

SNAPHOMZ – (RESULTS SUMMARY)

Objective

Build a mini Retrieval-Augmented Generation (RAG) system over a real-estate dataset to answer user questions using semantic search and a local language model.

Delivered

- ⇒ Made a Normalized Dataset with [“Price”, “Beds”, “Baths”, “City”, “Clean_remarks”]
- ⇒ Made a new column [“Price_per_sqft”].
- ⇒ Added a Price Band Label in the code according to the price [“Expensive”, “Medium”, “Low”].
- ⇒ EDA Analysis by making Histograms and Understood the Relations with the Columns.
- ⇒ Implemented the Regression and Classification models.
- ⇒ Done with Embedding for “Clean_remarks”, “City” and “Price_per_sqft” and Faiss_database_vector for implementing the Rag Chatbot.
- ⇒ Integrated the Model and LLM model (From Google: flan-t5-small) for text Generation.
- ⇒ Implemented the RAG-MINI Chabot with Ngrok as I’m using colab for faster Compliance.

index	task	model	r2	mae	accuracy	precision	recall	f1
0	price_regression	LinearRegression	0.9105682085643567	223281.20397557094	NaN	NaN	NaN	NaN
1	price_regression	RandomForestRegressor	0.9714639474910284	92444.25374999997	NaN	NaN	NaN	NaN
2	price_per_sqft_regression	LinearRegression	1.0	3.090880900556436e-13	NaN	NaN	NaN	NaN
3	price_per_sqft_regression	RandomForestRegressor	0.9990582923212398	3.0017336162051924	NaN	NaN	NaN	NaN
4	sqft_regression	LinearRegression	0.4318399420305148	840.0038298223159	NaN	NaN	NaN	NaN
5	sqft_regression	RandomForestRegressor	0.7082494341104445	429.05875000000003	NaN	NaN	NaN	NaN
6	price_band_classification	RandomForestClassifier	NaN	NaN	0.90825	0.9023892773892773	0.9002849002849004	0.8996624290741938
7	expensive_binary	RandomForestClassifier	NaN	NaN	0.90825	0.9	0.925	0.9039038039039038

The screenshot shows a web application interface for a RAG system. The main title is "RAG Mini Edition – Snaphomz". Below the title, a sub-header says "Ask anything from your dataset!". There is a text input field containing the query "Backyards near Dallas" and a "Search" button. Below the search results, there is a section titled "Matched Rows" which displays a table with three rows of data. The table has columns for id, address, city, state, price, beds, baths, sqft, and remarks. The first row shows an address in Dallas, TX with solar panels. The second row shows an address in Dallas, TX with quiet cul-de-sac. The third row shows an address in Dallas, TX with ev charger. At the bottom of the page, there is a button labeled "Answer" with a checkmark icon.