

# RAG imobiliar – selecție și estimare de „preț corect”

Proiectul implementează un sistem **Retrieval-Augmented Generation (RAG)** specializat pe piața imobiliară, având rolul de a selecta proprietăți relevante și de a estima un „**Preț Corect**” transparent. Sistemul hibrid combină **filtrarea pe metadate, regăsirea semantică și analiza cantitativă ponderată** pentru a clasifica prețul listat ca fiind UNDERPRICED, FAIR sau OVERPRICED.

## Sursa de Date și Indexare:

data\_preprocessing.py

Setul de date inițial a fost curățat prin eliminarea dupliilor și a înregistrărilor incomplete.

## Gestionarea Duplicatelor și a Valorilor Lipsă:

- S-a menținut eliminarea dupliilor pe baza coloanei **id** pentru a evita suprareprezentarea aceleiași proprietăți în setul de date.
- S-a aplicat o filtrare strictă (dropna) pe coloanele: price\_eur, size\_sqm, property\_type, city, neighborhood deoarece aceste atribute sunt obligatorii atât pentru indexarea semantică, cât și pentru calculele de preț.

## Conversia Tipului de Date:

- S-a confirmat conversia la tip numeric (pd.to\_numeric) pentru coloanele numerice. Folosirea argumentului errors="coerce" este utilizată pentru a transforma valorile non-numerice (ex: -, N/A) în NaN, permitând tratarea lor ulterioră.

## Verificări de Sănătate (Sanity Checks):

- S-au menținut filtrele de bun simț (ex: preț între \$5.000\$ și \$2.000.000\$ €; suprafață între \$10\$ și \$2.000\$ mp). **Justificare:** Acestea elimină erorile de introducere a datelor sau proprietățile extreme care ar distorsiona media prețului/mp.

## Crearea Caracteristicilor Derivate (Feature Engineering)

- **price\_per\_sqm**
- **age**
- **is\_new\_build**

- **distance\_score**

## Construcția Documentului pentru Embeddings:

Construcția textului (text\_for\_embedding) este benefică pentru performanța RAG, deoarece acest text este vectorizat și folosit pentru căutarea semantică (similitudinea).

**Concatenare Logica:** S-a menținut concatenarea atributelor esențiale (property\_type, neighborhood, city, size\_sqm, rooms, year\_built) urmate de textul liber din description.

Datele curățate și îmbogățite sunt salvate în fișierul properties\_clean.csv, care va fi sursa de intrare pentru modulul de indexare (build\_embeddings.py)

A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM					
1	d	tele	property_price_per_sqm	city	neighborhood	lat	lon	size_sqm	rooms	year_built	floor	max_floor	parking	heating	dist_to_center	dist_to_z	descrip_source	price_per_sqm	age	is_new	distance_text	embedding																			
2	1	Land ca lant	1E+09	Bucuresti Colentina	44,436	26,125	890	2	1982	yes	central	3	5	Proprietate https://	1170	0	6	2, land în Colentina, Bucuresti, 890 mp, nici camere, construit în anul 1982. Detali Proprietate în zona Colentina, suprafata 890 mp, pret 1041300 EUR.																							
3	2	House 2 houses	158943	Bucuresti Pipera	44,425	26,107	57	2	1976	1	10	gas	9	3	Proprietate https://	915	43	0	4,24 houses în Pipera, Bucuresti, 2,78 mp, 2 camere, construit în anul 1976. Detali Proprietate în zona Pipera, suprafata 37 mp, pret 102000 EUR.																						
4	3	Apartment apartment	108891	Bucuresti Mihai Viteazul	44,401	26,112	68	3	2005	no	electrică	10	3	Proprietate https://	12490	43	0	3,2, apartament în Mihai Viteazul, Bucuresti, 68 mp, 3 camere, construit în 2005. Detali Proprietate în zona Mihai Viteazul, suprafata 37 mp, pret 105200 EUR.																							
5	4	House 3 house	94520	Bucuresti Mihai Viteazul	44,401	26,112	68	3	2005	2	2021	9	10	temofică	7	14	Proprietate https://	1752	4	1	8,2 apartament în Mihai Viteazul, Bucuresti, 68 mp, 2,0 camere, construit în 2005. Detali Proprietate în zona Mihai Viteazul, suprafata 68 mp, pret 105072 EUR.																				
6	5	Apartment apartment	150672	Bucuresti Mihai Viteazul	44,406	26,107	86	2	2021	no	temofică	7	14	Proprietate https://	1752	4	1	8,2 apartament în Mihai Viteazul, Bucuresti, 86 mp, 2,0 camere, construit în 2021. Detali Proprietate în zona Mihai Viteazul, suprafata 86 mp, pret 105067 EUR.																							
7	7	Apartment apartment	123094	Bucuresti Drumul Taberei	44,426	26,106	102	2	2016	7	10	gas	10	6	Proprietate https://	1002	1	1	3,6, apartament în Drumul Taberei, Bucuresti, 102 mp, 2,0 camere, construit în 2016. Detali Proprietate în zona Drumul Taberei, suprafata 102 mp, pret 104000 EUR.																						
8	8	House 3 house	137923	Bucuresti Distor	44,427	26,117	107	3	1979	2	Proprietate https://	1285	46	0	5,8 houses în Distor, Bucuresti, 307 mp, 30 camere, construit în anul 1979. Detali Proprietate în zona Distor, suprafata 107 mp, pret 107323 EUR.																										
9	10	House 4 house	141840	Bucuresti Tatarasi	44,419	26,105	119	4	2023	yes	gas	4	12	Proprietate https://	152	2	1	10,8 houses în Tatarasi, Bucuresti, 119 mp, 4 camere, construit în 2023. Detali Proprietate în zona Tatarasi, suprafata 119 mp, pret 108400 EUR.																							
10	11	House 3 house	123882	Bucuresti Distor	44,427	26,106	95	2	2000	no	temofică	10	12	Proprietate https://	1452	25	0	4,2 houses în Distor, Bucuresti, 95 mp, 2 camere, construit în 2000. Detali Proprietate în zona Distor, suprafata 95 mp, pret 105225 EUR.																							
11	12	House 3 house	68680	Bucuresti Pipera	44,435	26,061	72	3	1994	yes	central	4	7	Proprietate https://	1090	41	0	12,8 houses în Pipera, Bucuresti, 72 mp, 30 camere, construit în 1994. Detali Proprietate în zona Pipera, suprafata 72 mp, pret 106541 EUR.																							
13	14	Apartment apartment	106891	Bucuresti Colentina	44,438	26,194	53	1	1999	0	10	temofică	11	10	Proprietate https://	937	26	0	7,2 apartament în Colentina, Bucuresti, 53 mp, 2,0 camere, construit în 1999. Detali Proprietate în zona Colentina, suprafata 53 mp, pret 105712 EUR.																						
15	16	House 4 house	104893	Bucuresti Distor	44,434	26,108	60	4	2007	yes	gas	20	4	Proprietate https://	1430	22	0	3,2 houses în Distor, Bucuresti, 60 mp, 20 camere, construit în 2007. Detali Proprietate în zona Distor, suprafata 60 mp, pret 104890 EUR.																							
17	18	House 2 house	33951	Bucuresti Distor	44,444	26,151	43	2	2015	yes	gas	7	6	Proprietate https://	1749	18	0	4,4 houses în Colentina, Bucuresti, 43 mp, 2,0 camere, construit în 2015. Detali Proprietate în zona Colentina, suprafata 43 mp, pret 104080 EUR.																							
19	20	House 2 house	60538	Bucuresti Pipera	44,424	26,105	46	3	1982	no	temofică	19	13	Proprietate https://	1391	30	0	14, houses în Pipera, Bucuresti, 46 mp, 30 camere, construit în anul 1982. Detali Proprietate în zona Pipera, suprafata 46 mp, pret 106544 EUR.																							
21	20	Apartment apartment	123882	Bucuresti Distor	44,437	26,167	48	1	1979	6	10	no	electrică	16	11	Proprietate https://	1873	46	0	1,11 apartamente în Bucuresti, 48 mp, 4,0 camere, construit în 1979. Detali Proprietate în zona Distor, suprafata 48 mp, pret 102272 EUR.																					
22	23	Land ca lant	123882	Bucuresti Tatarasi	44,419	26,107	950	1	1982	yes	gas	10	8	Proprietate https://	1714	65	0	1,4 apartamente în Tatarasi, Bucuresti, 950 mp, 10 camere, construit în 1982. Detali Proprietate în zona Tatarasi, suprafata 950 mp, pret 105918 EUR.																							
24	23	House 3 house	71045	Bucuresti Mihai Viteazul	44,436	26,088	61	4	1986	yes	temofică	13	8	Proprietate https://	1764	39	0	3,8 houses în Mihai Viteazul, Bucuresti, 61 mp, 30 camere, construit în 1986. Detali Proprietate în zona Mihai Viteazul, suprafata 61 mp, pret 106549 EUR.																							
25	24	Apartment apartment	68680	Bucuresti Distor	44,436	26,173	52	2	1991	5	10	no	temofică	5	12	Proprietate https://	1704	34	0	10,2 apartamente în Distor, Bucuresti, 52 mp, 2,0 camere, construit în 1991. Detali Proprietate în zona Distor, suprafata 52 mp, pret 106508 EUR.																					
26	27	Apartment apartment	204643	Bucuresti Distor	44,404	26,105	103	5	1981	7	10	no	central	17	13	Proprietate https://	1871	52	0	1,4 apartamente în Distor, Bucuresti, 103 mp, 2,0 camere, construit în 1981. Detali Proprietate în zona Distor, suprafata 103 mp, pret 104045 EUR.																					
27	28	House 4 house	145238	Bucuresti Distor	44,432	26,115	101	4	2003	yes	temofică	6	6	Proprietate https://	1438	22	0	11,8 apartamente în Distor, Bucuresti, 115 mp, 3,0 camere, construit în 2003. Detali Proprietate în zona Distor, suprafata 115 mp, pret 105128 EUR.																							
29	30	Apartment apartment	19884	Bucuresti Colentina	44,411	26,091	33	3	1982	3	10	no	temofică	13	15	Proprietate https://	988	63	0	4,2 apartamente în Colentina, Bucuresti, 33 mp, 30 camere, construit în 1982. Detali Proprietate în zona Colentina, suprafata 33 mp, pret 10988 EUR.																					
31	30	Apartment apartment	150672	Bucuresti Distor	44,409	26,072	48	1	1979	6	10	no	central	10	11	Proprietate https://	1040	44	0	5,5 apartamente în Distor, Bucuresti, 48 mp, 10 camere, construit în 1979. Detali Proprietate în zona Distor, suprafata 48 mp, pret 105044 EUR.																					
32	31	Apartment apartment	123882	Bucuresti Distor	44,437	26,167	66	1	1979	6	10	yes	gas	4	12	Proprietate https://	1848	65	0	1,12 apartamente în Distor, Bucuresti, 66 mp, 4,0 camere, construit în 1979. Detali Proprietate în zona Distor, suprafata 66 mp, pret 102272 EUR.																					
33	32	House 4 house	202950	Bucuresti Mihai Viteazul	44,421	26,161	10	4	1958	yes	gas	8	8	Proprietate https://	1845	27	0	3,8 houses în Mihai Viteazul, Bucuresti, 10 mp, 8 camere, construit în 1958. Detali Proprietate în zona Mihai Viteazul, suprafata 10 mp, pret 102950 EUR.																							
34	33	Apartment apartment	108891	Bucuresti Tatarasi	44,416	26,098	41	1	1979	8	10	yes	central	6	4	Proprietate https://	960	46	0	12,8 apartamente în Tatarasi, Bucuresti, 41 mp, 10 camere, construit în 1979. Detali Proprietate în zona Tatarasi, suprafata 41 mp, pret 103360 EUR.																					
35	36	Land ca lant	662605	Bucuresti Distor	44,424	26,182	445	3	2003	yes	gas	7	13	Proprietate https://	1489	65	0	6,6 land în Distor, Bucuresti, 445 mp, nan camere, construit în man. Detali Proprietate în zona Distor, suprafata 445 mp, pret 106265 EUR.																							
37	36	House 3 house	579020	Bucuresti Distor	44,424	26,107	117	3	2003	yes	gas	8	13	Proprietate https://	1051	22	0	4,4 houses în Distor, Bucuresti, 117 mp, 10 camere, construit în 2003. Detali Proprietate în zona Distor, suprafata 117 mp, pret 107255 EUR.																							
38	37	Land ca lant	579020	Bucuresti Distor	44,424	26,107	115	3	1980	4	10	yes	electrică	10	12	Proprietate https://	1332	45	0	4,4 houses în Distor, Bucuresti, 115 mp, 10 camere, construit în 1980. Detali Proprietate în zona Distor, suprafata 115 mp, pret 107255 EUR.																					
39	40	Apartment apartment	17831	Bucuresti Mihai Viteazul	44,411	26,067	71	2	2007	6	10	no	central	8	3	Proprietate https://	1073	18	0	12 apartamente în Pipera, Bucuresti, 71mp, 2,0 camere, construit în 2007. Detali Proprietate în zona Pipera, suprafata 71mp, pret 76783 EUR.																					
40	41	House 2 house	507203	Bucuresti Berceni	44,414	26,065	675	1	1987	2	Proprietate https://	1036	38	0	9,8 houses în Berceni, Bucuresti, 675 mp, 20 camere, construit în 1987. Detali Proprietate în zona Berceni, suprafata 675 mp, pret 107200 EUR.																										
41	41	Land ca lant	507203	Bucuresti Berceni	44,414	26,065	675	1	1987	2	Proprietate https://	1034	20	0	5,2 land în Berceni, Bucuresti, 675 mp, 20 camere, construit în 1987. Detali Proprietate în zona Berceni, suprafata 675 mp, pret 107200 EUR.																										
42	42	Apartment apartment	165120	Bucuresti Distor	44,404	26,059	9	3	2004	10	10	yes	gas	13	3	Proprietate https://	1668	21	0	5,4 apartamente în Dobrogea, Bucuresti, 9 mp, 30 camere, construit în 2004. Detali Proprietate în zona Dobrogea, suprafata 9 mp, pret 108120 EUR.																					
43	44	3 House 3 house	171337	Bucuresti Tatarasi	44,412	26,052	101	3	1995	4	10	no	temofică	16	8	Proprietate https://	1051	30	0	5,2 houses în Berceni, Bucuresti, 101 mp, 30 camere, construit în 1995. Detali Proprietate în zona Berceni, suprafata 101mp, pret 107397 EUR.																					
44	45	Land ca lant	407402	Bucuresti Mihai Viteazul	44,436	26,199	420	1	1977	4	10	yes	gas	19	14	Proprietate https://	970	48	0	3,4 land în Mihai Viteazul, Bucuresti, 420 mp, nan camere, constr																					

- **Construcția Textului de Indexare (build\_index\_text):** S-a adoptat o strategie de concatenare comprehensivă. Spre deosebire de textul simplu folosit doar pentru vizualizare sau descriere (din data\_preprocessing.py), acest text include și detalii cantitative critice (preț, etaj, distanță la metrou).

### Generarea Embeddings-urilor

Această etapă transformă textul din coloana index\_text în vectori numerici.

- **Modelul de Embeddings:** A fost ales **sentence-transformers/all-MiniLM-L6-v2**
- **Procesul de Codificare:** Funcția model.encode este aplicată pe întregul set de date, folosind un **batch\_size** de 64. Acest lucru optimizează utilizarea resurselor (GPU/CPU) și accelerează procesul.

### Popularea Vector Store-ului (ChromaDB)

Datele și vectorii sunt stocați într-o colecție ChromaDB pentru a permite căutarea rapidă și filtrarea eficientă.

- **Clientul Persistent:** S-a utilizat **chromadb.PersistentClient** cu calea **vector\_store**, asigurând că baza de date este salvată local și poate fi încărcată rapid la rulări ulterioare ale aplicației (retrieval.py și app.py).
- **Colecția (real\_estate\_properties):** S-a ales spațiul de similaritate cosine ("hnsw:space": "cosine"), care este cel mai comun și eficient pentru măsurarea similarității între vectorii generați.
- **Adăugarea Datelor:** Metoda collection.add este folosită pentru a încărca:
  1. ids: ID-urile proprietăților (df["id"]).
  2. embeddings: Vectorii numerici generați.
  3. documents: Textul original de indexare (pentru context).
  4. metadata: Setul de atrbute structurate esențiale pentru filtrare (ex: property\_type, city, neighborhood, price\_eur, size\_sqm).

Aceste metadate permit ca o interogare precum "*apartament, Titan, minim 50 mp*" să fie tradusă în: Filtru: property\_type='apartment' AND neighborhood='Titan' AND size\_sqm>=50 + Căutare Vectorială: pe textul "apartament Titan 50 mp".

## Motorul de Regăsire:

retrieval.py

Modulul îndeplinește funcția de a găsi proprietățile **cel mai relevante** (comparabile) pentru cererea utilizatorului. Utilizează o abordare **hibridă**: combină regăsirea semantică (vectorială) cu aplicarea filtrelor structurate și un algoritm de re-ranking bazat pe factori economici.

### Extracția Filtrelor Logice (extract\_filters)

Această funcție parsează cererea text liber a utilizatorului și extrage atributele structurate necesare pentru pre-filtrare și re-ranking.

### Logică Bazată pe Expresii Regulate (Regex):

- **Tipul Proprietății:** Detectează cuvinte cheie ca "apartament", "casa", "teren" pentru a seta filtrul property\_type.
- **Număr de Camere:** Folosește expresia regulată `r"(\d+)\s*cam"` pentru a extrage numărul exact de camere.
- **Buget (Preț Maxim):** Folosește `r"(\d{2,6})\s*euro"` pentru a identifica un buget specific, stocat ca price\_max.
- **Cartier/Zonă:** Caută cuvinte cheie prestabilite dintr-o listă (ex: "titán", "militari") pentru a seta filtrul neighborhood.

### Funcția Principală de Regăsire (get\_comparables)

Această funcție orchestrator folosește regăsirea vectorială, pregătind datele pentru re-ranking.

- **Generarea Embedding-ului Query:** Cererea utilizatorului este transformată în vector (q\_emb) folosind același model (all-MiniLM-L6-v2) utilizat la indexare.
- **Interogarea ChromaDB:** Se execută collection.query cu vectorul de interogare.
- **Pregătirea Datelor:** Rezultatele brute (ID-uri și embeddings) sunt combinate cu datele structurate din DataFrame-ul df (încărcat anterior) pentru a avea acces la price\_per\_sqm, neighborhood, etc.

### Algoritmul de Re-ranking (rerank\_results)

Această etapă adaugă inteligență economică peste similaritatea semantică pură. Scopul este de a ajusta similarity (scorul inițial) la un final\_score care reflectă mai bine relevanța pentru estimarea prețului.

- **Similaritatea de Bază (Cos-Similitudine):** Scorul de bază este calculat folosind similaritatea cosinus între vectorul de interogare și vectorul proprietății.
- **Ponderare Logică (Bonus/Penalizare): Apropierea de Cartier (Bonus)** -> Dacă proprietatea se află în același cartier cu cel cerut în filtru (ex: "Titan"), i se aplică un bonus semnificativ de 0.20. Dacă se află în alt cartier, i se aplică o penalizare de 0.10. Similaritatea geografică și locală are un impact major asupra prețului/mp, fiind mai importantă decât similaritatea strict textuală.
- **Penalizare Preț/mp vs. Buget (Relevanță Economică)** -> Se calculează un `price_per_sqm ideal` pe baza bugetului utilizatorului (`price_max`) și a suprafeței proprietății comparabile (`size_sqm`). Diferența absolută (`diff`) dintre `price_per_sqm` real al comparabilei și cel `ideal` este folosită pentru a aplica o penalizare, limitată la 0.15. Aceasta asigură că proprietățile regăsite nu sunt doar "similare" (ex: "studio vechi"), ci și similare din punct de vedere al prețului percepțuit de către utilizator.

## Sortarea Finală

Lista finală de comparabile este sortată descrescător pe baza `final_score` și este trunchiată la top-k (implicit k=10). Aceste rezultate sunt salvate în `comparables.json` pentru a fi folosite în modulul de preț.

Exemplu:

```
{
  "id": 762,
  "similarity": 0.6281761515674025,
  "final_score": 0.8279024673568761,
  "property_type": "apartment",
  "neighborhood": "Titan",
  "city": "Bucuresti",
  "price_eur": 60078,
  "size_sqm": 57,
  "price_per_sqm": 1054.0
}
```

## pricing\_model.py

Modulul pricing\_model.py este responsabil pentru analiza cantitativă a proprietăților regăsite și pentru generarea verdictului final (UNDERPRICED, FAIR, OVERPRICED).

- Încărcarea Comparabilelor (load\_comparables): Modulul începe prin încărcarea rezultatelor din etapa anterioară (retrieval.py), salvate în comparables.json. Aceste rezultate sunt o listă de proprietăți, fiecare având deja calculat **final\_score** (scorul de similitudine re-clasificat) și **price\_per\_sqm**.
- Calculul Prețului Corect Ponderat (compute\_fair\_price):
  1. Determinarea Suprafaței Țintă (target\_sqm) → Dacă suprafața proprietății analizate (target\_sqm) nu a fost specificată în cererea inițială a utilizatorului, funcția folosește ca valoare implicită **media suprafetelor comparabilelor regăsite**
  2. Media Ponderată a Prețului/mp → Prețul corect pe metru pătrat fair\_ppsqm este calculat ca o **medie ponderată**, unde ponderile sunt date de **final\_score-ul** din etapa de re-ranking. Aceasta asigură că proprietățile considerate mai relevante (similaritate semantică mare, același cartier, preț/mp apropiat de bugetul utilizatorului) au o influență mai mare asupra estimării finale.
  3. Calculul Prețului Total → Prețul Corect Total este determinat prin scalarea prețului unitar estimat la suprafața țintă.
- Logica de Clasificare și Etichetare (Verdictul): Verdictul este dat prin compararea prețului listat al proprietății analizate (target\_price) cu un interval de încredere calculat în jurul Prețului Corect Total.

Etichetă	Condiție (conform cerinței inițiale)
<b>UNDERPRICED</b>	Preț Listat < 90% din Preț Corect Total
<b>FAIR</b>	Preț Listat se încadrează între 90% și 110% din Preț Corect Total
<b>OVERPRICED</b>	Preț Listat > 110% din Preț Corect Total

- Generarea Raportului de leșire (evaluate\_property): Funcția principală orchestrează încărcarea, calculul și salvarea datelor într-un raport JSON (pricing\_output.json).

Exemplu:

```
• "estimation": {
    "fair_price": 60535.62,
    "fair_ppsqm": 1121.03,
    "confidence_interval": {
        "lower": 57508.84,
        "upper": 63562.4
    },
    "verdict": "FAIR",
    "target_sqm": 54
},
```

## Explicația și Interfața:

explanation\_module.py

Modulul este responsabil de faza de Generation (Generare) din arhitectura RAG. Aceasta ia rezultatul numeric și structurat din etapa de Retrieval & Pricing și îl transformă într-o justificare narativă, ușor de înțeles de către utilizator.

- **Rolul și Integrarea LLM-ului:** LLM-ul nu inventează date. Rolul său este strict de a rezuma și justifica concluzia pe baza contextului furnizat (comparabile și estimări), asigurând acuratețea rezultatului. Model Utilizat: Mistral (accesat local prin Ollama).
- **Construirea Contextului (Prompt Engineering):** Calitatea explicației depinde de cât de bine sunt structurate datele de intrare în prompt. Modulul extrage datele cheie din raportul de preț (pricing\_output.json). Prețul listat, prețul corect estimat și intervalul de încredere (fair\_min – fair\_max). **Verdictul final (label):** Aceasta este punctul central al explicației. **Lista Comparabilelor:** Sursa de credibilitate. Lista este formatată concis, incluzând ID-ul, cartierul, prețul/mp și scorul final, pentru a fi ușor de referențiat de către LLM.
- **Instructiunea Prompt-ului (Persona & Scop):**
  1. **Persona:** Asistent de evaluări imobiliare.
  2. **Atenționare:** Să nu inventeze informații (fara sa inventezi informatii).
  3. **Restricție:** Concizie (maxim 5 propoziții).
  4. **Sarcina:** Justificarea verdictului bazată pe datele oferite.

- **Execuția LLM (Rularea Locală):** Se utilizează subprocess.run pentru a apela utilitarul ollama din linia de comandă, trimițând promptul prin stdin (input=prompt.encode("utf-8")). Se specifică modelul mistral (ollama run mistral), care este un model puternic pentru sarcini de înțelegere și rezumare în limba română.
- **Generarea Ieșirii:** Rezultatul rulării LLM-ului este extras, decodat și inclus în obiectul final JSON, împreună cu un avertisment obligatoriu. **explanation\_text:** Textul narativ generat de LLM. **disclaimer:** "Proiect educațional; nu reprezintă consultanță imobiliară.", îndeplinind o cerință de etică și bune practice.

Exemplu:

{

"explanation\_text": "Proprietatea analizată are un preț corect estimat de 60.535,62 €, cu o margină de eroare de aproximativ ± 3.000 € (interval 57.508,84 - 63.562,4 €). Comparând-o cu alte proprietăți similare din sistemul RAG, scorul evaluat este de 0,70, care indică o valorificare Fair. Prețul pe metru pătrat al proprietății analizate variază de la 1.054,0 €/mp (casa cu 762 număr de titan) până la 1.289,0 €/mp (casa cu 1270 număr de titan), cu o medie de 1.033,0 €/mp. Această valoare este apropiată de media prețului pe metru pătrat al proprietăților comparabile cu scorul de evaluare Fair (scoruri cuprinse între 0,68 și 0,71). În concluzie, verdictul pentru proprietatea analizată este justificat prin comparația cu alte proprietăți similare din sistemul RAG și de medie scorului evaluat.",

"disclaimer": "Proiect educațional; nu reprezintă consultanță imobiliară."

}

app.py

Modulul app.py servește ca **Frontend și Orchestrator** al întregului sistem RAG Imobiliar. Utilizează **Streamlit** pentru a oferi o interfață web rapidă și interactivă, integrând celelalte module (retrieval.py, pricing\_model.py, explanation\_module.py).

- **Interfața Utilizator și Preluarea Input-ului:**
  1. **Sidebar (Filtre și Proprietate Tintă):** Colectează filtrele necesare pentru a construi cererea în limbaj natural (ex: property\_type, neighborhood, min\_size, max\_budget). Colectează datele proprietății analizate: title\_input, listed\_price\_input și size\_input. Butonul "Calculeaza preț corect" inițiază întregul pipeline, iar intrările sunt salvate în st.session\_state pentru persistență.

**2. Main Panel (Rezultate):** Afisează succesiv rezultatele celor trei etape principale.

- Fluxul de Execuție (Orchestrarea):**

La apăsarea butonului, app.py execută secvența RAG: Construcția Query-ului, Retrieval & Ranking, Calculul Prețului, Generarea Explicației

Transparența este asigurată prin afișarea tuturor datelor intermedii și finale: Tabelul Comparabilelor, Estimarea Prețului Corect, Verdictul și Explicația

- Vizualizarea Geografică (Folium):**

Secțiunea de hartă adaugă un context vizual critic: Date Geografice, Vizualizarea Prețului/mp.

- Exportul JSON:**

Pentru reproducibilitate și îndeplinirea cerinței, funcția build\_final\_output creează obiectul JSON final, care este oferit utilizatorului spre descărcare printr-un widget st.download\_button.

The screenshot shows the main panel of the RAG Imobiliar application. On the left, there is a sidebar with filtering options: 'Tip proprietate' (apartment), 'Cartier (ex: Titan)' (Titan), 'Suprafața minima (mp)' (40), 'Buget maxim (EUR)' (120000), and 'Numar comparabile (k)' (10). The main area has a title 'RAG Imobiliar – Estimare pret corect (demo)' and a subtitle 'Rulez: apartment Titan 54 mp buget 60000 euro'. Below this is a section titled 'Comparabile (după reranking)' containing a table with the following data:

	ID	property_type	neighborhood	price_eur	size_sqm	price_per_sqm	final_score
0	762	apartment	Titan	60078	57	1054	0.8279
1	1556	apartment	Titan	59565	57	1045	0.8186
2	1345	apartment	Titan	69464	76	914	0.7996
3	115	apartment	Titan	72028	44	1637	0.7936
4	1076	apartment	Titan	88160	80	1102	0.7915
5	733	apartment	Titan	51840	40	1296	0.7895
6	232	apartment	Titan	73718	58	1271	0.7734
7	838	apartment	Titan	72336	48	1507	0.767
8	843	apartment	Titan	71632	74	968	0.7669
9	31	apartment	Titan	79305	85	933	0.7647

Buget maxim (EUR)  
120000 - +

Numar comparabile (k)  
10

Input pentru proprietatea analizata (optional):  
Titlu proprietate  
Apartament 2 camere titan

Pret listat (EUR)  
60000 - +

Suprafata (mp)  
54 - +

**Calculeaza pret corect**

Verdict: UNDERPRICED

**Explicație**

Apartamentul cu două camere titan este subevaluat, având un preț listat de 60.000 €, în timp ce prețul corect estimat este de 63.273,53 €, cu o gamă de prețuri de 60.109,85–66.437,2 €. Evaluarea a fost făcută prin comparare cu alte apartamente similare din sistemul RAG. Acestea au prețuri medii pe metru pătrat (€/mp) cuprinse între 914 și 1637 €/mp, având suprafață cuprinsă între 44 și 85 mp și un scor de 0,76–0,83 în sistemul RAG. Prețul actual al apartamentului nu este corespunzător cu celelalte proprietăți similare din aceeași categorie, pentru care a fost clasificat ca fiind subevaluat (UNDERPRICED).

Proiect educational; nu reprezintă consultanță imobiliară.

Hartă 