

Who Am I & Who Is the Client?

- Role: Data Science Undergraduate Student
- Client: Erin Robinson
- Socially responsible real-estate investor
- Goal: Improve housing quality without luxury-driven gentrification

Dataset Description

- Dataset: King County Housing Data
- Location: Seattle & surrounding King County areas
- Size: ~21,000 house sales
- Time span: Multiple years of recorded transactions

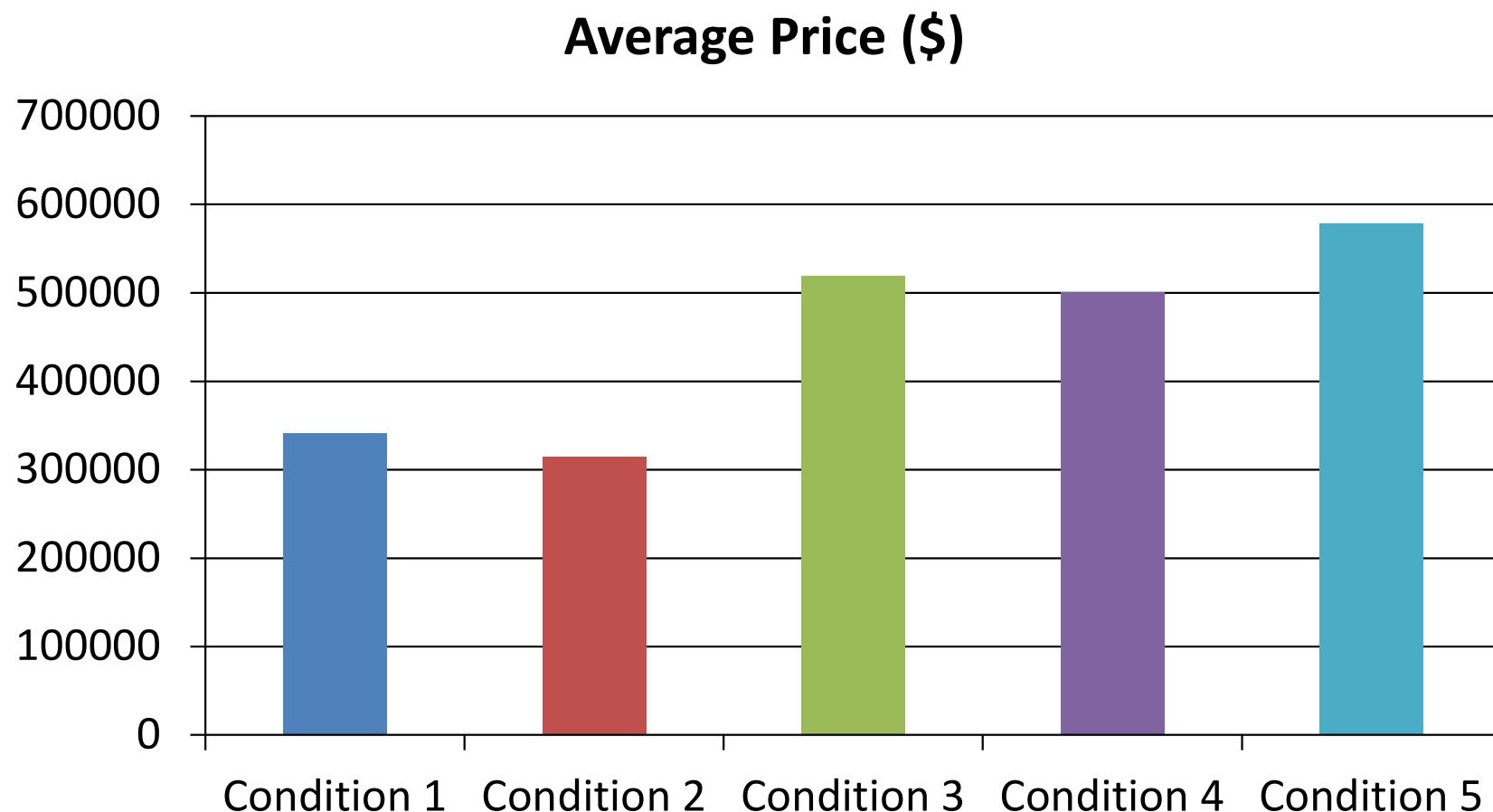
Data Overview

- Target variable: House price
- Key features: size, condition, location, renovation status
- Mostly numerical variables with minimal missing data
- 0 values in renovation year treated as not renovated

Research Questions & Hypotheses

- H1: Poor-condition houses offer strong renovation leverage
- H2: Some zipcodes are undervalued despite similar house characteristics
- H3: Houses farther from the city center are cheaper without major quality loss

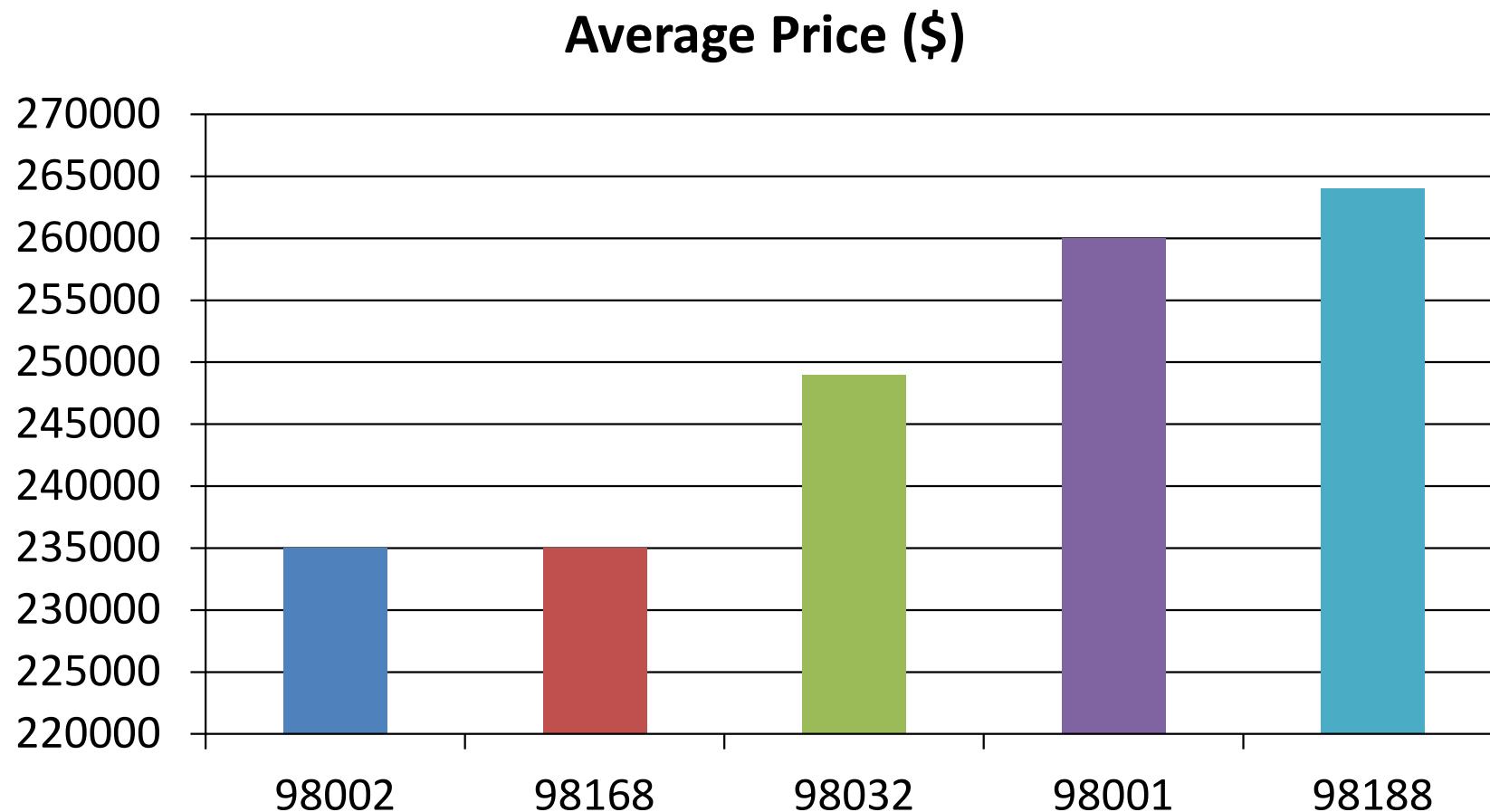
H1 Result: Average Price by House Condition



H1 Interpretation

- House condition has a weak relationship with price
- Price differences are small compared to renovation costs
- Supports Hypothesis 1 (renovation leverage exists)

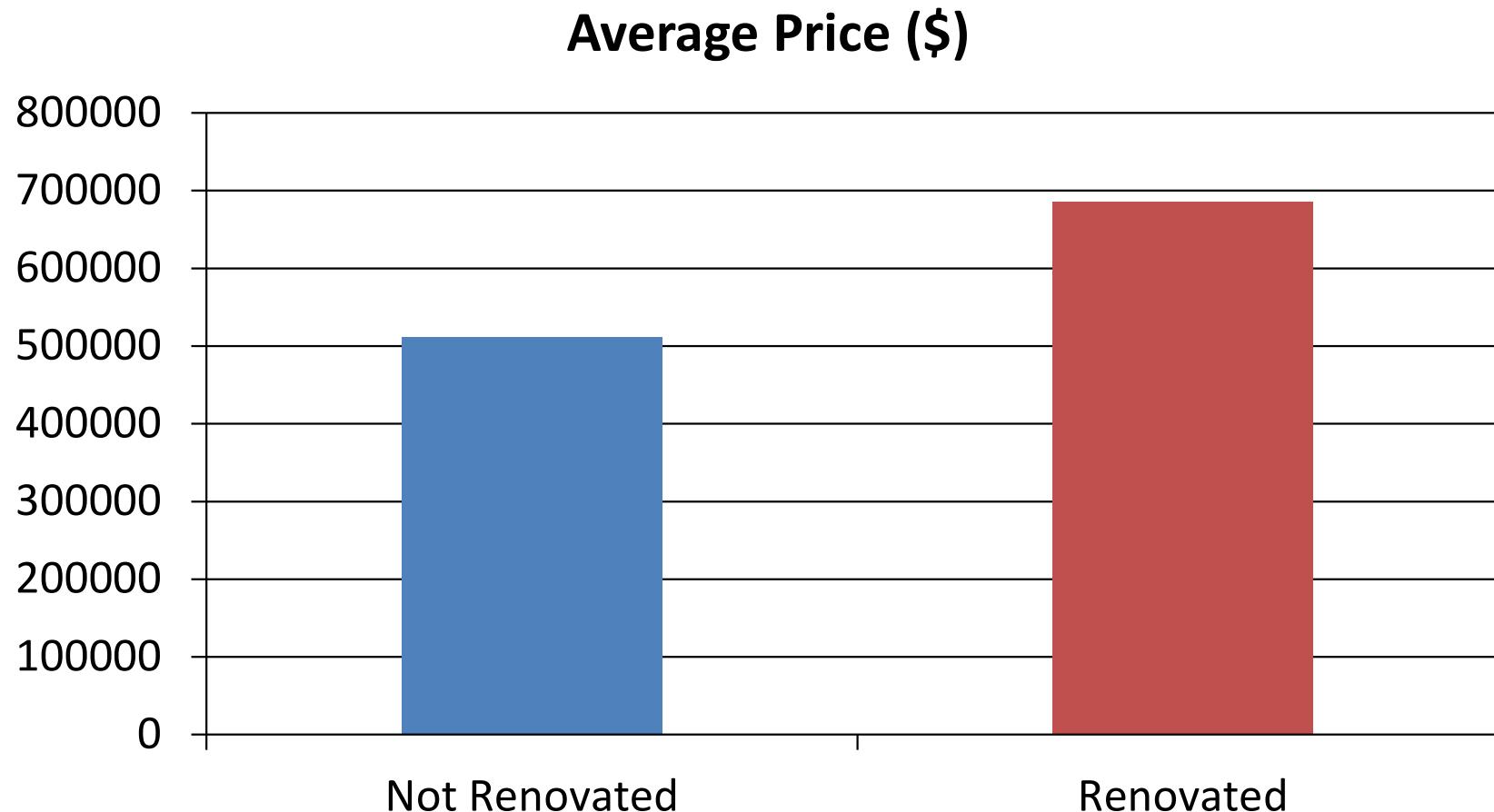
H2 Result: Undervalued Zipcodes (Median Price)



H2 Interpretation

- Several zipcodes show significantly lower median prices
- Price differences not fully explained by house quality
- Supports Hypothesis 2 (location effects are strong)

H3 Result: Renovated vs Non-Renovated Homes



H3 Interpretation

- Renovated houses sell for substantially higher prices
- Distance-based affordability does not imply poor housing quality
- Supports Hypothesis 3

Insights for the Client

- Renovation provides measurable financial uplift
- Zipcode selection is more important than house condition
- Southern King County presents ethical investment opportunities

Recommendations

- Target homes in condition 2–3
- Focus on 1,200–2,500 sqft houses
- Prioritize undervalued southern zipcodes
- Avoid luxury renovations and waterfront properties

Conclusion

- EDA supports all three hypotheses
- Data-driven strategy aligns profitability with social impact
- Further modeling could improve investment predictions

Correlation Matrix (Key Variables)

