# Project review -
# Prediction on property rental price

# Context of our study

- 1nn the Neighborhood = Online platform to rent your property for short stays.
- Only 2% of people investigating our website start to using our platform
- Product manager would like to increase this percentage
- Project to develop a tool that help people estimate how much they could rent their living space
- Success criteria  = Estimate the actual price of their renting with a 25$ range.

# Input information

- Regression problem where the objective is to estimate the price of a rented property based on its features

- The project manager provided a dataset of 8100 rented properties with the following features :
  - Latitude / Longitude
  - Numbers of bedrooms and bathrooms
  - Minimum night someone can book
  - Property type
  - Room type
  - Price (Target variable)

# Latitude and Longitude (San Franscisco)


From S to N

Higher price

From W to E

No significant increase

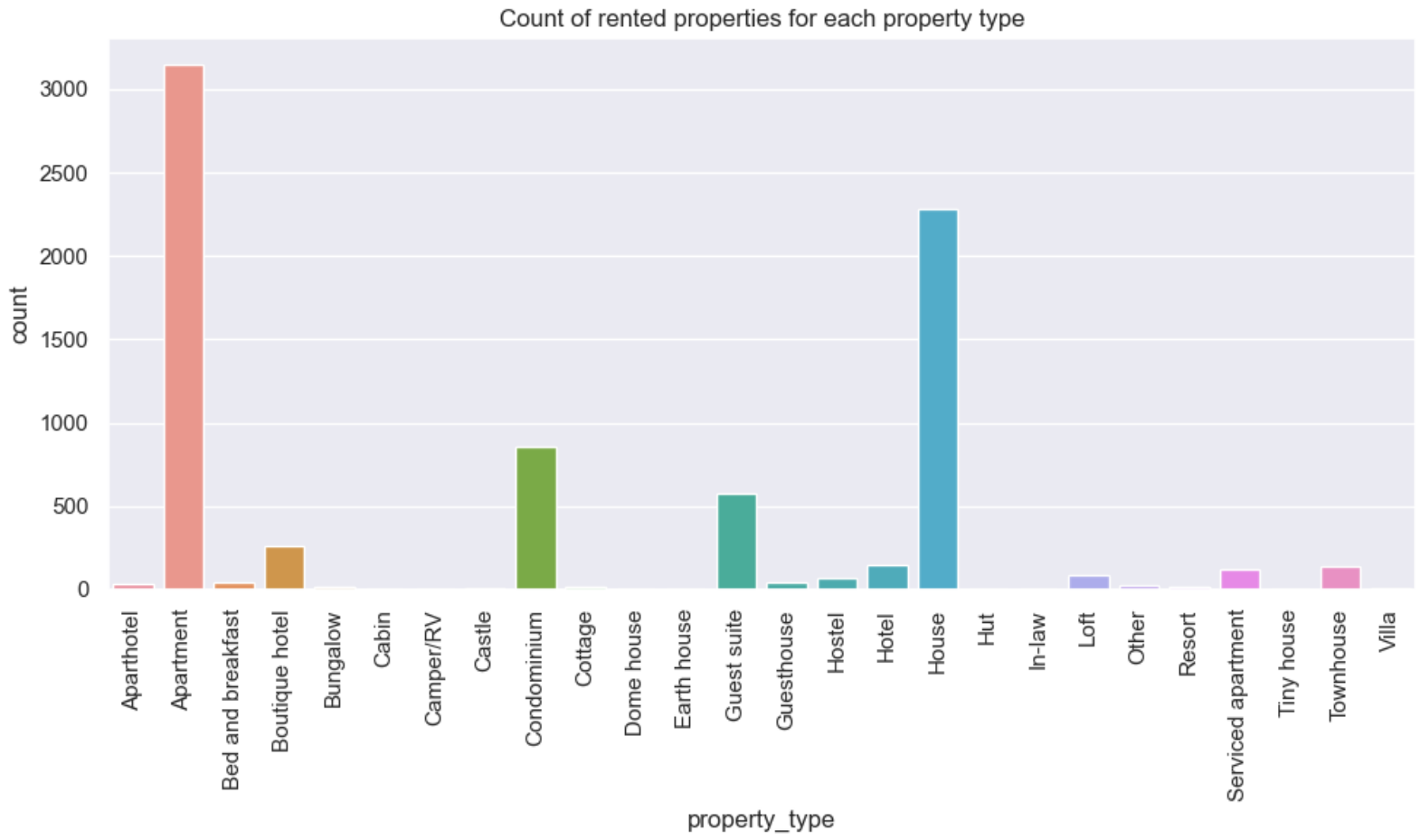
# Bathrooms and bedrooms



- More bedrooms → Higher price
- Inconsistency between price and number of bathrooms

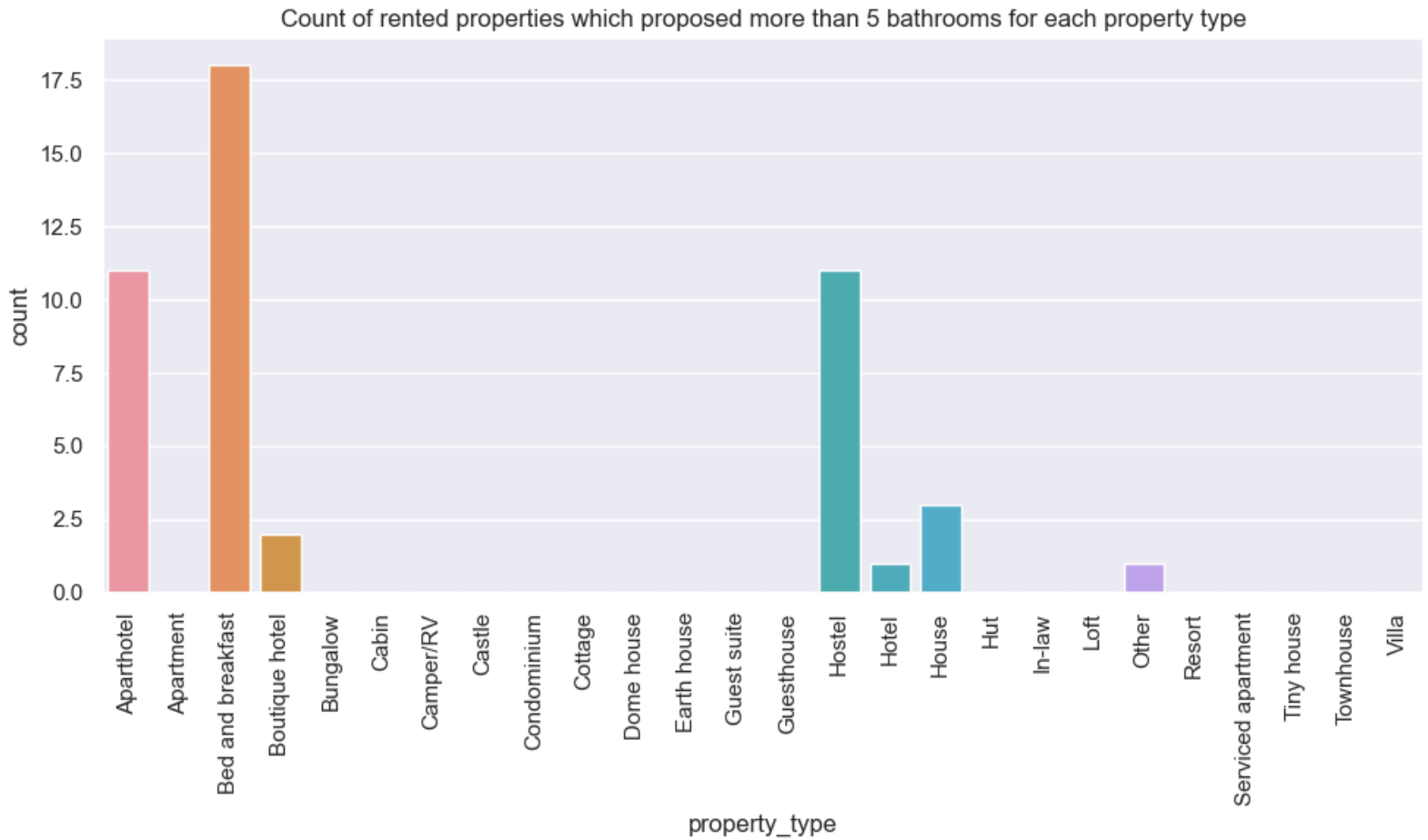
( > 5 bathrooms)

# Professional renters on our platform



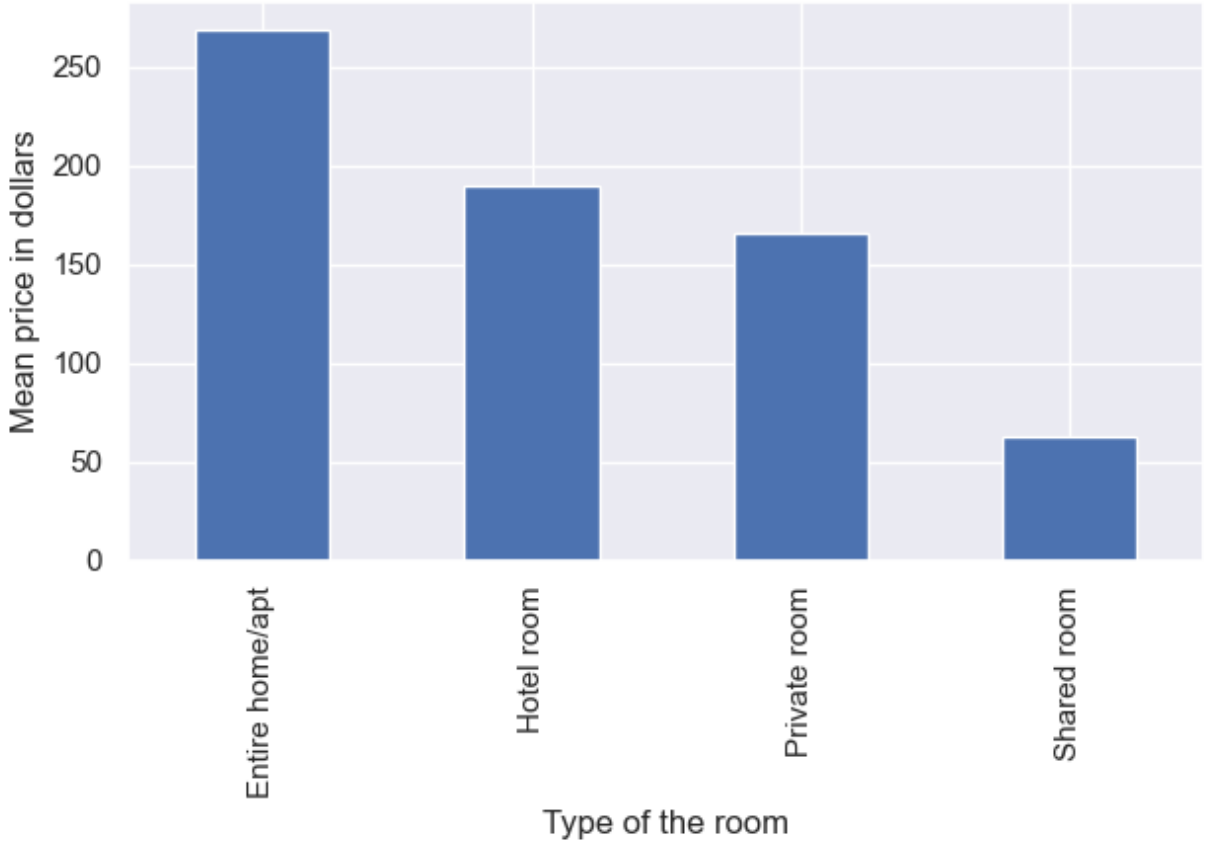

- Property_type variable too scattered / Too many categories

# Professional renters on our platform




- When property have more than 5 bathrooms it's a professional renter (Aparthotel, Bed and BreakFast, Boutique hotel, Hostel and Hotel)
- Professionnal renters consider all common bathrooms when they fill in their data

# Room type and minimum night features



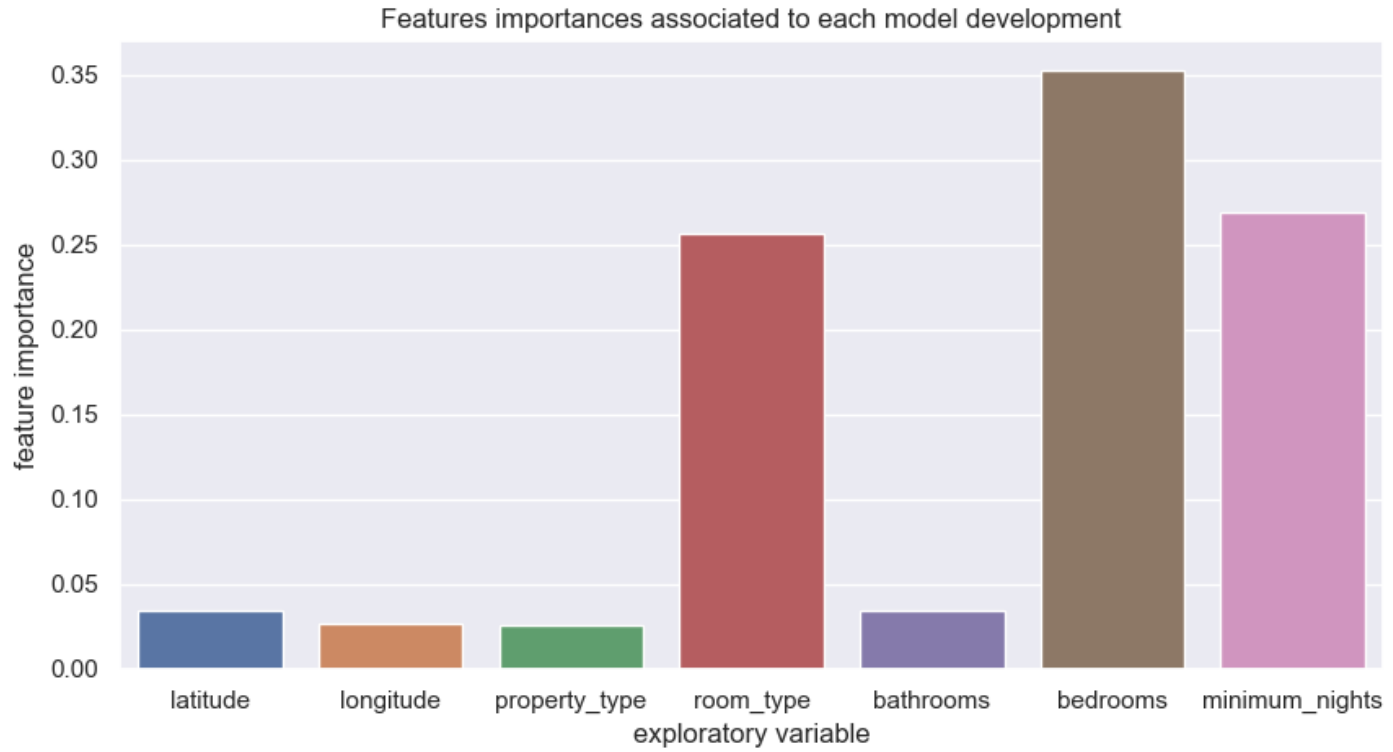

More privacy

→ Higher price

- No clear correlation between minimum booking nights and price

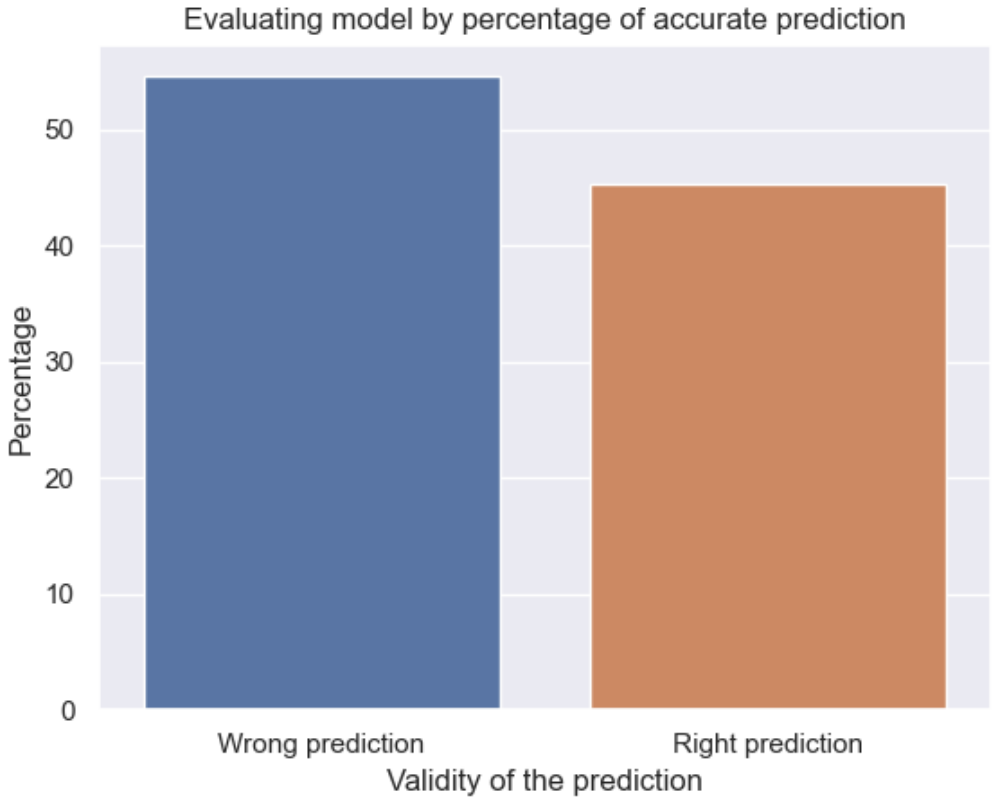
# Model Development

- Recall : Regression problem where the objective is to estimate the price of a rented property based on its features
- Best model = XGBoost Regressor with an **R2_value of 0.7 → 70%**
- Recall : R2_value is in a range from 0 to 1 and is commonly stated as **percentage from 0% (always wrong prediction) to 100% (perfect prediction)**



# Business criteria

- Success criteria = Estimate the actual price of their renting with a 25$ range.

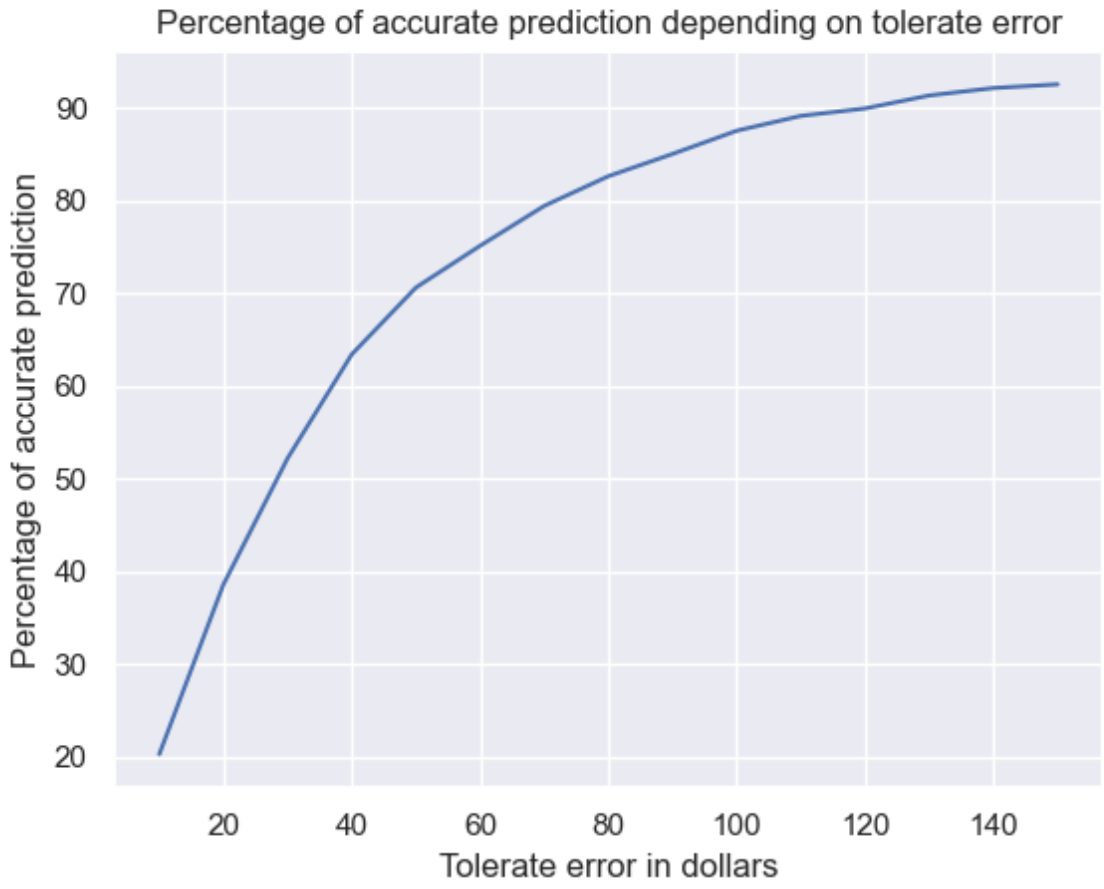



**51$**

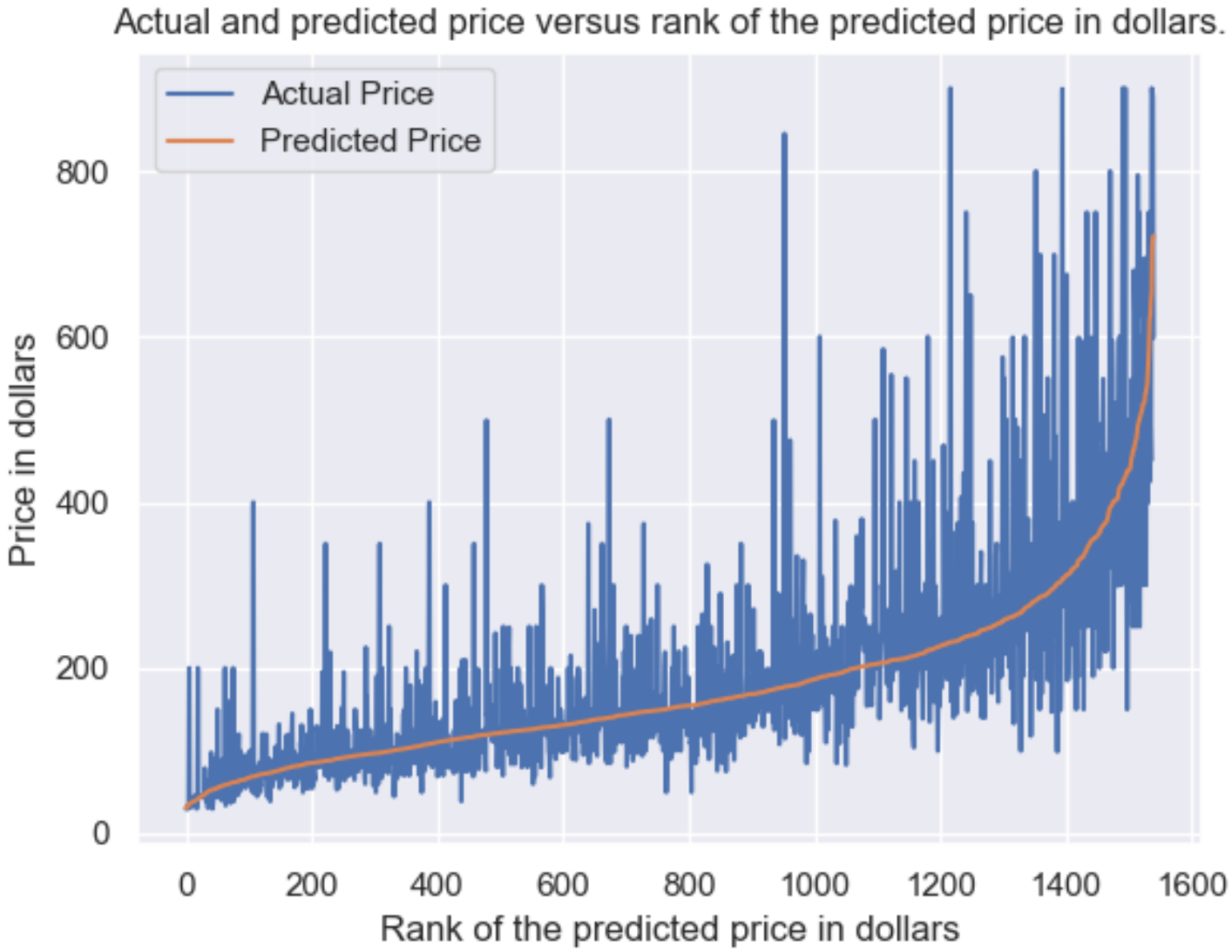
Mean prediction error

# Business criteria

- Lower success criteria to **60$** → Able to predict correctly **75%** of property price



# Explanation on model performance




For two equivalent properties
→ Our data is too scattered

# Recommandations

- **Assess data quality before collecting it.**
  - There are several inconsistancies in the dataset that could be corrected during data collection which will greatly increase our model performance.
- **Create more meaningful features**
  - For example, we could have in our model a variable that describe the state of use of the property.
- **Collect more data.**
  - During our analysis we try to drop the rows which correspond to inconsistancies.
  - This has always significantly decrease our model performance.
  - This implies we should collect more data in order to make our model more robust.

# Thanks you for your attention !