

German Foreclosure Auction Ranking

22th January 2024

Eric Jonas

Objective

Immo
Scout24

Trained on Immoscout
Rent Data



FORECLOSURE

Apartment
Foreclosure Data

Model to Predict Rent

Create Ranking of
Apartments



Problem

- Most information is stored in unstructured text
- Extracting structured information from these texts is challenging due to different layouts and text structure



Zwangsversteigerungstermine

Sie sind hier: Amtsgericht Bamberg in Bayern

Termine suchen

Übersicht

Hinweise für Bieter

Bieter Handbuch

Startseite

Impressum

Rechtliche Hinweise

Kontakt

Suche

0003 K 0056/2022

Art der Versteigerung:

Grundbuch:

Objekt/Lage:

Beschreibung:

Verkehrswert:

Verkehrswert in €:

Termin:

Ort der Versteigerung:

Gericht:

GeoServer:

amtliche Bekanntmachung

Exposee:

(letzte Aktualisierung: 05-09-2024 13:40)

Zwangsversteigerung zum Zwecke der Aufhebung der Gemeinschaft

Eingetragen im Grundbuch des Amtsgerichts Forchheim von Wolfsburg

Lfd.Nr.	Gemarkung	Flurstück	Wirtschaftsart	Anschrift	Hektar	Blatt
1	Wolfsberg	894	Ackerland	Steig	0,3680	227
2	Wolfsberg	895	Waldfläche	Steig	0,5080	227

Land- und forstwirtschaftliche Fläche: Steig, 91286 Obertrubach, Wolfsburg

Lfd. Nr. 1

Objektbeschreibung/Lage: *(lt. Angabe d. Sachverständigen):*
Land- und forstwirtschaftliche Fläche im Außenbereich. Fläche liegt im Landschaftsschutzgebiet Fränkische Schweiz-Veldensteiner Forst. Im Norden Grünland, im Süden Waldstreifen.

Verkehrswert: 4.000,00 €

Lfd. Nr. 2

Objektbeschreibung/Lage: *(lt. Angabe d. Sachverständigen):*
Land- und forstwirtschaftliche Fläche im Außenbereich. Fläche liegt im Landschaftsschutzgebiet Fränkische Schweiz-Veldensteiner Forst. Im Norden Wald- und Wegfläche, im Süden Grünland.

Verkehrswert: 6.100,00 €

Lfd.Nr. 1: 4.000,00 €

Lfd.Nr. 2: 6.100,00 €

Montag, 09. Dezember 2024, 09:30 Uhr

Amtsgericht Bamberg, Synagogenplatz 1, 96047 Bamberg, Sitzungssaal 101

[Internetseite des Gerichtes](#)

[Karten, Luftbilder](#)

[amtliche Bekanntmachung1.pdf](#) 311,98 kB

[Exposee1.pdf](#) 876,11 kB

Weitere Infos

Justizportal des Bundes und der Länder

Gemeinsames Registerportal der Länder

Amtsgericht Deggendorf
Abteilung für Vollstreckungssachen
Az.: 1 K 4923

Deggendorf, 18.10.2024

Verkehrswert: 17.000,00 €

Der Versteigerungsvermerk ist am 03.01.2024 in das Grundbuch eingetragen worden.

Die amtliche Bekanntmachung der Terminbestimmung erfolgt im Internet unter www.rvg-portal.de

Terminbestimmung:

Zum Zwecke der Aufhebung der Gemeinschaft soll am

Datum	Uhrzeit	Raum	Ort
Freitag, 03.01.2025	09:00 Uhr	E 29, Sitzungssaal	Amtsgericht Deggendorf, Amanstr. 17, 94469 Deggendorf

öffentlich versteigert werden:

Grundbucheintragung:

Eingetragen im Grundbuch des Amtsgerichts Deggendorf von Deggendorf
Miteigentumsanteil verbunden mit Sondereigentum

Fl.	MG-Anteil	Sondereigentums-Art	SE-Nr.	Blatt
Nr.				
1	4/1000	Treilparagensteilplatz	S 17	826/1
2	4/1000	Treilparagensteilplatz	S 16	826/6

an Grundbuch:

Gemarkung	Flurstück	Wirtschaftsart u. Lage	Anschrift	Hektar
Deggendorf	445	Gebäude- und Freifläche	Ostlicher Stadtgraben 37	0,0650
Deggendorf	445/3	Gebäude- und Freifläche	Nähe Fährdenweg	0,0040

Lfd. Nr. 1

Objektbeschreibung/Lage: *(lt. Angabe d. Sachverständigen):*

Treilparagensteilplatz in zentrumsnaher Lage

Objektsanschrift: Ostlicher Stadtgraben 37, 94469 Deggendorf.

Verkehrswert: 17.000,00 €

Lfd. Nr. 2

Objektbeschreibung/Lage: *(lt. Angabe d. Sachverständigen):*

Treilparagensteilplatz in zentrumsnaher Lage

Objektsanschrift: Ostlicher Stadtgraben 37, 94469 Deggendorf.

Eric Jonas

3

Methods

Rule Based

Define rules to extract information based on the structure of the text

- Regex
- Hierarchical rules
- **Cant** handle unseen layout structures
- We need to define the rules

Classification Pipelines

Classify sentences or paragraphs



Fine-grained predictions within a class

- Sequence models (RNNs, LSTMs, ...)
- NER, RE, EE models
- **Can** handle unseen layout structures
- We need labeled data

Template Filling with LLMs

Define template with needed information and use an LLM to fill the template based on provided information.

- Guided Generation
- Restricted Generation
- **Can** handle unseen layout structures
- We need to define the template

Template

```
class ForclosureObjectModel(BaseModel):
    flaeche: int = Field(description="Die Fläche des Objekts in Quadratmetern.")
    verkehrswert: int = Field(description="Verkehrswert des Objekts.")
    typ: Optional[VersteigerungTyp] = Field(description="Art der Immobilie (z.B. Wohnung, Haus oder etwas anderes).")
    baujahr: Optional[int] = Field(description="Baujahr der Immobilie.")
    raeume: Optional[int] = Field(description="Anzahl der Räume im Objekt.")
    raum_typen: List[RaumTyp] = Field(default_factory=list, description="""Liste der Raumtypen im Objekt (z.B. Wohnzimmer, Küche).
    Die Anzahl der Einträge in der Liste sollte der Gesamtzahl der Räume entsprechen.
    Mehrere Räume desselben Typs sollten jeweils einzeln aufgeführt werden.""")
    balkon: bool = Field(description="Gibt an, ob das Objekt einen Balkon hat.")
    garten: bool = Field(description="Hat das Objekt einen Garten.")
```

You, 5 days ago | 1 author (You)

```
class ForclosureModel(BaseModel):
    objekte: List[ForclosureObjectModel] = Field(description="Liste der Zwangsversteigerungsobjekte, die zu diesem Fall gehören.")
    gesamtverkehrswert: int = Field(description="Gesamtverkehrswert aller Zwangsversteigerungsobjekte.")
```

```
question_with_schema = (
    f"The following text lists examples for your task: {examples}.\n"
    f"You MUST answer using the following JSON schema: {ForclosureModel.model_json_schema()}. \n"
    f>Please extract information about the following PDF content: {pdf_text}."
)
```

LLM Evaluation

Model	Performance
Llama 3.2 (3B)	Poor results; generation time ~20 seconds.
Meta-Llama 3.1 (8B)	Excellent results; generation time ~40 seconds.
Mistral-Nemo (12.2B)	Excellent results; extremely slow (~3 minutes).
Mistral (7B)	Mediocre results; similar speed to Llama 3.1 (~40 seconds).
Mistral-NeMo-Minitron (8B)	Excellent results; generation time ~1.20 minute (faster with smaller quantizations but worse results).
Gemma 2 (9B)	Failed to generate valid JSON output.

LLM Evaluation

Column	0-shot	1-shot	3-shot	5-shot	Ground Truth Nulls	Notes
Flaeche	7	10	8	7	0	
Typ	7	7	7	9	0	Some misclassified examples.
Baujahr	10	10	9	10	4	
Heizsystem	3	10	10	10	9	Rare; will be excluded.
Anzahl Räume	4	6	6	9	5	Some misclassified examples.
Raum_Typen	3	6	8	8	0	
Balkon	8	9	10	10	0	
Garten	9	9	9	10	0	False positives in most settings except 0-shot and 5-shot.
Verkehrswert	9	10	10	10	0	
Gesamtverkehrswert	10	10	10	10	0	
Anzahl_Objekte	10	10	10	10	0	

Extraction Pipeline

Start

Scrape Zvg Portal

Download relevant
PDFs

Extract Information
from PDFs with LLM

Extract General
Foreclosure Data

Clean and Merge
Datasets

Extract Foreclosure
Object Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4637 entries, 0 to 4636
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype
---  -
0   link                 4637 non-null  object
1   bundesland_code      4637 non-null  object
2   aktenzeichen         4637 non-null  object
3   letzte_aktualisierung 4637 non-null  object
4   art_der_versteigerung 4637 non-null  object
5   grundbuch           4273 non-null  object
6   objekt_lage          4637 non-null  object
7   beschreibung         4637 non-null  object
8   verkehrswert         4637 non-null  object
9   termin              4637 non-null  int64
10  ort_der_versteigerung 4637 non-null  object
11  amtliche_bekanntmachung 4519 non-null  object
12  exposee              2111 non-null  object
13  gutachten           2353 non-null  object
14  foto                 1025 non-null  object
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3202 entries, 0 to 3201
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype
---  -
0   foreclosurecase_link 3202 non-null  object
1   flaeche              3202 non-null  int64
2   verkehrswert         3202 non-null  int64
3   typ                  3199 non-null  object
4   baujahr              1316 non-null  float64
5   raeume               530 non-null   float64
6   raum_typen           492 non-null   object
7   balkon               3202 non-null  bool
8   garten               3202 non-null  bool
9   objekt_lage          3202 non-null  object
10  aktenzeichen          3202 non-null  object
11  termin               3202 non-null  int64
12  art_der_versteigerung 3202 non-null  object
13  beschreibung          3202 non-null  object
14  bundesland_name       3202 non-null  object
```

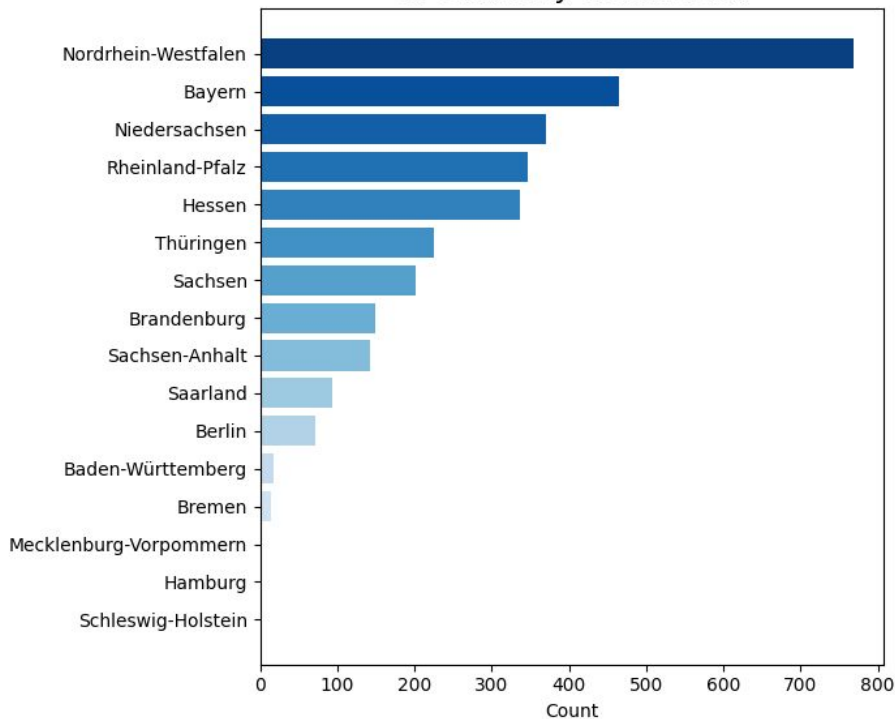
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3202 entries, 0 to 3201
Data columns (total 10 columns):
#   Column              Non-Null Count  Dtype
---  -
0   id                  3202 non-null  int64
1   foreclosurecase_link 3202 non-null  object
2   flaeche              3202 non-null  int64
3   verkehrswert         3202 non-null  int64
4   typ                  3199 non-null  object
5   baujahr              1316 non-null  float64
6   raeume               530 non-null   float64
7   raum_typen           492 non-null   object
8   balkon               3202 non-null  bool
9   garten               3202 non-null  bool
```


Foreclosure Auction Dataset - General

Zv-Count by Bundesland

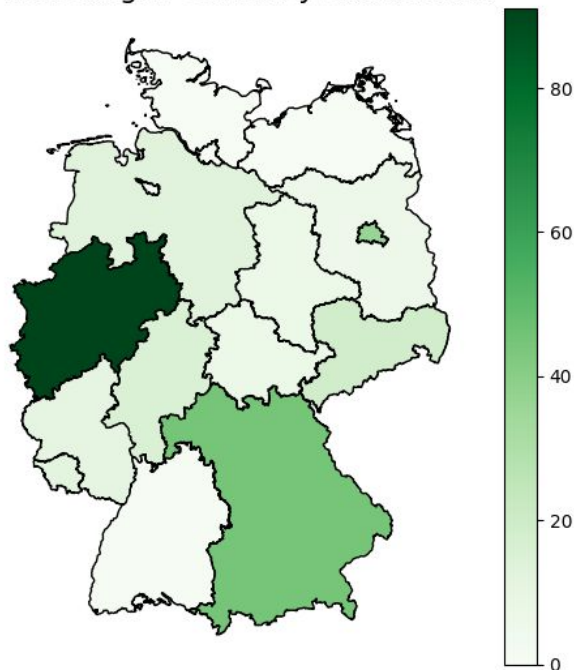


Zv-Count by Bundesland

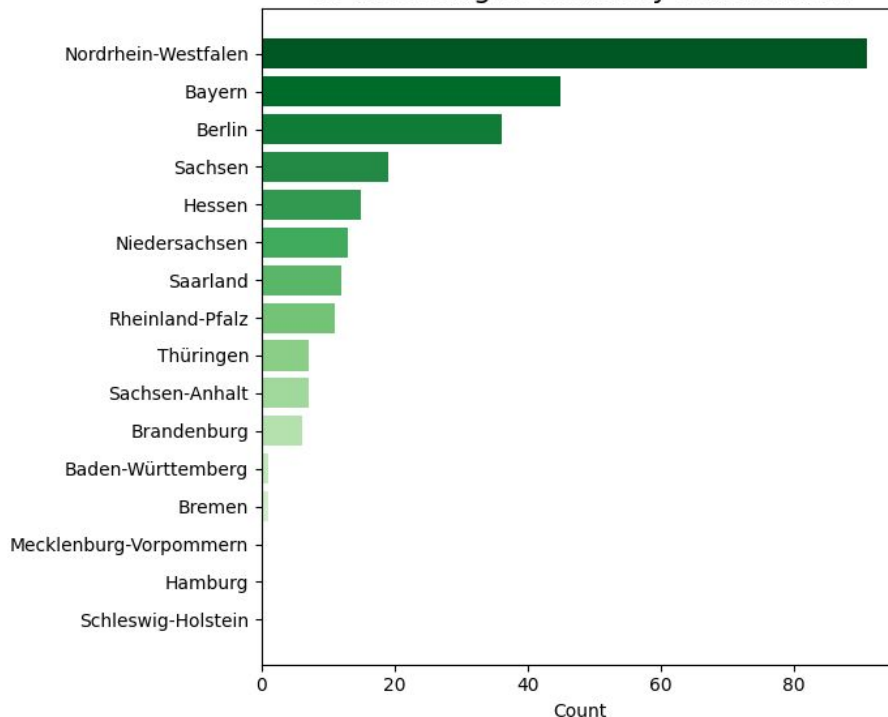


Foreclosure Auction Dataset - Apartments

Zv-Wohnungen-Count by Bundesland

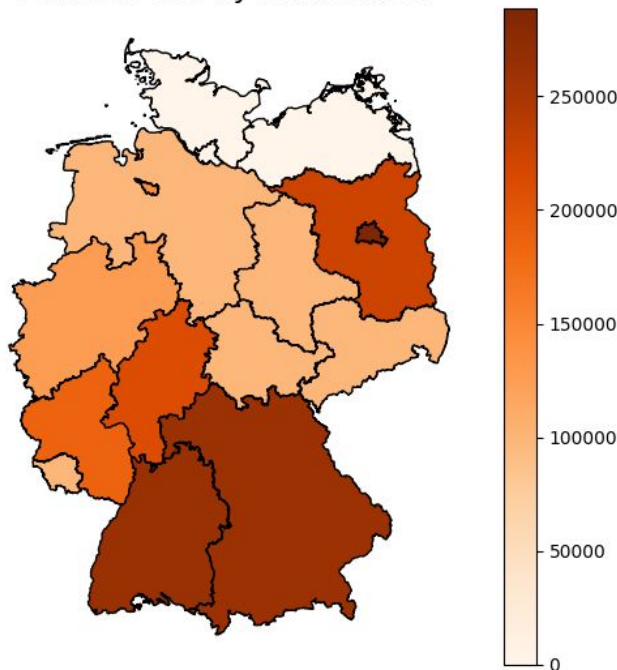


Zv-Wohnungen-Count by Bundesland

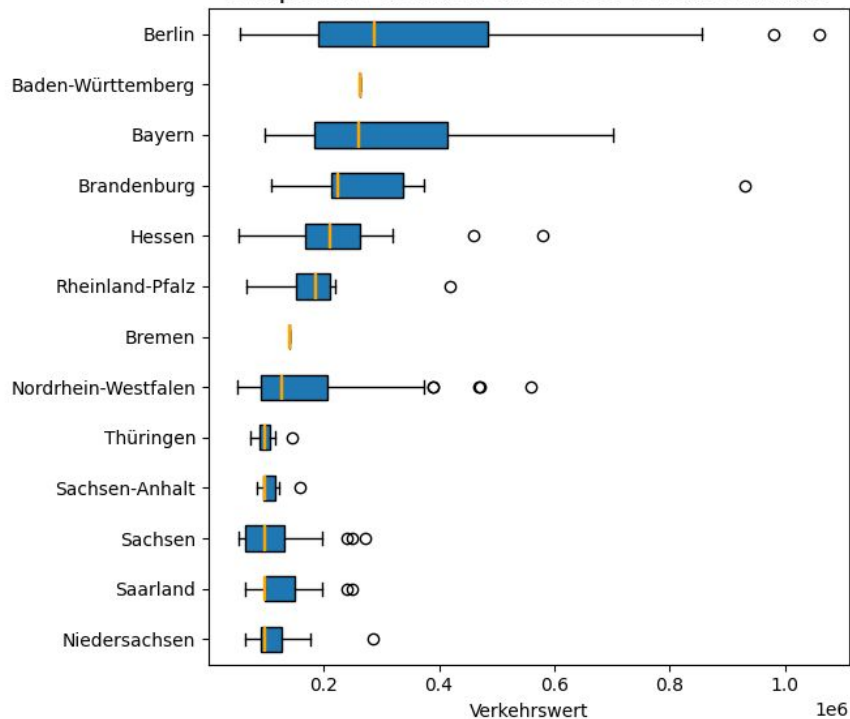


Foreclosure Auction Dataset - Apartments

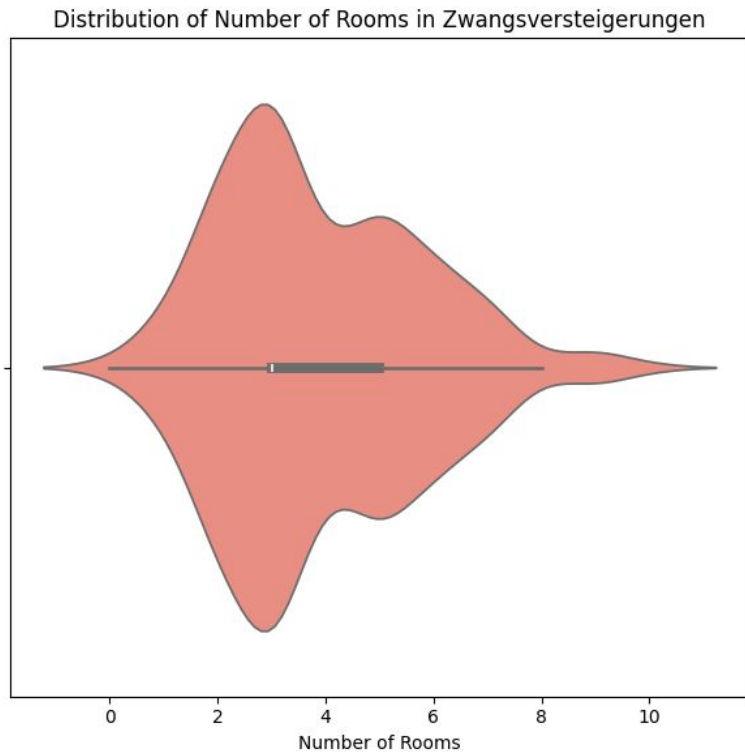
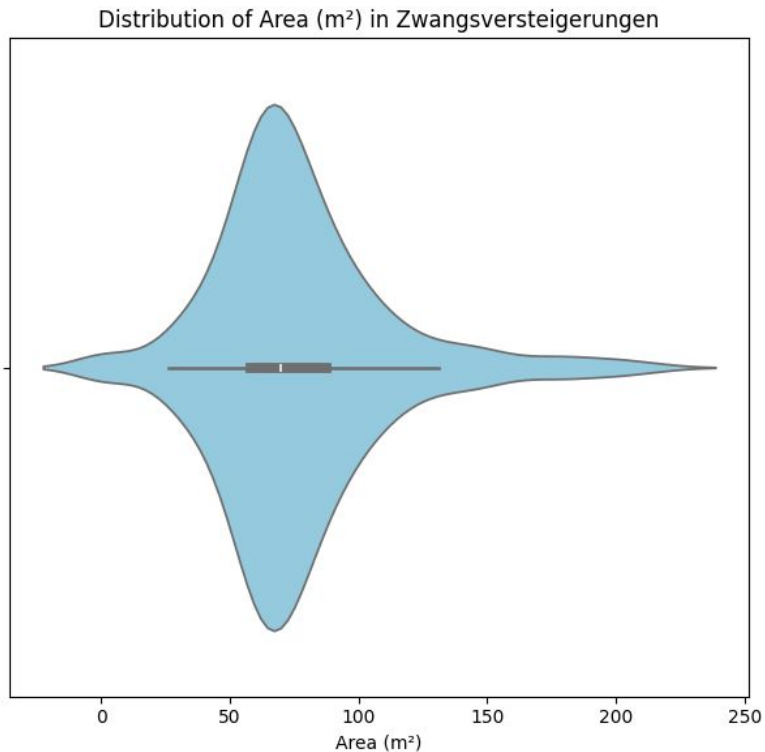
Median Price by Bundesland



Boxplot of Verkehrswert for Bundesländer



Foreclosure Auction Dataset - Apartments

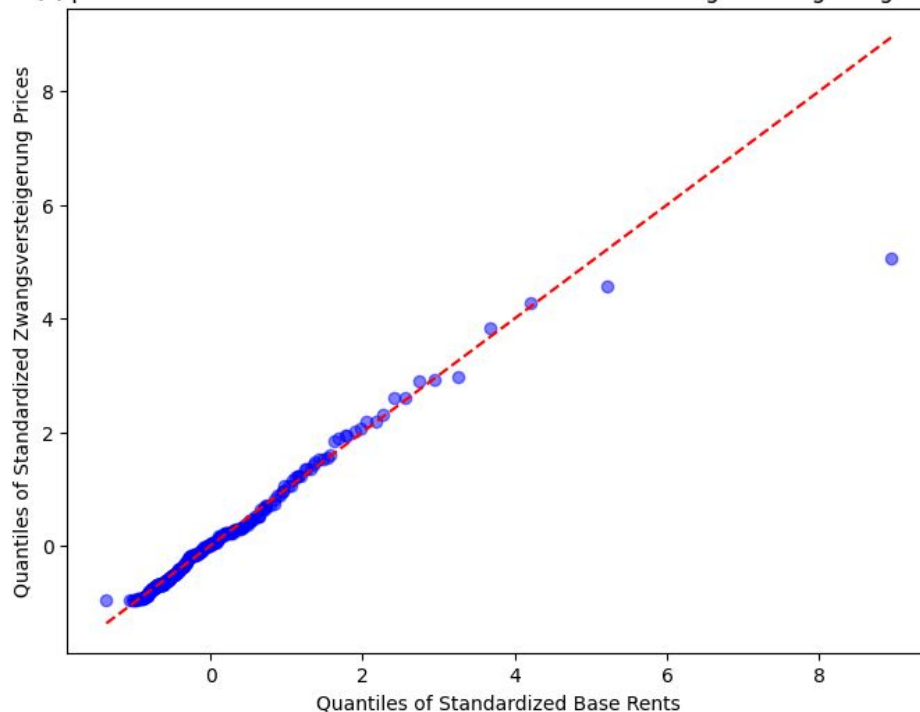


Immoscout Rent Dataset

Apartment Rental Offers in Germany

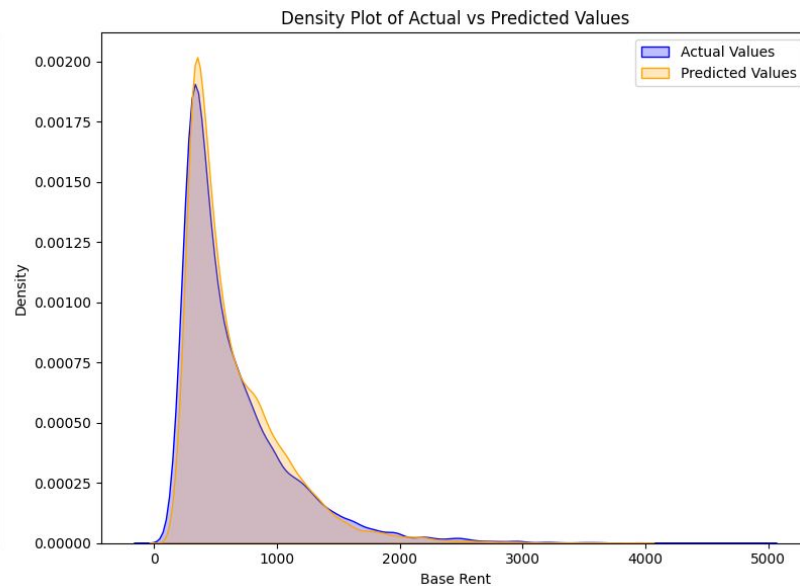
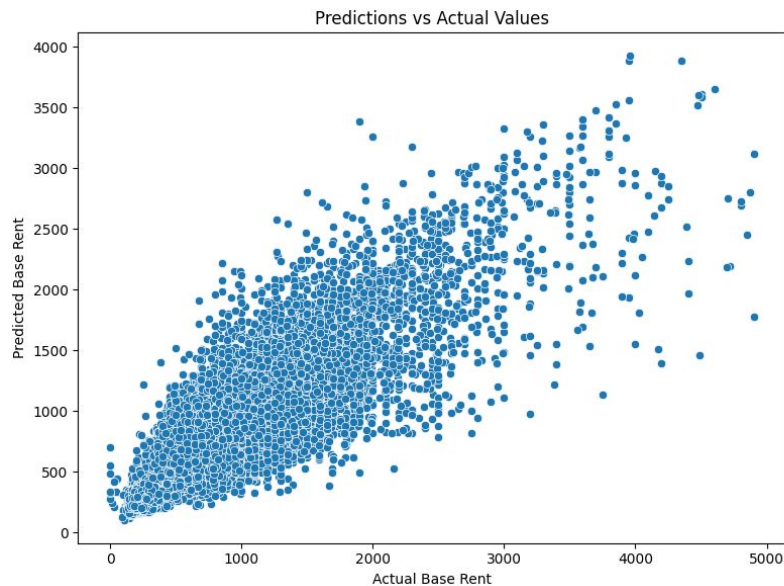
```
<class 'pandas.core.frame.DataFrame'>  
Index: 210654 entries, 0 to 268849  
Data columns (total 7 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                  
0    regio1                210654 non-null object    
1    balcony               210654 non-null bool      
2    baseRent              210654 non-null float64   
3    yearConstructed       210654 non-null float64   
4    noRooms               210654 non-null float64   
5    livingSpace           210654 non-null float64   
6    garden                210654 non-null bool
```

QQ plot of Standardized Base Rents vs Standardized Zwangsversteigerung Prices

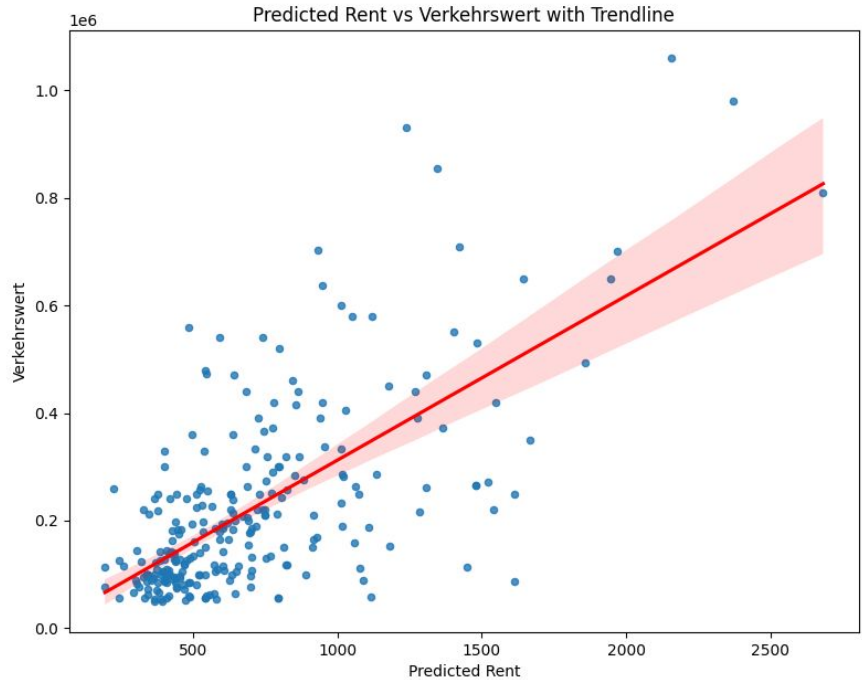
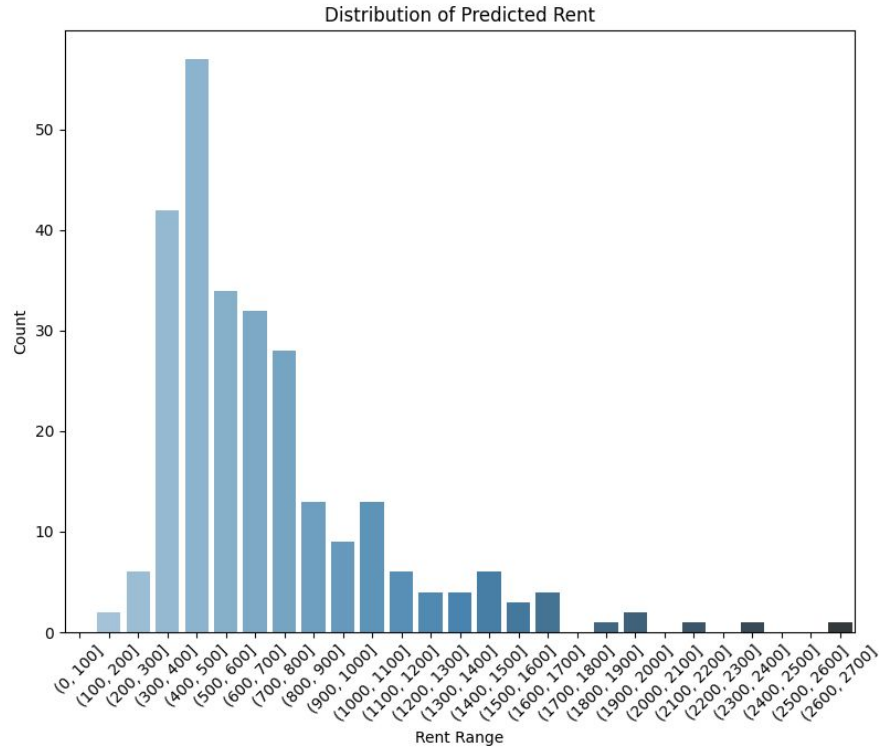


Model

Best Model: XGBRegressor
Mean Absolute Error: 130.03
Mean Squared Error: 46702.05

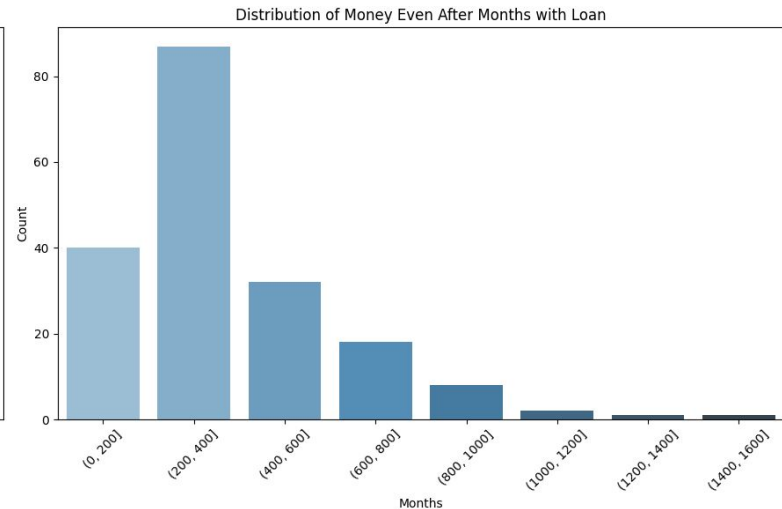
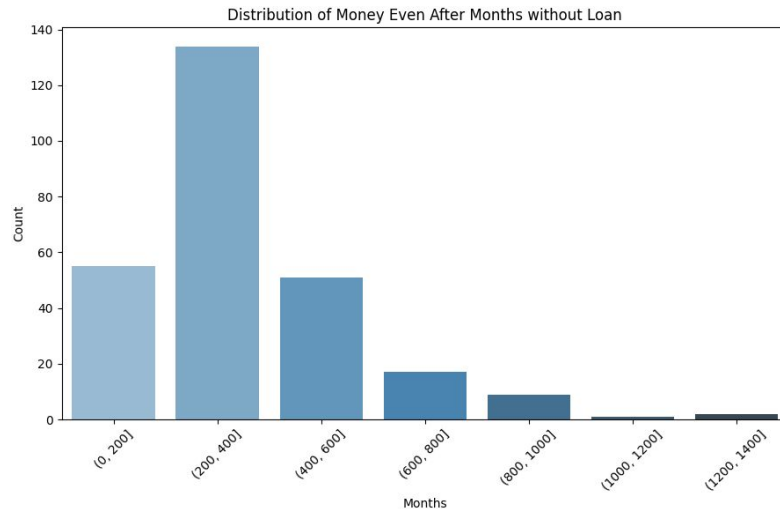


Evaluation - Foreclosure Rent Predictions



Evaluation - Ranking

- Ranking Based on Break-Even Point (does not consider market value changes)
- **Simple Break Even**
 - Based on market value and net rent (market value / net rent)
- **Complex Break Even**
 - Based on market value, net rent and user input
 - User can input his base capital, payment rates and loan interest rate



Evaluation - Example

- **User Inputs:**

- Base Capital: € 50,000
- Loan Payment: € 500/month + 90% of rent
- Interest Rate: 7 % per annum

- Apartment Link: [Foreclosure Case Link](#)
- Aktenzeichen: 0005 K 0121/ 2022
- Bundesland: Nordrhein-Westfalen
- Area (m²): 114
- Market Value: 58,000 €
- Year Built: 1895
- Predicted Rent: 1,115.61 €
- Monthly Loan Payment: 1,504.05 €
- Loan Duration: 5.42 months
- Total Loan Payment: 8,150.42 €
- Total Money Spent: 58,150.42 €
- Break-Even Point: 57.92 months (approximately 4.8 years)



Conclusion

- **LLM Output Limitations:** Current extraction isn't perfect; Larger models like ChatGPT-4 showed more promising results.
- **Data Source Challenges:** Key details (e.g., condition, location) are often in separate large documents like exposés, requiring text classification approaches.
- **Model Improvements:** Add parameters like district and flat condition to enhance predictions.
- **Prototype Assessment:** Tool offers a solid starting point for identifying promising apartments.
- **Replacement of Personal Review:** No, personal review of appraisals, exposés and related documents is still essential for final purchase decisions.

Questions?

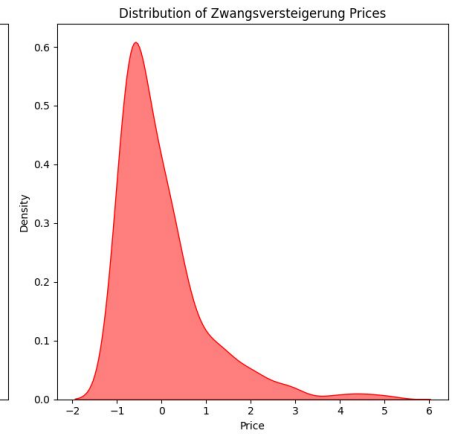
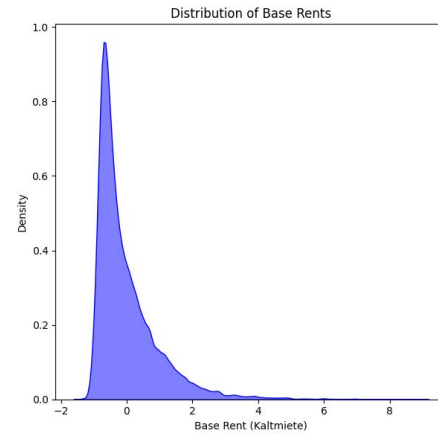
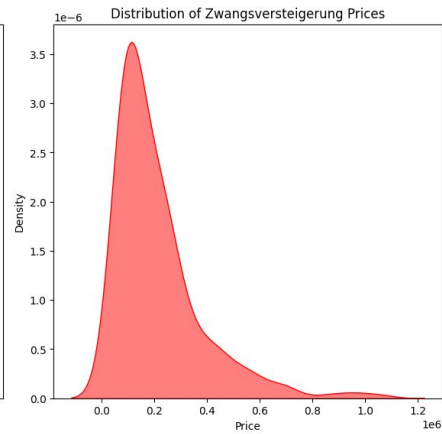
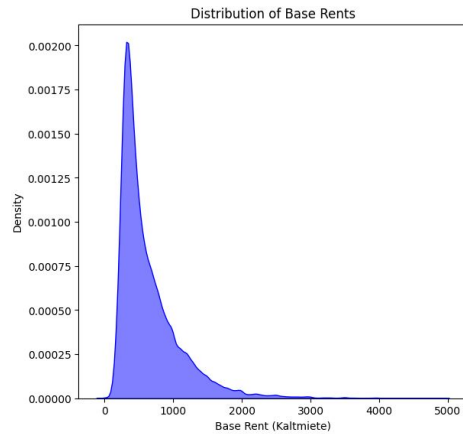
APPENDIX

https://github.com/EricCpy/German_Foreclosure_Auction_Analysis

Sources

1. GeeksforGeeks: [Information Extraction in NLP](#)
2. ZVG Portal: [Official Foreclosure Listings in Germany](#)
3. Journal of Pathology, March 2024, Daniel Truhn et al.: Extracting Structured Information from Unstructured Histopathology Reports Using GPT-4
4. Kaggle (CorrieBar): [Apartment Rental Offers in Germany](#)
5. Medium (MyChen76): [Practical Techniques to Constraint LLM Output in JSON Format](#)

Standardization Same Distribution



Evaluation - Ranking

- Ranking Based on break even point (increase/decrease of market value not considered)
- Simple break even: $\text{market value} / \text{net_rent}$
- Complex break even based on user needs with loan:
- User can input his `base_capital`, his `additional_payment_rate` (assumed he will pay the loan with rent and this additional rate), `interest_rate` for loan
- Break even with loan:
 - $\text{loan_duration} = \ln(R / (R - K * i)) / \ln(1 + i)$, R = monthly payment, K = loan_amount, i = interest_rate
 - $\text{Break_even} = \text{loan_duration} + (\text{additional_monthly_rate} * \text{loan_duration} + \text{base_capital}) / \text{net_rent}$