## Data description:

Recollected our data from a website called [aqarsas.sa](https://aqarsas.sa) and the website collected the data from Ministry of Justic , Ministry of Municipal and Rural Affairs and Housing, , and Riyadh Municipality. We contacted them and they kindly provided us with the last 10 years of data in Riyadh city , which is about 776800 records.

|  |  |  |
| --- | --- | --- |
| Feature name | Description | Example |
| Deal number | Integer from seven digits | 2041278 |
| Classification | Is it commercial or residential | سكني |
| Type | Is it agricultural land, apartment, house, plot of land, palace, aistiraha , building, villa, facility, commercial center, exhibition/shop | شقة |
| Price | The price of the deal in Saudi Riyal | 590000 |
| Area | The area of the real estate is square meter | 143.82 |
| Squared meter price | The price of one squared meter | 4102 |
| Date | The date of the deal in dd/mm/yyyy form | 27/01/2013 |
| District | Indicates watch district the real estate in | الرياض |
| City | Indicates which city the real estate in | الرياض |
| Neighborhood | Indicates in which neighborhood the real estate in | الياسمين |
| Plan | Indicates which plan the real estate in | 2413 |
| Land | Indicate the land of the rich state | 96/12 |

Table 1 Collected data characteristics.

## Processing the data

First we decided the features we are going to use to train our models. The classification the type the area the date the neighborhood the plan and the land. And our target feature is the price. We dropped the rest of the features. The deal number because it's a irrelevant the districts and the city also irrelevant since all of them are Riyadh. Squared meter price because it’s Highly Correlated with Price and Area. Then we made some changes on the rest as shown in the table below:

|  |  |  |
| --- | --- | --- |
| Feature name | Description | Changes |
| Classification | Is it commercial or residential | Commercial = 01  Residential = 02 |
| Type | Is it apartment, house, plot of land, villa | Apartment = 01  House = 02  plot of land = 03  villa = 04 |
| Area | The area of the real estate is square meter | Nonu |
| Date | The date of the deal in dd/mm/yyyy form | Nonu |
| Neighborhood | Indicates in which neighborhood the real estate in | 01 = البيان  02 = الرمال  03= الخير  04 = العارض  05 = القادسية  06 = الملقا  07 = المهديه  08 = الياسمين  09 = بنبان  10 = طويق  11 = لبن |
| Plan | Indicates which plan the real estate in | Nonu |
| Land | Indicate the land of the rich state | Nonu |

Table 2 features used in training and the changes we did on it.

As you can see in Table 2 we changed the classification the types and the neighborhoods to numbers for several reasons like:

Model Compatibility: Most machine learning algorithms work with numerical data. By converting text into numerical features, you make it possible to use a wide range of machine learning models effectively.

Feature Extraction: Numerical representations of text can capture essential patterns and relationships within the text data. Techniques like TF-IDF and word embeddings can help extract valuable information from text.

Efficiency: Numerical data is more efficient to work with in terms of computational resources and memory compared to raw text data.

Consistency: Numerical data is consistent and easier to handle across different machine learning libraries and frameworks.

For the neighborhoods we decided to focus only on 11 neighborhoods. Because they are the ones of the most underdeveloped neighborhoods.

## Models selection:

After researching in GP1 we found out that Decision Tree Regressor, Random Forest Regressor, Xgboost and GradientBoosting are the most used models and highly accurate for this kind of prediction. And we add some more to make sure we got the best result and the are :

### Turning:

### Testing:

### Result:

01 تجاري

02 سكني

01 شقة

02 بيت

03 قطعة ارض

04 فيلا

            01  البيان

            02 الرمال

            03 الخير

            04 العارض

            05 القادسية

            06 الملقا

            07 المهديه

            08 الياسمين

            09 بنبان

            10 طويق

            11 لبن