**Link to GitHub:** [**https://github.com/denocean/student-projects-repo**](https://github.com/denocean/student-projects-repo)

**Project Pitch: U.S. Housing Market**

**Executive Summary**

The U.S. housing market faces unprecedented challenges, including affordability crises, rising mortgage rates, and persistent inventory shortages. Policymakers, investors, and buyers often struggle to interpret fragmented datasets and dense reports, limiting effective decision-making.

This project proposes an AI-powered (“Chat GPT”) platform that integrates machine learning (ML) forecasting with large language model (LLM) summarization to deliver accessible, actionable housing market insights. By aggregating trusted datasets from Zillow, Freddie Mac, the U.S. Census Bureau, and FRED, the system will generate 1–5-year forecasts for home prices, affordability, and regional demand. These outputs will be summarized into clear, user-friendly narratives, supported by an interactive dashboard and conversational AI for natural language queries.

Expected outcomes include faster analysis, improved clarity, and broader accessibility of housing data. This solution bridges the gap between complex market forecasts and practical, stakeholder-ready insights.

**Problems in the U.S. Housing Market**

**Affordability Crisis**

* Home prices have outpaced income growth for over a decade.
* Mortgage rates nearly doubled from 2021 to 2023, driving monthly payments beyond reach for many families.
* Rent burdens continue rising, limiting pathways to homeownership.

**Inventory Shortages**

* The U.S. faces a deficit of more than 3.5 million homes (NAR estimate).
* High construction costs, zoning restrictions, and labor shortages exacerbate supply issues.

**Regional Disparities**

* Midwest markets remain stable, while West Coast and Sunbelt markets face extreme volatility.
* Stakeholders struggle to interpret these regional differences without clear tools.

**Data Fragmentation & Complexity**

* Housing data is spread across multiple sources (Zillow, Freddie Mac, FRED, Census).
* Reports are often technical and inaccessible to non-specialists.

**Market Volatility & Uncertainty**

* Inflation, interest rate hikes, and economic shocks create unpredictability.
* Forecasting tools provide numbers but rarely offer clear narratives for decision-makers.

**Stakeholders**

* Homebuyers & Renters – Seek affordability guidance and clarity on timing/location decisions.
* Investors & Developers – Require risk assessments and demand forecasts to guide acquisitions and construction.
* Policymakers & Government Agencies – Need accessible insights to inform zoning, housing policy, and affordability programs.
* Financial Institutions – Use forecasts to assess credit risks, lending portfolios, and rate strategies.
* Researchers & Analysts – Rely on accurate datasets for housing policy evaluation and modeling.
* Technology & Data Firms – Both competitors and collaborators (e.g., Zillow, Redfin) who may benefit from LLM-driven insights.

## **Scope**

* Data integration from public sources (Zillow, Freddie Mac, FRED, Census).
* ML models to forecast housing prices, affordability indices, and demand (1–5 years).
* LLM-driven summarization and conversational AI for natural language queries.
* Interactive dashboards with visualizations and query support.
* Risk and ethics considerations, including mitigation of bias and accessibility.

**Out-of-Scope**

* Proprietary data creation.
* Real estate transaction facilitation.
* Financial or legal advisory services.
* Long-term forecasts beyond 10 years.
* Guaranteed housing market outcomes.

**Success Metrics**

* **Forecast Accuracy** – ML models achieve ≤5% error when validated against historical housing data.
* **Summarization Quality** – LLM summaries rated ≥80% accurate and clear by housing experts.
* **User Adoption** – At least 75% of pilot users (buyers, investors, policymakers) find the platform usable and effective.
* **Decision-Making Impact** – Demonstrated influence on investment strategies, affordability planning, and purchase confidence.
* **Efficiency Gains** – 50% reduction in analysis time compared to static reports; query response times ≤3 seconds.
* yet—focus on proving the concept works.

**Minimal Artifact**

The minimal viable prototype will include:

* **Core Data Integration** – ETL pipeline aggregating Zillow, Freddie Mac, FRED, and Census data.
* **Basic Forecasting Module** – 1-year housing price predictions with visualization.
* **Summarization Output** – LLM-generated 1–2 paragraph summaries highlighting affordability and price trends.
* **Interactive Dashboard** – Simple charts/graphs with region selection and trend comparisons.

**Iterative Design Approach**

The platform will follow an **iterative, user-centered design** methodology:

* **Prototype Development** – Build MVP with forecasting, summarization, and minimal dashboard.
* **Stakeholder Testing** – Collect feedback from buyers, policymakers, and investors.
* **Incremental Refinement** – Enhance dashboard, expand forecasting horizons, and improve summarization.
* **Scalable Deployment** – Integrate conversational AI and advanced analytics for broader use.

This approach ensures risks are identified early, while feedback drives feature evolution.

**Evidence Base**

1. **Improved Predictive Performance**
   * A University of Florida study found ML significantly outperforms linear regression in real estate forecasting, reducing error by up to **68%** over intermediate and long-term horizons.
   * Source: [University of Florida – Machine Learning & Real Estate](https://warrington.ufl.edu/due-diligence/2025/03/19/machine-learning-big-data-predict-real-estate-returns/?utm_source=chatgpt.com)
2. **LLMs Enhance Accessibility**

* A 2025 study showed LLMs can democratize real estate appraisal, generating competitive and interpretable price estimates via in-context learning.
* This demonstrates LLMs’ capacity to **simplify technical housing data into accessible insights**.
* Source: [Arxiv – LLMs for Real Estate Appraisal](https://arxiv.org/html/2506.11812v1?utm_source=chatgpt.com)
* These findings confirm that combining ML and LLMs can **increase forecasting accuracy and improve usability**, directly addressing current housing market challenges.

**System Sketch**

A diagram of a data flow

AI-generated content may be incorrect.

**Risk Register:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Probability** | **Impact** | **Mitigation** |
| Inaccurate forecasts | Medium | High | Validate with historical data, retrain models |
| Data integration errors | Low | Medium | Automated ETL checks |
| LLM hallucination | Medium | High | Human review, prompt engineering |
| Dashboard performance | Low | Medium | Optimize queries and caching |
| Data privacy issues | Low | High | Use public data, anonymize sensitive info |
| Misinterpretation by stakeholders | Medium | Medium | Provide clear documentation |

**Conclusion**

The U.S. housing market suffers from affordability crises, inventory shortages, and fragmented data. Existing tools provide technical numbers but fail to communicate insights effectively to stakeholders.

This project’s AI-powered (“Chat GPT”) platform combines **ML-driven forecasts** with **LLM-driven narratives**, delivering a transparent, user-friendly, and actionable system. With a robust evidence base, clear success metrics, and an iterative development plan, the solution promises to improve decision-making for buyers, investors, policymakers, and financial institutions alike.

By making complex housing market data **accessible, accurate, and actionable,** this project can help stakeholders navigate uncertainty and plan effectively for the future.