



# Regression solutions in Real Estate

Project developed by:

- Leonor Drummond
- Jose Roldan
- Patrick Brunswyck
- Hoang Minh Nguyen



# Project Goals

Insight from  
10000 properties

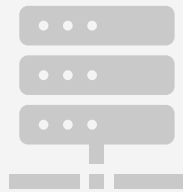
Defining a good  
regression model

Predicting house  
prices

# Table Content



Data cleaning



Data  
preprocessing



Defining the best  
regression type



What we could  
achieve



Recommendations

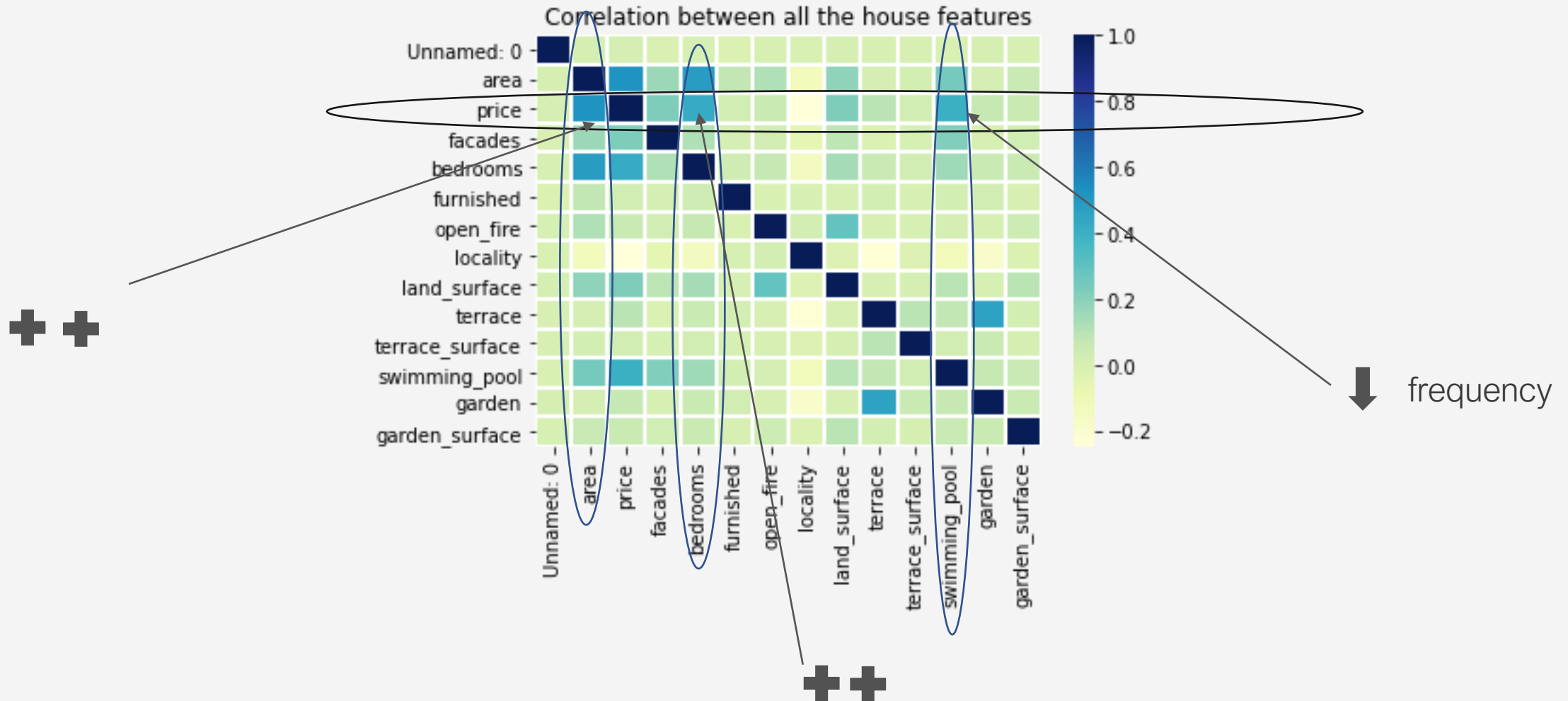


Conclusion

# Data Cleaning

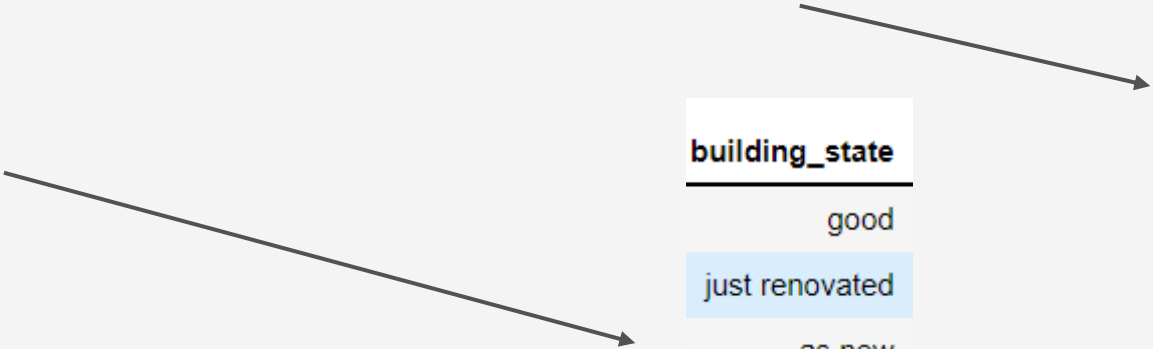
- Used the previous pre-cleaned data set;
- Focus on adapting data set for machine learning;
- Keeping the house features that are better for predictions;
- Dropped rows with outliers that don't have an impact on our model.

# Data Cleaning



# Data preprocessing

- Removing redundancy such as subtype property;
- Data engineering;
- Normalization;
- One Hot encoding is depreciated in linear regression.

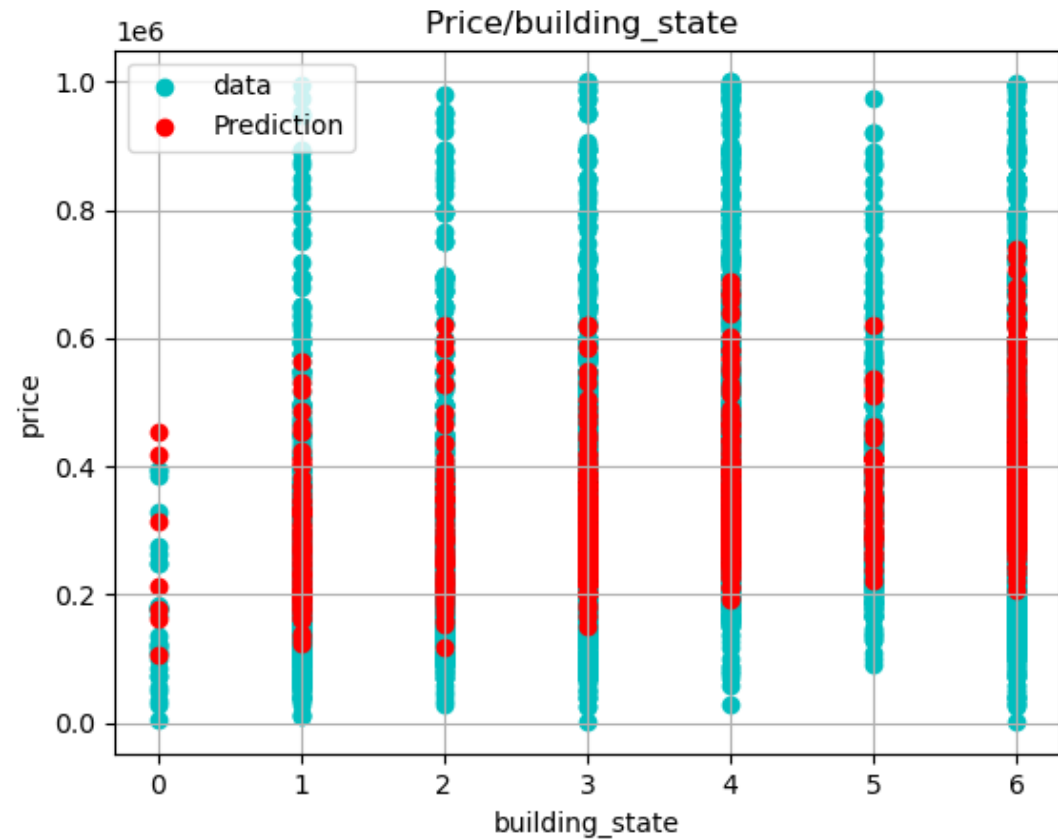


<u>building_state</u>
good
just renovated
as new
as new
good
...
just renovated
good
as new
as new
good

<u>property_type</u>	<u>property_subtype</u>
house	house
house	villa
house	villa
house	house
house	villa
...	...
house	house
house	house
house	house
house	house
house	house

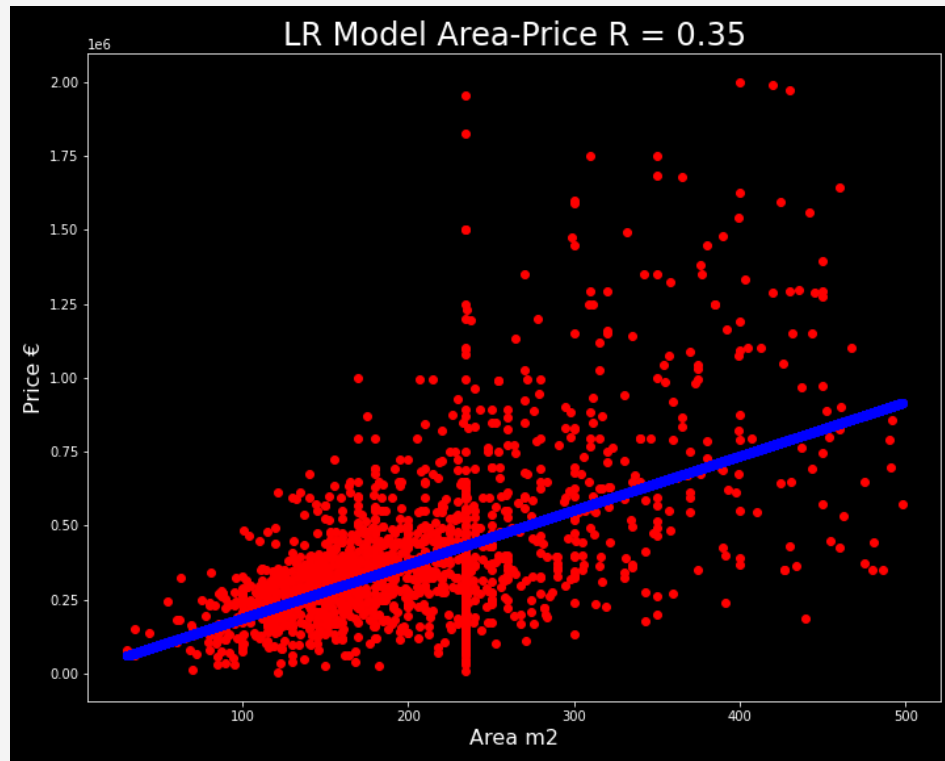
# Data preprocessing

- One hot encoding application;
- With bedrooms and area as features, model scored 32%;
- With building state model scored about 52%;
- One Hot encoding is depreciated in linear regression.



# Defining the best regression type

Linear Regression between price and area



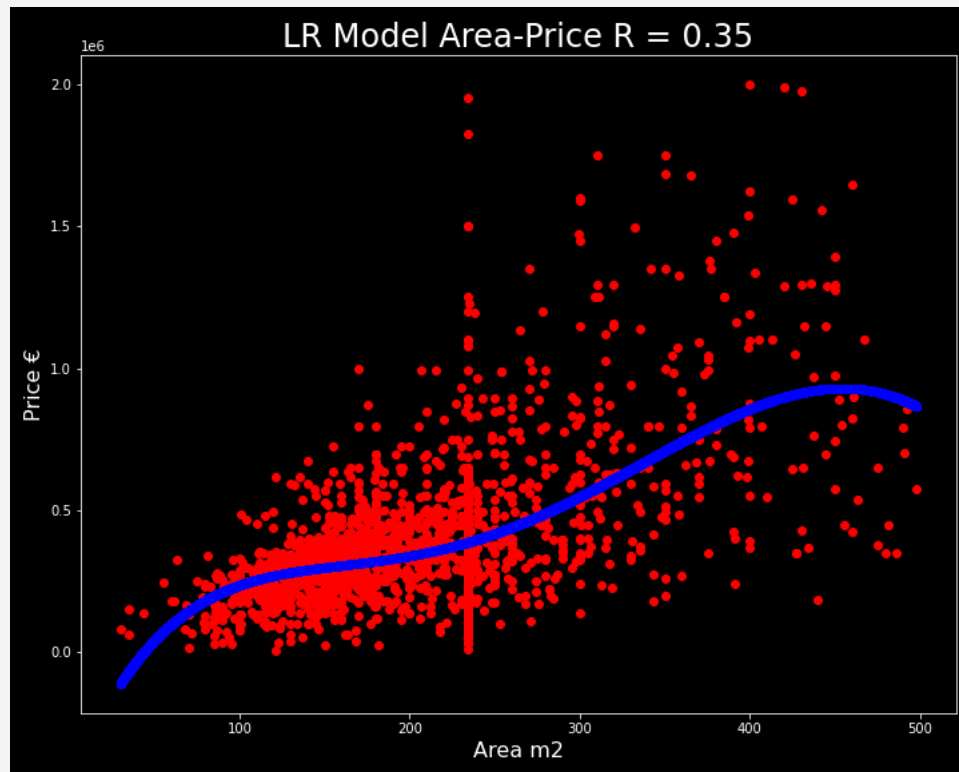
Linear Regression between price and bedrooms



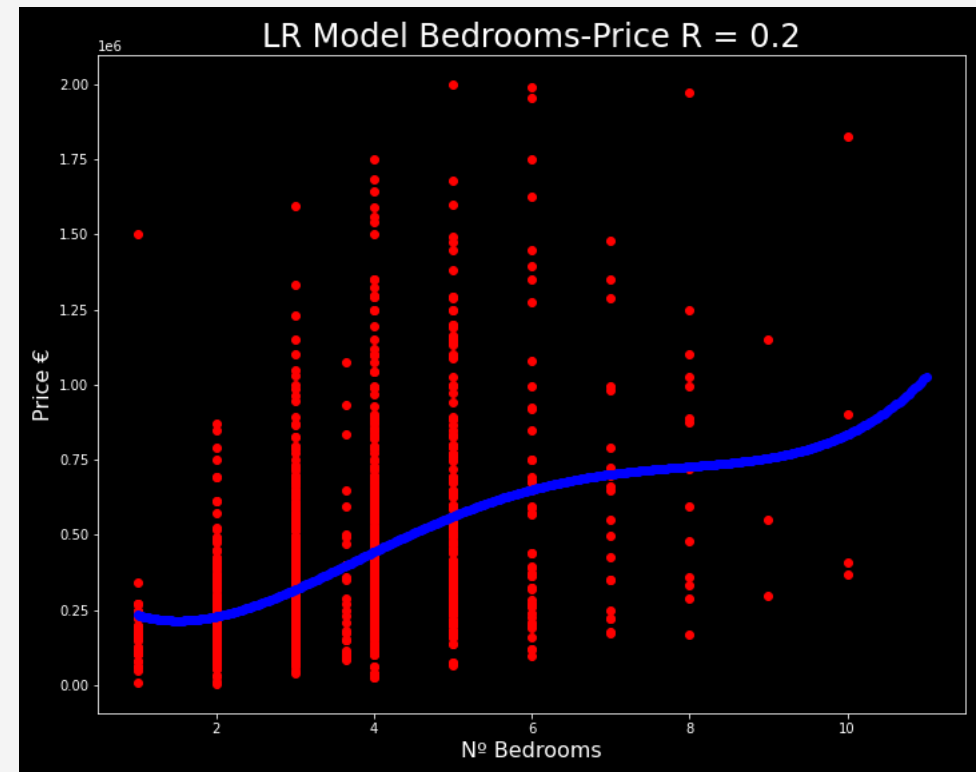


# Defining the best regression type

Linear Regression deg = 5 price and area



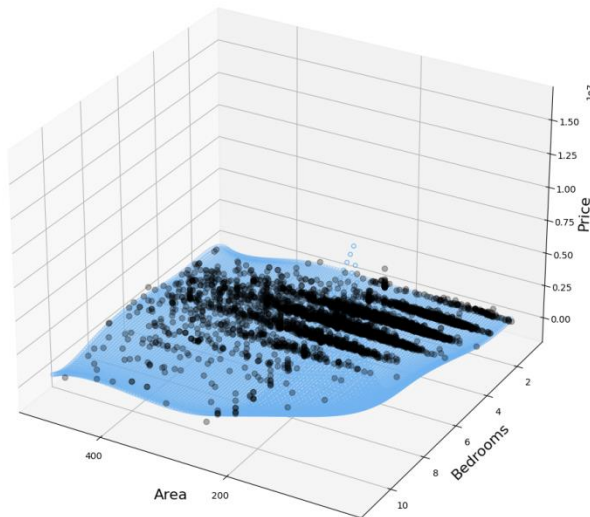
Linear Regression deg = 4 price and bedrooms



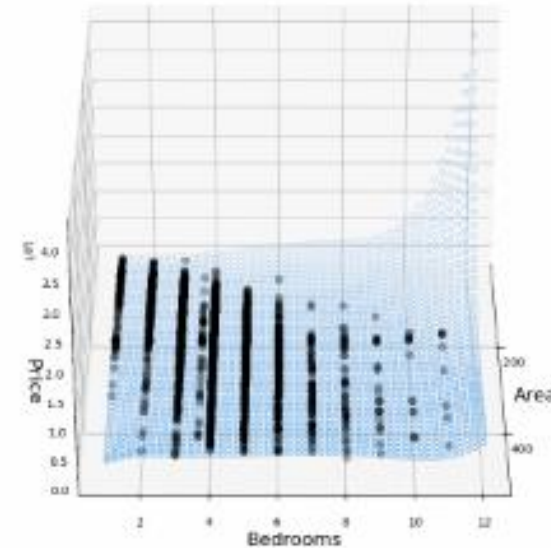
# Defining the best regression type

- Choosing our features:
  - bedrooms vs area vs state of building
- Linear or polynomial regression?
- Evaluating modeling scores

Price Prediction featuring Area & Bedrooms. Linear regression  $R^2 = 0.40$



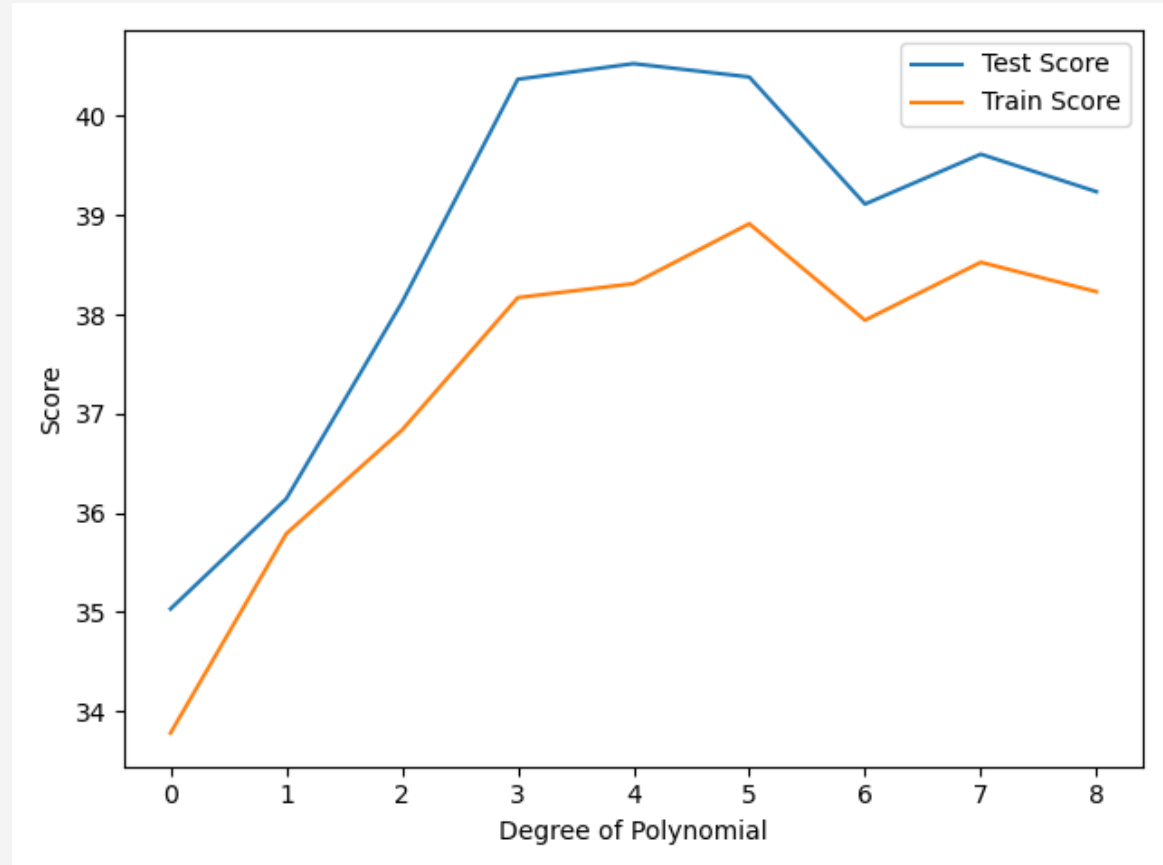
Price Prediction featuring Area & Bedrooms. Linear regression  $R^2 = 0.40$



imgflip.com

# What we could achieve

- Model chosen;
- Scores from the model;
- Features considered.
- Not all data is regressive;



# Recommendations

- Further studying different models for machine learning;
- Look for other solutions that can include other features;
- Bedrooms and area are two features that indeed are useful for price prediction.

# Conclusion

- What data preprocessing involves;
- Not all data is possible to adapt to linear regression
- One Hot Encode does not makes sense with linear regression;
- Let the data speak = different data types, different approaches! Not all is about model scores

Thank you for your trust and attention.