CSE6242 – Intelligent Real Estate Advisor – Team 114

1. Motivation / Introduction

1.1 What is the problem:

- Real estate markets are volatile and influenced by economic, demographic, and spatial factors.
- Investors require better tools to forecast trends, reduce risks, and maximize investment returns.
- We offer open-access tools for non-institutional investors to forecast U.S real estate trends.

1.2 Why is it important:

- Democratize access to real estate insights (private vs. institutional investors).
- Combines advanced analytics and intuitive dashboards for transparency and usability.

2. Approach

2.1 What was our approach, and how it works:

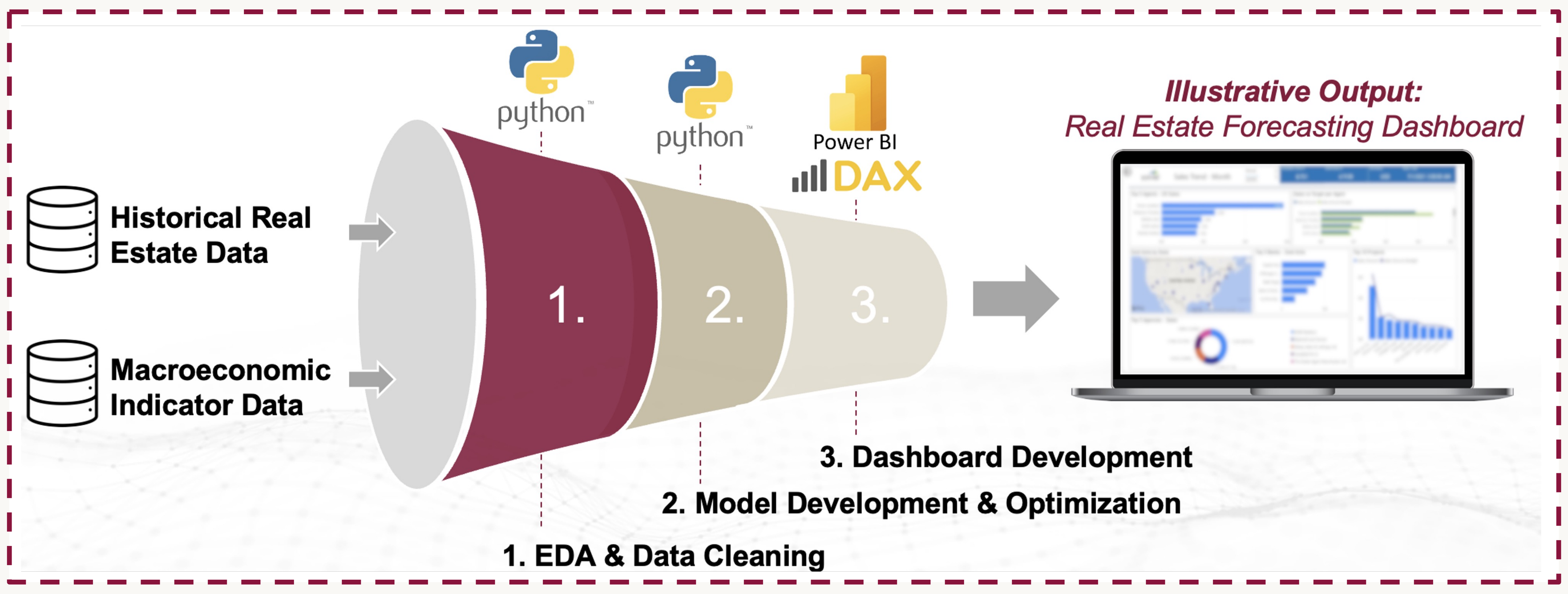
- Integrated historical supply, demand and price data, and macroeconomic.
- Engineered features (ratios, lagged values, moving averages) for improved modeling.
- Trained 1900+ county-specific time-series models (Ridge, XGBoost, Random Forest and Prophet).
- Visualized forecasted YoY price change and key indicators integrating public county level json data with PowerBI's mapping functionality.
- Built an interactive user-friendly dashboard for investment exploration.

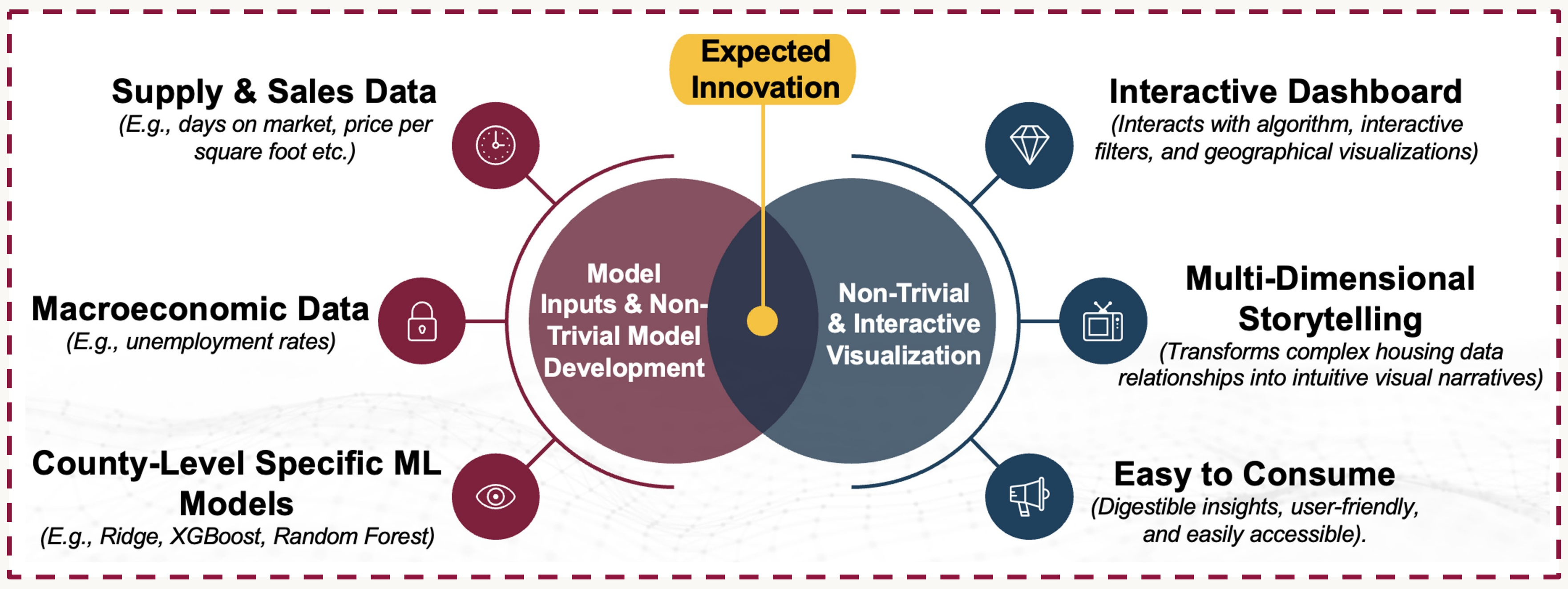
2.2 Why does it solve the problem:

- Open access for time-aware forecasts.
- Granular predictions at county-level support local decision-making.
- Intuitive and interactive dashboard supports non-technical users to easily consume insights.
- Ensemble models and socio-economic data capture complex housing dynamics.

2.3 What is new:

- Merged public housing and macroeconomic data at scale.
- Blind tested county-level specific ML models.
- Real-time interactive dashboard with drill-down, comparison, and prediction features.
- Bridged the gap between expert analysis and public tools.





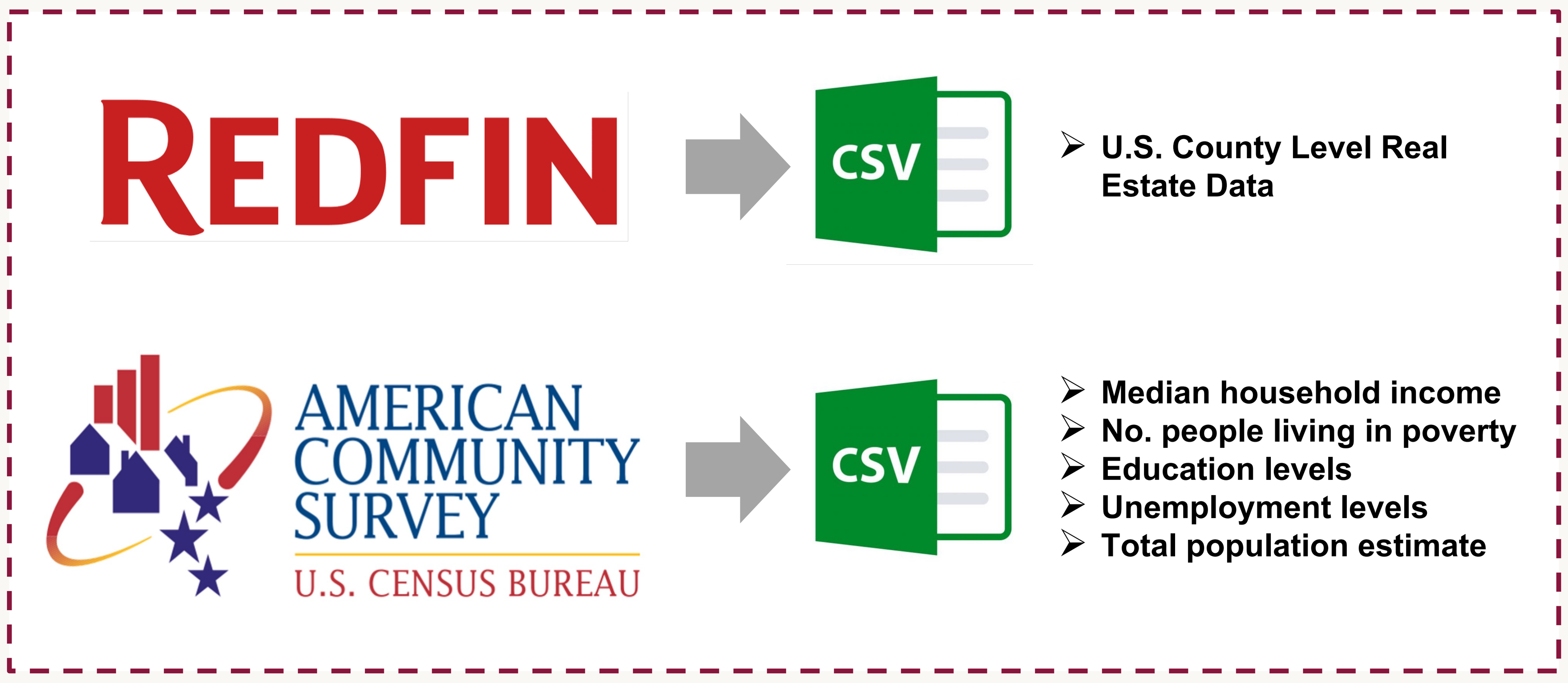
3. Data

3.1 How we collected data:

- Data sourced from Redfin, A real estate platform time-series data on the U.S. real estate market.
- Data sourced from the American Community Survey, run by the U.S. Census Bureau, a government survey offering annual socioeconomic data.

3.2 What are the data characteristics:

- 60-month time series across 1900+ U.S. counties.
- Target variable is YoY price change.
- Engineered 10+ features (price change %, inventory, unemployment, etc.).
- Handled missing data using interpolation and filtering for data quality.



4. Experiments / Results

4.1 How was the project evaluated:

- Time-based split: 4 years for training, 1 year for testing.
- Evaluated models using MAE, RMSE, and R².
- Cross-validation used for hyperparameter tuning.
- Focused on generalization to future (unseen) time points.

4.2 What are the results:

- 12-month price prediction models: $R^2 = 0.9$, RMSE = 81,382.11
- YoY price change harder to predict.
- Short-term (1-month) predictions: $R^2 = 0.80$, RMSE = 13%
- Long-term (12-month) predictions: $R^2 = 0.40$, RMSE = 22%
- Numerous UAT sessions conducted to test and refine dashboard features and useability.

4.3 How does our methods compare to other methods:

- Traditional models lack regional detail and macroeconomic context.
- Our hybrid approach improves accuracy.
- More transparent and interpretable than black-box neural networks.
- Dashboard interactivity improves investment exploration and dashboard useability over statis reports.