# House Sales in King Country, USA

Exploratory Data Analysis

Dimitra Gkogkou

26 April 2024

# Buyer

**Erin Robinson**

- Invest in poor neighborhood
- Buying & selling
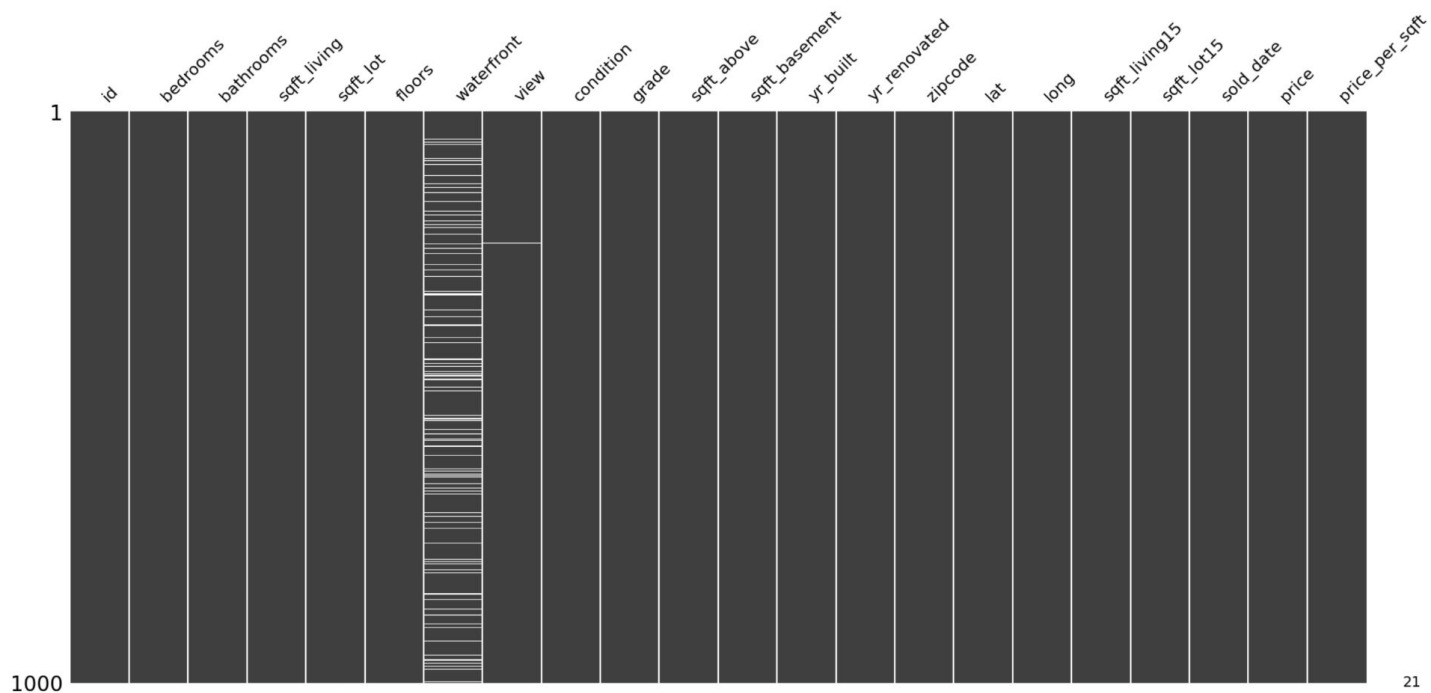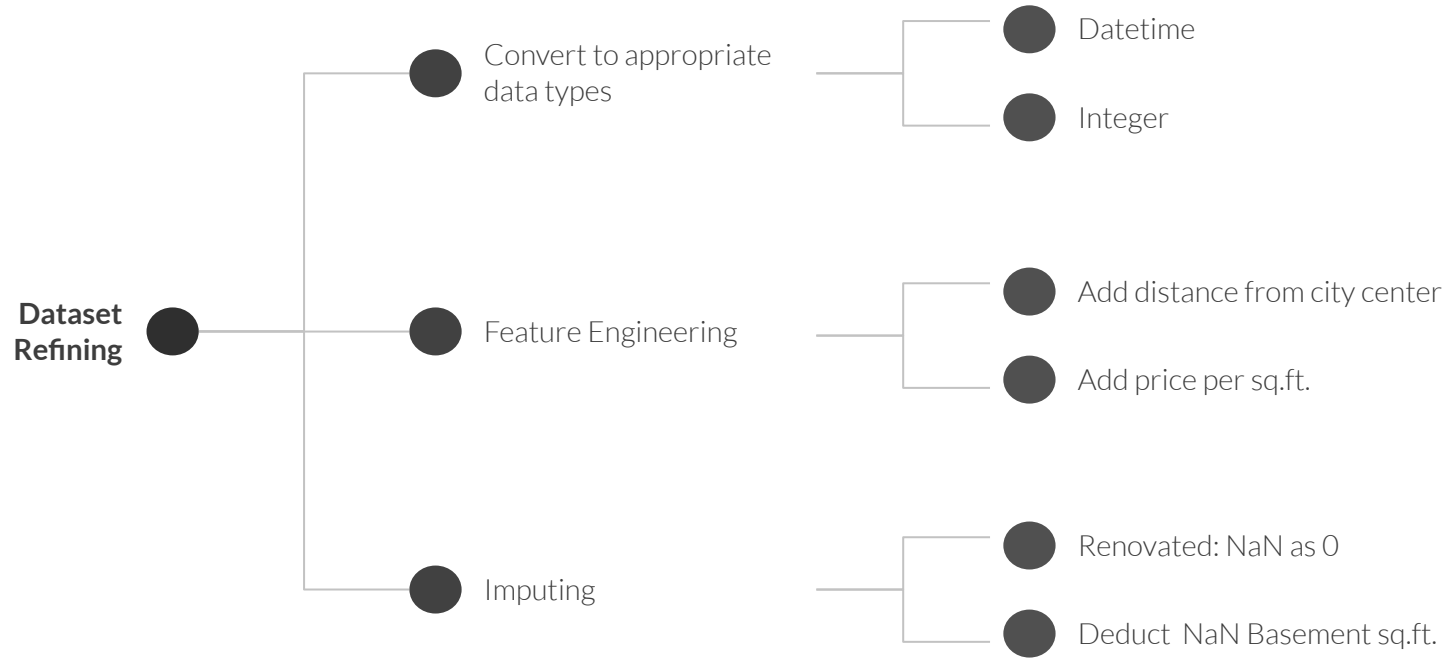- Make costs back  and little profit
- Socially responsible

# Objectives

- Identify poor income areas
- Be within size range of neighborhood
- Define reasonable profit margin

# Dataset

21,597 entries

```
Dataset          Convert to appropriate          Datetime
Refining   ●     data types            ●
                                              ●   Integer

                 ●     Feature Engineering  ●     Add distance from city center

                                              ●   Add price per sq.ft.

                 ●     Imputing              ●     Renovated: NaN as 0

                                              ●   Deduct  NaN Basement sq.ft.
```

Strategic decisions

Exclude View & Waterfront :
- Many missing data
- Weak price- waterfront correlation (numerical vs categorical Point - Biserial correlation = 0.27)
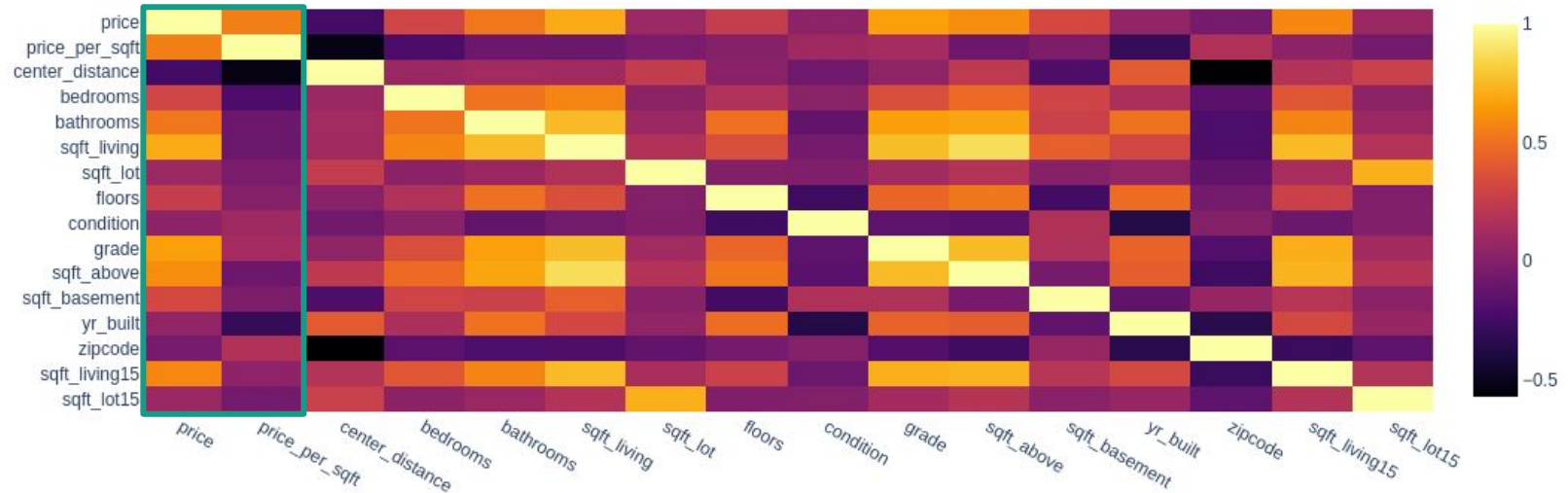- Irrelevant to client

# Hypotheses

Price defined by:

- Condition
- Distance from Seattle center
- Clustering of afluent vs poor areas by zip code
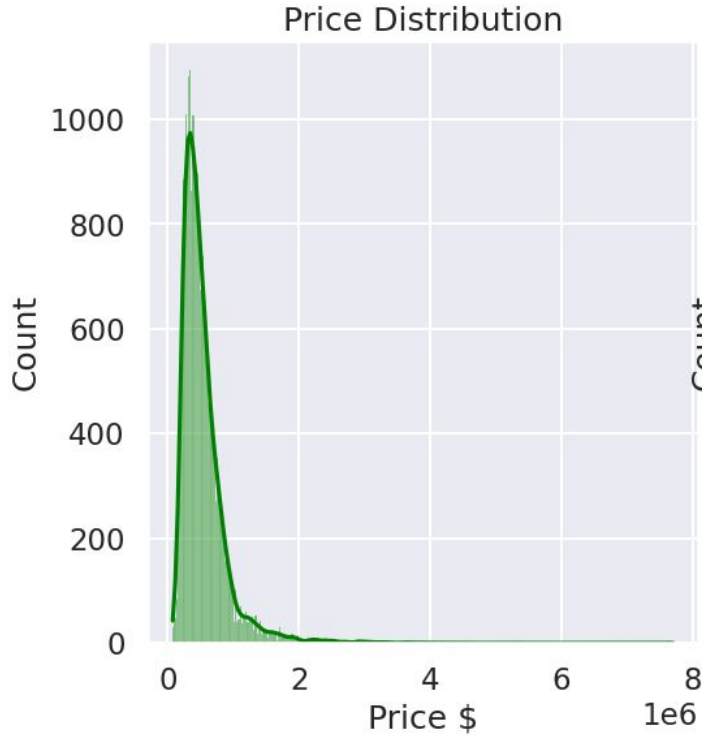
# Price defined by?



Column Correlations

Price correlations
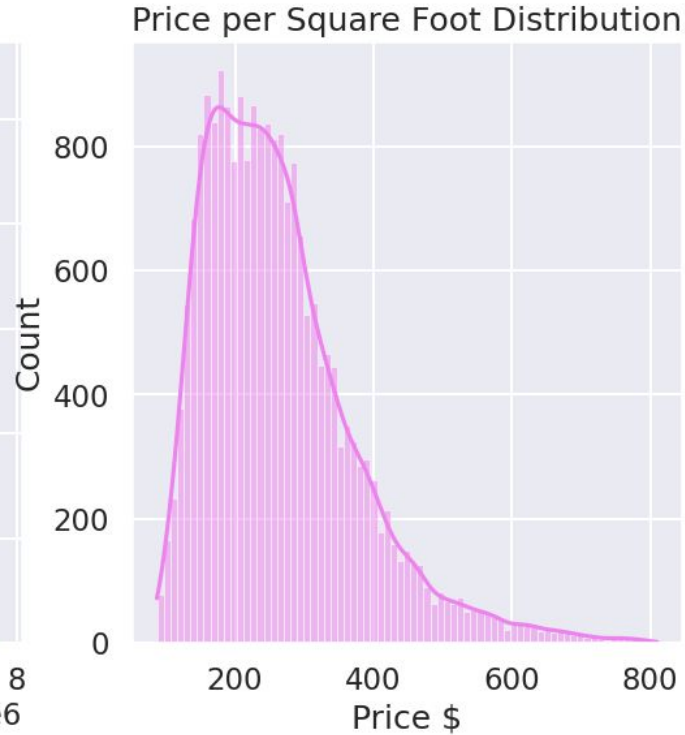bathrooms: 0.526    sqft_living: 0.702    grade: 0.668    sqft_above: 0.605    sqft_living15: 0.585

Price per sq.ft. correlations:
center_distance: 0.523

# Price vs Price per Square Foot

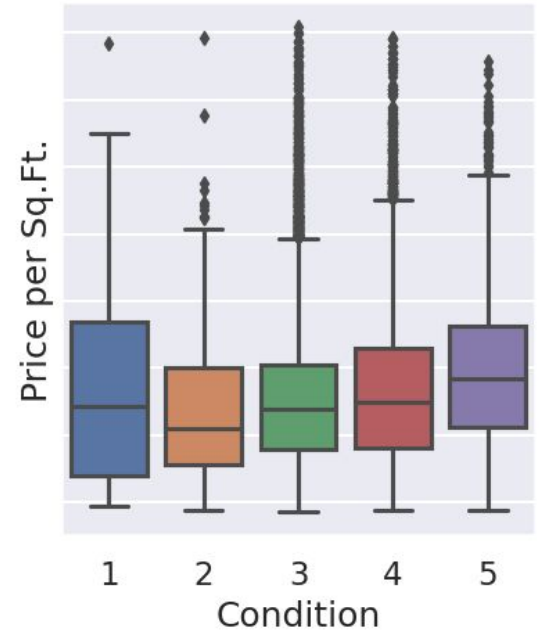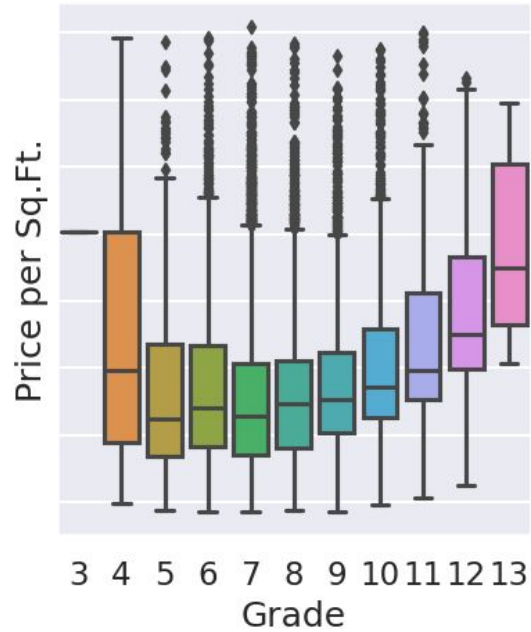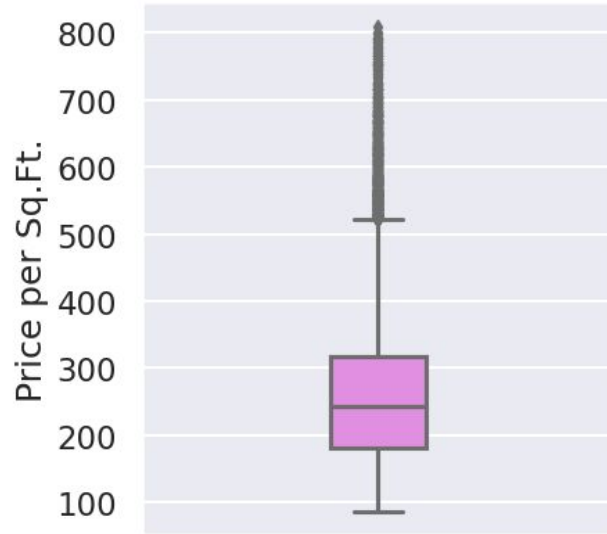

Price Distribution

Price per Square Foot Distribution

Large skew & kurtosis (4,35)
Extreme outliers

Moderate skew & kurtosis (1,2)
Closer to normal distribution

# Price per Square Foot Distributions
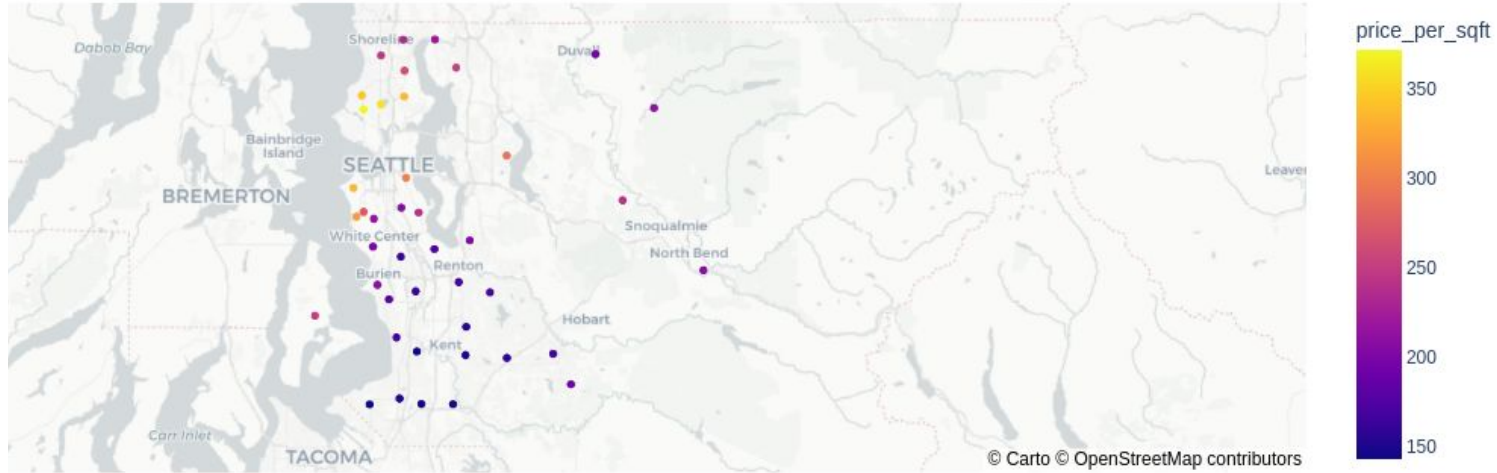
Use Median as a more accurate compensation for outliers

# Zip Code Querying

To identify areas of interest

Aggregate

- Price per Sq.Ft.
- Sq. Ft. Living Area
- Sq. Ft. Neighbours
- Grade

# Areas of interest price scale (Grade 7)



Further focus:
- Low price per sq.ft.
- Within 1.5 hours from city center (urban areas instead of rural)

South of Seattle, till Federal Way and East limits at Hobart.
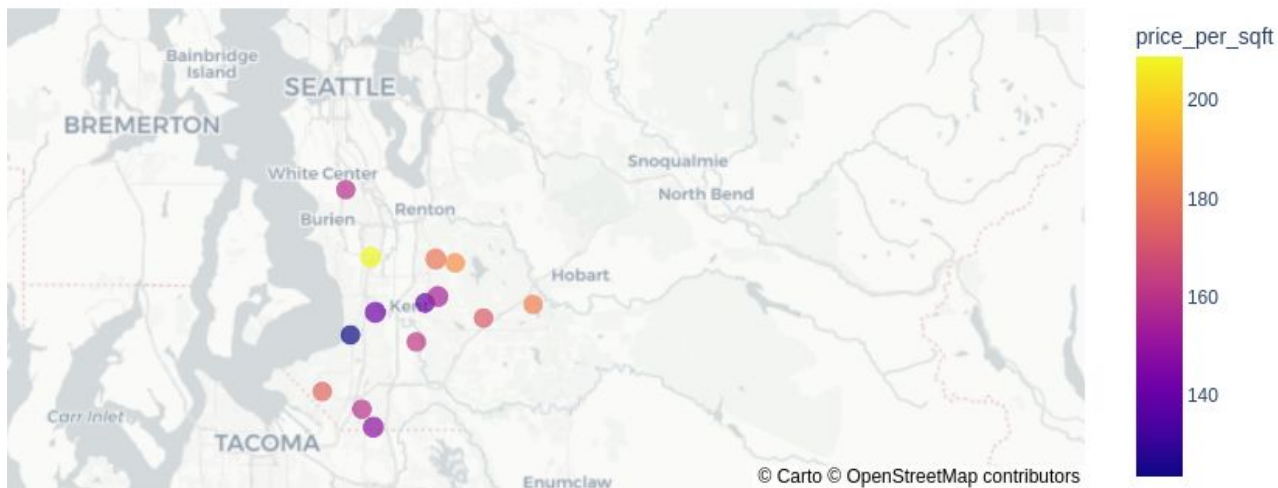
# Recommended listings

Criteria:

Renovation or built within last 20 years | Living area within IQR of neighbors | Price per square foot under the general median
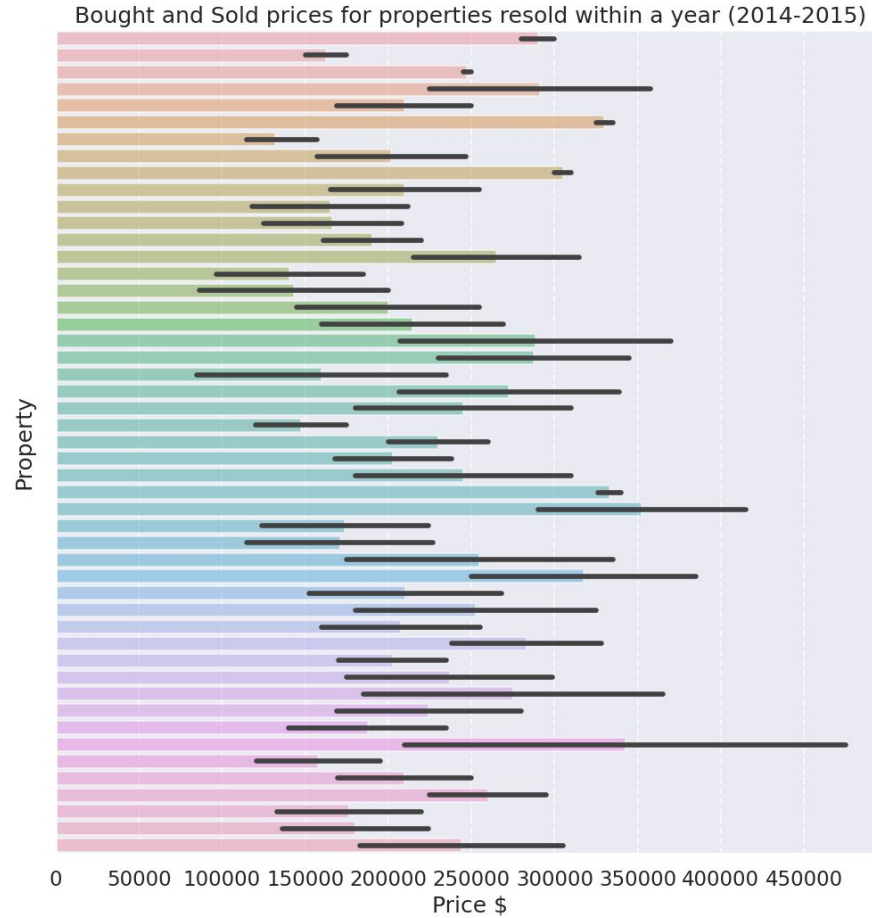
Results:

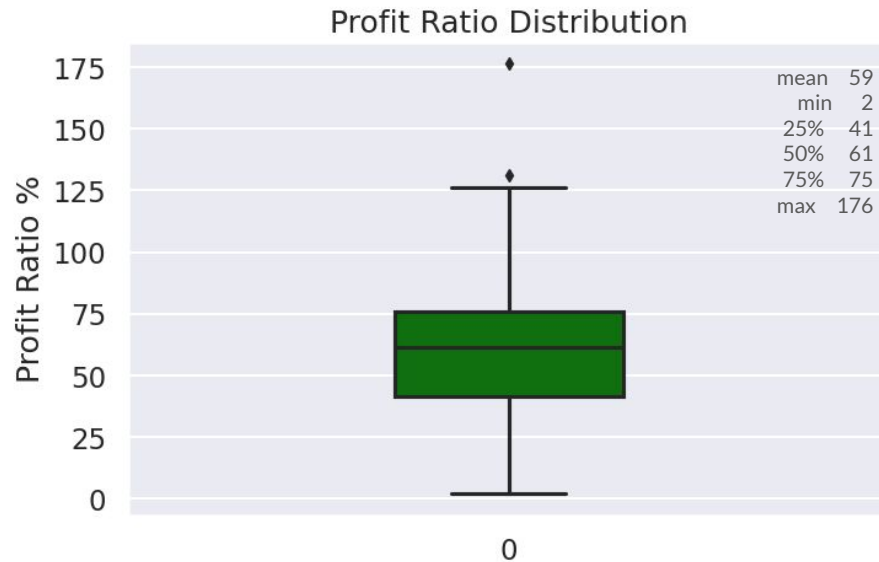436 listings

Map:

14 listings : One random pick per zip code

# Profit ratio?



Bought and Sold prices for properties resold within a year (2014-2015)

# Potential Speculative Market

**41% profit within fair limits**



Profit Ratio Distribution

| | |
|---|---|
| mean | 59 |
| min | 2 |
| 25% | 41 |
| 50% | 61 |
| 75% | 75 |
| max | 176 |

# Thank you!