Leo's House Hunt

Freguenet
**Dan Hong**
Derek Denning
Mike George

Motivation and Summary

Leo is becoming increasingly frustrated by our basic github questions and ready to retire from an illustrious career in data analytics. He's intrigued by Washington DC real estate and looking for where best to buy?

Leo is looking for a turn-key property at a modest price point.

Objective: Determine where and what Leo should buy for an optimal return in 10 years assuming current market conditions remain stable

Question: Is there a clear quadrant, ward, and house type that is poised to deliver superior investment returns over the next 10 years?
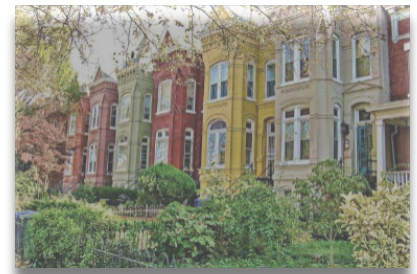
Questions & Data

We used the D.C. Residential Properties dataset from Kaggle as our sole dataset

We needed key data parameters that would provide insight into a quadrant and wards investment prospects. Some of the data was numerical (sale px, transaction count) while others were qualitative (quad, ward, condition). Some of the key functions we used to cut our data included .groupby, .mean, .count, .agg, .loc. Important parameters in our data include:

- Price
- Sale Date
- Location (quad/ward)

Year built
Bedroom/bath
Remodel Year

Building type
Square feet

Data Complication

Sale Date: was a key data point that the project was predicated on, and needed to be extracted from a string

Key Insights:

- Upwards of half of the 158,000 records didn't include price data

- Commercial properties included in the "Residential" dataset severely skewed our findings and needed to be removed.

- Data prior to 1992 and after 2017 was too sparse to draw any meaningful conclusion and needed to be removed
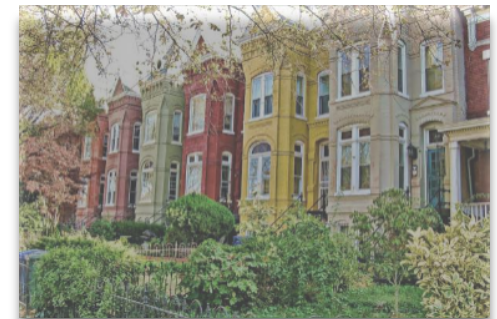
When all was said and done, we were left with about a third of the original data ~58,000 rows

DC Housing by Quadrant

Avg Sales Price by Quad & Year
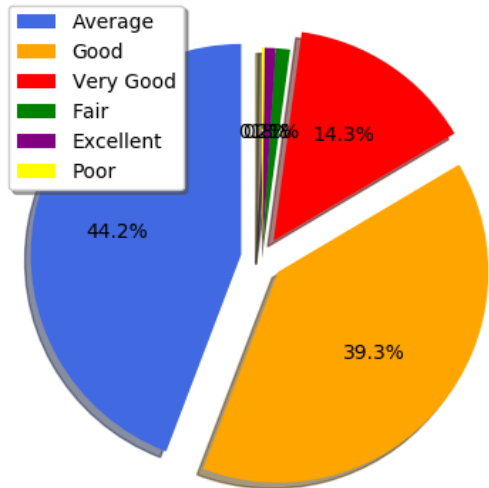

Number of Sales by Quad & Year
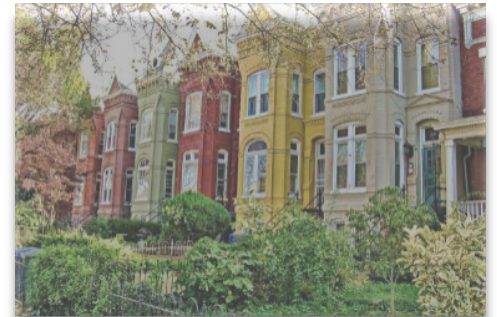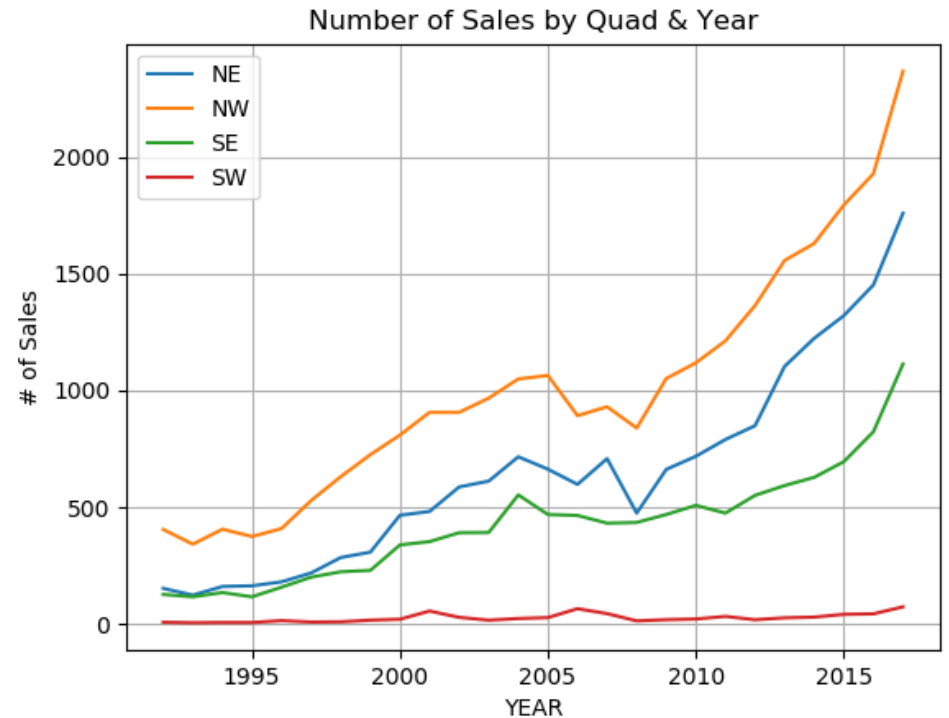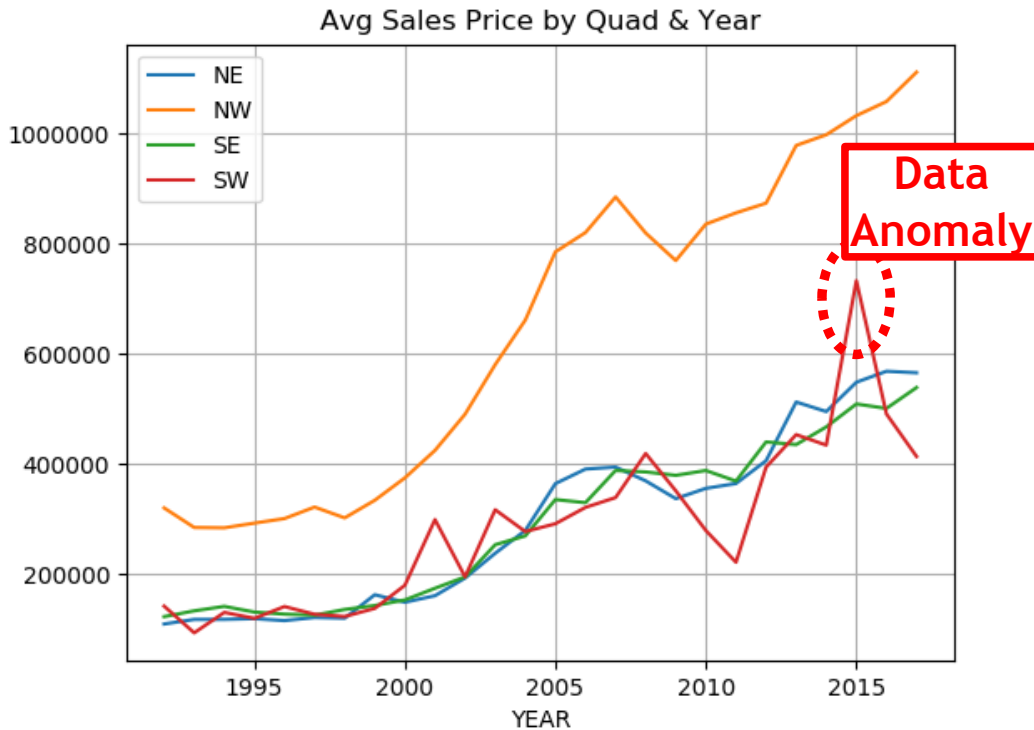
Finding: In addition to consistently having the highest average sale price by a wide margin, Northwest also has the highest number of transactions. Currently, Southwest has the lowest average selling price with the fewest quantity of homes sold
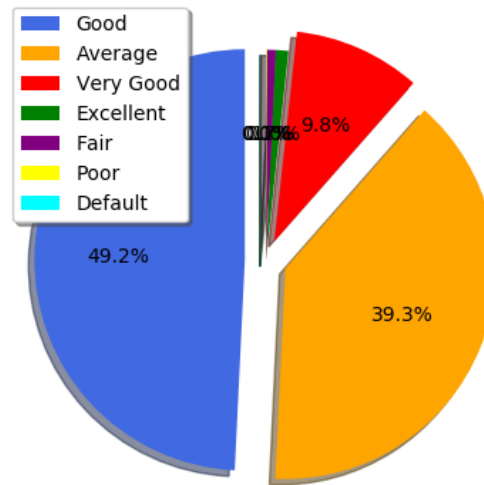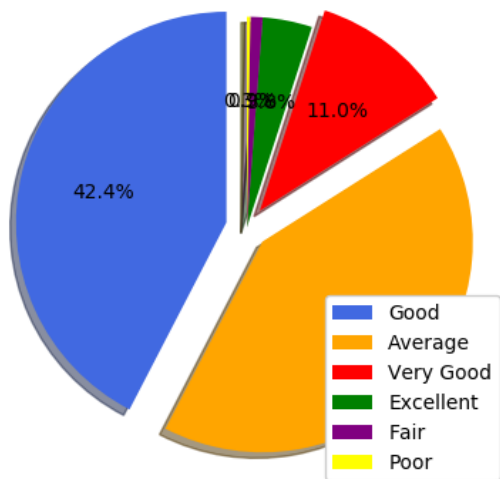
House Conditions by Quadrant
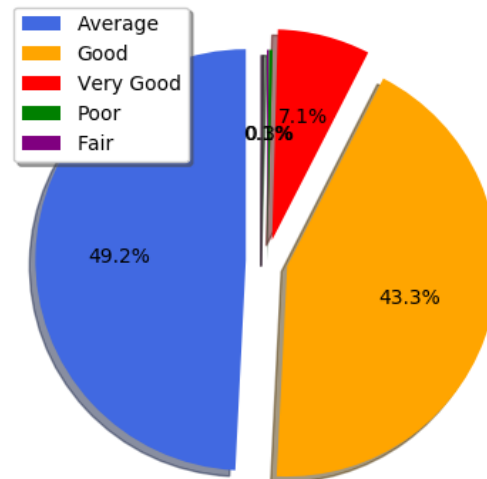
Conditions of Homes Sold in NE

- Average
- Good
- Very Good
- Fair
- Excellent
- Poor

44.2%
14.3%
39.3%

Conditions of Homes Sold in NW

- Good
- Average
- Very Good
- Excellent
- Fair
- Poor
- Default

49.2%
9.8%
39.3%

Conditions of Homes Sold in SE

- Good
- Average
- Very Good
- Excellent
- Fair
- Poor

42.4%
11.0%

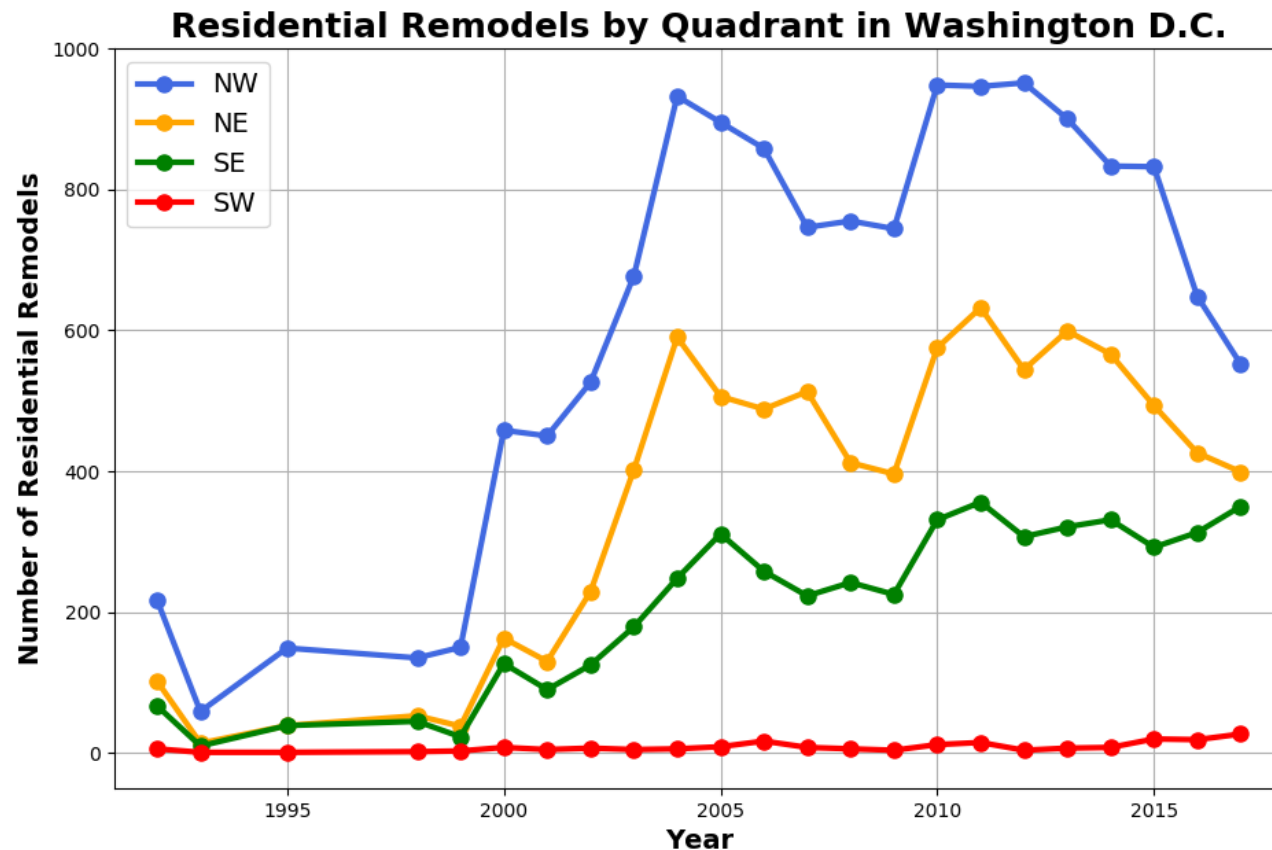### Conditions of Homes Sold in SW

- Average
- Good
- Very Good
- Poor
- Fair

49.2%
7.1%
43.3%

Finding: There is not a wide dispersion of home condition (quality) across quadrants allowing us to conclude any difference in price across quadrants isn't being driven by house condition
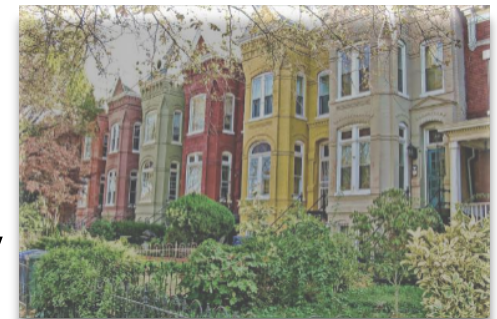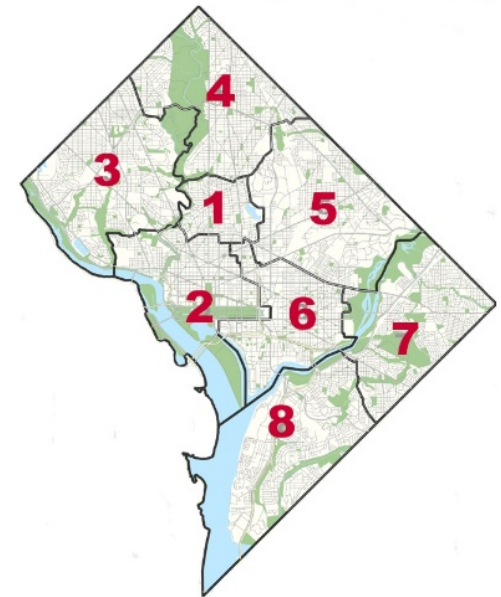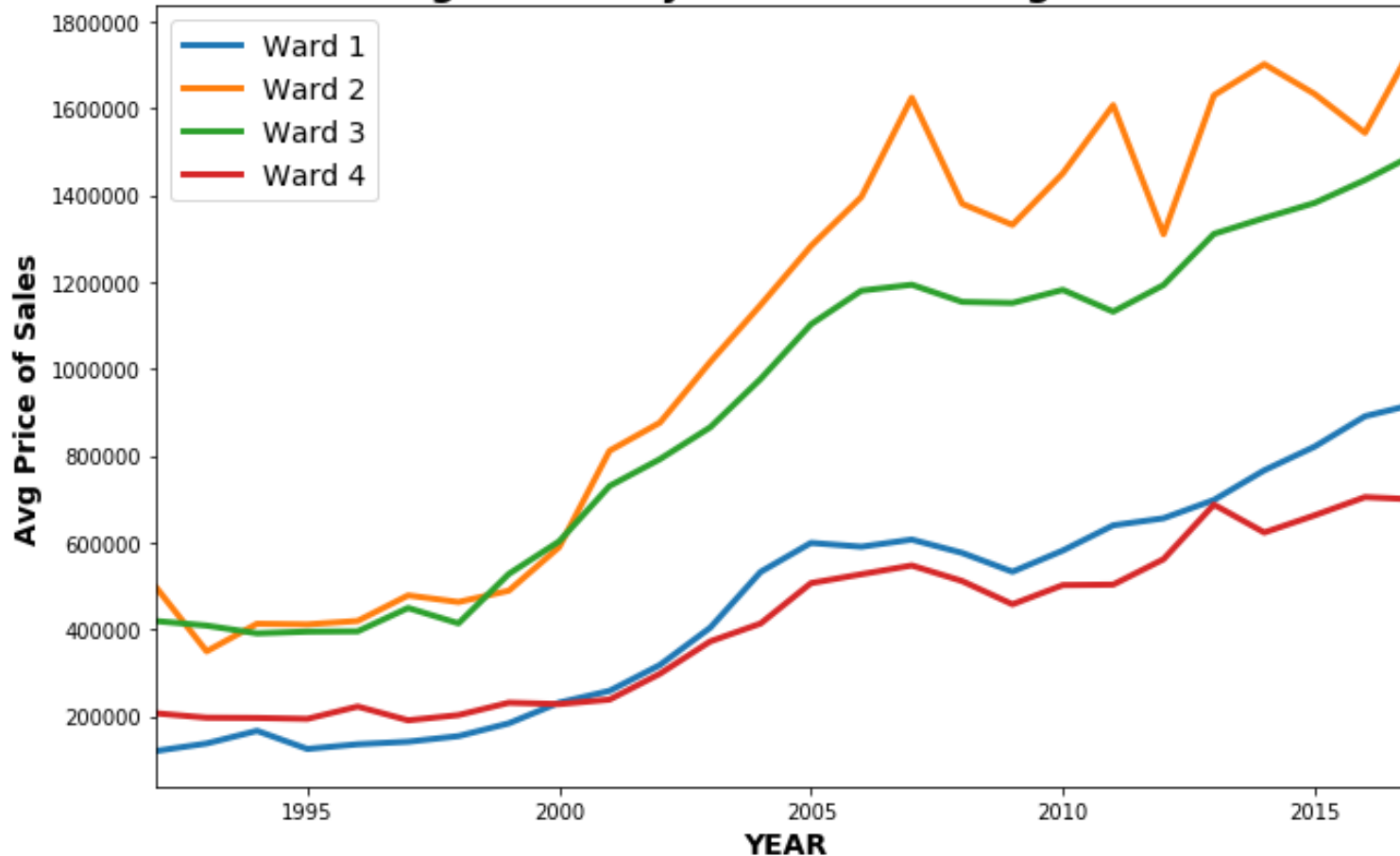
Remodels by Quadrant

Finding: Through time, more homes have been remodeled in Northwest while a relative few are remodeled in Southwest each year
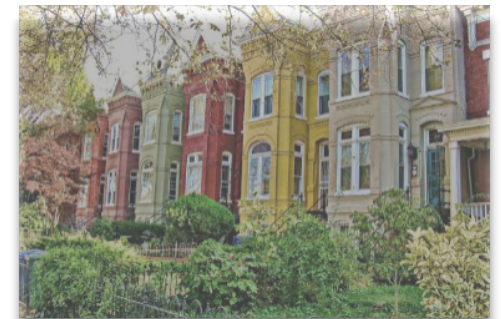
Conclusion: Northwest is a compelling opportunity

Sale Price by Northwest Wards


Avg Sale Px by NW Ward Through Time

Legend: Ward 1, Ward 2, Ward 3, Ward 4

Y-axis: Avg Price of Sales (200000 to 1800000)
X-axis: YEAR (1995 to 2015)

Map of DC wards showing wards 1–8

Finding: Within Northwest, Ward 2 has the highest, while Ward 4 has the lowest average selling price

Remodels by Northwest Wards

Residential Remodels by Ward in Washington D.C.
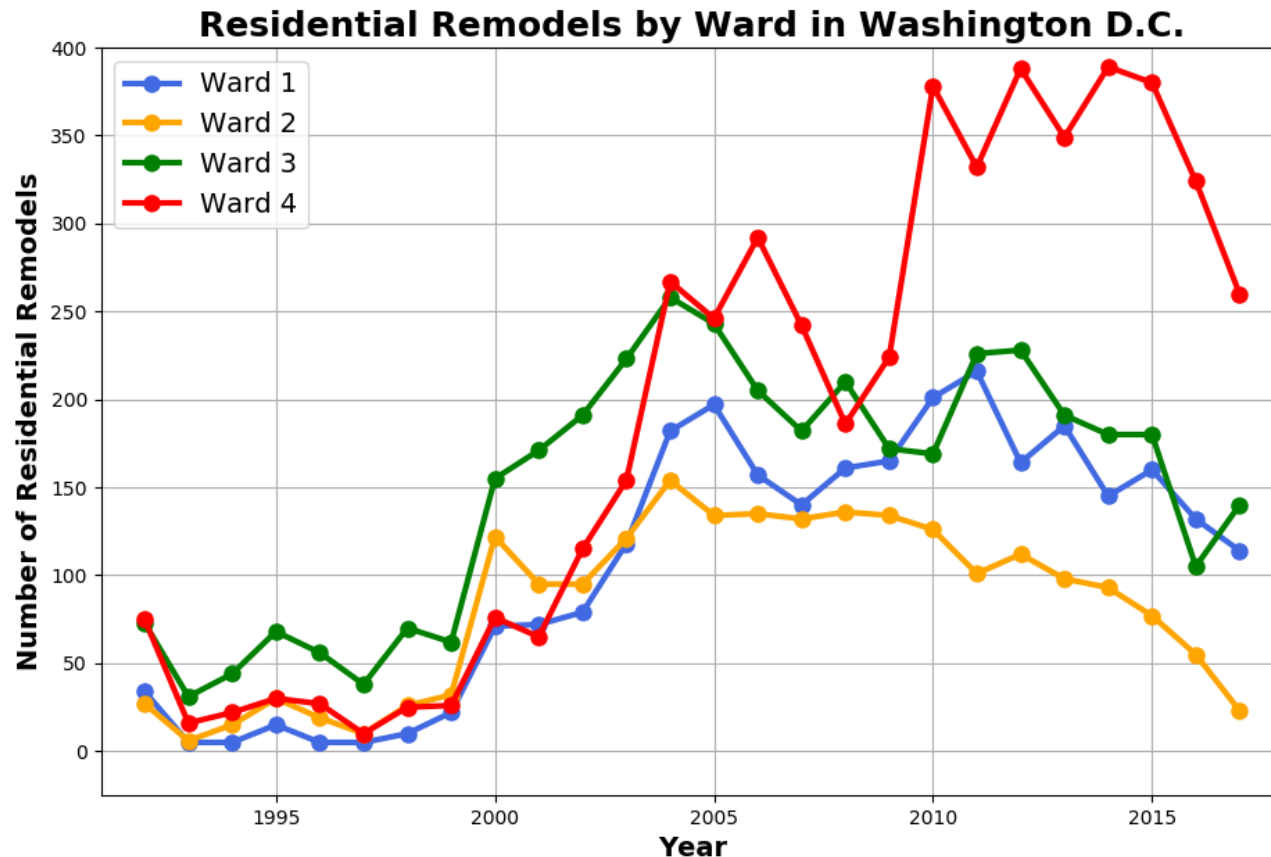
Finding: Since 2009, homes in Ward 4 are remodeled with greater frequency than any other ward in Northwest
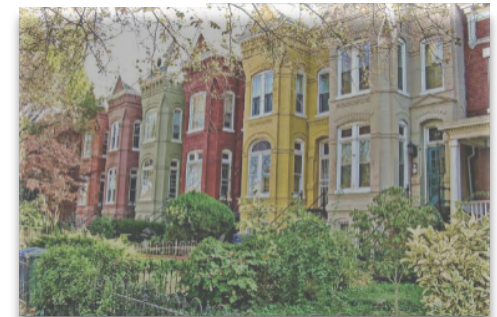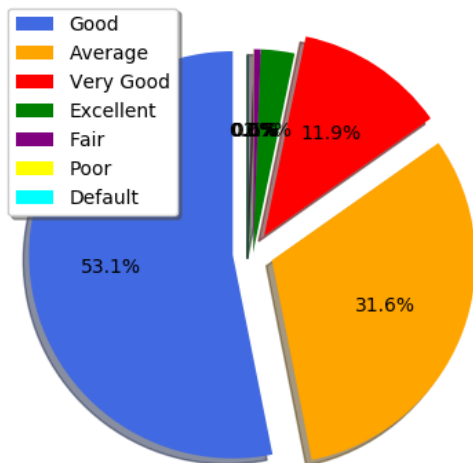
Home Conditions by Northwest Wards
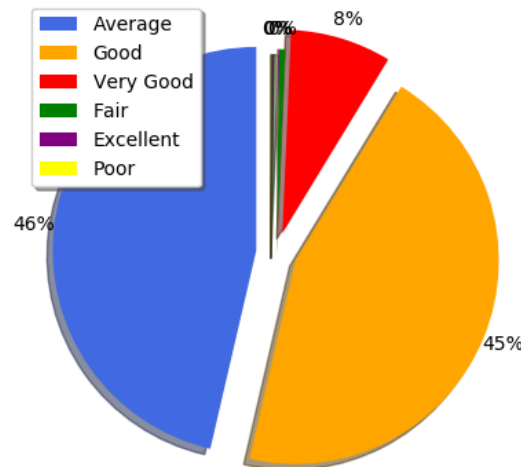
Conditions of Homes Sold in NW Ward 1

Legend:
- Good
- Average
- Very Good
- Fair
- Poor
- Excellent

52.1%
43.7%
3.5%
0.6%

Conditions of Homes Sold in NW Ward 2

Legend:
- Good
- Average
- Very Good
- Excellent
- Fair
- Poor
- Default

51.9%
27.8%
18.6%
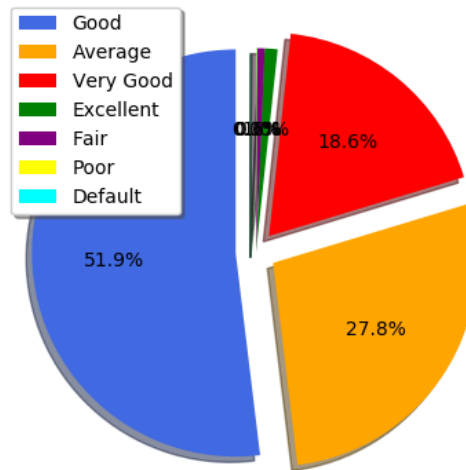
Conditions of Homes Sold in NW Ward 3

Legend:
- Good
- Average
- Very Good
- Excellent
- Fair
- Poor
- Default

53.1%
31.6%
11.9%

Conditions of Homes Sold in NW Ward 4

Legend:
- Average
- Good
- Very Good
- Fair
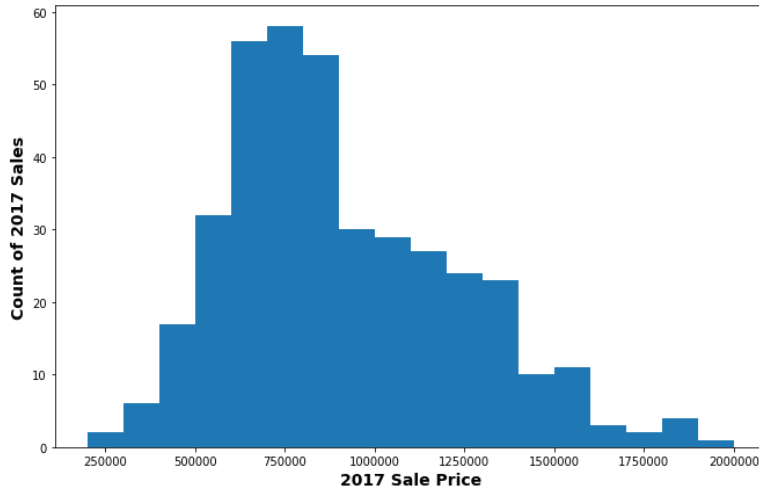- Excellent
- Poor

46%
45%
8%
0%

Finding: There is not a wide dispersion of home condition (quality) across wards allowing us conclude the perceived value of Ward 4 is not due to poor quality homes
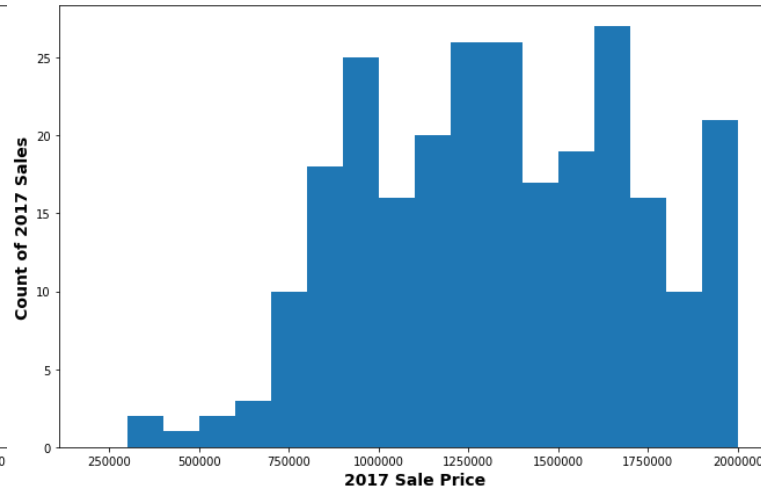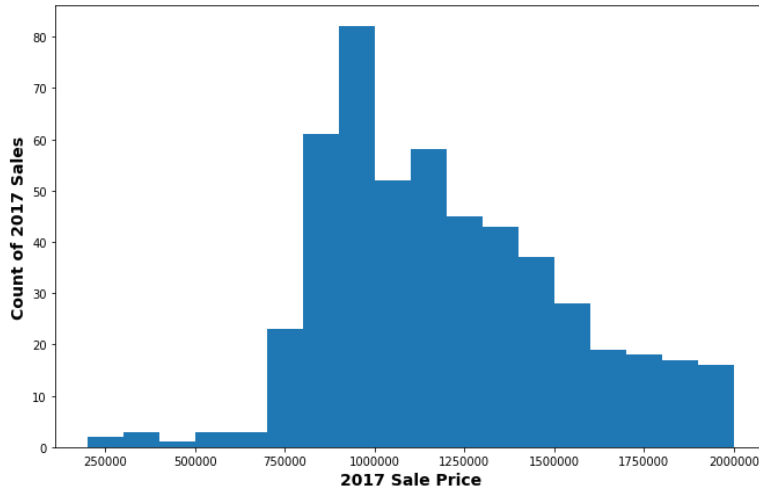
Sale Count by Price in Northwest Wards

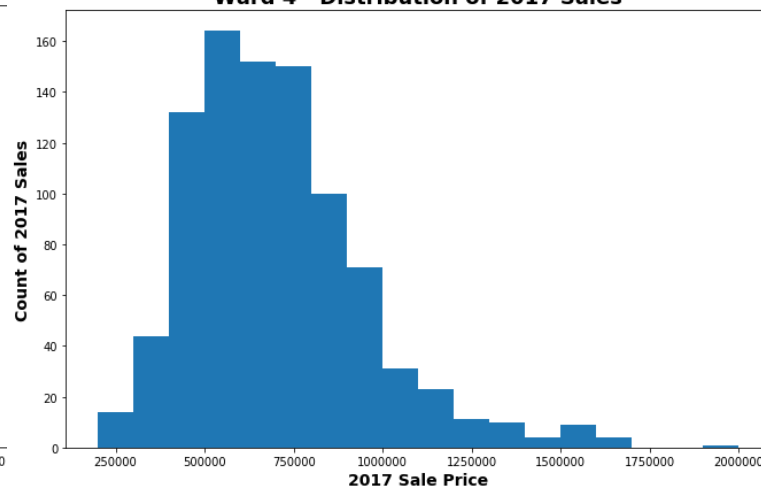**Finding:** In 2017, Ward 4 had the greatest quantity of homes within a more modest price point
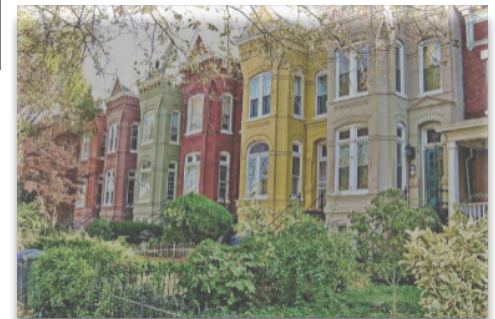
Sq. Footage by Northwest Wards

Finding: In 2017, Ward 4 had the greatest quantity of modest sized homes within a reasonable price point

Ward 1 - 2017 Price vs. Sq. Ft

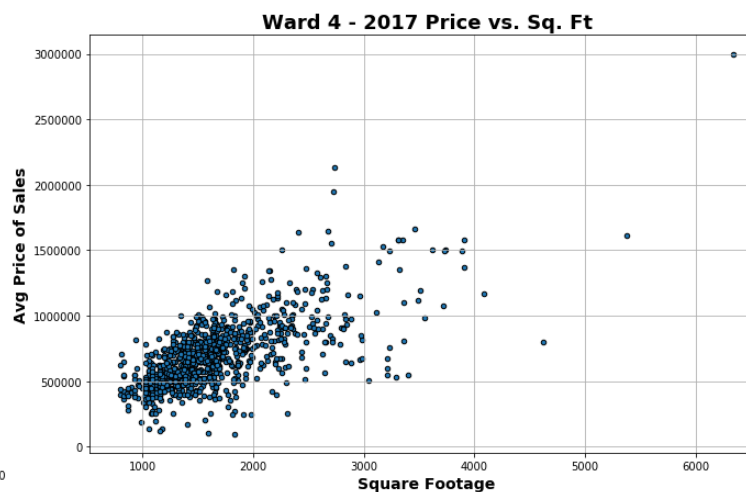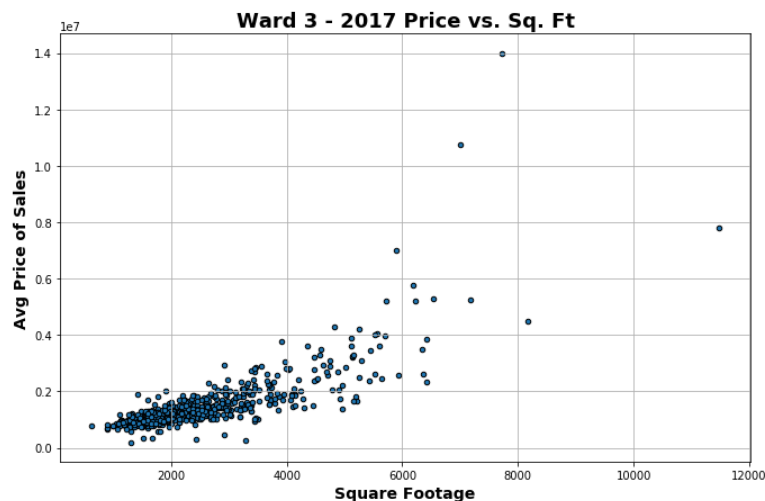Ward 2 - 2017 Price vs. Sq. Ft

Ward 3 - 2017 Price vs. Sq. Ft

Ward 4 - 2017 Price vs. Sq. Ft
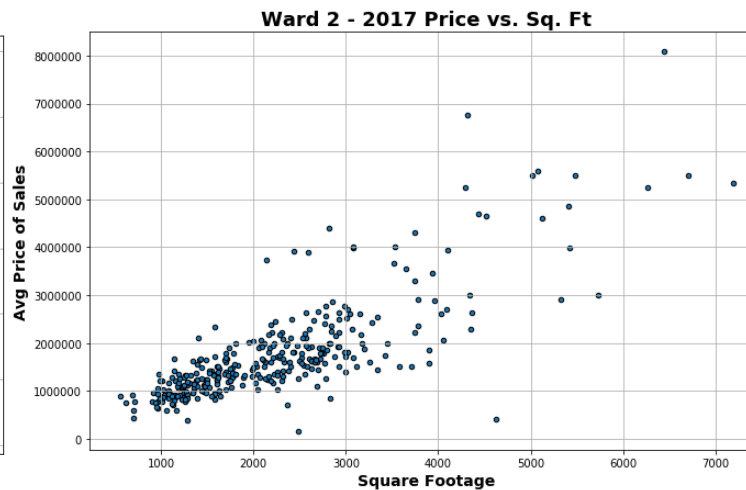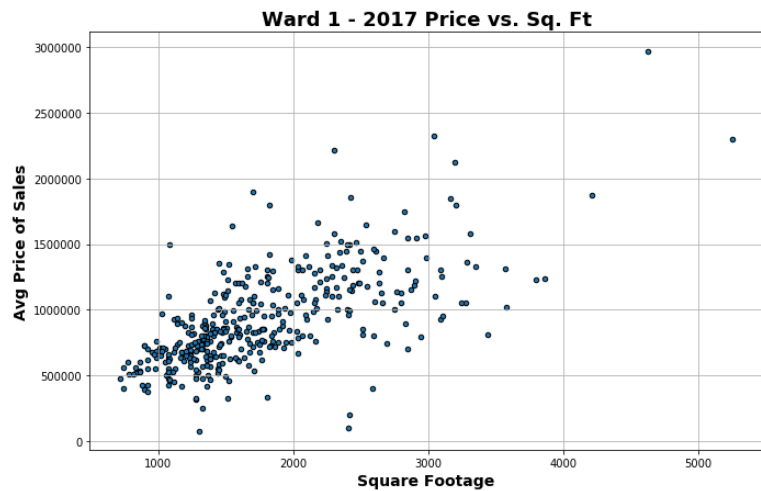
Conclusion: Ward 4 offers greatest value within the most expensive quadrant (Northwest) serving as a natural hedge against extreme price fluctuations

Additionally, the price pt. is not a function of house condition
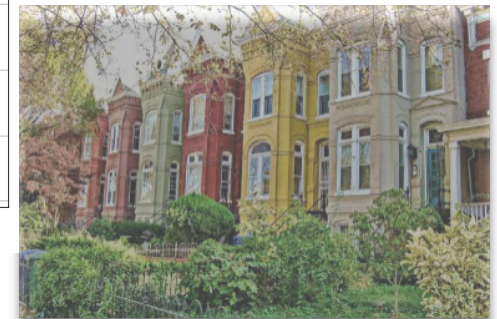
What to buy in Ward 4?

Cumulative Price Change (Ward 4, 2007-2017)

Finding: Over the last decade, a 2 bedroom, 1 bath has delivered far superior returns than any other bed/bath combination

Extrapolating this trend forward, we anticipate valuations of 2bed/1bath to coalesce around $468k this year, and $680k by 2029
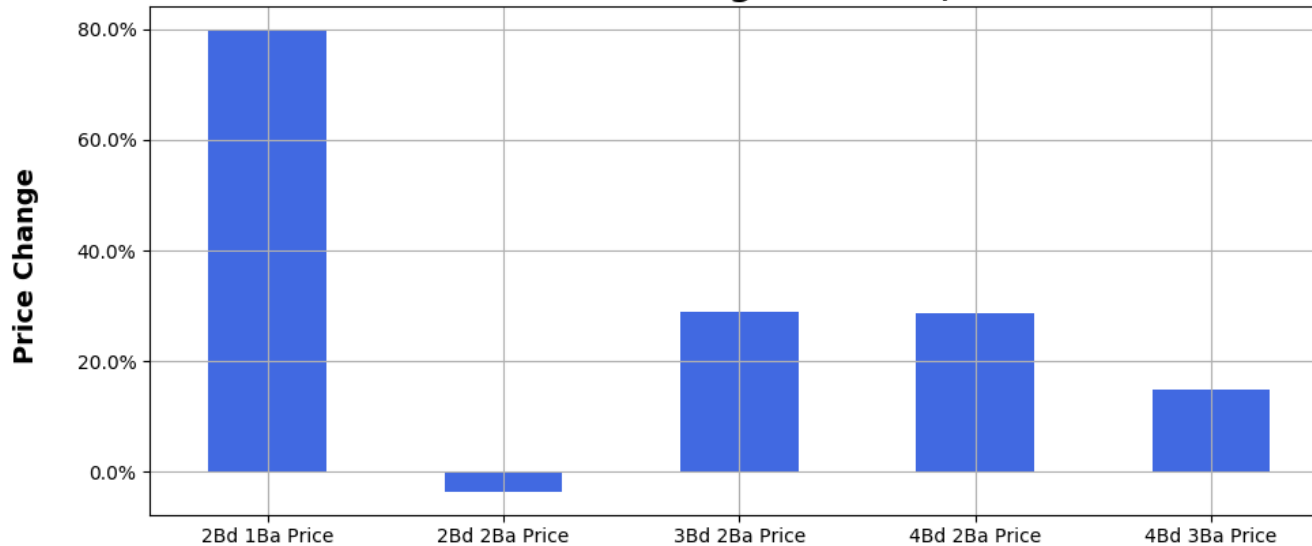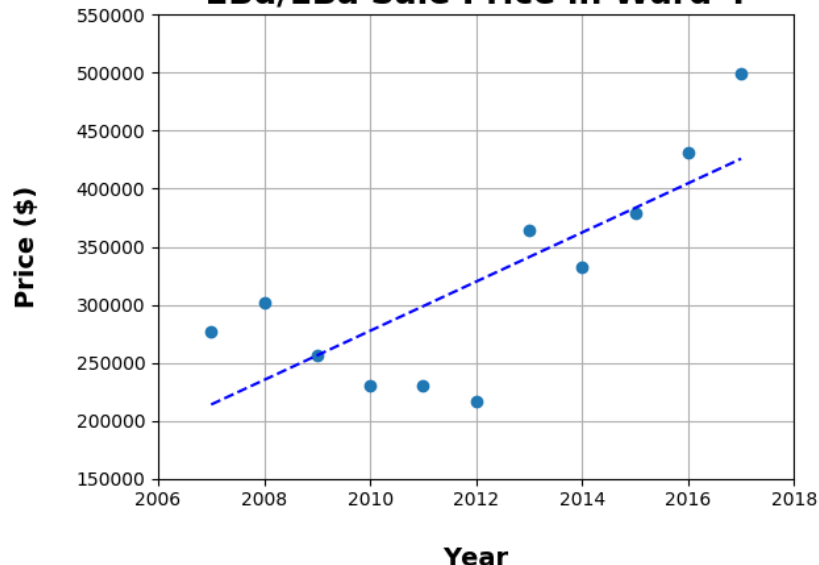
2Bd/1Ba Sale Price in Ward 4


```
In [31]: year1 = 2019.0
         year1_price = s * year1 + yint
         print("The average sale price in ward 4 in 2019 will be " +
               str(s * year1 + yint) + ".")

         year2 = 2029.0
         year2_price = s * year2 + yint
         print("The average sale price in ward 4 in 2029 will be " +
               str(s * year2 + yint) + ".")

         profit = year2_price - year1_price
         print(profit)

The average sale price in ward 4 in 2019 will be 468243.60238095373.
The average sale price in ward 4 in 2029 will be 680144.7625541091.
211901.16017315537
```

Difficulty and Resolutions

After an initial glance of the data set, the group had a general idea of our proposed objective. However, with over 150,000 rows and over 20 columns of data, we were not sure whether the data had the capacity or which data to use to answer our objectives.

There were several difficulties in data cleaning between group members (such as extracting year our from a full date) and coordinating the generation of useful plots between members was challenging. This resulted in making several plots that were of no use to our defined objective.

To remedy these challenges, we scoped the data to be used in our analyses to quadrant and ward, which made our analysis much more manageable.

The group also had frequent communication (class, slack, checking out branches) to know what everyone was working on, and ensure that all of us were marching towards the goal of answering our objective and having presentable data analysis.

Additional Questions to Explore

It would have been interesting to look at the impact additional parameters such as acreage, year built, # of fireplaces have on price

We would have liked to overlay additional datasets such as crime and walkability scores to ensure Leo remains safe and gets his steps in.