

# Exploratory Data Analysis

Deep-dive into Seattle's Real Estate Scene

# Exploratory Data Analysis for King County Historic Housing



## Introduction:

- ◆ Briefing: EDA tailored for Jacob Phillips' research to find the perfect historic house in Seattle.



## Research Question:

- ◆ *"How can we identify historic houses with 4 or more bathrooms and large lots suitable for installing a tennis court and/or pool, located near golfing amenities, and not situated on waterfronts, in the Seattle area?"*
- ◆ *If such a house cannot be found, what are the options for acquiring two adjacent properties that together meet these criteria?"*



## Objective:

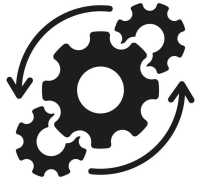
- ◆ Assist in finding a property that meets unique criteria: Historic charm, ample bathrooms, extensive outdoor space, near golf facilities, and away from waterfronts.

# Exploratory Journey and Hypotheses



## Hypotheses to Explore:

- ◆ **H1:** Historic houses are more likely to have larger lots, which could accommodate Jacob's desire for a tennis court and/or pool.
- ◆ **H2:** Areas with lower average lot sizes might have a higher density of houses, potentially increasing the likelihood of finding two adjacent properties for sale.
- ◆ **H3:** Historic houses with recent renovations are more likely to meet modern standards, including the desired number of bathrooms.



## Methodological Approach:

- ◆ **Data Retrieval:** From PostgreSQL for a comprehensive analysis foundation.
- ◆ **Data Preparation:** Cleaning and preprocessing for dataset relevance and reliability.
- ◆ **Exploratory Data Analysis:** Identifying trends, patterns, and outliers.
- ◆ **Statistical Analysis & Visualization:** Testing hypotheses and drawing conclusions to guide the property search effectively.

# Evaluation of Lot Sizes for Historic Homes

## Introduction to Hypothesis 1:

- Hypothesis: Historic houses are more likely to have larger lots, which could accommodate Jacob's desire for a tennis court and/or pool.
- Definition of Historic: Homes built before 1965 (Source: U.S. National Register)
- Definition of large lot: has to fit a tennis court (approx. 2'800 sqft) and a pool (approx. 5'000sqft), plus extra space, in total 20'000+ sqft is to be considered large.

## Statistical Analysis Overview:

- T-test comparing lot sizes of historic vs. non-historic homes.
- Correlation analysis between year built and lot size.

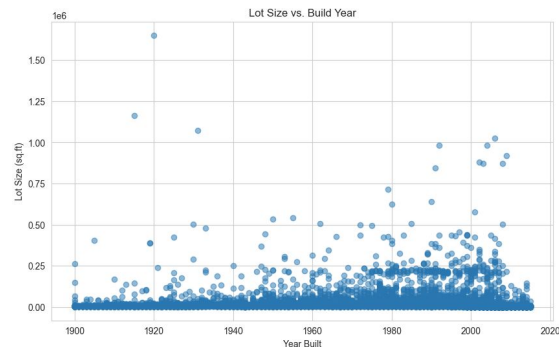
## Key Findings:

- T-test: Significant difference with a T-statistic of -10.98, indicating historic homes have smaller lots on average (p-value < 0.05).
- Correlation: Weak positive relationship (coefficient: 0.0529) between year built and lot size, suggesting newer homes have slightly larger lots (p-value < 0.05).

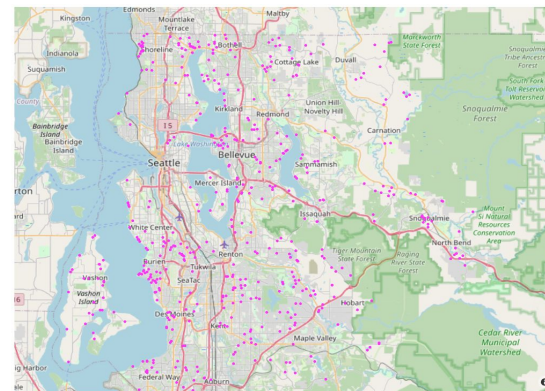
## Conclusion:

- Contrary to our hypothesis, historic houses are generally smaller in lot size compared to newer homes.
- Newer homes may offer more potential for amenities like tennis courts or pools.

Seattle's Historic Houses: Mapping Large Lot Distributions by Build Year



Geographical distribution of historic houses with large lots (>20,000 sqft) in Seattle.



# Evaluation - Lot Size and Housing Density (1/2)

## Introduction to Hypothesis 2:

Hypothesis: In areas with smaller average lot sizes, we expect a higher density of houses, increasing the chances of finding two adjacent properties for sale.

## Statistical Analysis Overview:

- Analysis of variance (ANOVA) conducted to assess differences in lot sizes (sqft\_lot15) across different zip codes.
- Examination of potential correlations between average lot sizes and housing density.

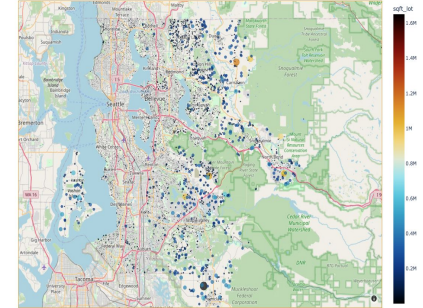
## Key Findings:

- ANOVA Results: Confirmed significant differences in average lot sizes by zip code.
- Implications for Housing Density: Suggests smaller lots may lead to higher density of houses, based on average lot sizes.
- Adjacent Property Availability: Hypothesizes that higher density could correlate with increased likelihood of adjacent properties for sale.

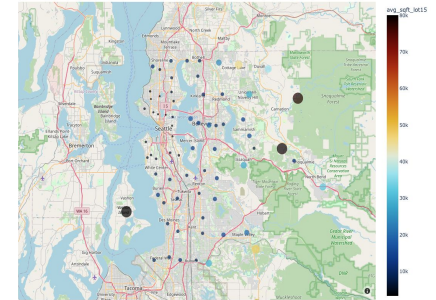
## Conclusion:

- Preliminary Support: Statistical significance in lot size variation lends indirect support to the hypothesis.
- Further Research Needed: Direct measurement of housing density and analysis of real estate listings are necessary to conclusively evaluate the hypothesis.

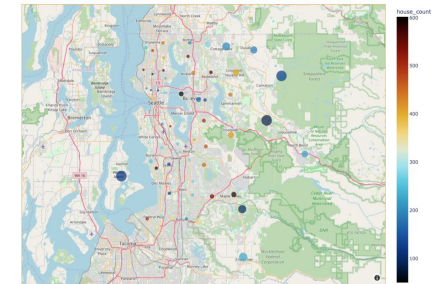
**Fig.1:** Seattle's Property Lot Size Overview



**Fig.2:** Comparative Analysis of Average Lot Sizes Across Seattle's Zip Codes

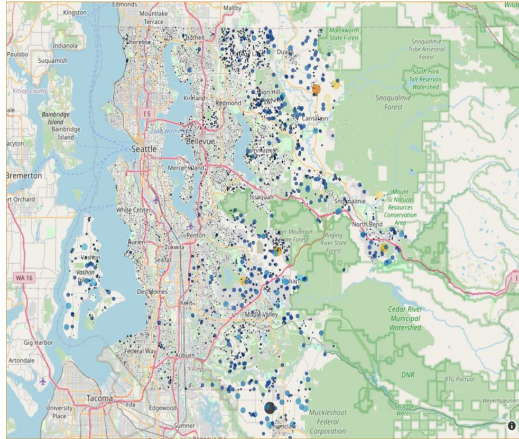


**Fig. 3:** Correlation Between Lot Sizes and Housing Density in Seattle by Zip Code

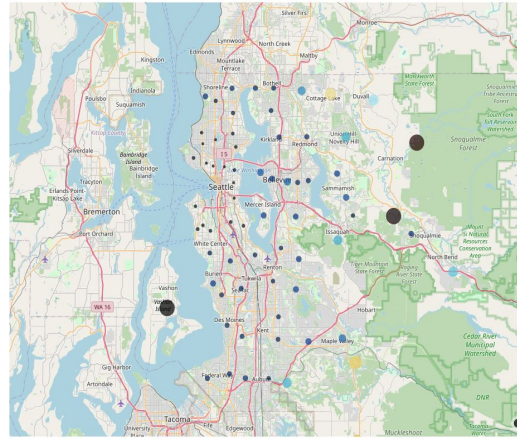


# Evaluation - Lot Size and Housing Density (2/2)

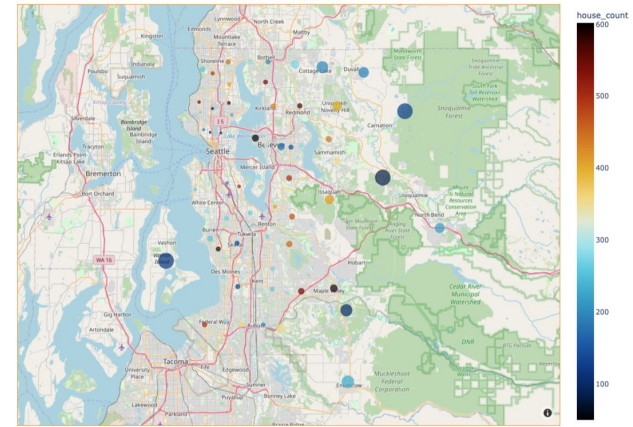
**Fig.1:** Seattle's Property Lot Size Overview



**Fig.2:** Comparative Analysis of Average Lot Sizes Across Seattle's Zip Codes



**Fig. 3:** Correlation Between Lot Sizes and Housing Density in Seattle by Zip Code



# Evaluation - Renovations and Modern Bathroom Standards

## Introduction to Hypothesis 3:

- Hypothesis: Historic houses with recent renovations are expected to be closer to modern bathroom standards.
- Modern Standards Definition for Jacob: 4 bathrooms are considered the modern standard by Jacob for accommodating contemporary needs.

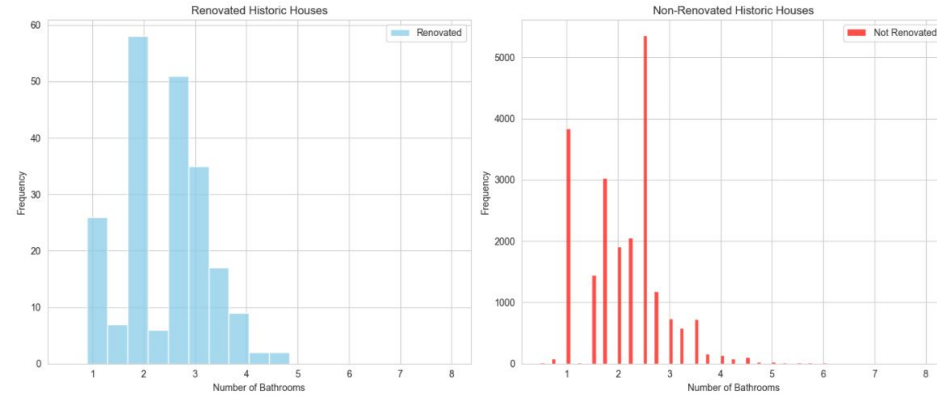
## Statistical Analysis Overview:

- Comparison of the average number of bathrooms in renovated versus non-renovated historic houses.
- T-test to assess the statistical significance of the observed differences.

## Key Findings:

- Renovated Historic Houses: Average of 2.40 bathrooms, a notable increase from non-renovated counterparts (1.6).
- Comparison with Modern Standards: Renovated historic houses are closer but still below the modern standard of 4 bathrooms.
- Statistical Analysis Results: Strongly significant T-statistic (12.999) with a compelling P-value ( $5.109884472511518e-29$ ), indicating a genuine difference not attributable to chance.

**Fig.4:** Distribution of the number of bathrooms in renovated versus non-renovated historic houses.



## Conclusion:

- Renovations significantly increase bathroom numbers, pushing historic homes towards modern standards but often falling short of the defined criterion.
- To fully meet modern standards, historic houses might require more extensive renovations specifically aimed at adding bathrooms.

# Summary of Hypotheses Evaluation

H1: Historic houses are more likely to have larger lots, which could accommodate Jacob's desire for a tennis court and/or pool.

H2: Areas with lower average lot sizes might have a higher density of houses, potentially increasing the likelihood of finding two adjacent properties for sale.

H3: Historic houses with recent renovations are more likely to meet modern standards, including the desired number of bathrooms.

Answers to hypotheses:

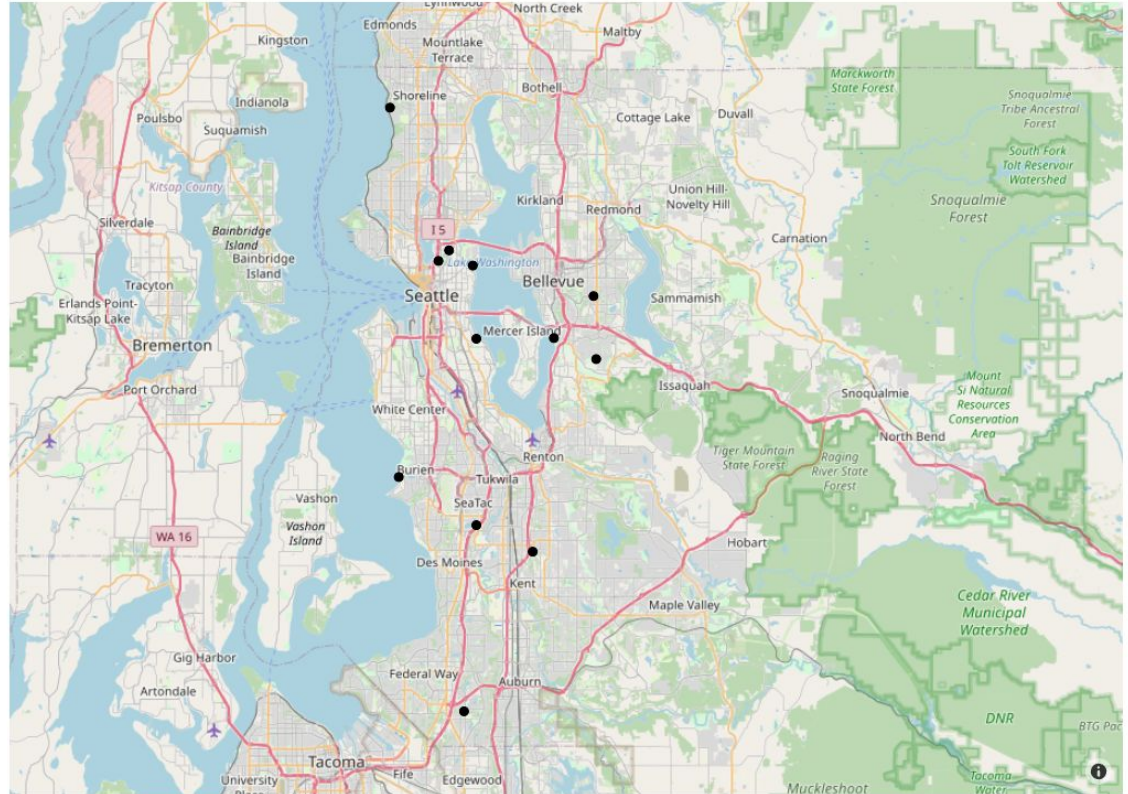
1. **Size of Historic Lots:** Data reveals that historic houses typically have smaller lots than newer homes.
2. **Amenities in Newer Homes:** Newer constructions are more likely to have the space for amenities like tennis courts or pools.
3. **Lot Size Variance by Location:** ANOVA highlights significant differences in lot sizes across Seattle's zip codes.
4. **Lot Size and Density Link:** Smaller average lot sizes are indicative of higher housing density.
5. **Sale of Adjacent Properties:** Areas with higher housing density may have a greater chance of adjacent properties being on the market.
6. **Bathrooms in Renovated Homes:** Renovated historic homes average 2.40 bathrooms, surpassing the 1.65 in non-renovated ones.
7. **Jacob's Modern Bathroom Standard:** Even with renovations, historic homes don't typically reach the modern standard of 4 bathrooms.
8. **Impact of Renovations:** While renovations lead to more bathrooms, they may still not meet the highest contemporary expectations.



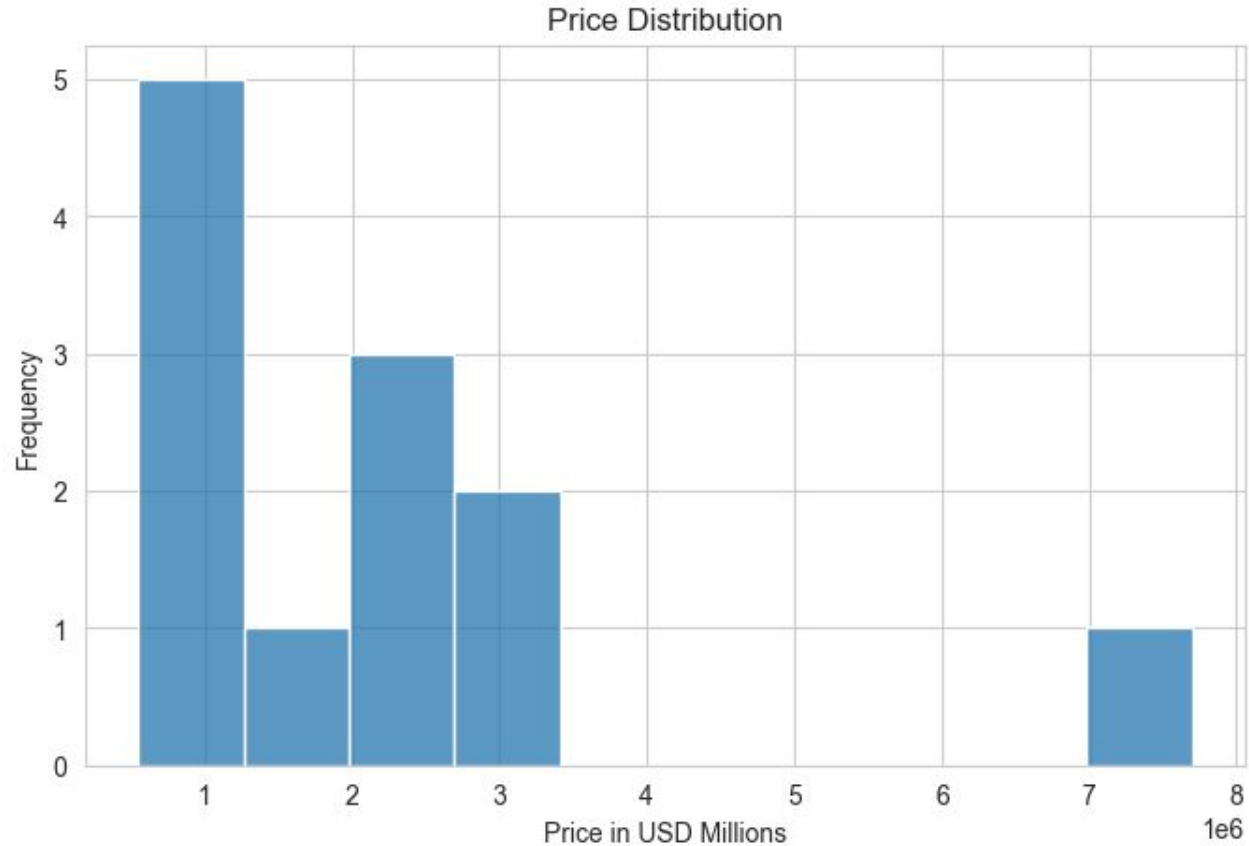
**12 potential properties, all meeting the criteria: (a) historic (>1965), (b) 4+ bathrooms, (c) lots with 20'000+ sq ft and (d) no waterfront". 6 are in the King County Area.**

The houses are located in:

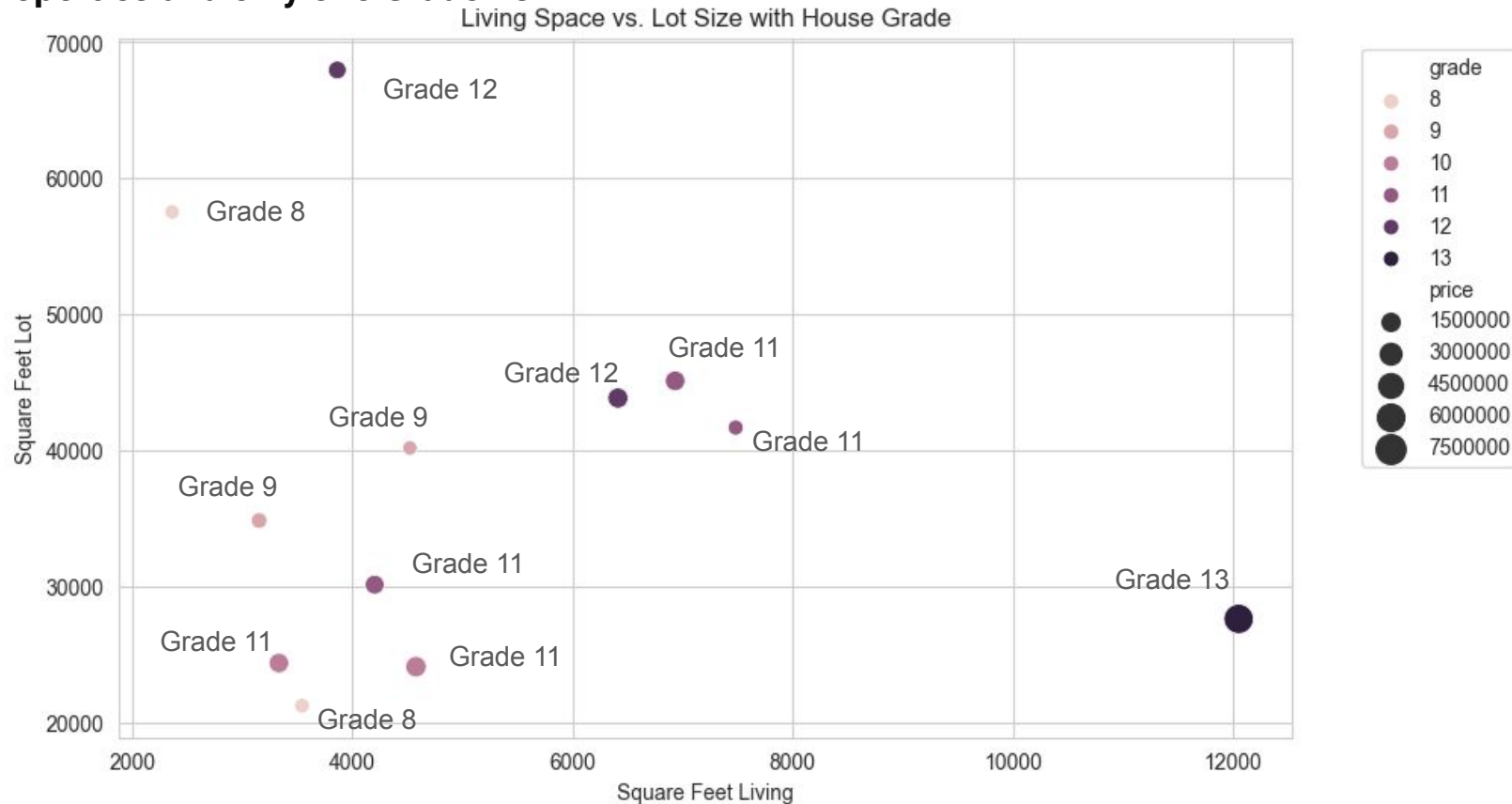
1. Greater Seattle Area, 98006
2. Greater Seattle Area, 98003
3. **King County, 98166**
4. **King County, 98144**
5. **King County, 98102**
6. Greater Seattle Area, 98007
7. **King County, 98112**
8. **King County, 98177**
9. **King County, 98188**
10. Greater Seattle Area, 98031



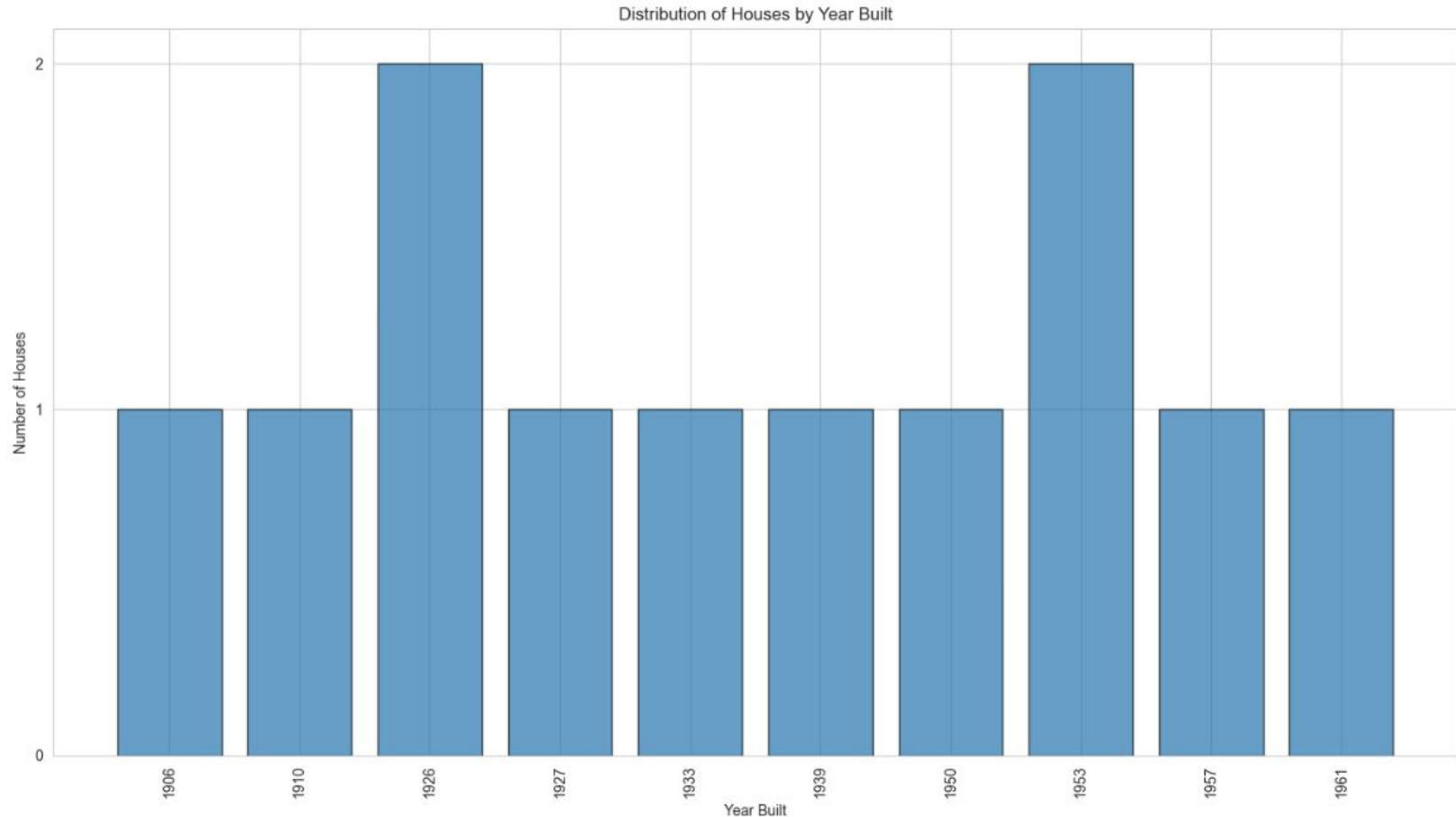
**Price overview for houses meeting the criteria ranges mainly between 0.5 and 3.5M USD, with an exception of 7.7M USD**



**The Lot and Living sq ft, categorized by house grade Vis., displays a high presence of Grade 11 properties and only one Grade 13**



**The properties can be almost evenly split between pre-war houses (7) and post-war (5)**



# Overview of the correlation characteristics of houses meeting the criteria

Only relevant correlations have been highlighted:

1. Homes with higher prices are likely to have higher grades
2. Homes with higher prices are more likely to have high sqft for the living area, but do not necessarily have a larger lot
3. Higher grade houses are very likely to have many bathrooms

