Which neighborhood fits you best?

Project 4
Real Estate Pricing Forecasting

Group 2: Anh Pham, Kathy Nguyen, Miguel Soriano, Robert Morris

Problem

Why finding the right home is challenging

- Complex influential factors
- Identifying affordable properties within homebuyers' criteria
- Understanding market trends

Solution

Smarter way to buy: Data-driven decisions

- Using AI to predict real estate home prices
- Categorize neighborhoods based on user defined preferences
- Develop interactive community dashboards
- Maintaining 75% accuracy within predictions

Objective and Methodology

Objective

- Forecast home prices based on historical data and external factors
- Provide a user-friendly tool for personalized recommendations
- Ensure accuracy and user satisfaction

Methodology

- **Data cleaning** – Python Panda
- **Model Training** – Scikit-learn
- **Visualization** – Python Matplotlib, Tableau
- **User Interface** – HTML/CSS, JavaScript, MongoDB Database

Data Sources and Cleaning

Data sources: from Kaggle

- **Primary Dataset:**
  - Housing Price Dataset (more than 21.613 entries)
- **Supplementary Datasets:**
  - Crime Rates by City and State
  - Average income by Zip Code
  - US City Zip Code Data

Data cleaning:

- **Data Cleaning and Preparation**: Remove inconsistencies, convert data to correct data types
- **Merge Datasets:** Merge some supplementary dataset to get strong features collerating to price
- **Feature Engineerin**g: Create new factors, detect outliers

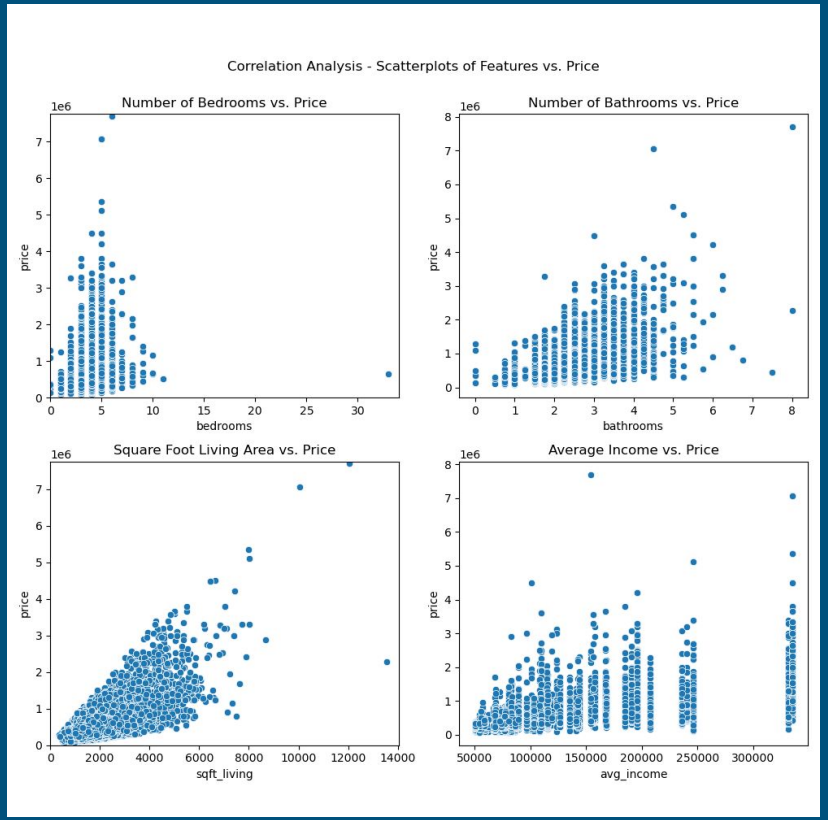
Model Building and Analysis

Model building:

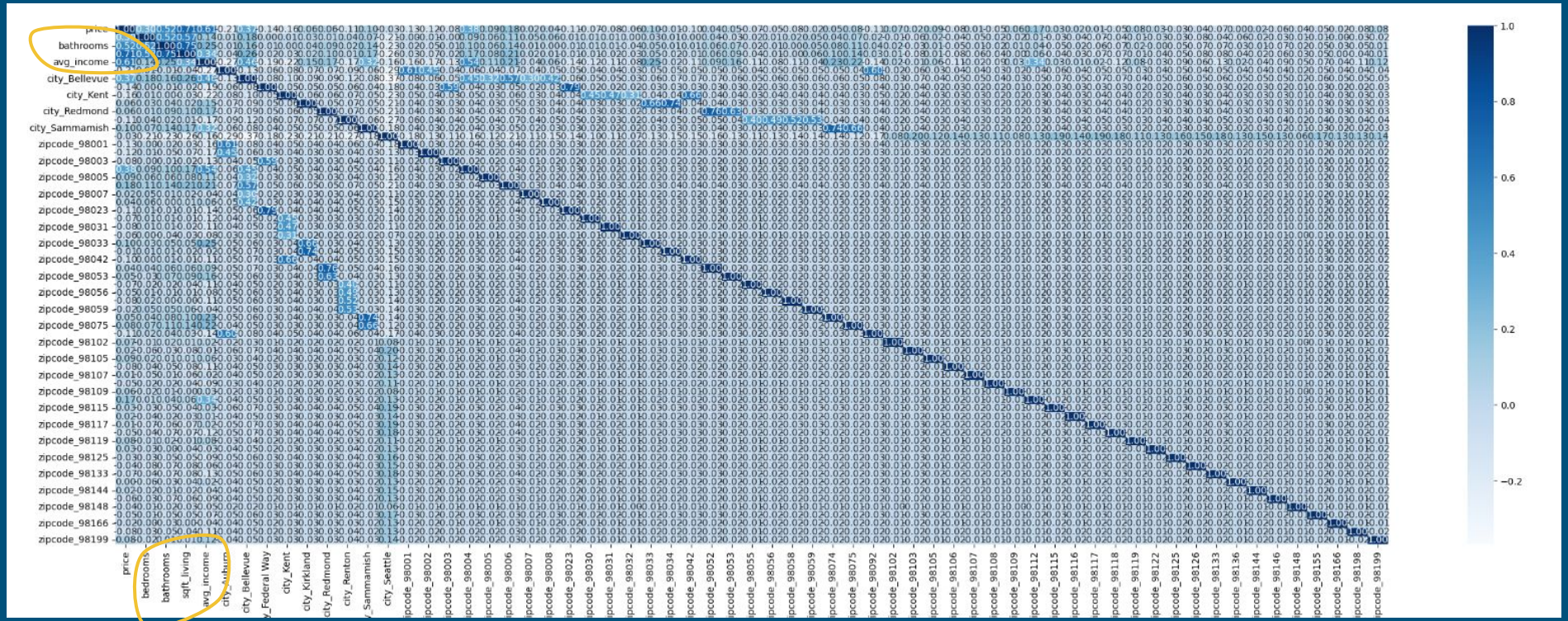
- **First test:** with 18 features -> Identify strong/weak features
- **Second test**: with 4 features
- **Select the best model**: Use GridSearchCV

Model	best_score
linear_regression	0.744857
lasso	0.744803
decision_tree	0.671300

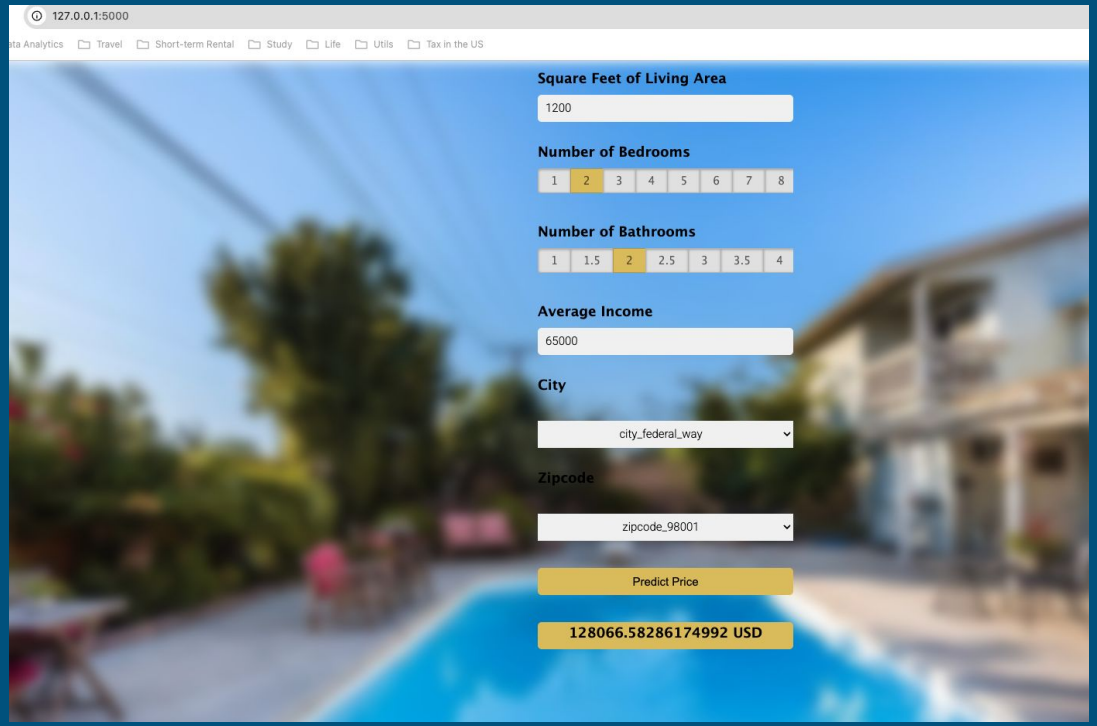
⇒ Model for prediction**: Linear Regression**



Correlation Analysis



Dashboard Demo



Kathy

~75 %

Model Performance

- Achieved ~75% accuracy using Linear Regression
- Tested 18 features, narrowed down to top 4 via correlation analysis
- Used GridSearchCV to confirm best model: Linear Regression

Conclusion

💡 Key Takeaways

- Data-driven models simplify complex home-buying decisions
- Neighborhood insights based on user preferences and market data
- Transparent, accurate predictions build user trust

🔍 Final Thought

Smarter real estate choices start with the right data.

Questions?