REAL ESTATE PRİCE PREDİCTİON SYSTEM

Domain: Real Estate Analytics / PropTech

Powered by Omnia AI Lab

Objective

To build a predictive model that accurately estimates real estate prices, helping stakeholders optimize property investments and pricing strategies.

Business Problem

Real estate companies struggle to maintain pricing accuracy due to market fluctuations and reliance on subjective manual valuation. This leads to over/under-pricing, revenue loss, and slow decision-making.

A data-driven predictive system ensures accurate, consistent, and efficient property valuation.

Data Collection & Preparation

Key Questions

- What internal and external data sources are available (listings, transactions, locations, amenities)?
- How should missing, duplicate, or inconsistent records be handled?
- Are there seasonal or macroeconomic factors influencing prices?
- Which features most impact property value (e.g., size, location, condition)?

Tasks

- Collect historical property listings & transactions
- Enrich with location and infrastructure data (POIs, schools, transit)
- Handle missing values, outliers, and currency normalization
- Engineer features:
  - price_per_m2
  - building_age
  - distance_to_city_center
  - neighborhood_demand_index

Data Collection & Preparation

Deliverables

- Cleaned dataset + data schema
- Feature engineering notebook
- Data documentation summary

Exploratory Data Analysis (EDA)

Objective:

Understand key price drivers and spatial/temporal patterns.

Key Questions

- What is the price distribution across locations?
- Which factors correlate most with price?
- Are there anomalies or market outliers?
- How do prices vary by year or season?

Tasks

- Visualize correlations and distributions (scatter, heatmaps, boxplots)
- Identify top predictive variables
- Detect outliers and market anomalies
- Summarize business-relevant insights

Exploratory Data Analysis (EDA)

Deliverables

- EDA report (plots + insights)
- Interactive Power BI dashboard

Model & Evaluation

Objective:

Build and evaluate predictive models for property pricing.

Key Questions

- Which regression algorithm best fits (linear, XGBoost, or RandomForest)?
- How do we measure accuracy and interpretability?
- Is the model robust across regions and property types?
- How to handle overfitting or multicollinearity?

Tasks

- Train baseline (Linear Regression) → optimized models (XGBoost)
- Perform cross-validation and hyperparameter tuning
- Evaluate with:
  - RMSE, MAE, R^2
  - Residual analysis
- Use SHAP/feature importance to interpret drivers

Model & Evaluation

Deliverables

- Trained model + evaluation report
- Model explainability visuals (feature impact plots)

Deployment & Integration

Objective:

Operationalize the model for real-time or batch price predictions.

Key Questions

- How will agents or clients access the predictions?
- Should deployment be via API, dashboard, or integrated CRM?
- How will new listings be automatically scored?
- How to maintain data freshness?

Tasks

- Wrap the model into FastAPI endpoint /predict_price
- Containerize with Docker for scalability
- Build a Streamlit / Power BI dashboard for visualization
- Schedule auto-retraining with Cron / AWS Lambda

Deployment & Integration

Deliverables

- FastAPI model endpoint
- Dashboard UI for property price visualization
- Deployment documentation

Insights for Business

- What drives prices (location, age, demand)
- Key neighborhoods
- Price patterns over time

Monitoring & Continuous Improvement

Objective: Ensure the model stays accurate as market conditions evolve.

Key Questions

- How often should the model be retrained with new market data?
- What metrics indicate model drift or market shifts?
- How can analysts flag inaccurate predictions for feedback?
- How to log and analyze prediction errors over time?

Tasks

- Track prediction accuracy monthly
- Set up model drift dashboards (Prometheus/Grafana)
- Create feedback loop for retraining with new listings
- Monitor data pipeline health and input consistency

Monitoring & Continuous Improvement

Deliverables

- Model performance tracking dashboard
- Retraining schedule & automation script
- Governance report on accuracy trends

Deliverables

Phase	Deliverable
Business Understanding	Problem statement + valuation KPI document
Data Preparation	Clean dataset + metadata + feature schema
EDA	Visual analysis report (location, demand, seasonality)
Modeling	Regression model notebook + trained XGBoost model
Explainability	SHAP plots + feature importance report
Deployment	FastAPI /predict_price endpoint + Docker image
Dashboard	Streamlit/Power BI valuation dashboard
Monitoring	Monthly drift report + retraining workflow

Thank You!