

# KG-based Housing Market Analysis

---

**Ziping Hu, Zepei Zhao**, Mohammad Rostami, Filip Ilievski, Basel Shbita

# Understanding the housing market is important, but difficult

**Important** for anyone who wants to rent or buy/sell

**Difficult because it requires:**

- Data collection from various websites
- Narrowing down search based on geographical or economical constraints
- Iterative searching over the collected information

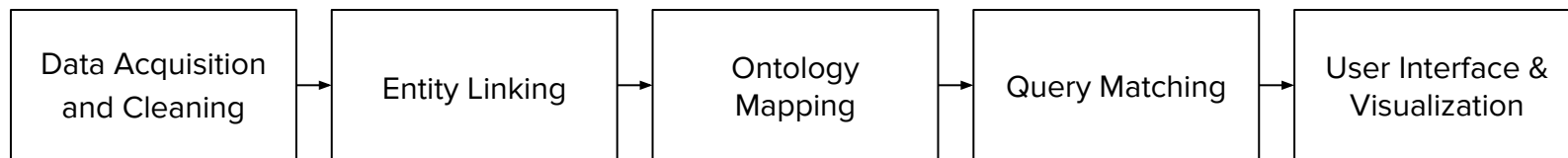
→ Highly **time-consuming** and **laborious**

Becomes **infeasible** when the **data size is large**

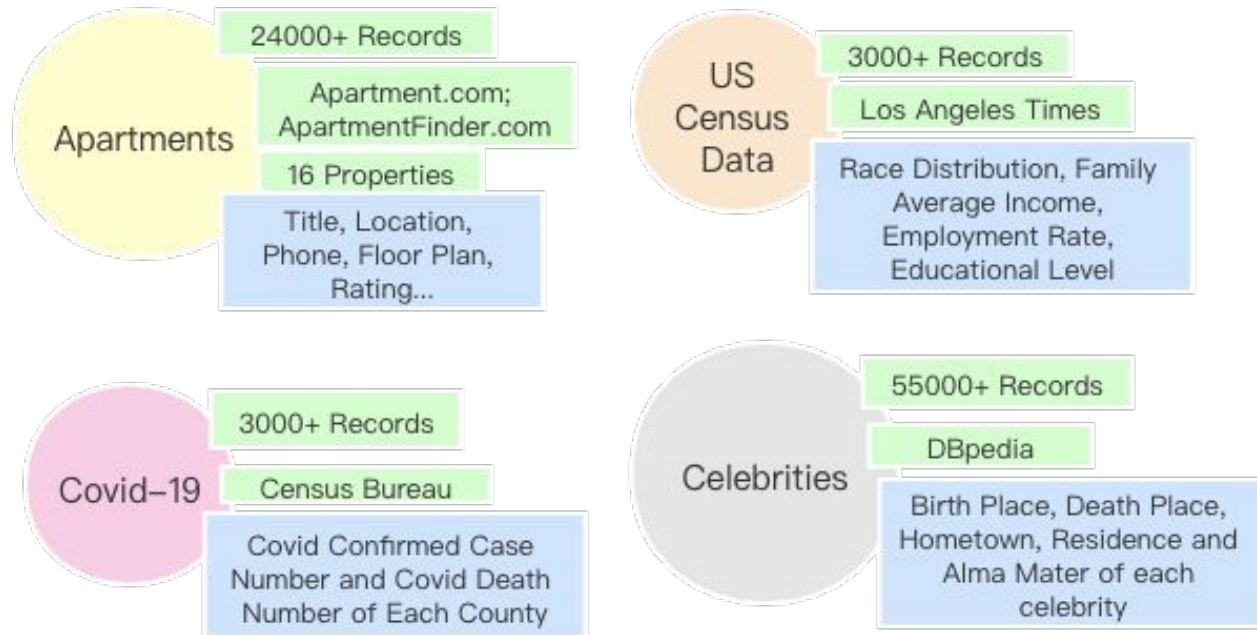
# Our Solution: KG-based tool for flexible analysis

- A **KG construction pipeline** for integrating information
- A **tool** given the geography of interest (California)
- Two **use-cases**:
  - Rental recommendations
  - Spatio-temporal dynamics of the market

# KG Construction Method



# 1. Data Acquisition and Cleaning



## 2. Entity Linking

### Apartments between different websites

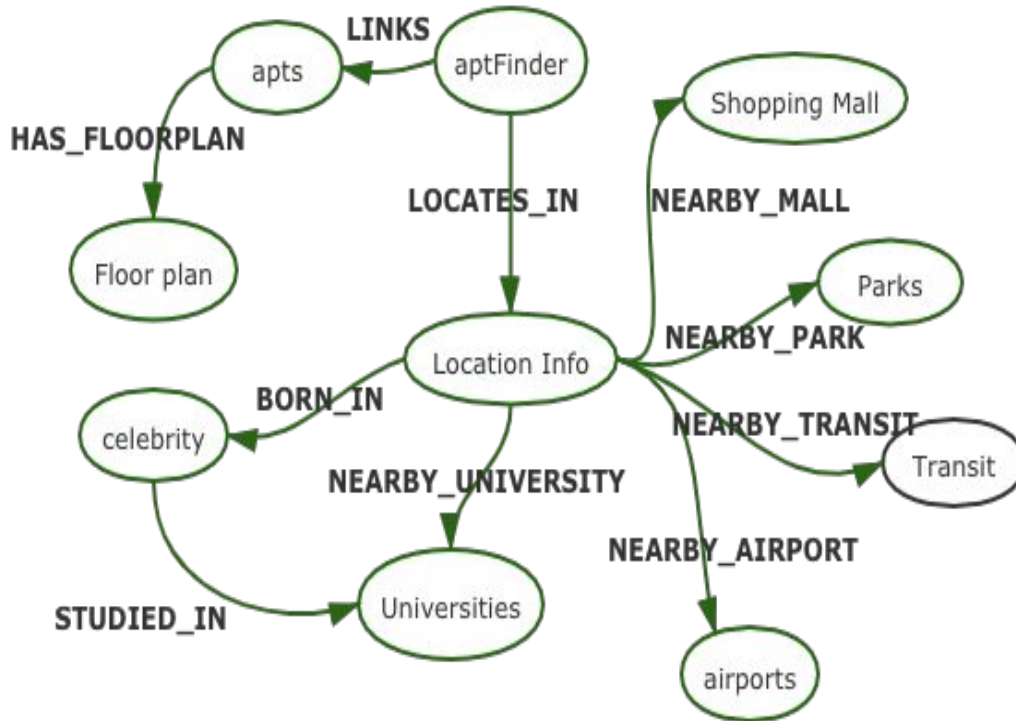
$$\text{Similarity} = \alpha * \text{SimLocation} + (1 - \alpha) * \text{SimTitle}$$

**Zipcode-based blocking improves time from 30mins to 30s per county!**

The screenshot shows the Apartments.com website interface. At the top, there is a navigation bar with a back arrow, a menu icon, the text 'Menu', a globe icon with 'Español', and the 'Apartments.com' logo. Below this, the listing title 'Sienna on Sanborn' is displayed with a 'Verified' badge. The address '1026-1036 Sanborn Ave, Los Angeles, CA 90029' is shown, followed by a truncated location name 'Northeast Lo...'. There are five green stars and '(0 Reviews)' with a checkmark. Below the listing, there is a section with a back arrow, the 'AF' logo, and a search bar containing 'Los Angeles, CA'. Two green boxes with labels are used for annotations: one labeled 'Title' points to the listing title, and another labeled 'Location' points to the address. Below the search bar, the text 'Sienna on Sanborn Apartments' is shown, followed by the address '1026-1036 Sanborn Ave, Los Angeles, CA 90029' and the location 'Northeast Los Angeles'.

### 3. Ontology Mapping

**46K nodes**  
**160+ relationships**



## 4a. Query Matching with **Keywords**

### Keywords

Q: “An apartment near the **shopping center**, with high **security**. I can go for a walk in the nearby **park** on weekends. And the people here are highly **educated** and relatively **rich**.”

Q: “I want to rent a **new** apartment where I can easily make **friends** with **young** neighbors. It's better if many **Asian** lives here.”

**shopping**

number of nearby shopping malls

**security**

confirmed cases & deaths of covid-19; crime rate

category	measurement
<b>education</b>	the proportion of high school degree & bachelor; number of nearby universities
<b>economy</b>	employ rate; mean income of families
age of apartment	built_in time
<b>social</b>	median age; the proportion of married
race distribution	the proportion of Asian/black/white



## 4b. Query Matching through **Entity resolution**

### ER & Rule-based Matching

Q: “An apartment near the residence of **Nylon Pink**, with high security. It’s better if I can go for a walk in the **Los Encinos State Historic Park** on weekends. And the people here are highly educated and relatively rich.”

Q: “An apartment near **6201 Hollywood Blvd.**”

### Using SpaCy

#### Celebrity

Capitalized/Entity: PERSON

#### Spot

Capitalized/Entity: ORG,LOC,FAC

#### Address

Capitalized/pattern: CARDINAL,LOC,GRE...

## 4c. Query Matching by **topic modeling**

### Topic Modeling: LDA

Q: “I want to find an apartment near Hollywood, with a big swimming pool and a big kitchen. I’ll appreciate if the apartment has a well equipped onsite gym. Be close to downtown. Modern designed. I like relaxing in my private pool after a long day....”

**Topic ID**  
+  
**words & weights**

#### Getting Raw Data

apartment.describe

#### Data Pre-processing

tokenization, removing stop words, lemmatizing, stemming, filtering

#### Training Models

Bag of Words / TF-IDF

#### Evaluation

Average score

#### Save Results & Trained model

## 5a. Use case: Rental recommendations

County Name

Special Requirement

Pet Policy ☒ Pet Allowed

Parking ☐ Parking Contained

Beds

Baths

Sqft

Price Range

Lease Term

Features Choose

Floorplan Features	Apartment Features
Air Conditioning	Breakfast/Coffee Concierge
Balcony	Car Charging Station
Cable Ready	Clubhouse

Submit

Reset

### Golden Valley Luxury Apartments [Apartment Finder](#) [Apartments](#)

Address: 600 Hosking Ave, Bakersfield, CA Zipcode: 93307 Phone Number: 844-425-7080 Lowest\_price: 1350 Rating: 5.0

Cat Policy: Not Allowed Dog Policy: Allowed Parking Policy: Allowed Shortest Lease Term: 12 Builtin Time: Unknown Reviews: 7

**Covid Deaths: 428**

**Covid Confirmation: 34973**

Bachelor Rate: 0.04 Marriage Rate: 0.49 Jail: 0.02 Employment Rate: 0.59 White Ratio: 0.71 Black Ratio: 0.06 Asian Ratio: 0.01

**Description** Bakersfield Apartments at their BEST! Gated luxury community of spacious 2- and 3-bedroom townhome-style apartments. Large gourmet kitchens with granite counters, ceramic tile floors, excellent appliances included, stainless double-bowl sink, and dining space in kitchen! Great living area with laminate wood floors and pre-wired for digital cable/Internet/phone. Ceiling fans throughout, efficient 13-SEER HVAC and tankless water heaters. 1059-1236 square feet. No noise upstairs - your unit is 2-story! Centered around a large green park, with a cardio gym, clubhouse, basketball court, soccer field, BBQ's, and picnic benches. Separate dog park! Management on-site, and roving courtesy patrol to maintain peace and quiet. Model apartments available to view with front office. We look forward to seeing you! Easy Online Application at [www.GVLuxury.com](http://www.GVLuxury.com)

**Rent Special** Staying Home? We offer online virtual tours! We are also scheduling in-person, self-guided tours by appointment! E-mail us at [managers@heritagepmr.com](mailto:managers@heritagepmr.com) for more information! Our office doors are closed, but we remain available during our regular office hours and we look forward to helping you move in!

High Speed Internet Access | Washer/Dryer | Washer/Dryer Hookup | Air Conditioning | Heating | Ceiling Fans | Cable Ready | Floorplan Tub/Shower | Wheelchair Accessible | Dishwasher | Disposal | Granite Countertops | Eat-in Kitchen | Kitchen | Microwave | Oven | Amenities Range | Refrigerator | Hardwood Floors | Carpet | Tile Floors | Dining Room | Vaulted Ceiling | Walk-In Closets | Double Pane Windows

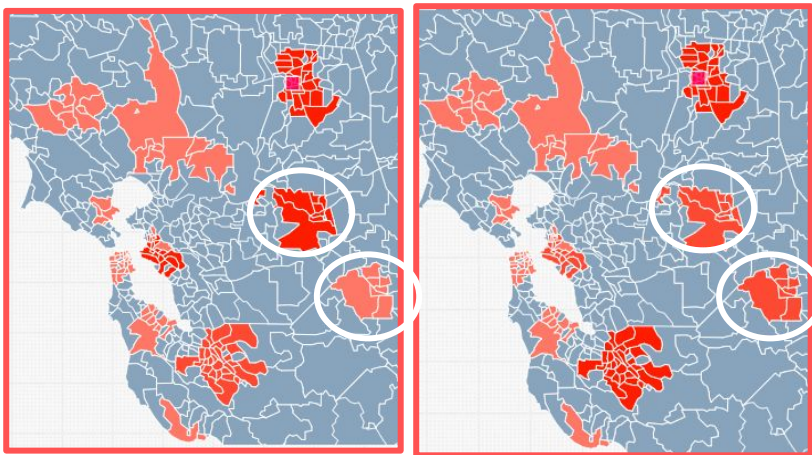
Community Amenities Laundry Facilities | Furnished Units Available | Pet Play Area | Business Center | Clubhouse | Lounge | Corporate Suites | Sundeck | Cabana | Picnic Area | Fitness Center | Spa | Pool | Playground | Basketball Court |

**User Evaluation:** Compare the top 50 results from 10 users with standard results for 9 scenarios (see paper for details)

## 5b. Use case: Spatio-temporal dynamics of the market

- Track and analyze data changes between different counties, states, or countries
- Easily and continuously expand data over time
- Answer questions related to marketing or social science

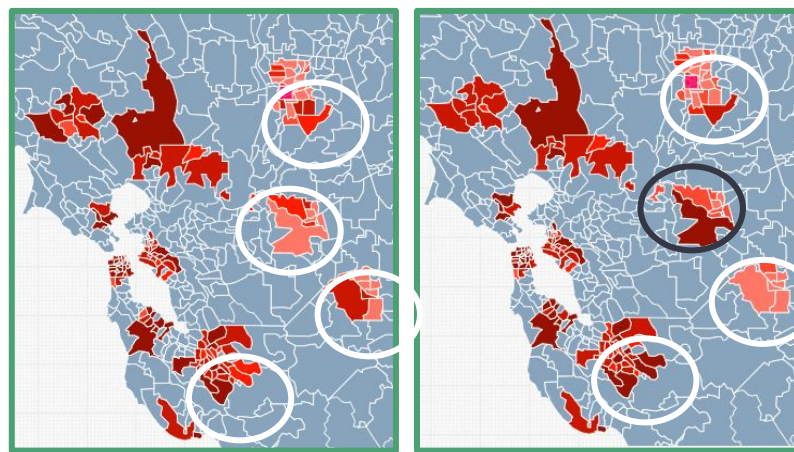
Example (deaths):



Nov 2020

Feb 2021

Example (prices):



Nov 2020

Feb 2021

**KG-based  
method**

**Data Integration**

**Personalized Recommendation**

**Housing Market Analysis**

**Reuse and Expand Easily!**

**Novel use cases**

*Impact of regional education and climate on China's housing prices*

# Thank You

---