BANGALORE HOUSE PRICE PREDICTION

Project Overview

Objection: Predict house price in Bangalore using machine learning

Dataset: bengaluru_house_prices.csv

Rows: 13,320

Feratures: Area type, location, size, total_sqft, balcony, price

Data Cleaning

- Dropped columns: area_type, society, balcony, availability
.
- Removed rows with missing values
- Converted size to numerical bhk
- Handled total_sqft ranges by averaging values

Feature Engineering

- Created price_per_sqft = price × 100000 / total_sqft
- Reduced location dimensionality: grouped rare locations as "other"
- Removed outliers:
- Properties with total_sqft/bhk < 300
- Extreme price_per_sqft values per location
- 2 BHK priced higher than 3 BHK in same location

Visual Analysis

- Scatter plots for 2 BHK vs 3 BHK in Rajaji Nagar & Hebbal
- Histograms for price_per_sqft and bath
- Identified anomalies (e.g., 8 BHK with 600 sqft

Final Dataset

- Final shape: 7,291 rows × 7 columns
- Columns: location, total_sqft, bath, bhk, price
- One-hot encoded location column
- Dropped size and price_per_sqft

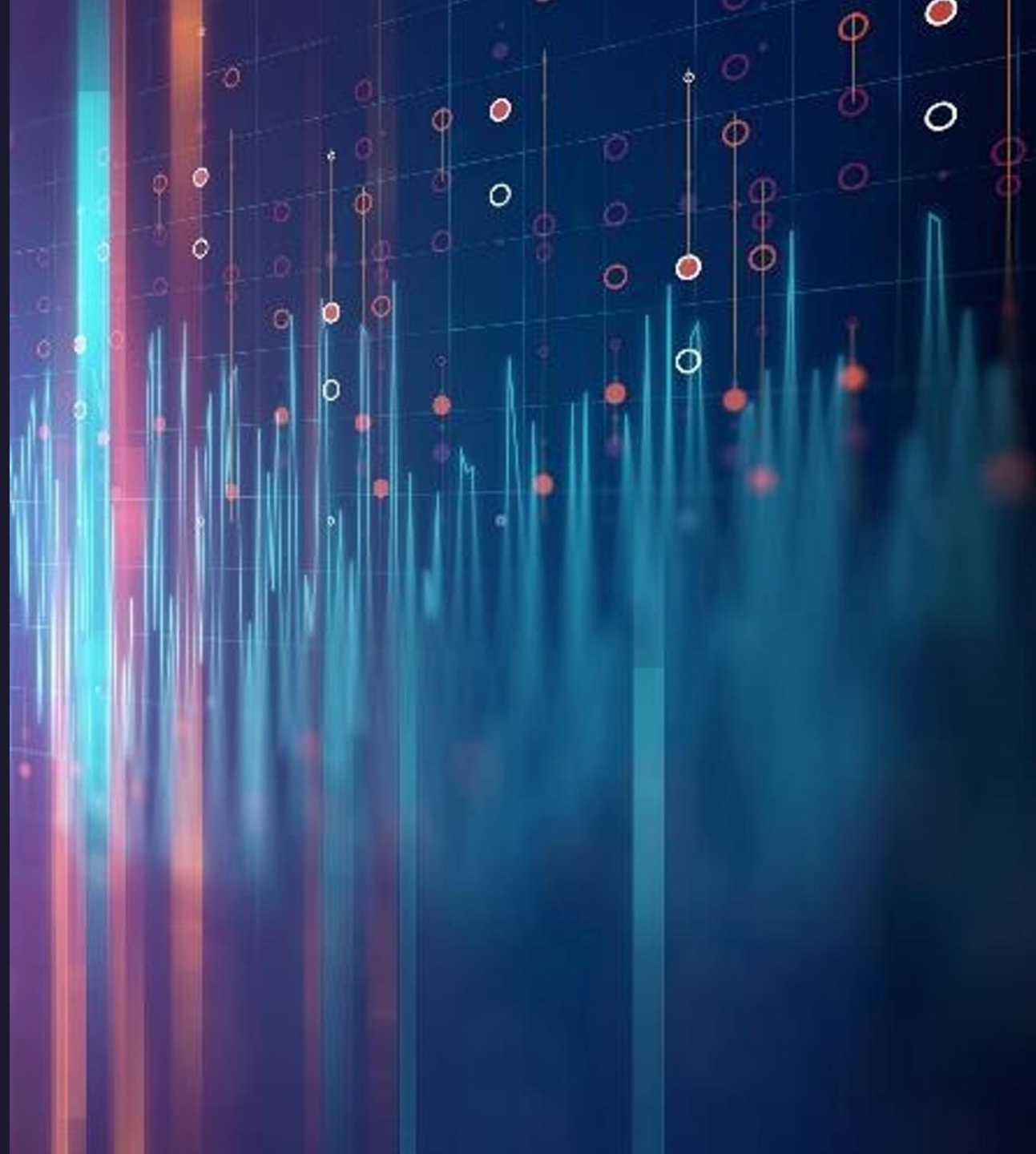
Model Building

- Features: X = all columns except price
- Target: y = price

- Train-test split: 80/20
- Model: Linear Regression
- Accuracy: **85.57%**

Key Takeaways

- Importance of thorough data cleaning
- Feature engineering boosts model performance
- Outlier removal improves prediction accuracy
- Dimensionality reduction simplifies modeling



Conclusion

- Successfully built a predictive model for Bangalore house prices
- Future work: Try other models (e.g., Random Forest, XGBoost)
- Consider integrating real-time data for dynamic pricing

Thank you