

FARHAN PAGARKAR

Data Analyst | Python | Machine Learning | AI/LLM | Predictive Modeling

+91 9321409230 | pagarkarfarhan@gmail.com | LinkedIn: [Farhan Pagarkar](#) | Thane, Maharashtra

PROFESSIONAL SUMMARY

Data Analyst with 1+ year of production experience using Python, SQL, statistics, machine learning, and AI/LLM workflows to solve analytical and product problems. Built predictive models, production ML APIs, LLM content pipelines, and large-scale data workflows across 113M+ MLS records.

CORE SKILLS

Programming: Python, SQL, Pandas, NumPy

Machine Learning: Scikit-learn, XGBoost, CatBoost, KMeans, KNN, Naive Bayes, Predictive Modeling, Feature Engineering

AI / NLP: OpenAI API, Gemini API, LLMs, Generative AI, NLP

Data: Large-Scale Data Processing, Data Cleaning, Data Quality, ETL, Elasticsearch, Entity Resolution

Frameworks / Tools: FastAPI, REST APIs, Git, Jupyter, Google Colab, Excel, Power BI, Tableau

PROFESSIONAL EXPERIENCE

Data Analyst | Houzeo (PropTech)

May 2025 - Present | Thane, Maharashtra

- Owned production Elasticsearch datasets spanning 113M+ MLS property records (~4TB) and built Python workflows for data processing, quality, and high-volume updates.
- Built Python/Elasticsearch pipelines using bulk operations and scan to synchronize property data, updating ~598K documents per cycle.
- Developed quantitative listing-attribution logic using geospatial, polygon, price-distribution, and textual signals to classify unassigned listings.
- Led and mentored a team of 2 interns, reviewing analytical work and coordinating delivery across ongoing data initiatives.
- Led requirements and project execution for two live Buildings and Community features, translating analytical needs into product outcomes.
- Built data-quality remediation strategies and performed funnel analysis across Android and iOS products to support retention decisions.

Data Analyst Intern | Excellerate

Feb 2023 - Apr 2023 | Remote

- Led a cross-functional team of 5+ members across multiple time zones, delivering analytics projects on schedule.
- Received Star Performer recognition and a \$1,000 scholarship from Saint Louis University for exceptional contribution.

PROJECTS

AI Property Description & Summary Generation | Python | OpenAI | Gemini | Elasticsearch | Pydantic

- Built an LLM-powered content pipeline that aggregates building and listing data from Elasticsearch and generates structured building descriptions and buyer-oriented AI summaries.
- Implemented configurable OpenAI/Gemini API integration with token tracking, JSON parsing, logging, error handling, and concurrent batch generation for up to 5 buildings at a time.

Property Price Prediction Model | Python | XGBoost | CatBoost | Scikit-learn | FastAPI

- Built a property price prediction model across 11 Miami, Florida cities, achieving R2 of 0.8882 and MAE of approximately \$23K.
- Engineered spatial and statistical features including KMeans clustering, KNN average price, and Naive Bayes ZIP-code probability.

MLS Data Pipeline & Entity Resolution | Python | Elasticsearch | Painless | elasticsearch-py

- Built scalable workflows to extract, clean, and match MLS listings to residential communities across the US using geospatial and textual signals.

EDUCATION

B.Sc. in Data Science & Artificial Intelligence | R.J. College, Mumbai University

2023 - 2026 | Mumbai, Maharashtra

Relevant Coursework: Machine Learning, Natural Language Processing, Statistics, Big Data

ACHIEVEMENTS & CERTIFICATIONS

- Star Performer Award - Excellerate (2023): Awarded a \$1,000 scholarship from Saint Louis University for outstanding performance during internship.
- Data Science & AI Certification - IDEMI, Government of India (MSME Ministry).