

REAL ESTATE

Houses price predictions

By Ahmed Aloliwi





TABLE OF CONTENTS

01

Goal of this project

02

Selecting the data source

03

Data cleaning and
exploration

04

Testing and training the
models

05

Tools and techniques

06

Conclusion



Goal of this project

The main objective of this project is to help us create a ML model to predict housing unit prices.

Dubai has been chosen as one of the most developed cities in the region. The existence of such models will be reflected in creating confidence in the real estate market and reassurance with the buyer and seller from the existence of illogical and understandable fluctuations in real estate prices.



Dubai

REAL ESTATE

Selecting the data source



- I tried looking for data on different sites like Github and Kaggle I chose a semi-clean data from Kaggle site that includes about 2000 and after cleaning the data I have 1906 and 38 columns.

Out[3]:

	id	title	price	location	latitude	longitude	type	size	no_of_bedrooms	no_of_bathrooms	completion_status	amenities
0	5528049	Price Reduced Middle Floor Nice Sea View PJ	2700000	Dubai, Palm Jumeirah	25.113208	55.138932	Apartment	1,079 sqft'in / ...	1	2.0	Off-plan	Partly furnished, Balcony, Barbecue Area, Cent...
1	6008529	2BR Type D Sea and Burj Al Arab Views PJ	2850000	Dubai, Palm Jumeirah, Shoreline Apartments	25.106809	55.151201	Apartment	1,582 sqft'in / ...	2	2.0	Ready	Furnished, Balcony, Built in Wardrobes, Centra...
2	6034542	AMAZING LAKE VIEW / 3 BR w/ Maid's Rm Apt @ JU...	1150000	Dubai, Jumeirah Lake Towers, Lake Almas West	25.063302	55.137728	Apartment	1,951 sqft'in / ...	3 + Maid	5.0	NaN	Unfurnished, Balcony, Built in Wardrobes, Cove...



For the cleaning stage:

- Removing unwanted columns.
- Removing duplicates.

For the exploration stage:

- I looking to price analysis.

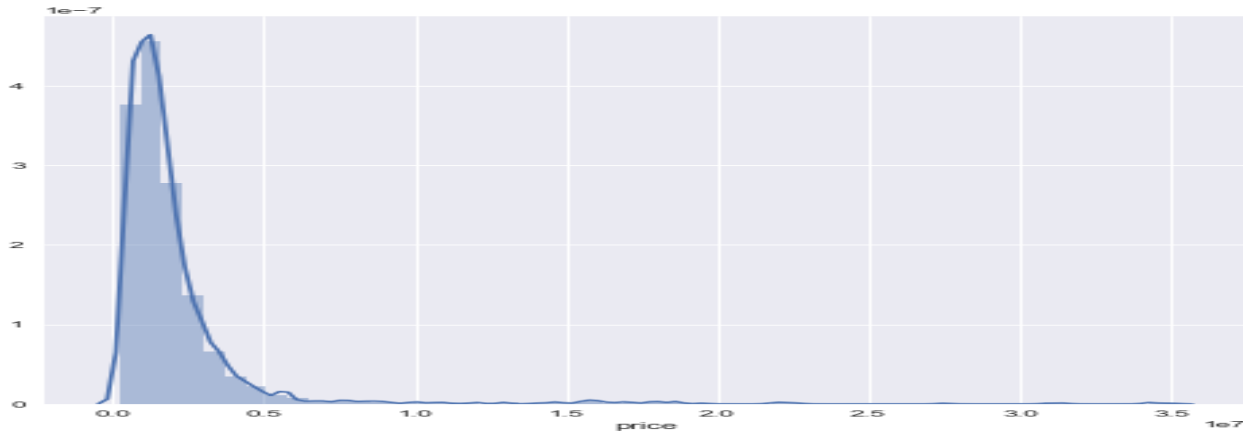
```
➤ count      1.905000e+03
➤ mean       2.085830e+06
➤ std        2.913200e+06
➤ min        2.200000e+05
➤ 25%        8.900000e+05|
➤ 50%        1.400000e+06
➤ 75%        2.200000e+06
➤ max        3.500000e+07
➤ Name: price, dtype: float64
```



Looking at descriptive analysis for property pricing in our dataset, we can say that:

- Property prices ranging from AED 220,000 to 35 Millions
- We have 1906 total properties in our dataset
- Mean price of our apartment is little over 2 Million AED

Our price distribution looks like:





- I trained and tested three models and finally I select **Randomforestregressor** because it has high R2 score which was closest to one was 0.9923263147255801 compared to other models.

Model	R2 score
GradientBoostingRegressor	0.991161083837371
RandomForestRegressor	0.9923263147255801
XGBRegressor	0.988561795411537



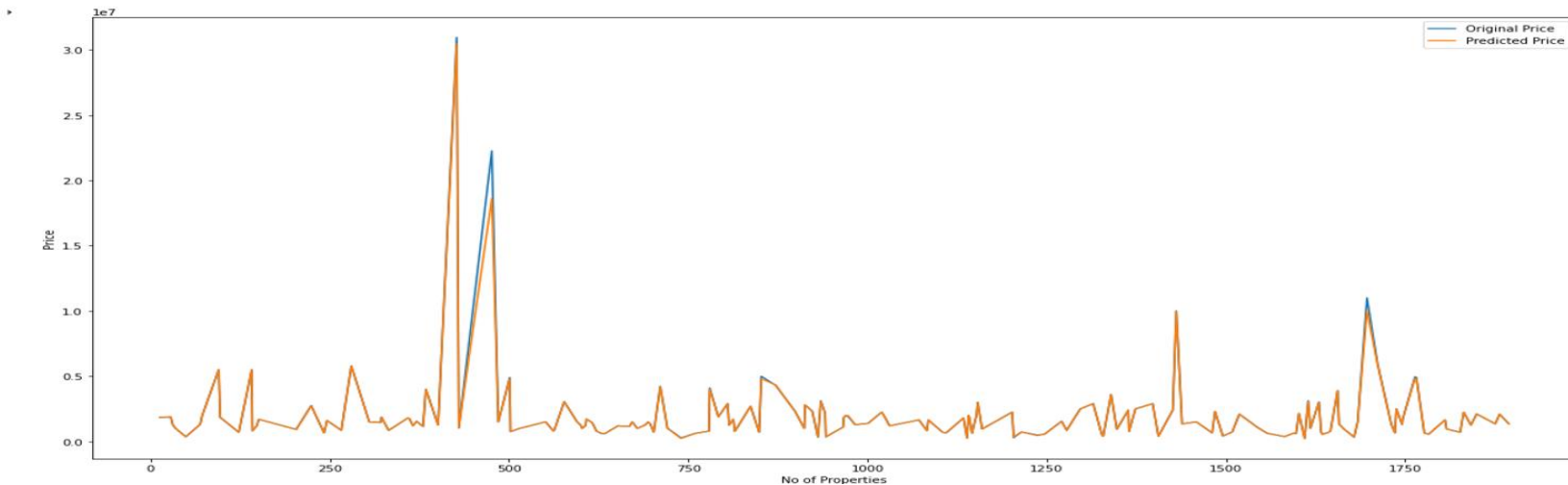
I used **python packages for data science models** like (sklearn, xgboost, etc.).

- I used Google cloud Colab to write my python script.
- Numpy and Pandas for data manipulation.
- Scikit-learn for modeling.
- Matplotlib and Seaborn for plotting.

Conclusion



- I think if I work with more and up-to-date data, it will help us get more accurate real estate price predictions



Thank You

