Section 1 Project

Analysis of the King County housing market

FLATIRON SCHOOL

Executive summary

- The business case
- What the model does and doesn't do
- The model in action
- Possible next steps

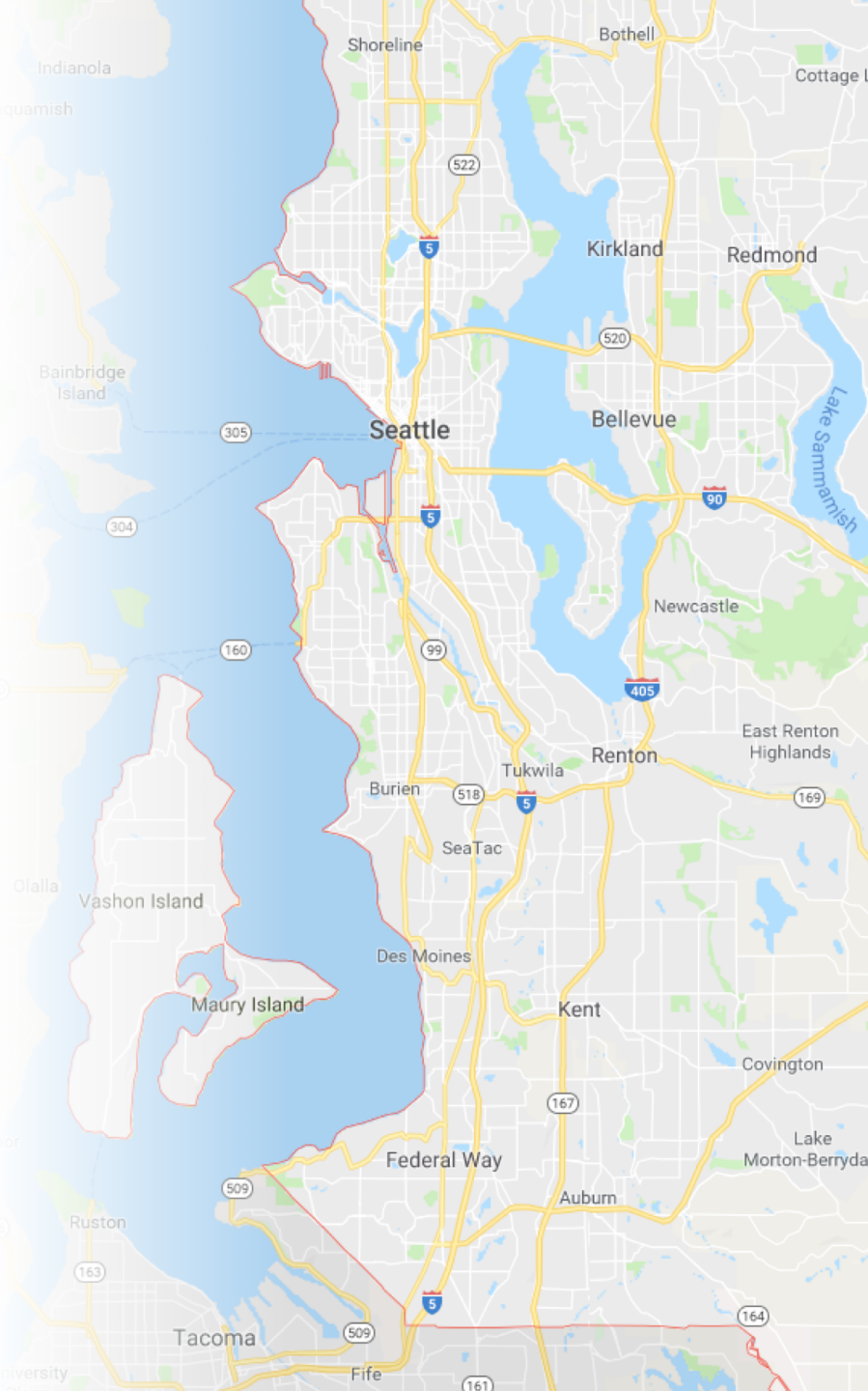
Appendix contains additional analysis

The business case

The King County housing market is worth over $11bn annually, with 60 properties being sold every day.

Being able to predict a property's value based on its characteristics allows a real estate consultancy to:

- Identify and invest in undervalued property
- Evaluate the profitability of renovation work done to properties
- Ensure that property is not sold below its intrinsic value



What the model does and doesn't do

The presented model:

- Predicts the sale price of a property, based on its characteristics
- Considers properties in Western King County
- Is appropriate for houses worth between $100k and $1m, and fewer than 6 bedrooms
- Excludes any houses designated Grade 4 or below[1]

1) Grades defined by King County Government. Properties Grade 4 or below are those that do not meet current building codes

How the model works

Bedrooms

- 1 Bedrooms: +$0.0
- 2 Bedrooms: +$58k
- 3 Bedrooms: +$92k
- 4 Bedrooms: +$116k
- 5 Bedrooms: +$135k

Grade[1]

If property is of Grade:

- 5. -$275k
- 6. -$119k
- 7. +$12k
- 8. +$126k
- 9. +$227k
- 10. +$317k
- 11. +$398k
- 12. +$473k

Size of lot[2]

1% increase in lot size results in $312 increase in property price.

Example lot sizes:

- 10k ft^2. +$288k
- 20k ft^2 . +$3097k
- 30k ft^2 . +$322k
- 40k ft^2 . +$331k
- 50k ft^2 . +$338k

Location[3]

If property in:

- Area A. +$52k
- Area B. +$66k
- Area C. +$74k
- Area D. -$125k
- Area E. -$87k
- Area F. -$160k
- Area G. +$51k
- Area H. -$117k

Build year

If property built between:

- 1899-1919. +$93k
- 1920-1939. +$80k
- 1940-1959. +$0k
- 1960-1979. -$49k
- 1980-1999. -$29k

Property condition[4]

- Condition 2. -$122k
- Condition 3. -$56k
- Condition 4. -$32k
- Condition 5. +3k

Date of transaction

If property sold in:

- Q1. +$12.4k
- Q2. +$11.8k

Waterside

If property is waterside:
+$271k

Basement

If property has a basement:
+$29k

1) Property grade defined by King County government. Classification by construction quality which refers to the types of materials used and the quality of workmanship.
2) Lot size has a logarithmic relationship with property price, with each additional square foot adding marginally fewer dollars to the property value.
3) Locations given as * A - Seattle, Shoreline, Lake Forest Park * B - Kirkland, Kenmore, Bothell, Redmond, Woodinville * C - Bellevue, Mercer Island, Newcastle * D - Renton, Kent * E - Burien, Normandy Park, Des Moines, SeaTac, Tukwila, Vashon Island * F - Federal Way, Auburn, Algona, Milton, Pacific * G - Sammamish, Issaquah, Carnation, Duvall * H - Covington, Maple Valley, Black Diamond, Enumclaw
4) Condition defined by King County government, relative to age and grade: 2 = Fair- Badly worn. Much repair needed. 3 = Average- Some evidence of deferred maintenance and normal obsolescence with age in that a few minor repairs are needed, 4 = Good- No obvious maintenance required but neither is everything new. 5= Very Good- All items well maintained, many having been overhauled and repaired as they have shown signs of wear.

The model in action 1 – Normandy Park Waterfront

Bedrooms	4
Grade	8
Lot size	21,500 ft^2
Zipcode	98166 (Area E)
Build year	1979

Condition	3
Sale date	April
Waterside	Yes
Basement	No

Current property value:
$787,644

The case: The homeowners want to build a basement to improve their property's value, and need our consultancy's advice!

One builder will construct a basement for $20k.

A second builder has offered to do the same basement, but also do renovation work that would improve the house's condition to 4, all for $50k.

Which company should they go with?

Builder 1 house value = $816,897
... after builder fees = $796,897 ✓

Builder 2 house value = $841,663
... after builder fees = $791,663

The model in action 2 – Vashon Island Cabin

Bedrooms	1
Grade	6
Lot size	11,200 ft^2
Location	98070 (Area E)
Build year	1983

Condition	2
Sale date	February
Waterside	No
Basement	No

Current property value:
$119,817

The case: Our company has the option to buy the property from its current owner for $130k.

We know that the neighbour would be willing to sell us 5,000ft^2 of adjacent land for $10k.

We estimate that doing the renovation work to bring the property to condition 3 would cost $20k.

House value with additional land = $131,346
Given the initial $130k purchase, the most we would pay for the extra land is $1,300

House value with condition 3 = $155,346
Given the initial $130k purchase, the most we would pay for the renovation is $15,300

The model in action 3 – Belleview Family Home

Bedrooms	5
Grade	10
Lot size	32,500 ft^2
Zipcode	98006 (Area C)
Build year	1965

Condition	5
Sale date	August
Waterside	No
Basement	Yes

Current property value:
$884,431

The case: The occupants are downsizing, and want to use the money to buy a 2 bedroom house in Duvall[1], and use any remaining money to buy a 2 bed house for their daughter in Seattle.

They won't to downgrade on their current condition but can live with grade 8 and no basement. What kind of property could the daughter get?

Estimate Duvall house cost = $580k
Leaving $304k for daughter's property

By using the model, we can calculate that $304k will buy a 2 bedroom, 80s/90s property in Seattle, though it will be grade 7, condition 4, and only be 2,400 ft^2

1) Average lot size of a 2-bed property in Duvall (zip code 98019) is 54,480 ft^2

Possible next steps

To further improve the model, there are other things that could be investigated:

- Create more versions of the dataset, using different decisions and assumptions during the initial data munging
- Calculated more KPIs from existing dataset in order to identify more outliers
- Use more granular zip code sorting
- Go beyond linear regression, and investigate non-parametric models to improve predictive capabilities

Thank you for your time

Further detail on the model, and additional analysis of the dataset is included in further pages

Full equation for model, and analysis of model fit

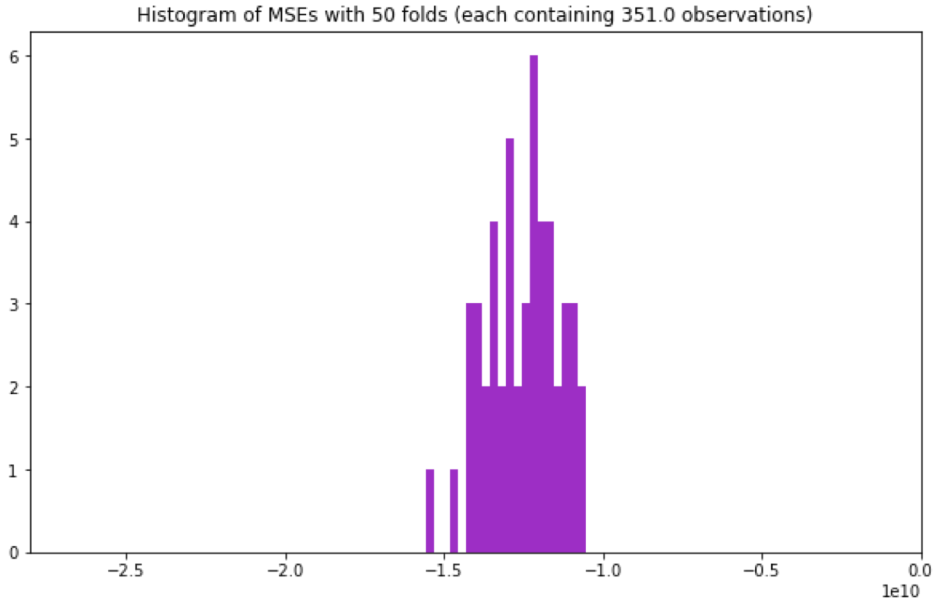
Estimated house price =

+ $83,766.8 * log(#bedrooms)
+ $85,4318.3 * log(grade)
+ $31,234.3 * log(lot size in ft^2)
+ $270,635.4 if house is waterside
+ $29252.4 if house has basement
+ If condition 2: -$35,528.3, if condition 4: $24,766.1, if condition 5: $53,317.9,
+ If sale in Q1: +$12,423.4, if sale in Q2: +$11,812.4,
+ If house built in 1899-1920: +$93,013.9, 1920-1940: +$79,781.1, 1960-1980: -$48,900.0, $1980-2000: -$28,907.6,
+ If house in zip area A: +$51,516.3, zip area B: +$65,576.7, zip area C: +$74,431.3, zip area D: -$124,510.2, zip area E: -$87,154.4, zip area F: -$159,891.6, zip area G: +$50,994.7, zip area H: -$117,246.4
+ -$1,650,121 (constant term)

Model has: $R^2 = 0.67$, and adjusted $R^2 = 0.67$

The model thus explains over two thirds of the dataset's variation from the mean price

The mean squared errors, using a 50-fold cross validation method, are within a relatively tight range (see histogram). This suggests that the model is not 'over-fitted' to the data, and is could be used with new observations.



Map of King County, with areas demarcated

A. Seattle, Shoreline, Lake Forest Park

B. Kirkland, Kenmore, Bothell, Redmond, Woodinville

C. Bellevue, Mercer Island, Newcastle

D. Renton, Kent

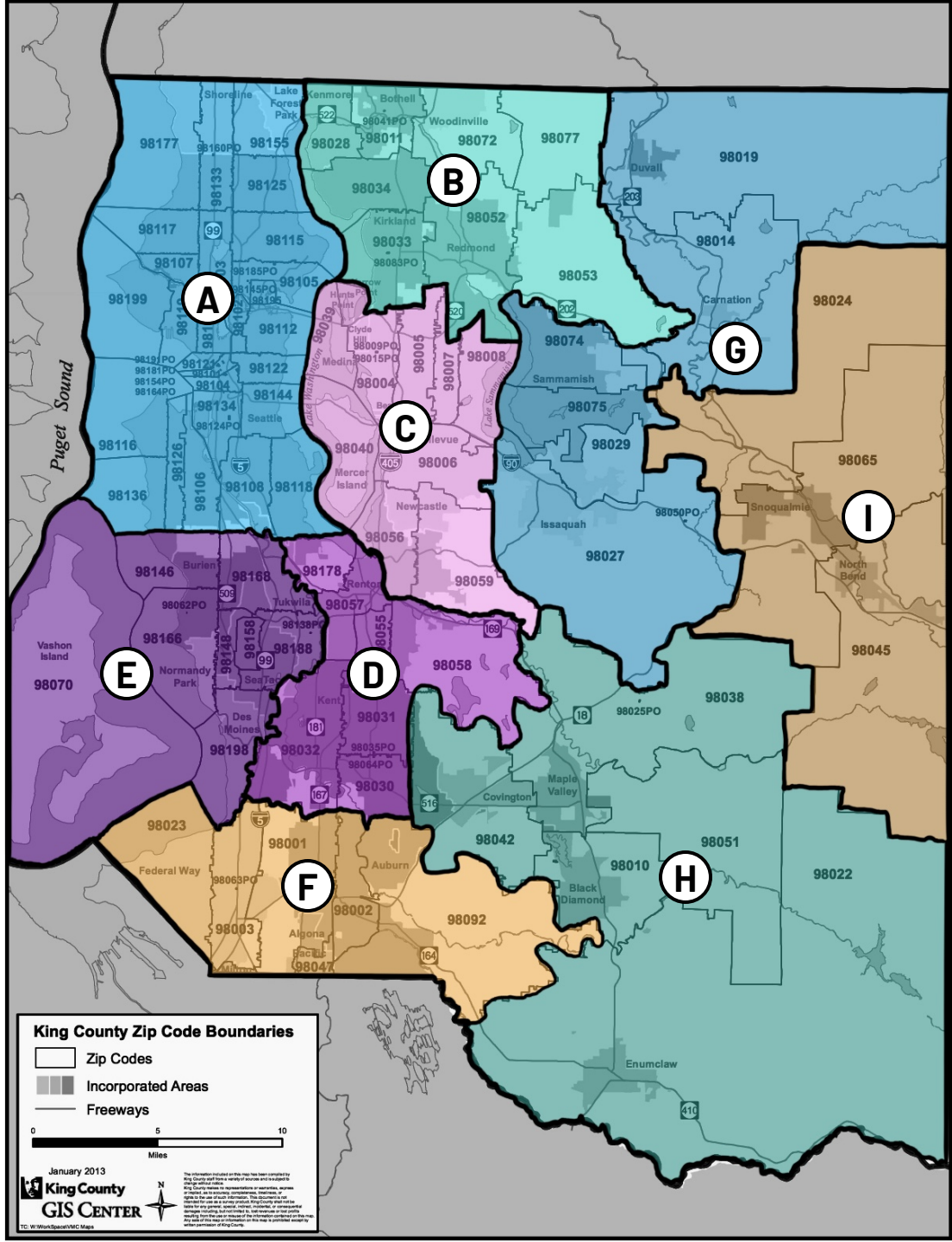
E. Burien, Normandy Park, Des Moines, SeaTac, Tukwila, Vashon Island

F. Federal Way, Auburn, Algona, Milton, Pacific

G. Sammamish, Issaquah, Carnation, Duvall

H. Covington, Maple Valley, Black Diamond, Enumclaw

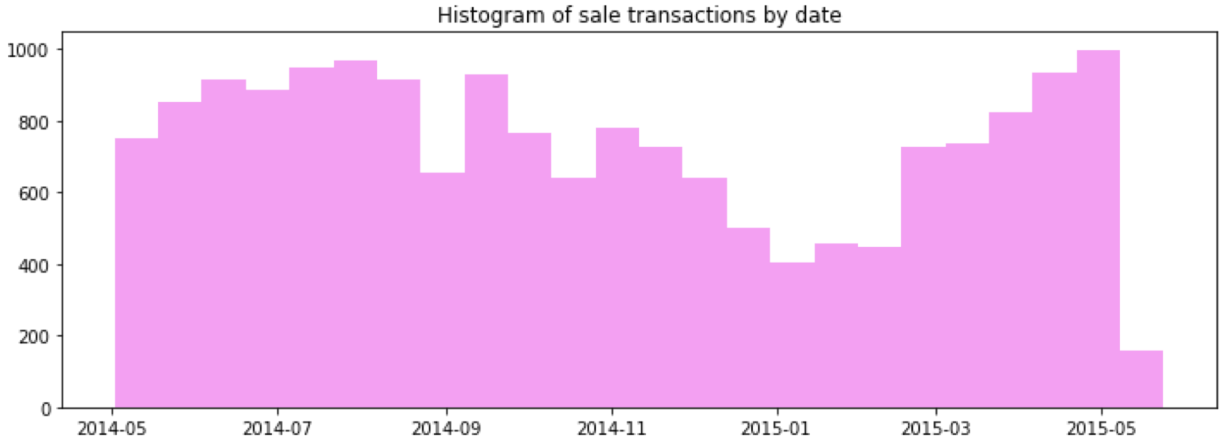
I. Snoqualmie, North Bend

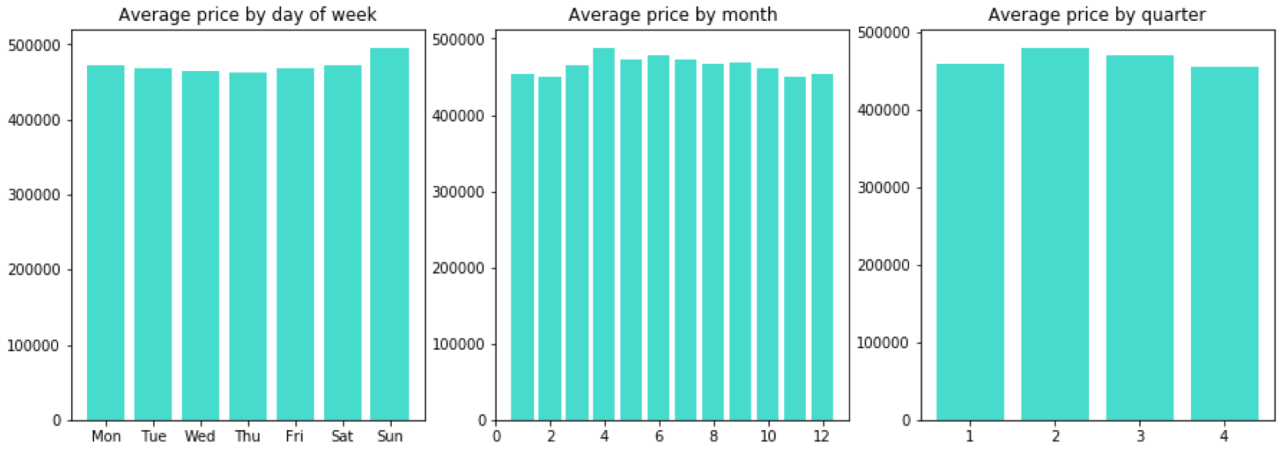


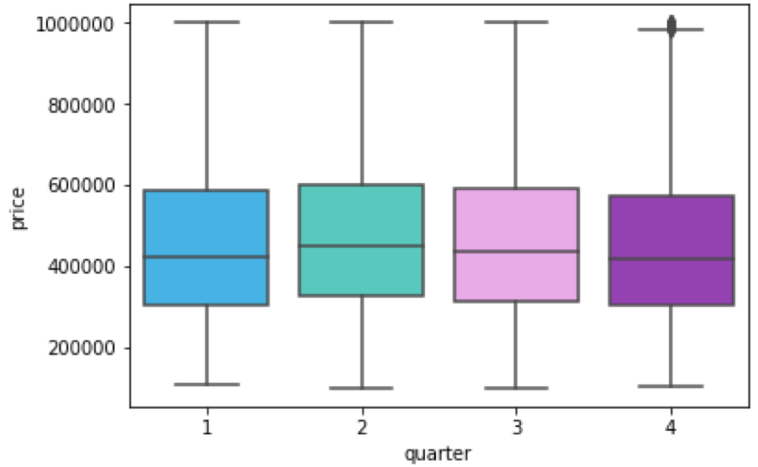
Additional analysis – seasonal effect on house sales

Number of sales does appear to be seasonal – the market slows down to half its peak in the first part of Q1, perhaps to be expected given the Christmas holidays. We would need further years of data to see if this trend persisted.

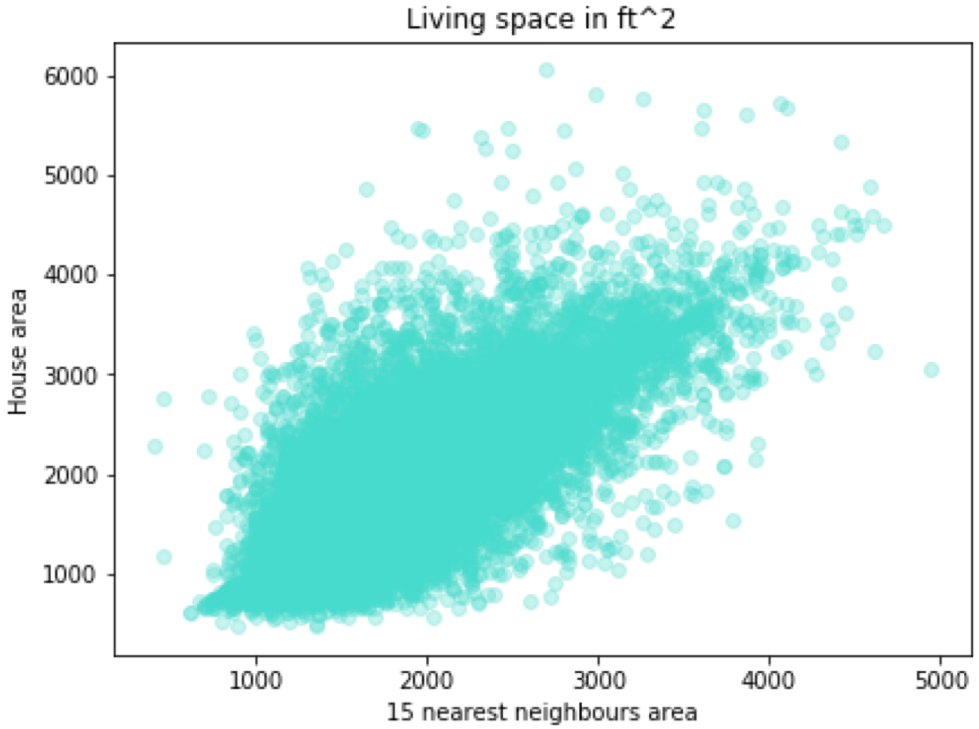
The average price fluctuates slightly by time of year as well. Q2 has the highest mean value sales, though the boxplot (bottom right) suggests that this is driven by a relative lack of cheaper houses.

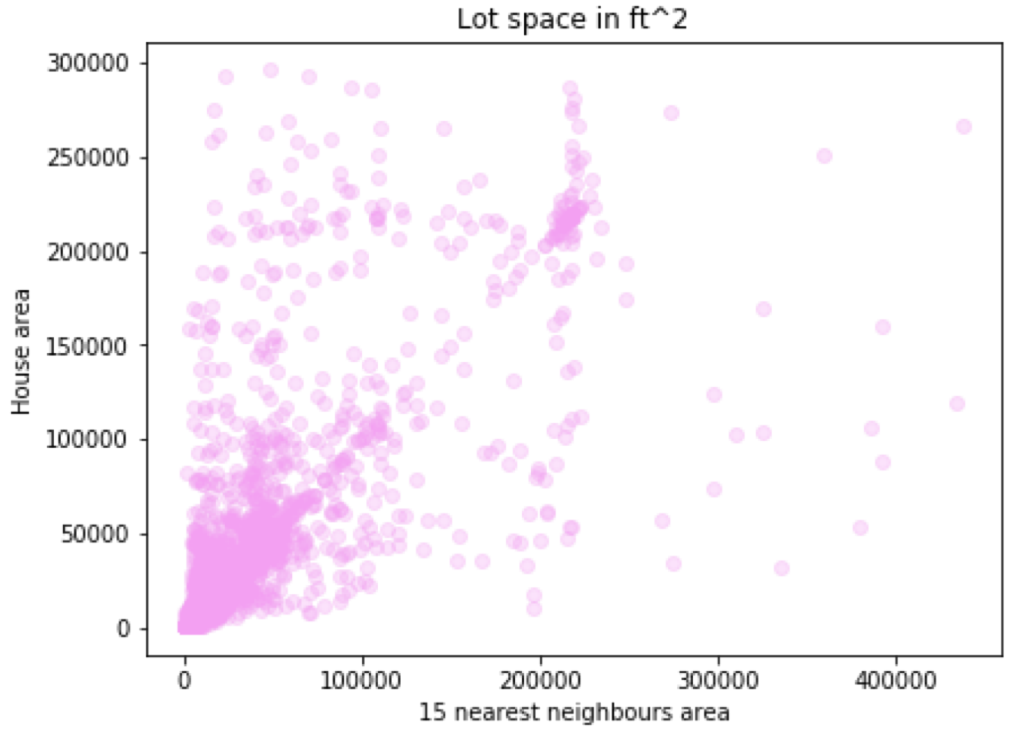






Additional analysis – 15 nearest neighbours





There is a relationship between the living space of a house and its 15 neighbouring properties. However, it is not a particularly strong correlation (R^2 is 0.54).

There is little correlation in lot space of neighbouring houses – suggesting that whilst houses close to each other are built to at least similar specifications, the plots of land are not necessarily apportioned equally between neighbours.