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

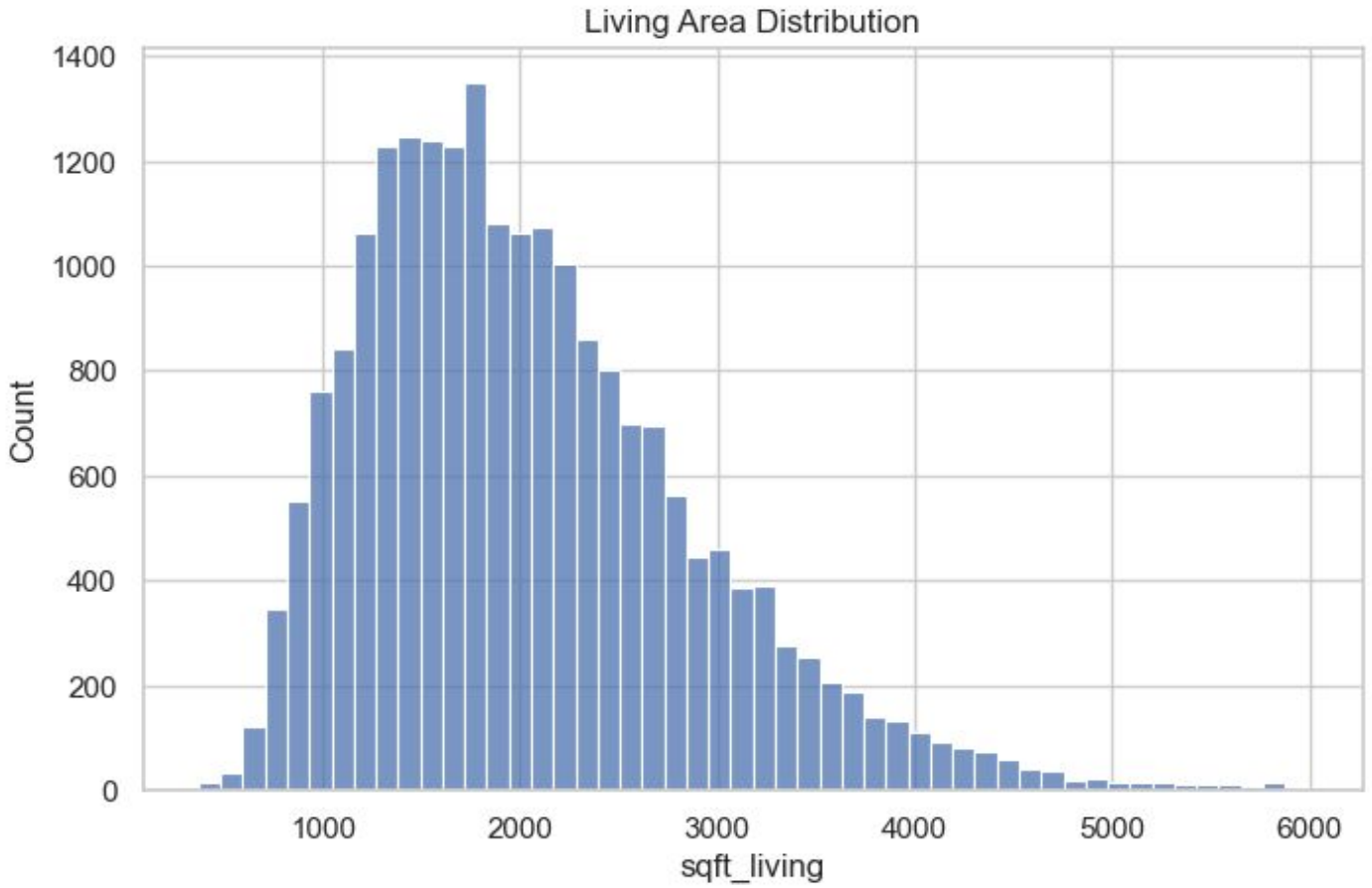
Research Questions & Hypotheses

- H1: Poor-condition houses offer strong renovation leverage
- H2: Some zipcodes are undervalued despite similar house characteristics
- H3: Latitude & longitude reveal undervalued housing areas away from Seattle center.

Observation:

Typical homes: 1,000–2,500 sqft

Large mansions are mostly excluded → filtered out



Exploring Distributions
Target Variable – Price

Observation:
Strong right skew
Majority of homes priced below approx.$750 000
Suitable for value-based investing

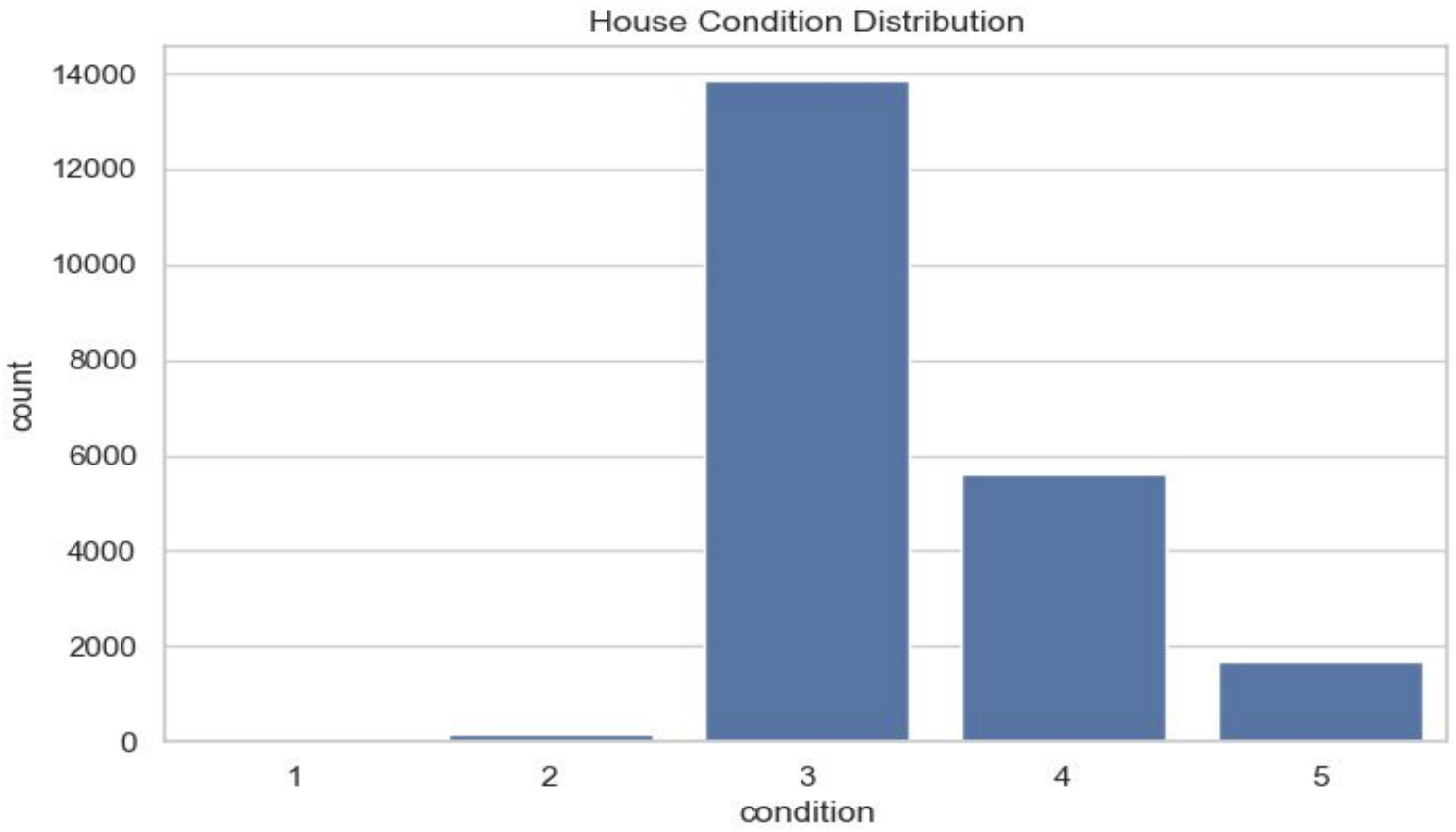
Renovated vs Non-Renovated Houses

Renovated vs Non-Renovated Houses

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

Observation:
Most houses rated 3 (average)

Few of them in very poor or excellent condition
one can see a good Opportunity to upgrade house condition cheaply



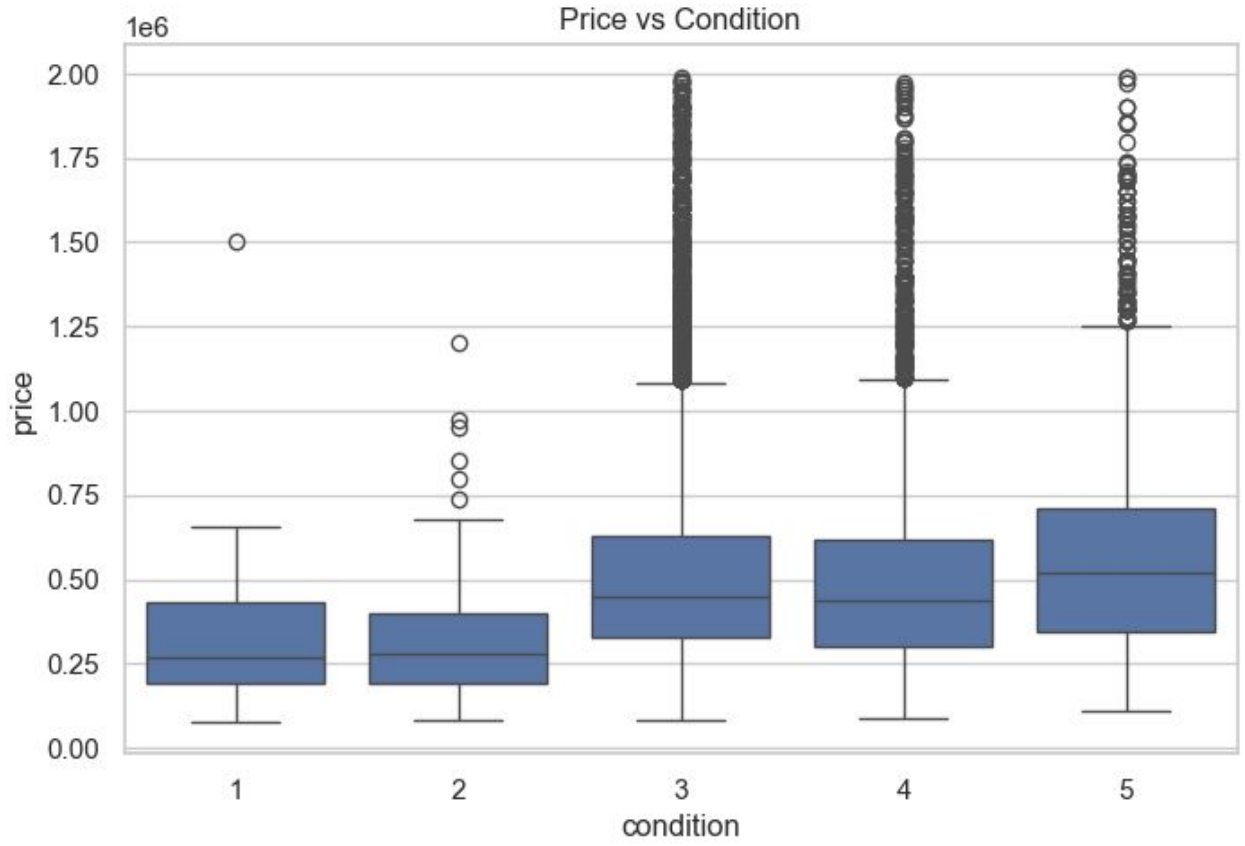
Hypothesis Analysis

Condition vs Price

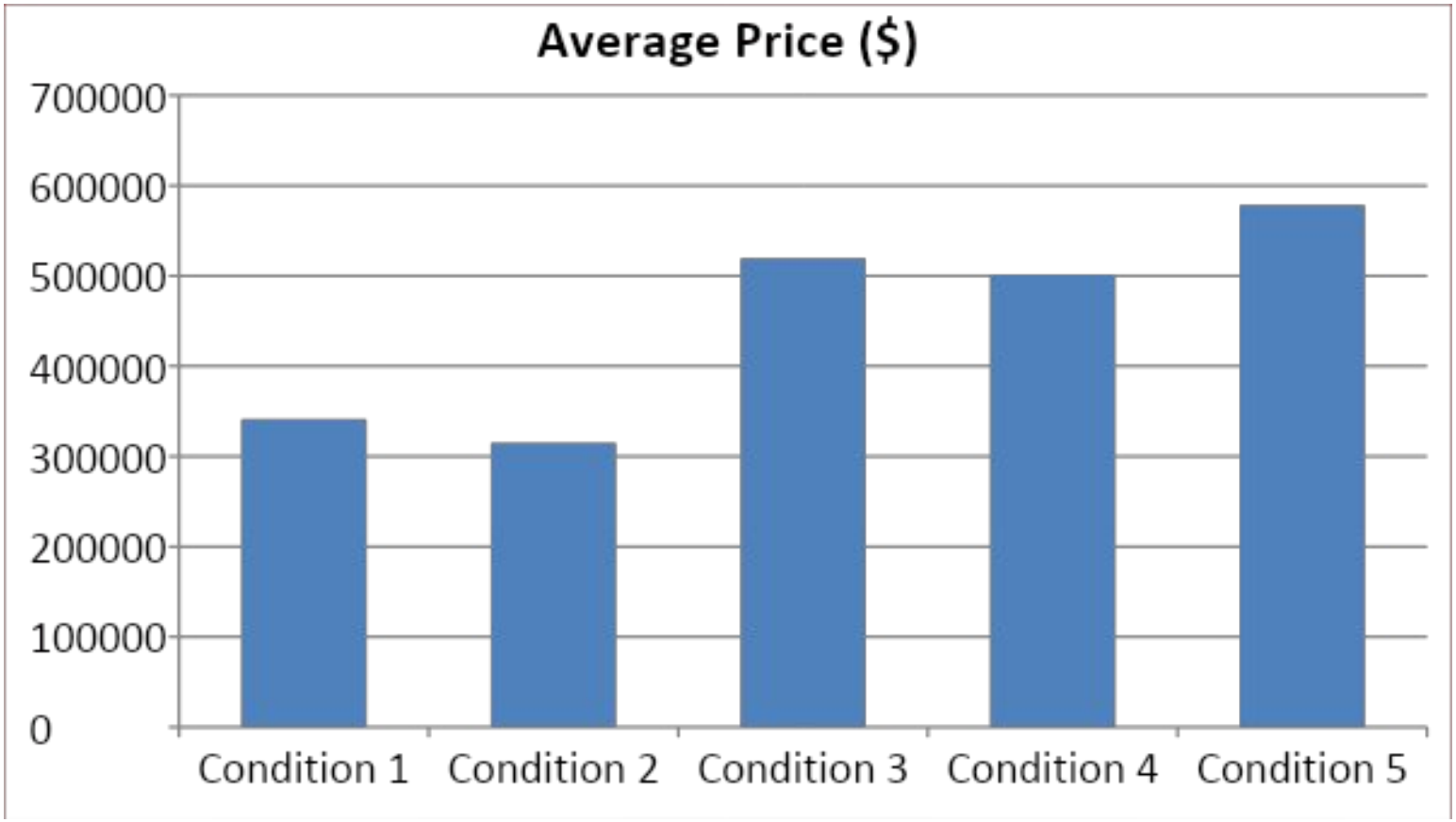
Insight:

Price does not rise sharply with condition

Renovations can increase resale value more than cost



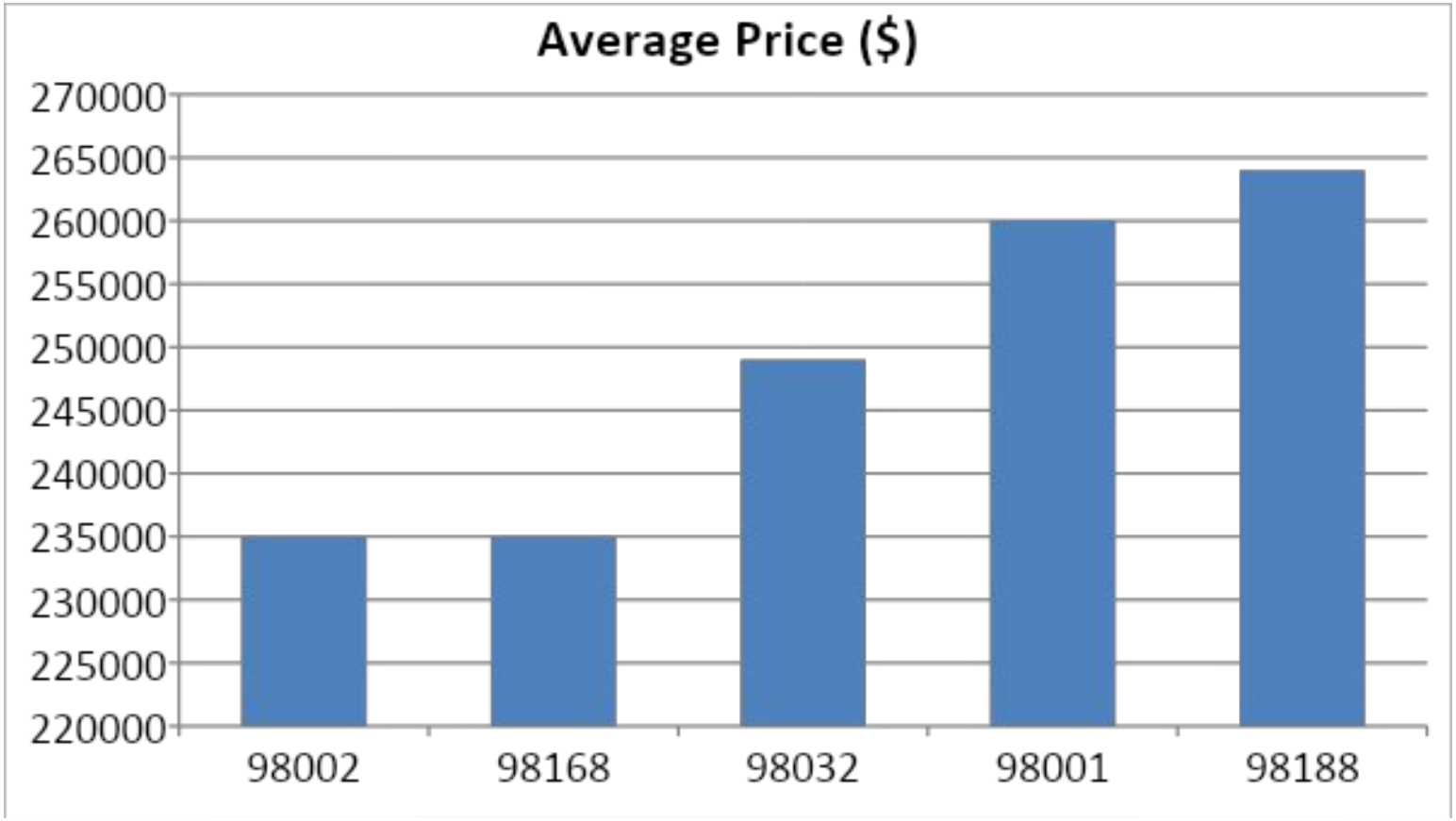
H1 Result: Average Price by House Condition



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

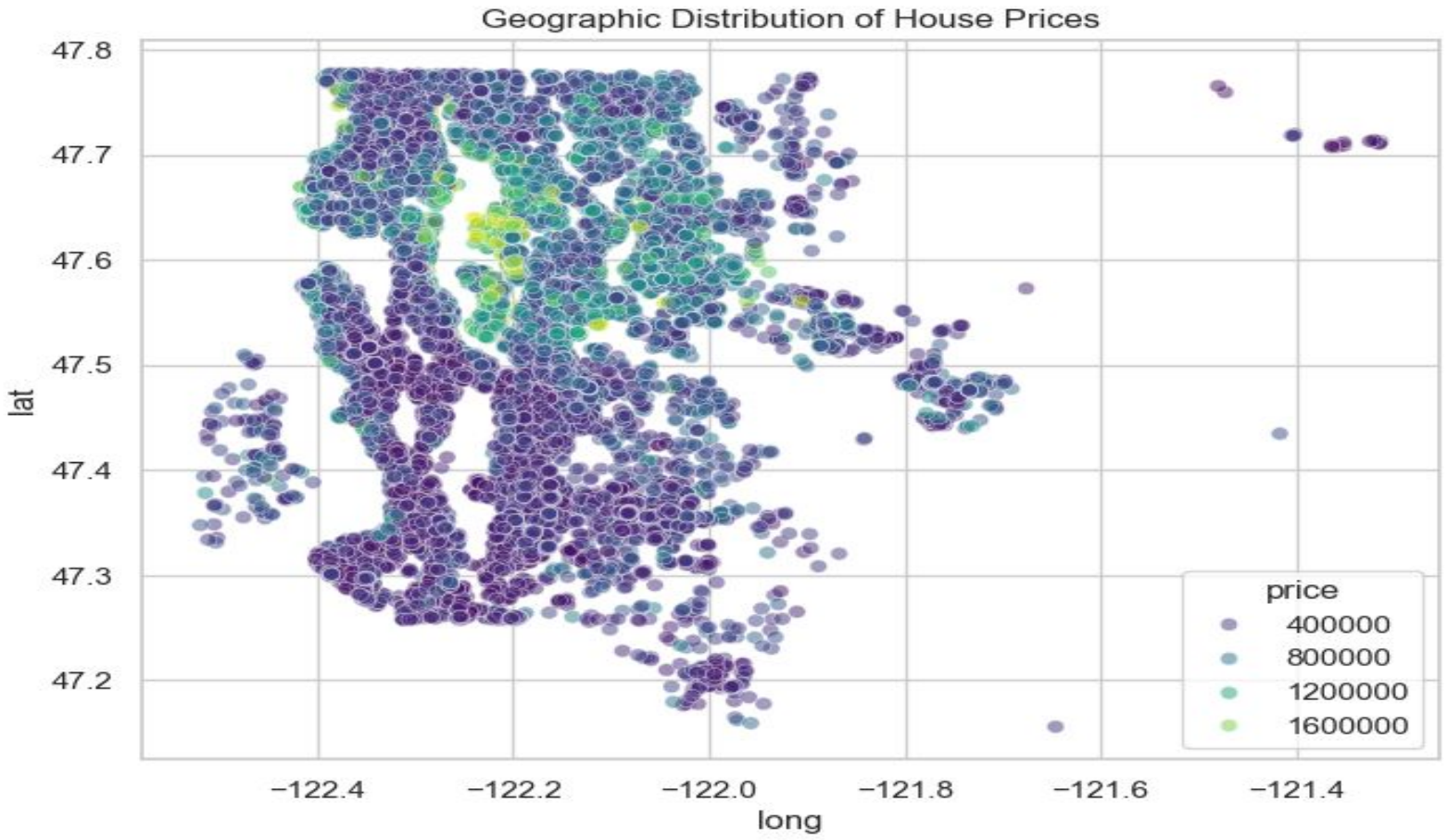
- Several zipcodes show significantly lower median prices
- Price differences not fully explained by house quality
- Supports Hypothesis 2 (location effects are strong)

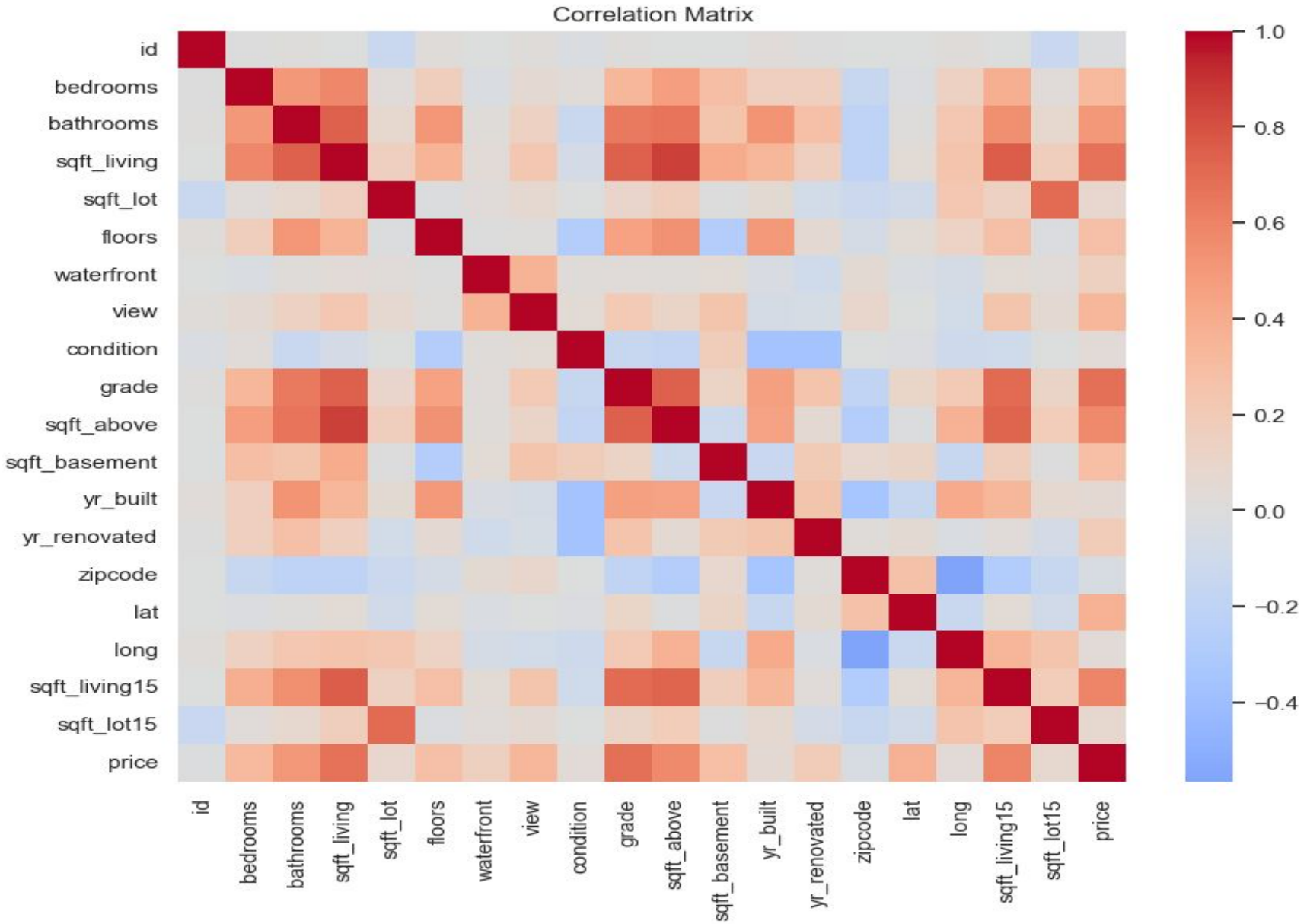
Geographical Insight: Supports Hypothesis 3

Lower prices cluster south & southeast of Seattle

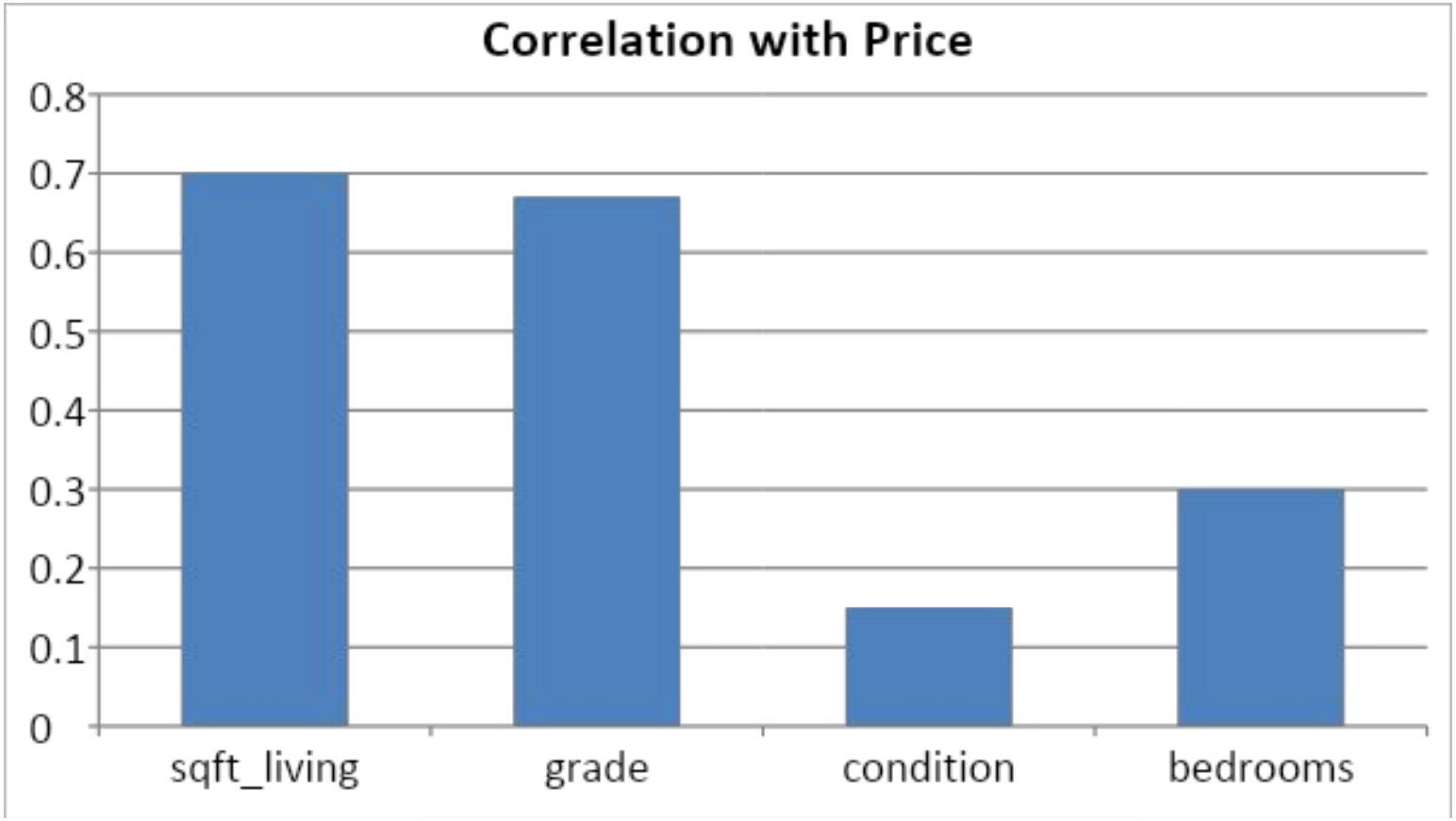
These areas still have:

1. Similar house sizes
2. Comparable conditions
3. Best locations for ethical flipping




Correlation Matrix
id
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
yr_built
yr_renovated
zipcode
lat
long
sqft_living15
sqft_lot15
price
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Correlation Matrix (Key Variables)



Key Findings:

sqft_living ↔ price → strong positive

grade ↔ price → very strong

condition ↔ price → weak
→ Condition upgrades = cheap leverage

Final Insights

Insight 1 – Renovation Leverage
House condition has a weak relationship with price, making not very expensive renovations economically efficient.

Insight 2 – Neighborhood Effects
Zipcode influences price more strongly than many physical house features.

Insight 3 – Geographic OpportunitySouthern King County offers affordable housing oportunities with
acceptable living standards, and this is also well connected with social responsibility goals.

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