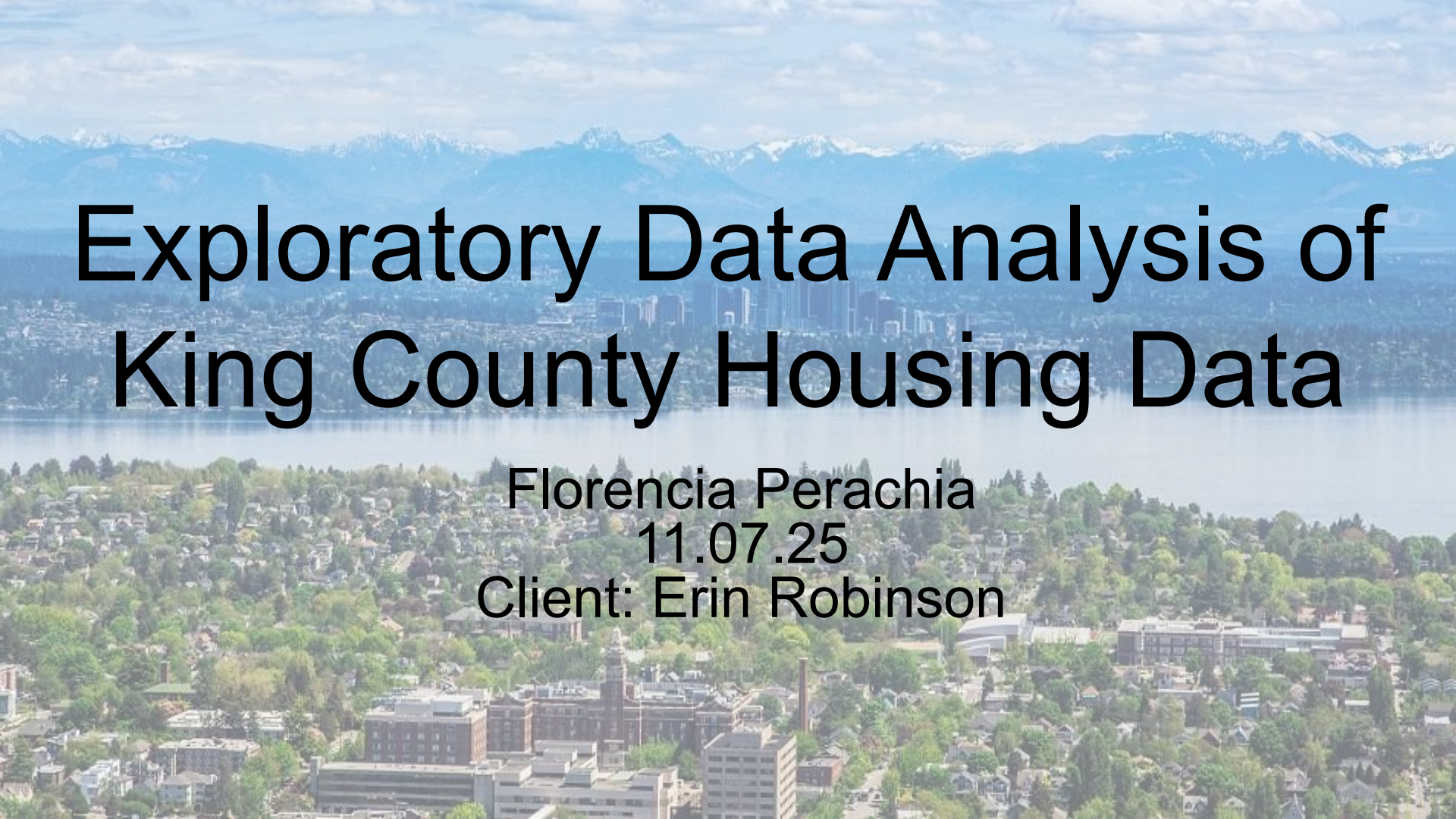


Exploratory Data Analysis of King County Housing Data

Florencia Perachia
11.07.25
Client: Erin Robinson

About the client:

- **Socially conscious** investor
- Interest: **affordable homes** in **under-resourced areas** that can be **improved and resold**, **balancing impact and return**

Goal of the EDA project:

- **Find** poorer areas,
- **Spot** undervalued homes
- **Assess** how small renovations can boost value
- **Keeping housing affordable**.

Data overview:

- **Dataset Size:** 21,597 entries
- **Key Features:** Price, square footage, location (ZIP, lat/long), condition (1-5), grade (1-13), year built/renovated

Preprocessing Steps:

- Handled **missing values** (e.g. set 0 for unrenovated)
- Decided to **keep extreme values** (Interest in poor houses)

Initial Hypotheses

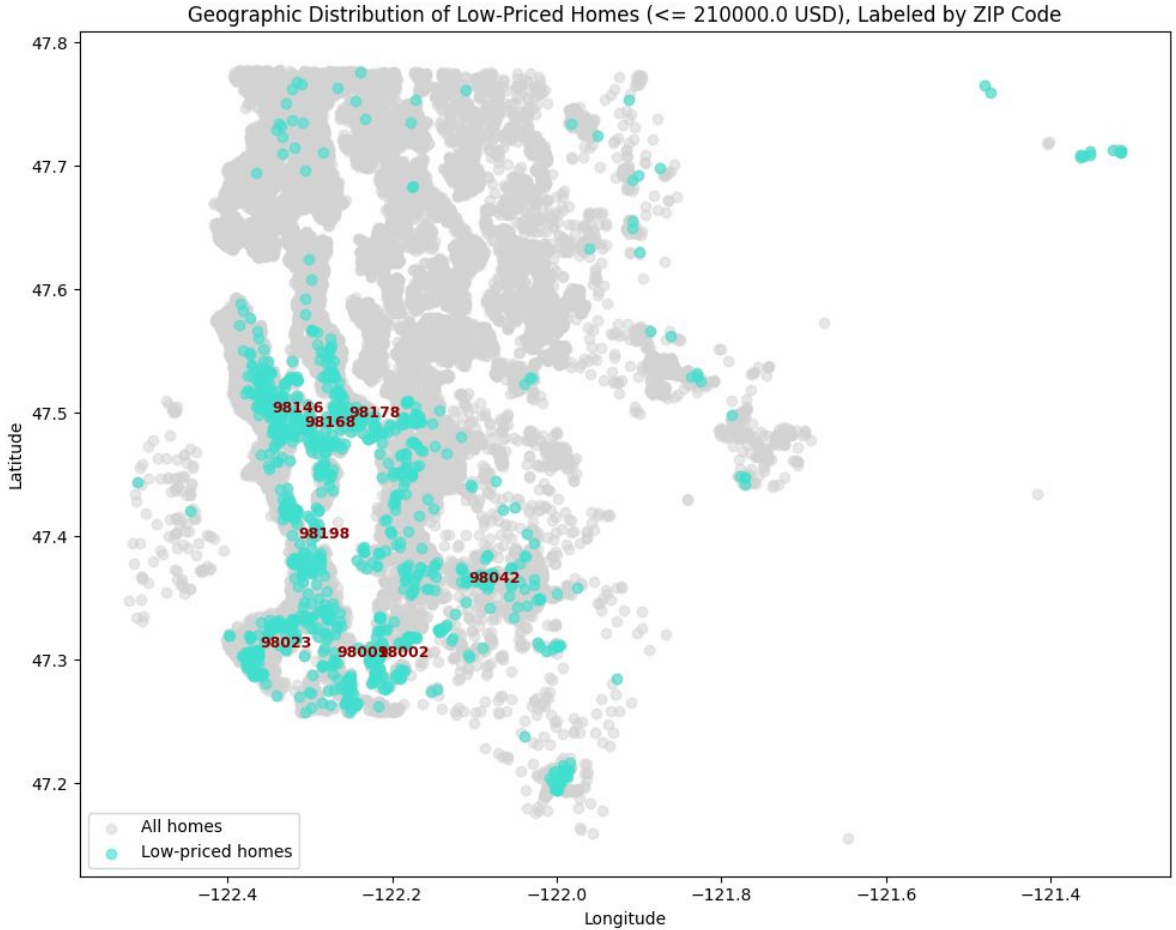
1st hypothesis: **"Low-priced homes cluster in specific geographic areas"**

<u>Poor house:</u>

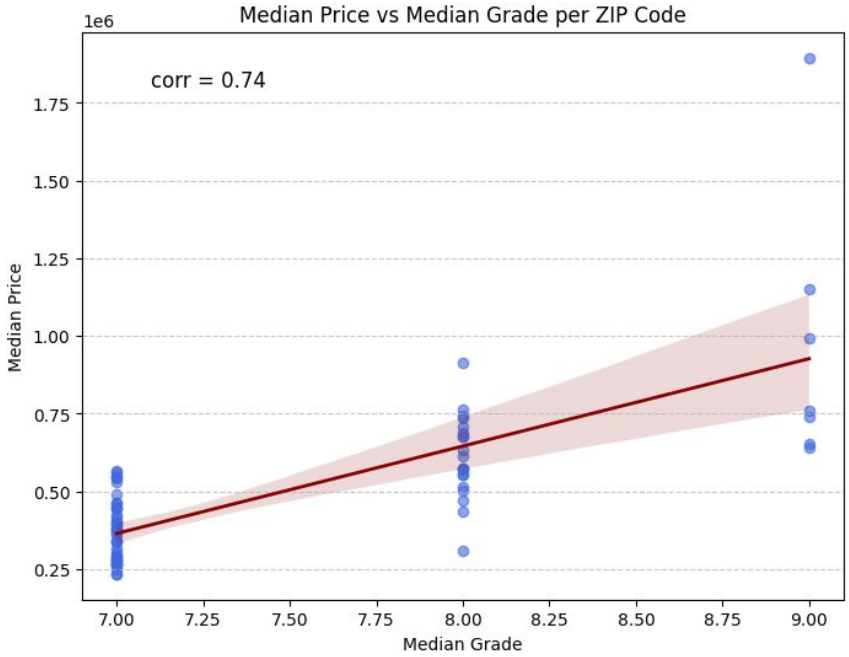
Prices below the 5th percentile

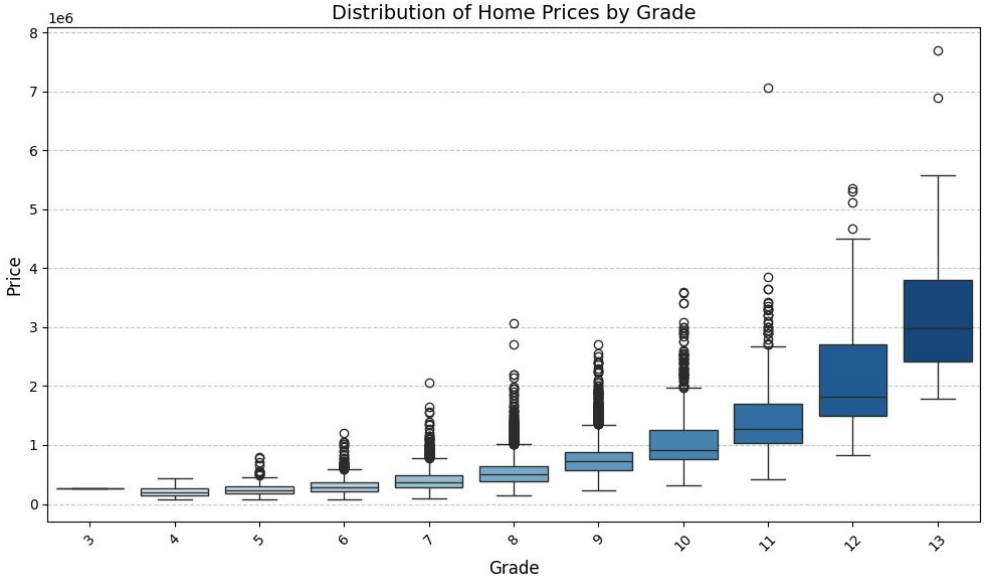
<= 210,000 USD

ZIPcodes:
poor houses > 50

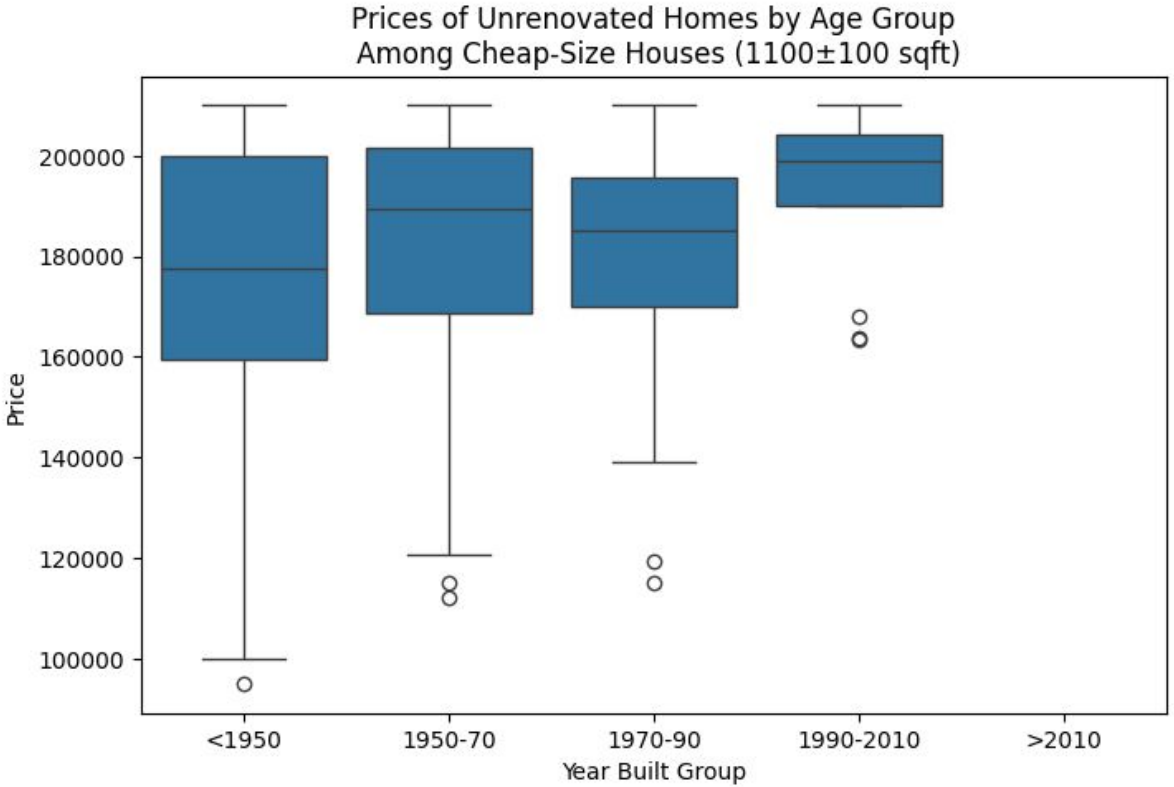


2nd Hypothesis: **"The median home price in a neighborhood is correlated with the overall grade of its houses."**





3rd Hypothesis: **" Among the cheapest and non-renovated homes, older houses are cheaper than newer ones of similar size."**



Questions from Erin
My insights

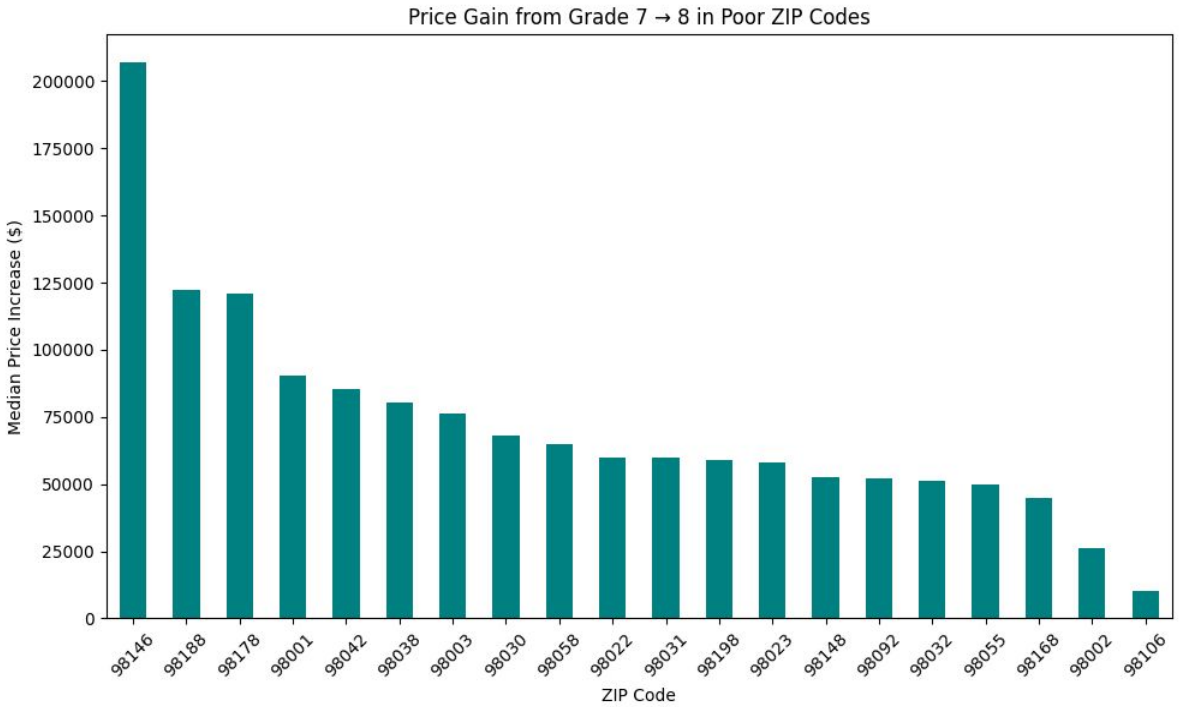
Question 1: **Which neighborhoods should I target? Remember, I am interested in poor neighborhoods**

Question 2: **Within the poor neighborhoods identified, are there homes that appear undervalued compared to others nearby and could be improved affordably?**

- 50% Undervalued
-never renovated
-Grade <=7

house_id	zipcode	grade	price_per_sqft	median_ppsqft
1823049202	98146	7	90.674	201.821
723049156	98146	6	87.647	201.821
3883800011	98146	6	95.349	201.821
723049326	98146	6	89.701	201.821
2113700360	98106	7	100.962	215.909
6453300306	98106	7	96.544	215.909
723049158	98146	5	92.466	201.821
6121800050	98148	7	89.862	181.423

Question 3: **How would small improvements increase home value? I do not want to make prices too high for the community (Only for the poor neighborhoods)**



Average price increase (grade 7→8) in poor ZIPs: 72,034 USD

Remember that poor houses < 210,000 USD

Question 4: **Which neighborhoods do need help? Are there ZIP codes with low quality homes?**

Low quality houses :

- Grade <= 5
- Condition <= 3
- Never renovated

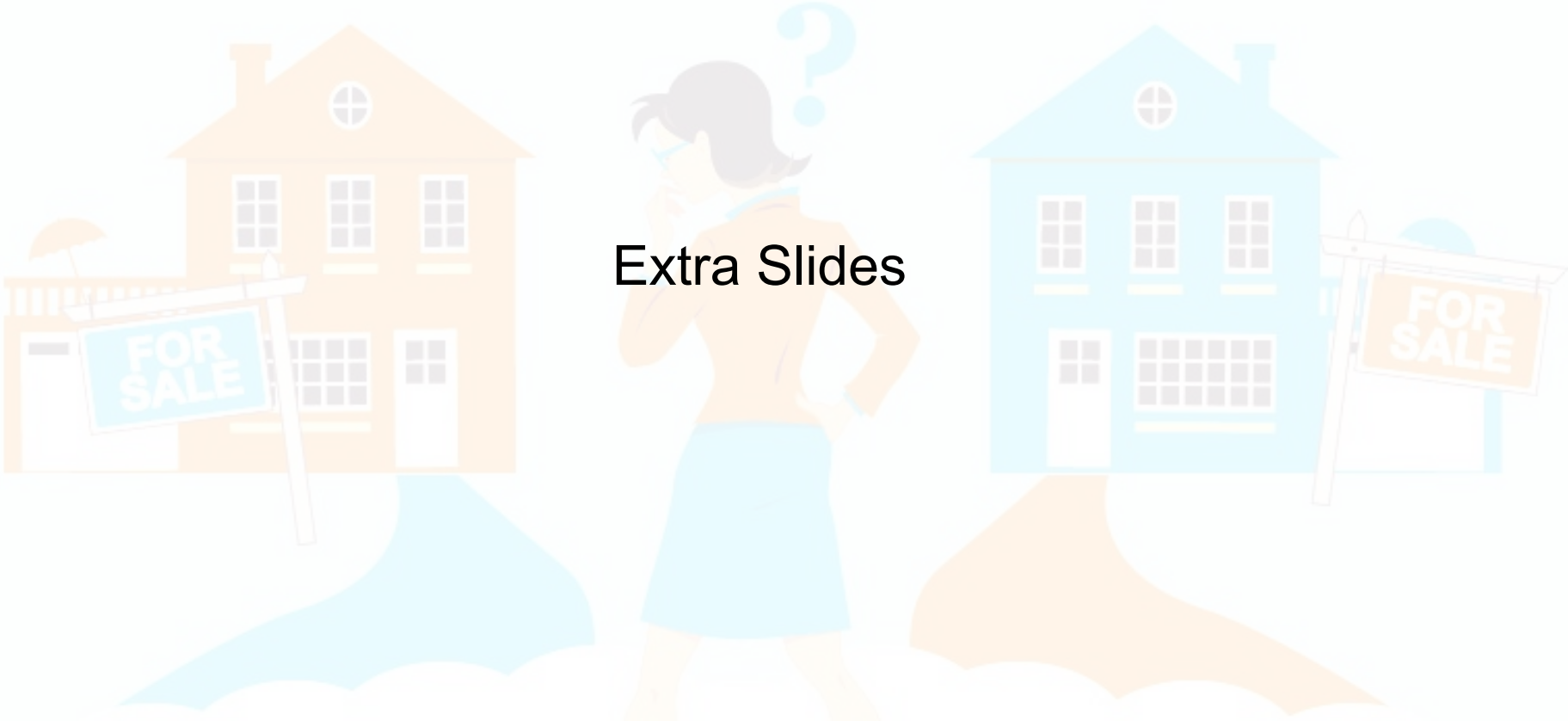
zipcode	total_homes	low_quality_homes	pct_low_quality
98014	124	8.000	6.452
98168	269	12.000	4.461
98146	288	11.000	3.819
98148	57	2.000	3.509
98055	268	8.000	2.985
98106	335	9.000	2.687
98108	186	4.000	2.151
98065	308	6.000	1.948
98070	117	2.000	1.709
98188	136	2.000	1.471

Conclusions:

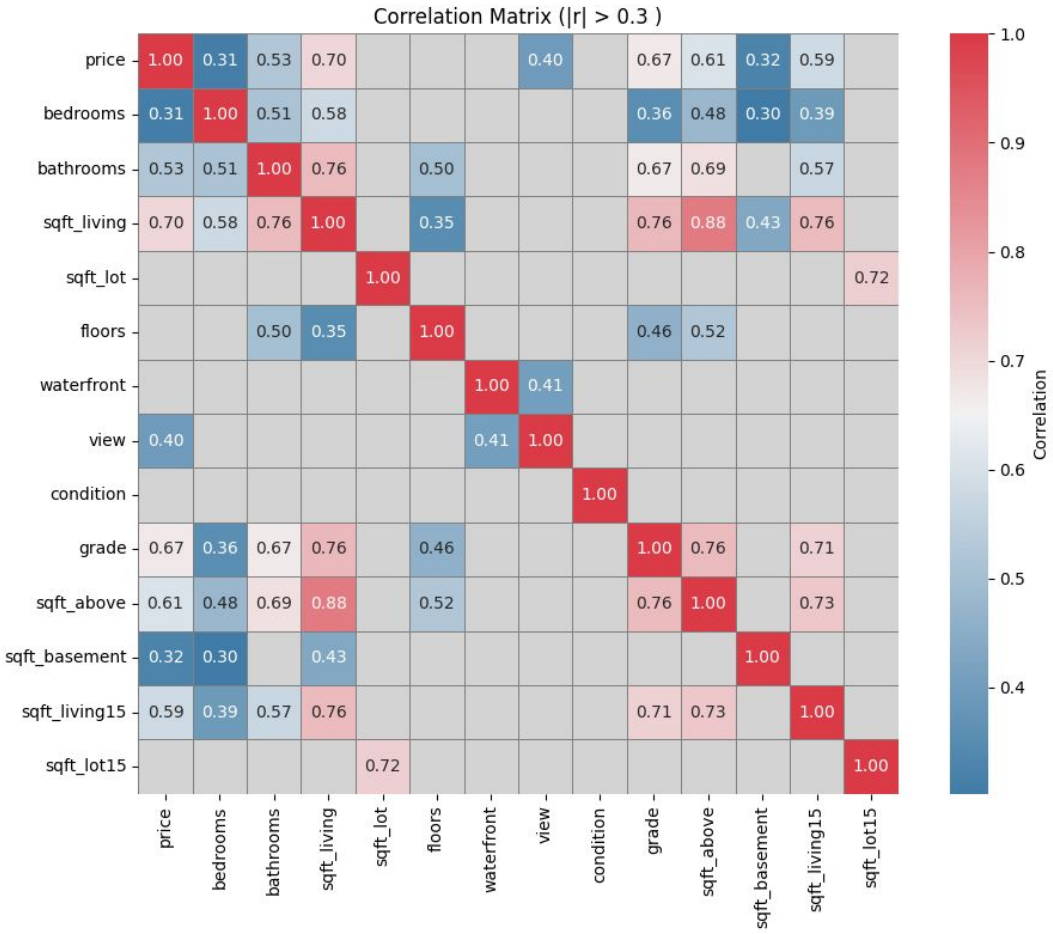
- Erin should focus on ZIP codes with **many low-priced homes**
- Target **undervalued, unrenovated houses with low grades**
- Prioritize areas where **renovations (e.g., grade 7 → 8)** bring **modest price increases**

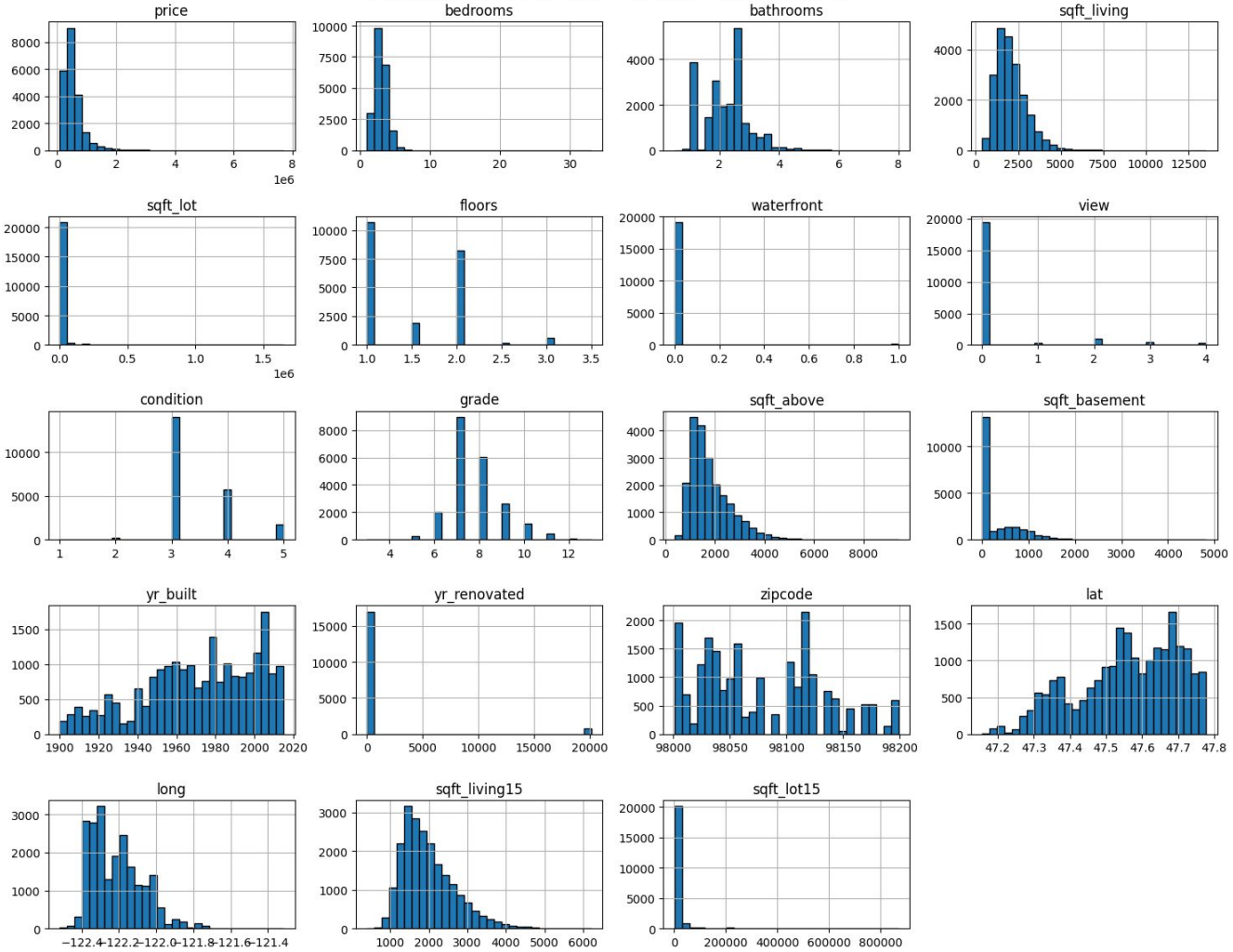
This balances **investment potential** with **affordable housing goals**

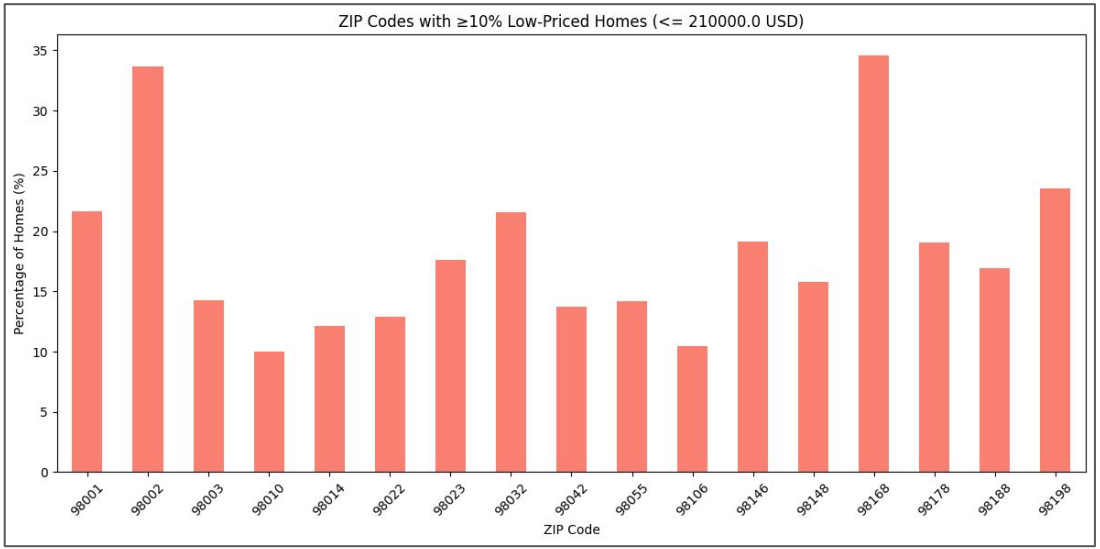
Thank you !

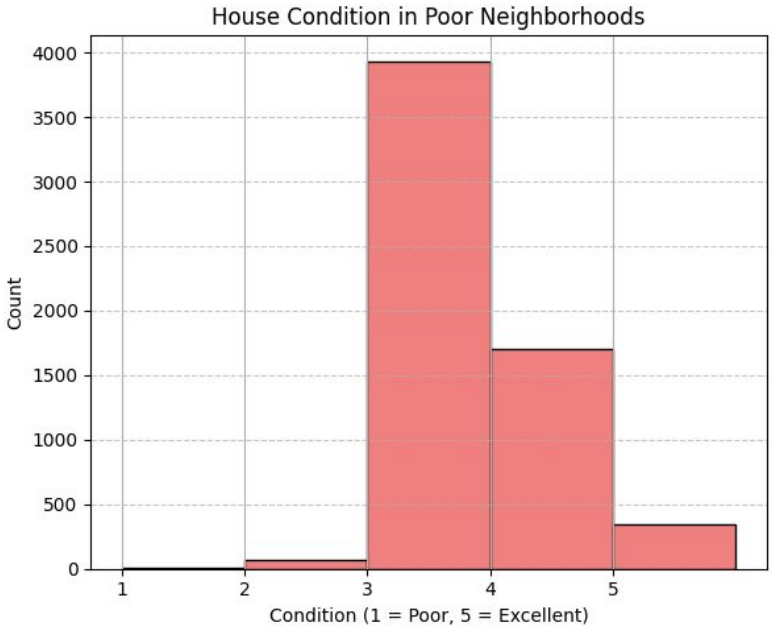


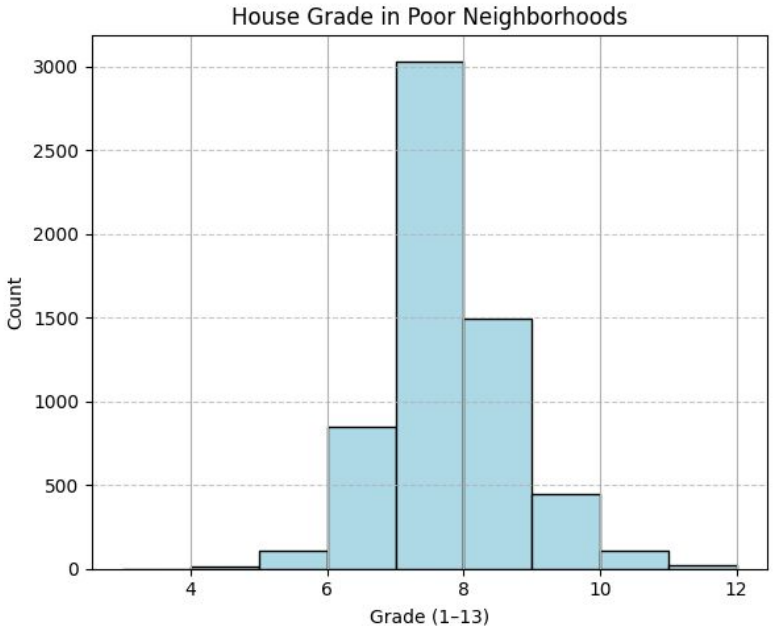
Extra Slides


Correlation Matrix (|r| > 0.3)
price 1.00 0.31 0.53 0.70 0.40 0.67 0.61 0.32 0.59
bedrooms 0.31 1.00 0.51 0.58 0.36 0.48 0.30 0.39
bathrooms 0.53 0.51 1.00 0.76 0.50 0.67 0.69 0.57
sqft_living 0.70 0.58 0.76 1.00 0.35 0.76 0.88 0.43 0.76
sqft_lot 1.00 0.72
floors 0.50 0.35 1.00 0.46 0.52
waterfront 1.00 0.41
view 0.40 0.41 1.00
condition 1.00
grade 0.67 0.36 0.67 0.76 0.46 1.00 0.76 0.71
sqft_above 0.61 0.48 0.69 0.88 0.52 0.76 1.00 0.73
sqft_basement 0.32 0.30 0.43 1.00
sqft_living15 0.59 0.39 0.57 0.76 0.71 0.73 1.00
sqft_lot15 0.72 1.00
price
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
sqft_living15
sqft_lot15
Correlation
1.0
0.9
0.8
0.7
0.6
0.5
0.4


Distributions of Continuous Variables (Excluding IDs)
price
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
1e6


ZIP Codes with ≥10% Low-Priced Homes (<= 210000.0 USD)
Percentage of Homes (%)
0
5
10
15
20
25
30
35
98001
98002
98003
98010
98014
98022
98023
98032
98042
98055
98106
98146
98148
98168
98178
98188
98198
ZIP Code


House Condition in Poor Neighborhoods
Count
Condition (1 = Poor, 5 = Excellent)
0
500
1000
1500
2000
2500
3000
3500
4000
1
2
3
4
5


House Grade in Poor Neighborhoods
Count
Grade (1–13)
0
500
1000
1500
2000
2500
3000
4
6
8
10
12