Project Proposal ↗

Prepared by :

AYA ALI

Prepared for :

DEPI

Table of Contents

Project Description

"

This project aims to predict housing price inflation (a CPI-like indicator for the real estate market) using the HouseTS dataset.

By analyzing historical housing data — including sale prices, listing prices, demographics, and service availability — the project builds a predictive model to forecast future inflation trends in the housing sector.

The results can help policymakers, investors, and real estate agencies understand market dynamics and anticipate housing cost increases.

Objectives

01

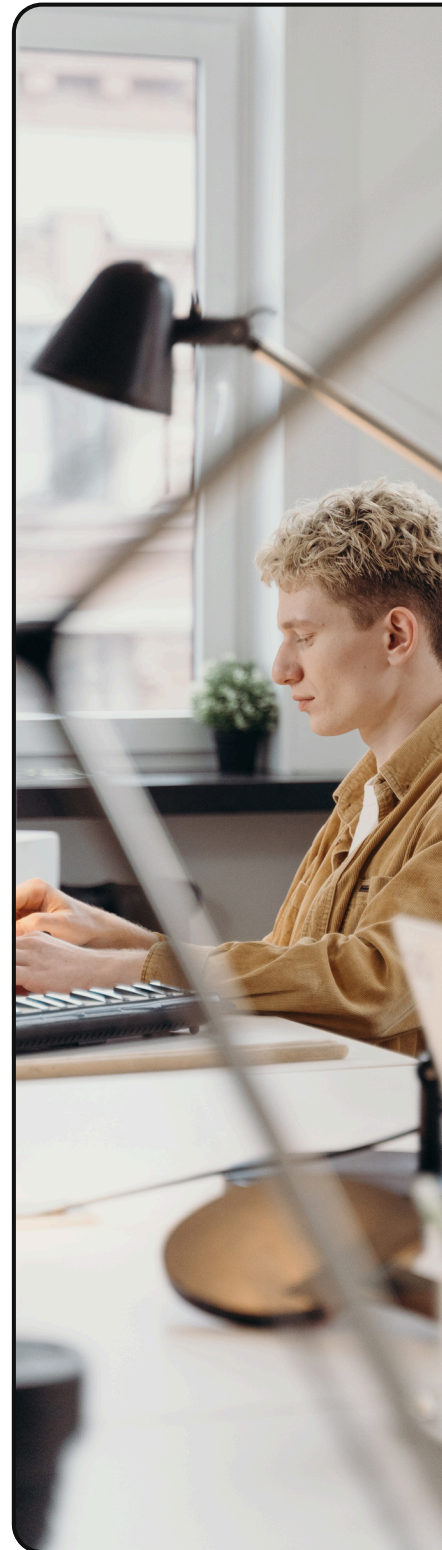Analyze housing market data to identify key price trends and patterns.

02

Build a predictive model for forecasting housing inflation (CPI-like).

03

Evaluate the model using accuracy, latency, and error metrics.

04

Deploy the model as an API to provide real-time housing inflation forecasts.

Dataset Description

- Name: HouseTS Dataset (Kaggle, 2024)

- Coverage: 6,000 ZIP codes across 30 major U.S. cities

- Period: March 2012 – December 2023 (monthly data)

Key Features:

- median_sale_price (target variable – House CPI)

- median_list_price, median_ppsf (price metrics)

- inventory, new_listings, homes_sold (market indicators)

- bank, school, hospital, park, restaurant, mall (location services)

Group Members & Roles

Name	Role	Responsibilities
Aya Ali	Project Leader (ML)	• Review and validate all team submissions. • Build and train the Housing Inflation Prediction Model. • Evaluate and improve model performance.
Alaa orabi	Data Cleaning	• Collect and combine raw datasets from multiple sources. • Handle missing, duplicated, and inconsistent data. • Prepare a clean dataset ready for analysis and modeling. • Prepare and deliver the final presentation and project submission.
Muhamad Assem	Feature Engineering	• Extract key features from raw data. • Create additional derived variables • Prepare and deliver the final presentation and project submission.
Malek Rafat	Data Analyst (EDA)	• Perform exploratory data analysis (EDA) to uncover insights.
Pola Saeed	Dashboard Developer	• Develop an interactive dashboard using Streamlit, or Power BI. • Visualize key metrics, model results, and housing trends. • Design a user-friendly and visually appealing interface. • Connect the dashboard to the model outputs for real-time display.
Mohamed mahmoud	API	• Develop a simple API to serve the trained model predictions. • Build an endpoint for sending input data and returning forecast results. • Integrate the API with the dashboard for real-time prediction display.

Methodology

01

Programming Language

- Python

02

Libraries

- Pandas, NumPy – for data manipulation and preprocessing
- Scikit-learn – for machine learning modeling and evaluation
- Matplotlib, Plotly – for data visualization and exploratory analysis

03

Modeling Techniques:

- Linear Regression, Random Forest Regressor, Gradient Boosting (XGBoost / LightGBM), Support Vector Regressor (SVR)

04

Dashboard & Visualization:

- Power BI or Tableau (for interactive dashboards)
- Jupyter Notebook (for analysis and documentation)

05

Version Control:

- GitHub

Timeline

Milestone	Description
Data Understanding & Cleaning	Explore dataset, handle missing values
Exploratory Data Analysis (EDA) feature engineering	Visualize housing trends feature engineering
Model Development Evaluation	Train ML models for prediction Evaluate accuracy, error rate, and latency
Deployment	API
Final Report & Presentation	Submit project report with results

Key Performance Indicators (KPIs)

1. Data Quality

Metric	Target
Percentage of missing values handled	100%
Data accuracy after preprocessing	>95%
Dataset diversity (representation of cities and zip codes)	>90%

2. Model Performance

Metric	Target
Model accuracy (R^2 equivalent)	>70%
Model prediction speed (Latency)	<200 ms
Error rate	<30%

3. Deployment & Scalability

Metric	Target
API uptime	>99%
Response time per request	<300 ms
Real-time processing speed (if streaming updates)	~10 records/sec

4. Business Impact & Practical Use

Metric	Target
Reduction in manual analysis effort	>70%
Expected cost savings	~50% (via automation and forecasting)
User satisfaction	>90% (based on feedback survey)

Let's Work Together ↗

Contact Us

📞 +123-456-7890

✉ hello@reallygreatsite.com

🌐 www.reallygreatsite.com