction is a critical task in the housing market, enabling buyers, sellers, and investors to make informed decisions. This dataset provides key attributes that influence property prices, including transaction date, house age, distance to the nearest MRT station, number of nearby convenience stores, and geographical coordinates (latitude and longitude). The target variable is the house price per unit area, which reflects property valuation in the market. By leveraging this dataset, we aim to build a predictive model that identifies relationships between these factors and property prices. Such a model can provide valuable insights into market trends, support decision-making for potential investments, and enhance transparency in real estate transactions. This analysis highlights the significance of location, accessibility, and amenities in determining property values, offering a comprehensive approach to understanding the housing market.

4o

## .

* Problem

## The real estate market is dynamic and influenced by various factors such as location, accessibility, age of properties, and proximity to amenities. Accurately predicting house prices is a challenging task due to the complexity and interplay of these variables. This dataset presents the opportunity to address this challenge by utilizing key features, including transaction date, house age, distance to the nearest MRT station, number of convenience stores, and geographical coordinates, to predict the price per unit area of properties. The problem lies in identifying and quantifying the relationships between these factors and property prices to develop a reliable predictive model. Solving this problem will not only aid stakeholders in making data-driven decisions but also help reduce risks in real estate investments and improve market efficiency.

* Data set

Source : Kaggle.com

Link: <https://drive.google.com/file/d/1SC0LZR4uwlzGEaZJjSXTSANyYzIehr5p/view?usp=sharing>

## 

## 

## 

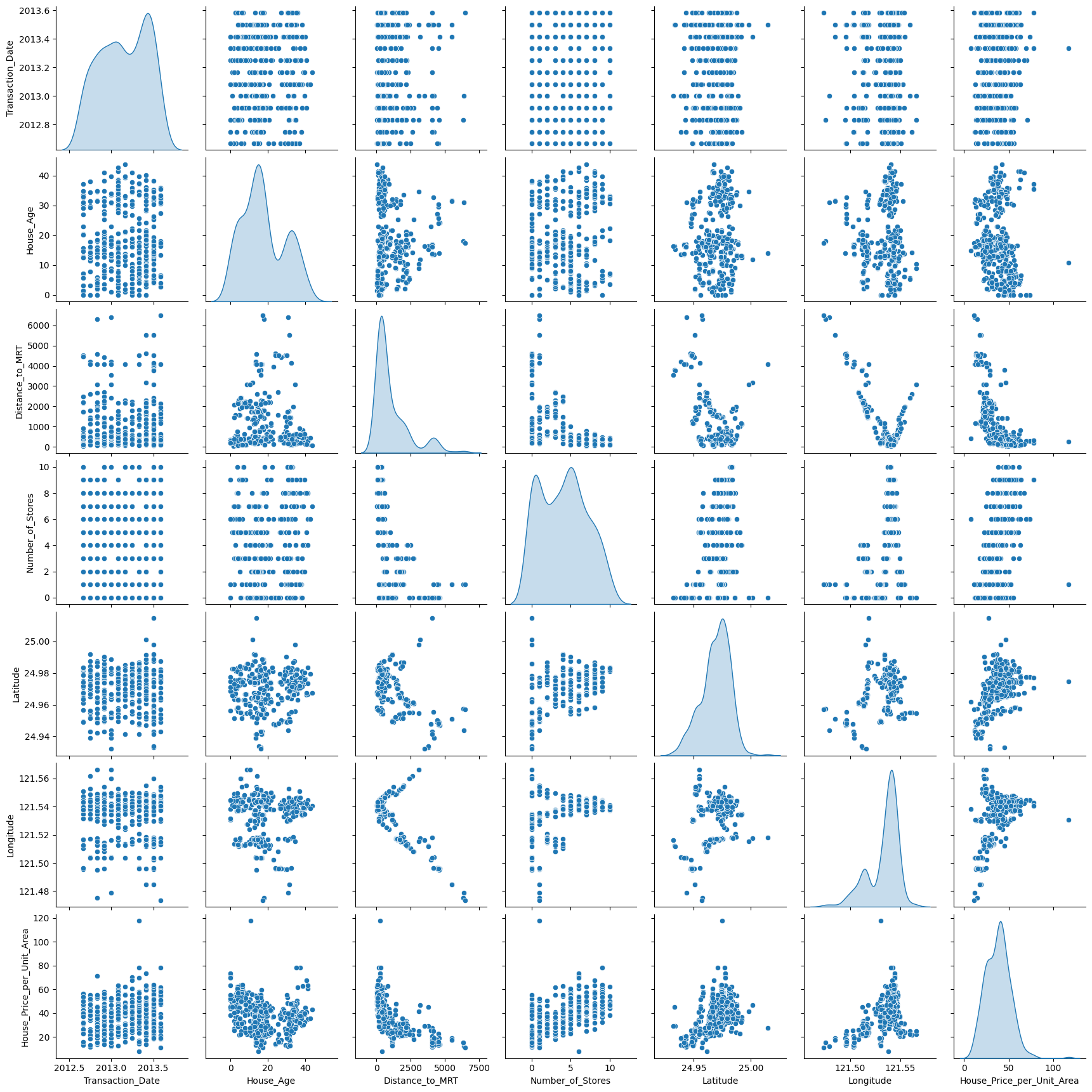
## Attribute Information

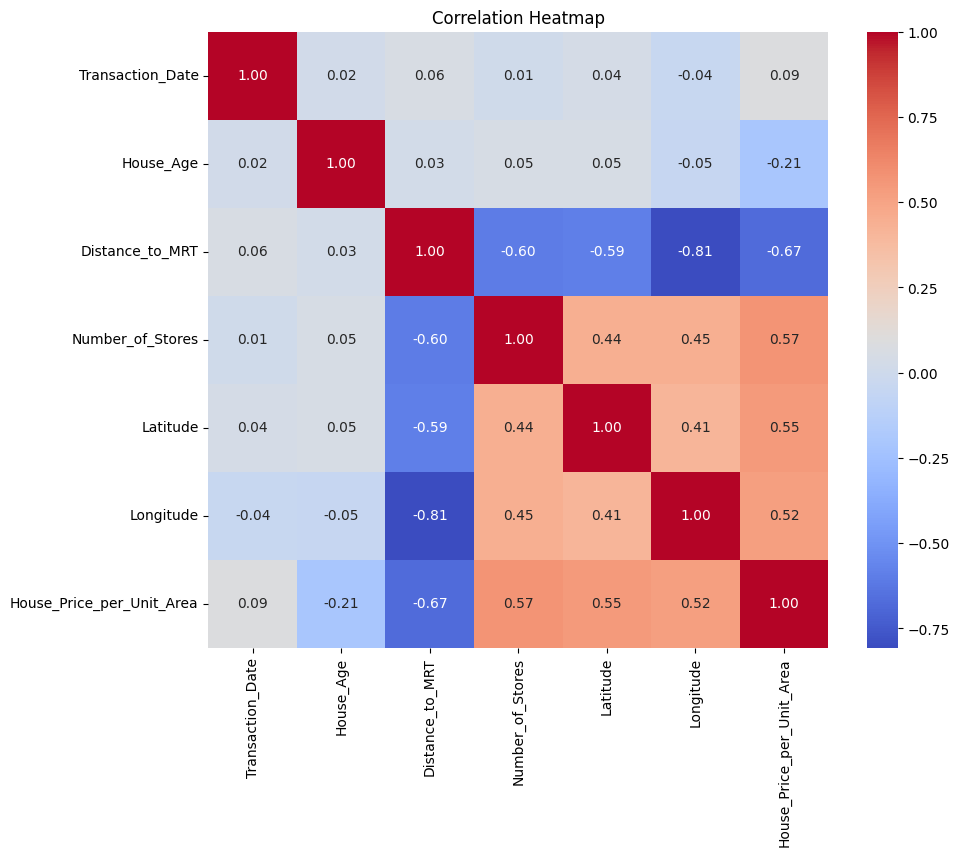
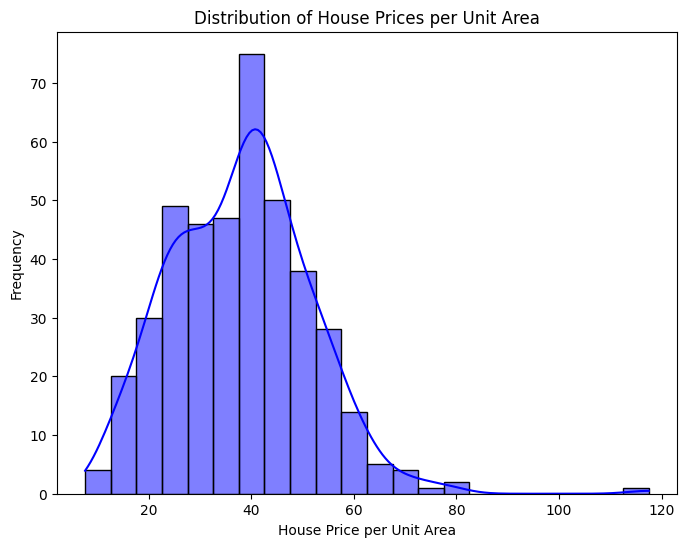
The dataset contains the following attributes:

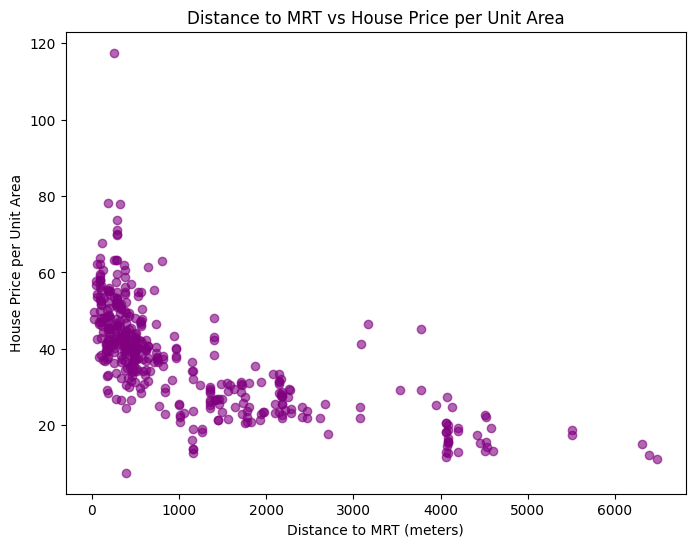
1. **No**: Serial number or record index.
2. **X1 transaction date**: The date of the transaction, represented as a decimal year (e.g., 2012.917 corresponds to late 2012).
3. **X2 house age**: The age of the house in years at the time of the transaction.
4. **X3 distance to the nearest MRT station**: The distance from the property to the nearest Mass Rapid Transit (MRT) station in meters.
5. **X4 number of convenience stores**: The number of convenience stores located within the vicinity of the property.
6. **X5 latitude**: The latitude coordinate of the property’s location.
7. **X6 longitude**: The longitude coordinate of the property’s location.
8. **Y house price of unit area**: The house price per unit area, measured in thousands of currency units (target variable).

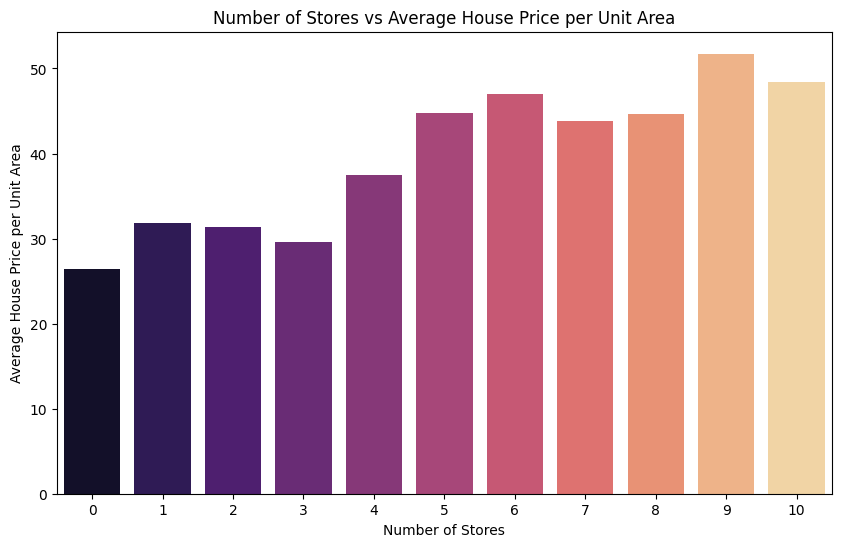
This dataset provides a combination of temporal, spatial, and structural features, making it suitable for studying the factors that influence real estate pricing and for building predictive models.

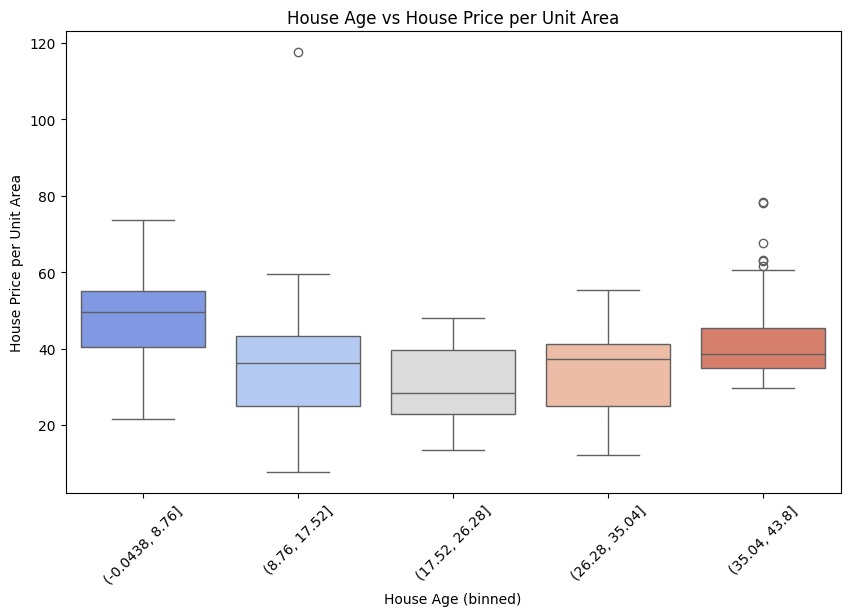
* Data visualization

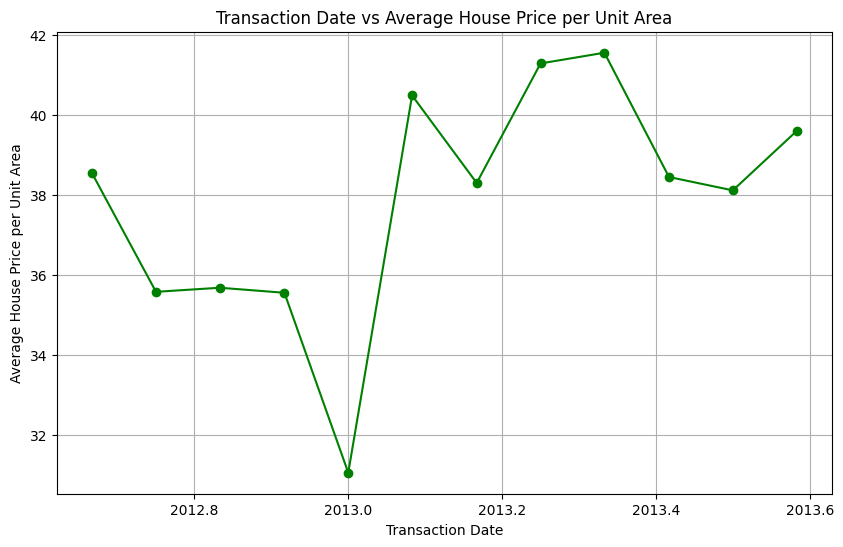


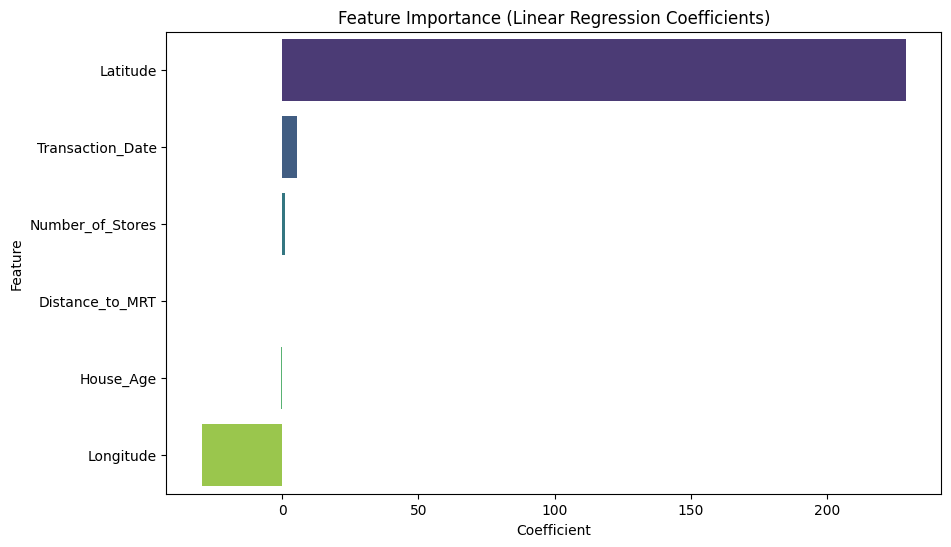
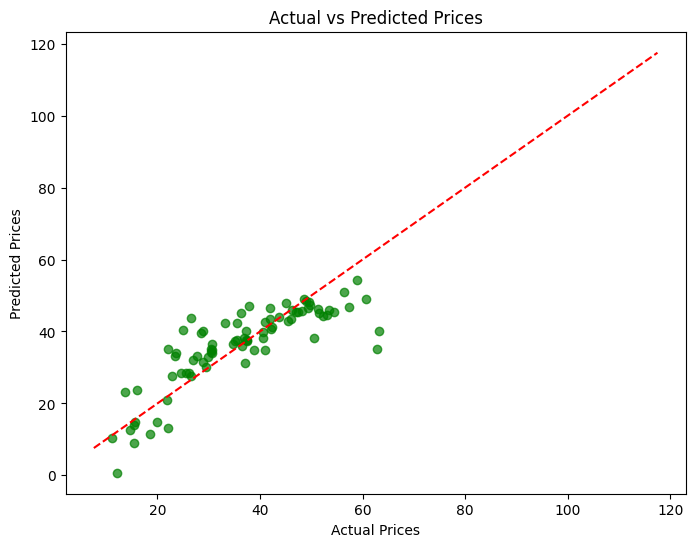












* Model

# Linear Regression

* Results
* Mean Absolute Error (MAE): 5.31
* Mean Squared Error (MSE): 53.51
* Root Mean Squared Error (RMSE): 7.31
* R-squared (R2): 0.68
* Github:

<https://github.com/DanaHa90321/Gaza-Sky-Geeks/blob/main/Real_estate_prediction.ipynb>