

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 duplicatelor și a înregistrărilor incomplete.

Gestionarea Duplicatelor și a Valorilor Lipsă:

- S-a menținut eliminarea duplicatelor pe baza coloanei **id** pentru a evita supra-reprezentarea 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, permițând tratarea lor ulterioară.

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 textuliber 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)

	M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI EJ EK EL EM EN EO EP EQ ER ES ET EU EV EW EX EY EZ FA FB FC FD FE FF FG FH FI FJ FK FL FM FN FO FP FQ FR FS FT FU FV FW FX FY FZ GA GB GC GD GE GF GG GH GI GJ GK GL GM GN GO GP GQ GR GS GT GU GV GW GX GY GZ HA HB HC HD HE HF HG HH HI HJ HK HL HM HN HO HP HQ HR HS HT HU HV HW HX HY HZ IA IB IC ID IE IF IG IH II IJ IK IL IM IN IO IP IQ IR IS IT IU IV IW IX IY IZ JA JB JC JD JE JF JG JH JI JJ JK JL JM JN JO JP JQ JR JS JT JU JV JW JX JY JZ KA KB KC KD KE KF KG KH KI KJ KL KM KN KO KP KQ KR KS KT KU KV KW KX KY KZ LA LB LC LD LE LF LG LH LI LJ LK LL LM LN LO LP LQ LR LS LT LU LV LW LX LY LZ MA MB MC MD ME MF MG MH MI MJ MK ML MN MO MP MQ MR MS MT MU MV MW MX MY MZ NA NB NC ND NE NF NG NH NI NJ NK NL NO NP NQ NR NS NT NU NV NW NX NY NZ OA OB OC OD OE OF OG OH OI OJ OK OL OM ON OP OQ OR OS OT OU OV OW OX OY OZ PA PB PC PD PE PF PG PH PI PJ PK PL PM PN PO PP PQ PR PS PT PU PV PW PX PY PZ QA QB QC QD QE QF QG QH QI QJ QK QL QM QN QO QQ QR QS QT QU QV QW QX QY QZ RA RB RC RD RE RF RG RH RI RJ RK RL RM RN RO RP RQ RR RS RT RU RV RW RX RY RZ SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SR SS ST SU SV SW SX SY SZ TA TB TC TD TE TF TG TH TI TJ TK TL TM TN TO TP TQ TR TS TT TU TV TW TX TY TZ UA UB UC UD UE UF UG UH UI UJ UK UL UM UN UO UP UQ UR US UT UU UV UW UX UY UZ VA VB VC VD VE VF VG VH VI VJ VK VL VM VN VO VP VQ VR VS VT VU VW VX VY VZ WA WB WC WD WE WF WG WH WI WJ WK WL WM WN WO WP WQ WR WS WT WU WV WW WX WY WZ XA XB XC XD XE XF XG XH XI XJ XK XL XM XN XO XP XQ XR XS XT XU XV XW XX XY XZ YA YB YC YD YE YF YG YH YI YJ YK YL YM YN YO YP YQ YR YS YT YU YV YW YX YY YZ ZA ZB ZC ZD ZE ZF ZG ZH ZI ZJ ZK ZL ZM ZN ZO ZP ZQ ZR ZS ZT ZU ZV ZW ZX ZY ZZ																											id	title	property_type	price_per_sqm	city	neighborhood	lat	lon	size_sqm	rooms	year_built	max_floor	floor	heating	dist_to_r	dist_to_g	descrip	source	price_per_sqm	age	is_new	distance	text_for_embedding
1	Land ca land	16406	Bucuresti	Colentina	44.436	26.125	890							no		10	10	Proprietate	https://	1170		0	6.2	land in Colentina, Bucuresti, 690 mp, nan camere, construit in nan. Detalii: Proprietate in zona Colentina, suprafata 690 mp, preț 1041000 EUR.																										
2	House 2 house	155849	Bucuresti	Pipera	44.425	26.107	97	2	1982					yes	central		3	5	Proprietate	https://	807	43	0	14.2	house in Pipera, Bucuresti, 97 mp, 3.0 camere, construit in 1982.0. Detalii: Proprietate in zona Pipera, suprafata 97 mp, preț 155849 EUR.																									
3	Apartment apartment	115284	Bucuresti	Dorobani	44.43	26.193	78	1	1976	1	10	no	10	gas		3	6	Proprietate	https://	1478	43	0	10.2	apartment in Dorobani, Bucuresti, 78 mp, 1.0 camere, construit in 1976.0. Detalii: Proprietate in zona Dorobani, suprafata 78 mp, preț 115284 EUR.																										
4	House 3 house	94520	Bucuresti	Militari	44.401	26.112	68	3	2005					no	electricity		10	3	Proprietate	https://	1390	20	0	8.4	house in Militari, Bucuresti, 68 mp, 3.0 camere, construit in 2005.0. Detalii: Proprietate in zona Militari, suprafata 68 mp, preț 94520 EUR.																									
5	Apartment apartment	159572	Bucuresti	Militari	44.402	26.107	86	2	2021	9	10	no	10	gas		7	14	Proprietate	https://	1752	4	1	6.2	apartment in Militari, Bucuresti, 86 mp, 2.0 camere, construit in 2021.0. Detalii: Proprietate in zona Militari, suprafata 86 mp, preț 159572 EUR.																										
6	House 4 house	14080	Bucuresti	Dumfries	44.406	26.056	86	4	1984					no	termostic		7	4	Proprietate	https://	930	41	0	12.2	house in Dumfries, Bucuresti, 86 mp, 4.0 camere, construit in 1984.0. Detalii: Proprietate in zona Dumfries, suprafata 86 mp, preț 14080 EUR.																									
7	Apartment apartment	132304	Bucuresti	Pipera	44.426	26.024	102	4	2016	7	10	no	10	gas		10	5	Proprietate	https://	1302	7	1	3.6	apartment in Pipera, Bucuresti, 102 mp, 4.0 camere, construit in 2016.0. Detalii: Proprietate in zona Pipera, suprafata 102 mp, preț 132304 EUR.																										
8	House 3 house	137923	Bucuresti	Dorobani	44.427	26.117	107	3	1979					yes	termostic		19	13	Proprietate	https://	1329	46	0	5.8	house in Dorobani, Bucuresti, 107 mp, 3.0 camere, construit in 1979.0. Detalii: Proprietate in zona Dorobani, suprafata 107 mp, preț 137923 EUR.																									
9	House 4 house	141948	Bucuresti	Tran	44.419	26.195	119	4	2023					yes	gas		4	12	Proprietate	https://	1592	2	1	10.8	house in Tran, Bucuresti, 119 mp, 4.0 camere, construit in 2023.0. Detalii: Proprietate in zona Tran, suprafata 119 mp, preț 141948 EUR.																									
10	House 2 house	67275	Bucuresti	Dorobani	44.413	26.093	45	2	2000					no	termostic		15	12	Proprietate	https://	1455	25	0	4.2	house in Dorobani, Bucuresti, 45 mp, 2.0 camere, construit in 2000.0. Detalii: Proprietate in zona Dorobani, suprafata 45 mp, preț 67275 EUR.																									
11	House 2 house	65560	Bucuresti	Tran	44.413	26.091	72	3	1984					yes	central		4	7	Proprietate	https://	1590	41	0	12.8	house in Tran, Bucuresti, 72 mp, 3.0 camere, construit in 1984.0. Detalii: Proprietate in zona Tran, suprafata 72 mp, preț 65560 EUR.																									
12	House 2 house	106878	Bucuresti	Colentina	44.435	26.194	114	2	1999					no	termostic		10	7	Proprietate	https://	937	26	0	9.2	house in Colentina, Bucuresti, 114 mp, 2.0 camere, construit in 1999.0. Detalii: Proprietate in zona Colentina, suprafata 114 mp, preț 106878 EUR.																									
13	Apartment apartment	15572	Bucuresti	Dorobani	44.434	26.075	53	1	1994	0	10	yes	10	gas		11	10	Proprietate	https://	308	41	0	7.4	apartment in Dorobani, Bucuresti, 53 mp, 1.0 camere, construit in 1994.0. Detalii: Proprietate in zona Dorobani, suprafata 53 mp, preț 15572 EUR.																										
14	House 3 house	155851	Bucuresti	Tran	44.434	26.103	109	3	2003					yes	electricity		16	10	Proprietate	https://	1439	22	0	3.2	house in Tran, Bucuresti, 109 mp, 3.0 camere, construit in 2003.0. Detalii: Proprietate in zona Tran, suprafata 109 mp, preț 155851 EUR.																									
15	House 4 house	134880	Bucuresti	Colentina	44.434	26.196	60	4	2007					yes	electricity		7	6	Proprietate	https://	1748	18	0	4.4	house in Colentina, Bucuresti, 60 mp, 4.0 camere, construit in 2007.0. Detalii: Proprietate in zona Colentina, suprafata 60 mp, preț 134880 EUR.																									
16	House 2 house	39517	Bucuresti	Dorobani	44.444	26.151	43	2	2016					yes	gas		7	6	Proprietate	https://	979	10	1	11.4	house in Dorobani, Bucuresti, 43 mp, 2.0 camere, construit in 2016.0. Detalii: Proprietate in zona Dorobani, suprafata 43 mp, preț 39517 EUR.																									
17	House 3 house	60536	Bucuresti	Pipera	44.406	26.139	46	3	1995					no	termostic		19	10	Proprietate	https://	1316	30	0	14	house in Pipera, Bucuresti, 46 mp, 3.0 camere, construit in 1995.0. Detalii: Proprietate in zona Pipera, suprafata 46 mp, preț 60536 EUR.																									
18	House 3 house	155844	Bucuresti	Dumfries	44.44	26.052	113	3	1975					yes	gas		4	7	Proprietate	https://	1388	50	0	12.8	house in Dumfries, Bucuresti, 113 mp, 3.0 camere, construit in 1975.0. Detalii: Proprietate in zona Dumfries, suprafata 113 mp, preț 155844 EUR.																									
19	Apartment apartment	52372	Bucuresti	Berceni	44.403	26.072	48	4	1980	3	10	yes	10	gas		3	10	Proprietate	https://	1714	65	0	11	apartment in Berceni, Bucuresti, 48 mp, 4.0 camere, construit in 1980.0. Detalii: Proprietate in zona Berceni, suprafata 48 mp, preț 52372 EUR.																										
20	Apartment apartment	123882	Bucuresti	Dorobani	44.437	26.167	66	1	1979	6	10	no	10	gas		16	11	Proprietate	https://	8777	46	0	4	apartment in Dorobani, Bucuresti, 66 mp, 1.0 camere, construit in 1979.0. Detalii: Proprietate in zona Dorobani, suprafata 66 mp, preț 123882 EUR.																										
21	Land ca land	83452	Bucuresti	Tran	44.44	26.052	308							yes			6	6	Proprietate	https://	979			10.8	land in Tran, Bucuresti, 308 mp, nan camere, construit in nan. Detalii: Proprietate in zona Tran, suprafata 308 mp, preț 83452 EUR.																									
22	House 4 house	71004	Bucuresti	Pipera	44.443	26.056	61	4	1986					yes	termostic		19	7	Proprietate	https://	184	39	0	3.8	house in Pipera, Bucuresti, 61 mp, 4.0 camere, construit in 1986.0. Detalii: Proprietate in zona Pipera, suprafata 61 mp, preț 71004 EUR.																									
23	House 3 house	153458	Bucuresti	Militari	44.439	26.195	32	3	2016					no	termostic		5	8	Proprietate	https://	1668	9	1	11.8	house in Militari, Bucuresti, 32 mp, 3.0 camere, construit in 2016.0. Detalii: Proprietate in zona Militari, suprafata 32 mp, preț 153458 EUR.																									
24	Apartment apartment	68608	Bucuresti	Dorobani	44.429	26.173	52	2	1991	5	10	yes	10	gas		5	12	Proprietate	https://	1704	34	0	10.2	apartment in Dorobani, Bucuresti, 52 mp, 2.0 camere, construit in 1991.0. Detalii: Proprietate in zona Dorobani, suprafata 52 mp, preț 68608 EUR.																										
25	Apartment apartment	67182	Bucuresti	Dorobani	44.433	26.1	73	3	1991	7	10	no	10	gas		19	5	Proprietate	https://	1704	34	0	4.6	apartment in Dorobani, Bucuresti, 73 mp, 3.0 camere, construit in 1991.0. Detalii: Proprietate in zona Dorobani, suprafata 73 mp, preț 67182 EUR.																										
26	Apartment apartment	206451	Bucuresti	Tran	44.404	26.095	113	3	1973	3	10	no	10	gas		8	4	Proprietate	https://	1827	52	0	11.6	apartment in Tran, Bucuresti, 113 mp, 3.0 camere, construit in 1973.0. Detalii: Proprietate in zona Tran, suprafata 113 mp, preț 206451 EUR.																										
27	House 4 house	145238	Bucuresti	Dumfries	44.432	26.16	101	4	2003					no	termostic		6	16	Proprietate	https://	1438	22	0	12	house in Dumfries, Bucuresti, 101 mp, 4.0 camere, construit in 2003.0. Detalii: Proprietate in zona Dumfries, suprafata 101 mp, preț 145238 EUR.																									
28	Apartment apartment	31984	Bucuresti	Colentina	44.411	26.091	93	3	1982	3	10	no	10	gas		13	15	Proprietate	https://	988	63	0	4.2	apartment in Colentina, Bucuresti, 93 mp, 3.0 camere, construit in 1982.0. Detalii: Proprietate in zona Colentina, suprafata 93 mp, preț 31984 EUR.																										
29	Apartment apartment	193550	Bucuresti	Militari	44.431	26.117	98	1	1981	2	10	yes	10	gas		10	11	Proprietate	https://	1975	44	0	5.8	apartment in Militari, Bucuresti, 98 mp, 1.0 camere, construit in 1981.0. Detalii: Proprietate in zona Militari, suprafata 98 mp, preț 193550 EUR.																										
30	Apartment apartment	147686	Bucuresti	Dorobani	44.434	26.167	75	4	2017	3	10	no	10	gas		6	3	Proprietate	https://	1846	8	1	13.2	apartment in Dorobani, Bucuresti, 75 mp, 4.0 camere, construit in 2017.0. Detalii: Proprietate in zona Dorobani, suprafata 75 mp, preț 147686 EUR.																										
31	Apartment apartment	79305	Bucuresti	Tran	44.424	26.189	85	2	1982	2	10	no	10	gas		12	4	Proprietate	https://	933	43	0	3.2	apartment in Tran, Bucuresti, 85 mp, 2.0 camere, construit in 1982.0. Detalii: Proprietate in zona Tran, suprafata 85 mp, preț 79305 EUR.																										
32	House 4 house	202350	Bucuresti	Militari	44.421	26.191	110	4	1996					yes	gas		8	9	Proprietate	https://	1845	27	0	9.6	house in Militari, Bucuresti, 110 mp, 4.0 camere, construit in 1996.0. Detalii: Proprietate in zona Militari, suprafata 110 mp, preț 202350 EUR.																									
33	Land ca land	60570	Bucuresti	Dorobani	44.438	26.181	594							yes			11	10	Proprietate	https://	1073			7.4	land in Dorobani, Bucuresti, 594 mp, nan camere, construit in nan. Detalii: Proprietate in zona Dorobani, suprafata 594 mp, preț 60570 EUR.																									
34	Apartment apartment	39380	Bucuresti	Dumfries	44.414	26.056	41	1	1979	8	10	yes	10	gas		6	4	Proprietate	https://	960	46	0	12.8	apartment in Dumfries, Bucuresti, 41 mp, 1.0 camere, construit in 1979.0. Detalii: Proprietate in zona Dumfries, suprafata 41 mp, preț 39380 EUR.																										
35	Land ca land	66205	Bucuresti	Dorobani	44.424	26.182	445							yes			7	13	Proprietate	https://	1489			8.6	land in Dorobani, Bucuresti, 445 mp, nan camere, construit in nan. Detalii: Proprietate in zona Dorobani, suprafata 445 mp, preț 66205 EUR.																									
36	House 3 house	217269	Bucuresti	Dorobani	44.418	26.093	117	3	2003					yes	gas		19	8	Proprietate	https://	1897	22	0	4	house in Dorobani, Bucuresti, 117 mp, 3.0 camere, construit in 2003.0. Detalii: Proprietate in zona Dorobani, suprafata 117 mp, preț 217269 EUR.																									
37	Apartment apartment	149345	Bucuresti	Dumfries	44.411	26.151	115	3	1980	4	10	no	10	gas		18	12	Proprietate	https://	1303	45	0	2.4	apartment in Dumfries, Bucuresti, 115 mp, 3.0 camere, construit in 1980.0. Detalii: Proprietate in zona Dumfries, suprafata 115 mp, preț 149345 EUR.																										
38	Land ca land	47070	Bucuresti	Tran	44.415	26.071	410							yes			15	14	Proprietate	https://	1007			3.4	land in Tran, Bucuresti, 410 mp, nan camere, construit in nan. Detalii: Proprietate in zona Tran, suprafata 410 mp, preț 47070 EUR.																									
39	Apartment apartment	70163	Bucuresti	Pipera	44.411	26.067	71	2	2007	6	10	no	10	gas		6	3	Proprietate	https://	1073	18	0	12	apartment in Pipera, Bucuresti, 71 mp, 2.0 camere, construit in 2007.0. Detalii: Proprietate in zona Pipera, suprafata 71 mp, preț 70163 EUR.																										
40	House 2 house	150345	Bucuresti	Militari	44.417	26.169	117	2	1987					yes	termostic		6	12	Proprietate	https://	1285	38	0	3.6	house in Militari, Bucuresti, 117 mp, 2.0 camere, construit in 1987.0. Detalii: Proprietate in zona Militari, suprafata 117 mp, preț 150345 EUR.																									
41	Land ca land	507200	Bucuresti	Berceni	44.4	26.095	676							no			20	2	Proprietate	https://	124			3.8	land in Berceni, Bucuresti, 676 mp, nan camere, construit in nan. Detalii: Proprietate in zona Berceni, suprafata 676 mp, preț 507200 EUR.																									
42	Apartment apartment	19020	Bucuresti	Dorobani	44.401	26.059	90	3	2004	10	10	no	10	gas		20	2	Proprietate	https://	1868	21	0	5.4	apartment in Dorobani, Bucuresti, 90 mp, 3.0 camere, construit in 2004.0. Detalii: Proprietate in zona Dorobani, suprafata 90 mp, preț 19020 EUR.																										
43	House 3 house	171337	Bucuresti	Berceni	44.412	26.052	101	3	1995					no	termostic		16	8	Proprietate	https://	1837	30	0	5.2	house in Berceni, Bucuresti, 101 mp, 3.0 camere, construit in 1995.0. Detalii: Proprietate in zona Berceni, suprafata 101 mp, preț 171337 EUR.																									
44	Apartment apartment	148603	Bucuresti	Pipera	44.425	26.162	91	1	1977	4	10	yes	10	gas		19	8	Proprietate	https://	823	48	0	3.4	apartment in Pipera, Bucuresti, 91 mp, 1.0 camere, construit in 1977.0. Detalii: Proprietate in zona Pipera, suprafata 91 mp, preț 148603 EUR.																										
45	Land ca land	407400	Bucuresti	Militari	44.431	26.199	420																																											

- **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 cosinus ("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. metadatas: Setul de attribute structurate esențiale pentru filtrare (ex: property_type, city, neighborhood, price_eur, size_sqm).

Aceste metadata permit ca o interogare precum *"apartment, Titan, minim 50 mp"* să fie tradusă în: Filtru: property_type='apartment' AND neighborhood='Titan' AND size_sqm>=50 + Căutare Vectorială: pe textul "apartment 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: "titan", "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): Aproximarea 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 perceput 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), salvează 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 Suprafeței Țintă (target_sqm) → Dacă suprafața proprietății analizate (target_sqm) nu a fost specificată în cererea inițială a utilizatorului, funcția folosește ca valoare implicită **media suprafețelor 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)
UNDERPRICED	Preț Listat < 90% din Preț Corect Total
FAIR	Preț Listat se încadrează între 90% și 110% din Preț Corect Total
OVERPRICED	Preț Listat > 110% din Preț Corect Total

- Generarea Raportului de leșire (evaluate_property): Funcția principală orcheștrează 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. Acesta 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):** Acesta 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.
- **Instrucțiunea 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 pret 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 conclusion, 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 Țintă):** 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): Afișează 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 intermediare 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 displays the 'RAG Imobiliar — Estimare pret corect (demo)' application. On the left, a sidebar contains filters: 'Tip proprietate' set to 'apartment', 'Cartier (ex: Titan)' set to 'Titan', 'Suprafata minima (mp)' set to 40, and 'Buget maxim (EUR)' set to 120000. Below these is a slider for 'Numar comparabile (k)' set to 10. The main panel features a title, a search rule 'Rulez: apartment Titan 54 mp buget 60000 euro', and a table titled 'Comparabile (dupa reranking)'. The table lists 10 properties with columns for id, property_type, neighborhood, price_eur, size_sqm, price_per_sqm, and final_score.

	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 preț corect

Deploy

Verdict: UNDERPRICED

Explicație

Apartamentul cu două camere titan este subevaluat, având un preț listat de 60.000 €, în timp ce pretul 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. Pretul actual al apartamentului nu este corespunzător cu celelalte proprietăți similare din aceeași categorie, pentru care a fost clasificat ca fiind subevaluat (UNDERPRICED).

Proiect educațional; nu reprezintă consultanță imobiliară.

Hartă

