EDA “King County”

Real estate market

Objectives

Within the scope of this project the real estate trade data (2014-05 to 2015-05) from King County (WA) were inspected and suggestions for the client were given.

The client was chosen from the list and some details were provided, namely:

Erin Robinson, Buyer:

- **Invest in poor neighborhood**
- **buying & selling**
- costs back + little profit
- *socially responsible*

Questions

My client intends to **buy** a real estate as an investment, which implies the **selling** of the property at some point in time. The process of purchase requires answering several questions:

1. What is preferable location of the property (Hint given: *poor neighborhood*)
2. What is the price range of the property? (Hint given: *poor neighborhood*)
3. When the purchase is expected to occur? (**assumption**: within next year)

The same true for the sell:

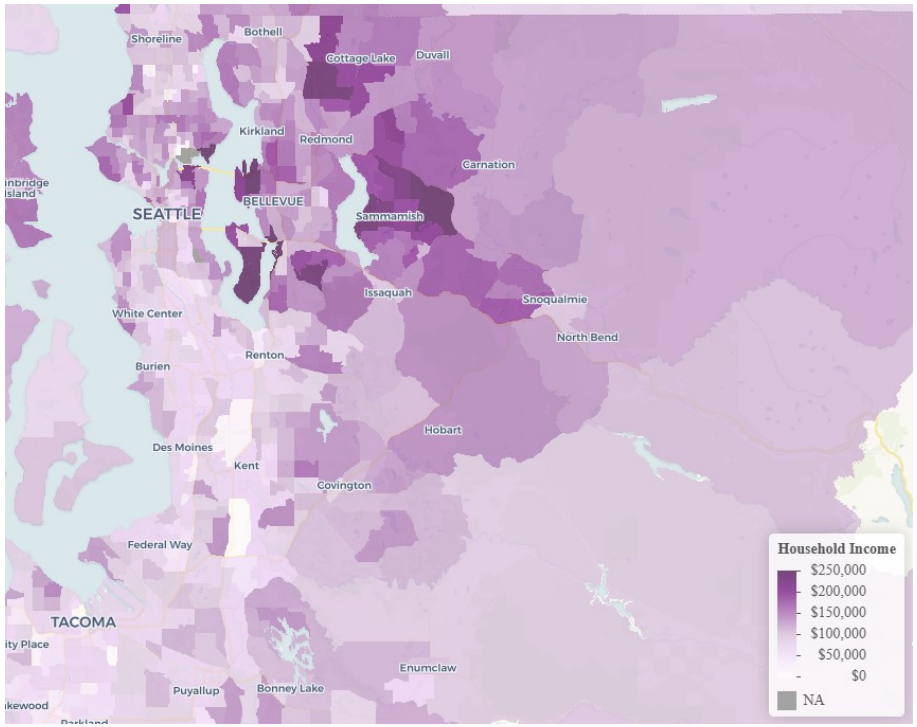
1. What is the expected return time range? (Hint given: data time range)

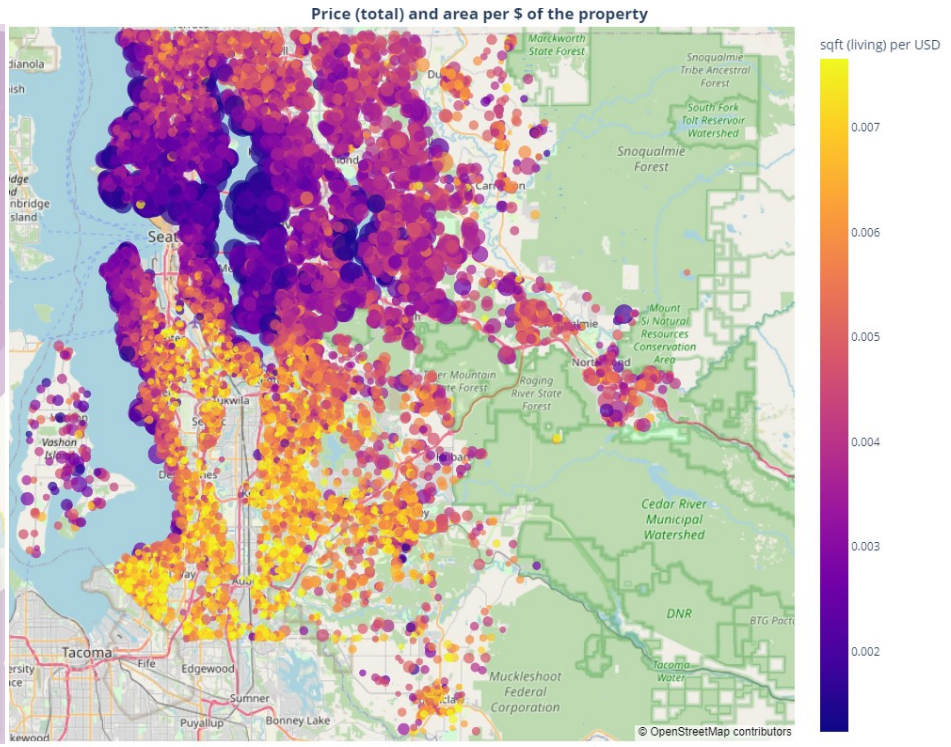
The aim of this project is to provide the client with a short **list** (ca. 10) of properties that are **for sale at the** moment, are within client's requirements, and provide insights on the price seasonality (for selling).

Questions 1&2: location & price

- Total price paid presumably **correlates** with the household income (***partial assumption***).
- King County has median household income of $106,300.
- Median house price in the county is $540,000.

Due to the lack of data on the household income of the buyers, I would **assume** the “poor neighborhood” house prices to be below half of the median.

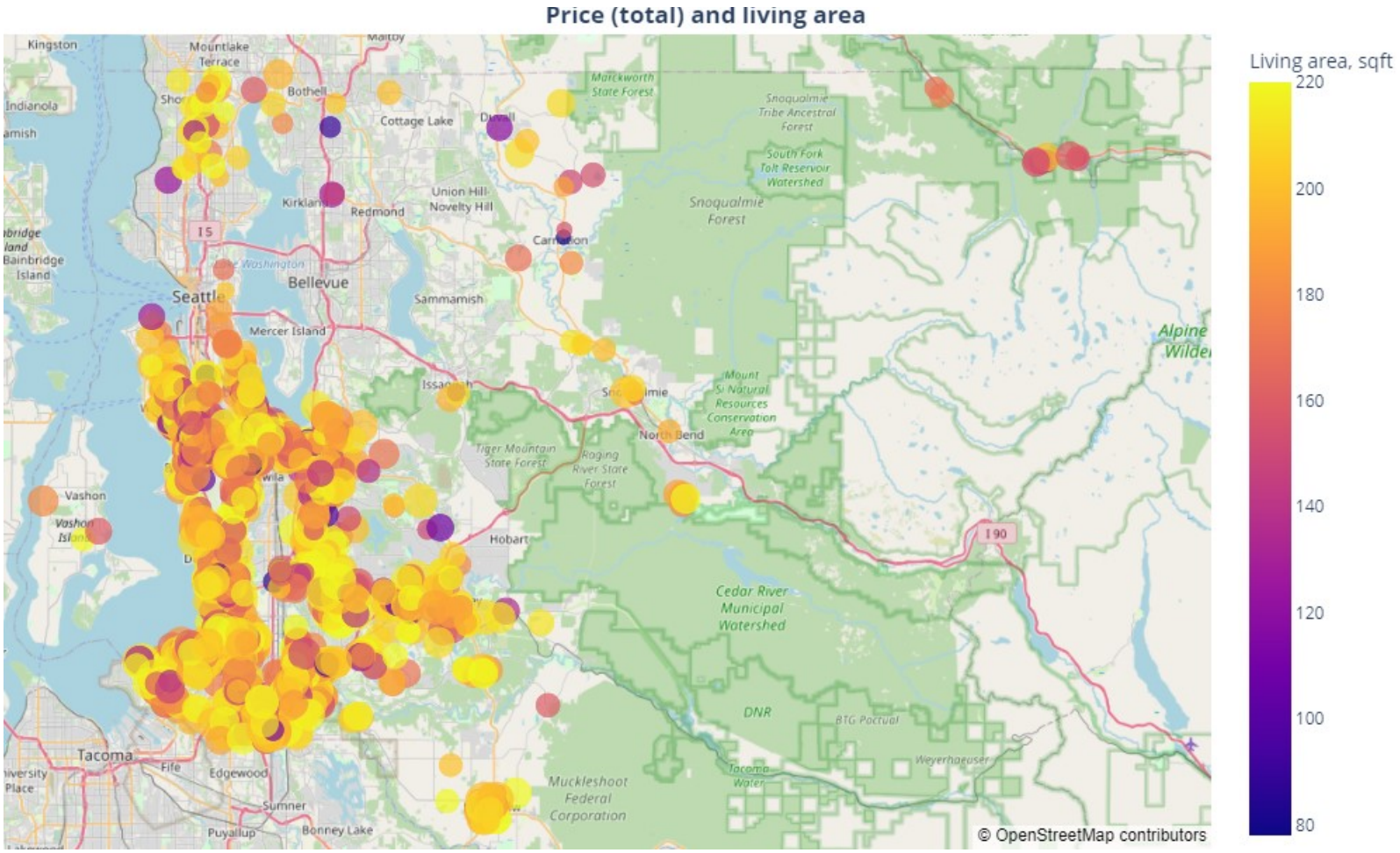




https://www.psrc.org/our-work/equity/equity-tracker/economy/household-income

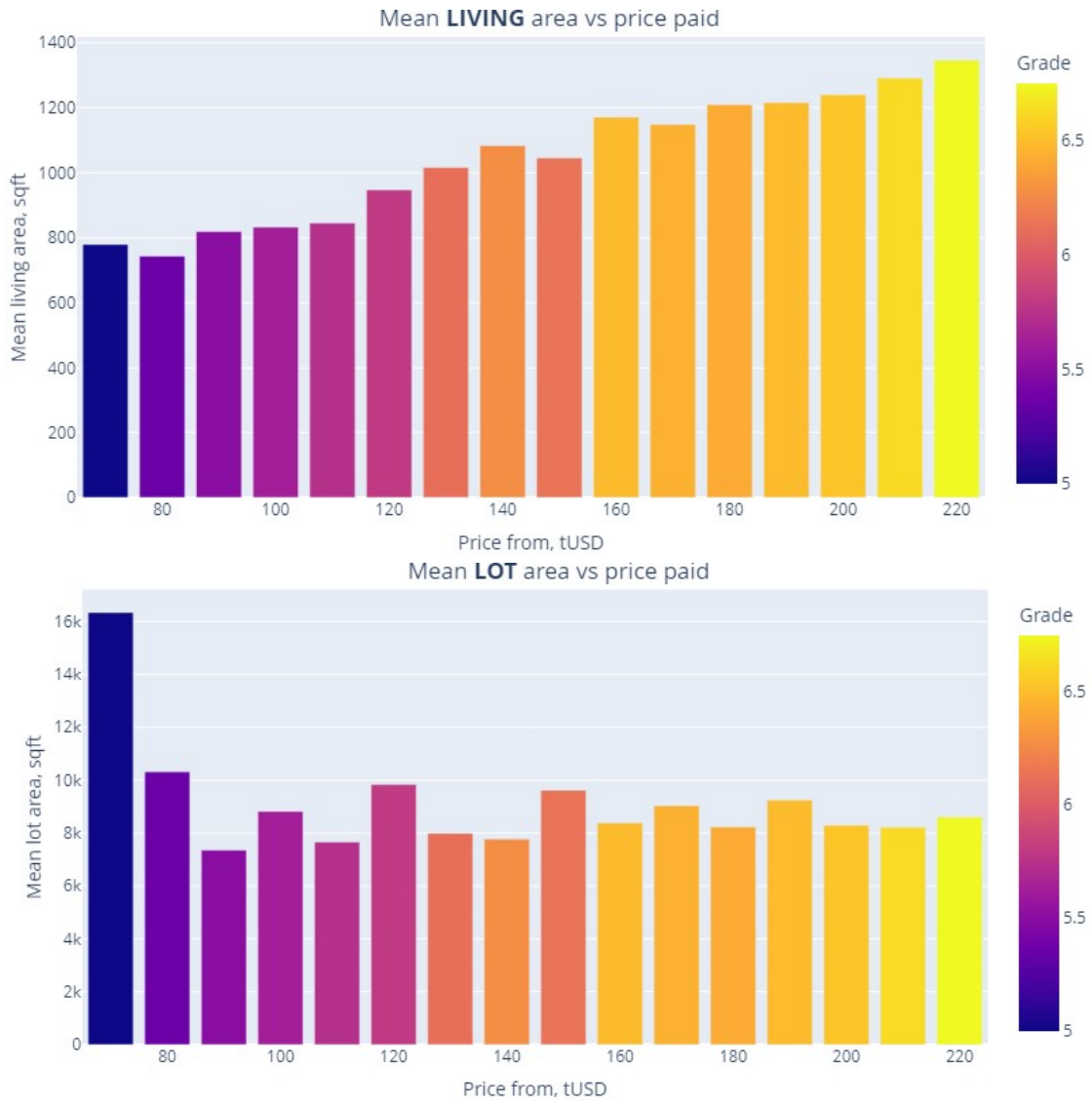
Questions 1&2: location & price

By narrowing down the price range I got significant reduction in data (to 1415 ID's). This narrowed data doesn't seem to have significant deviation in living area.



Questions 1&2: location & price

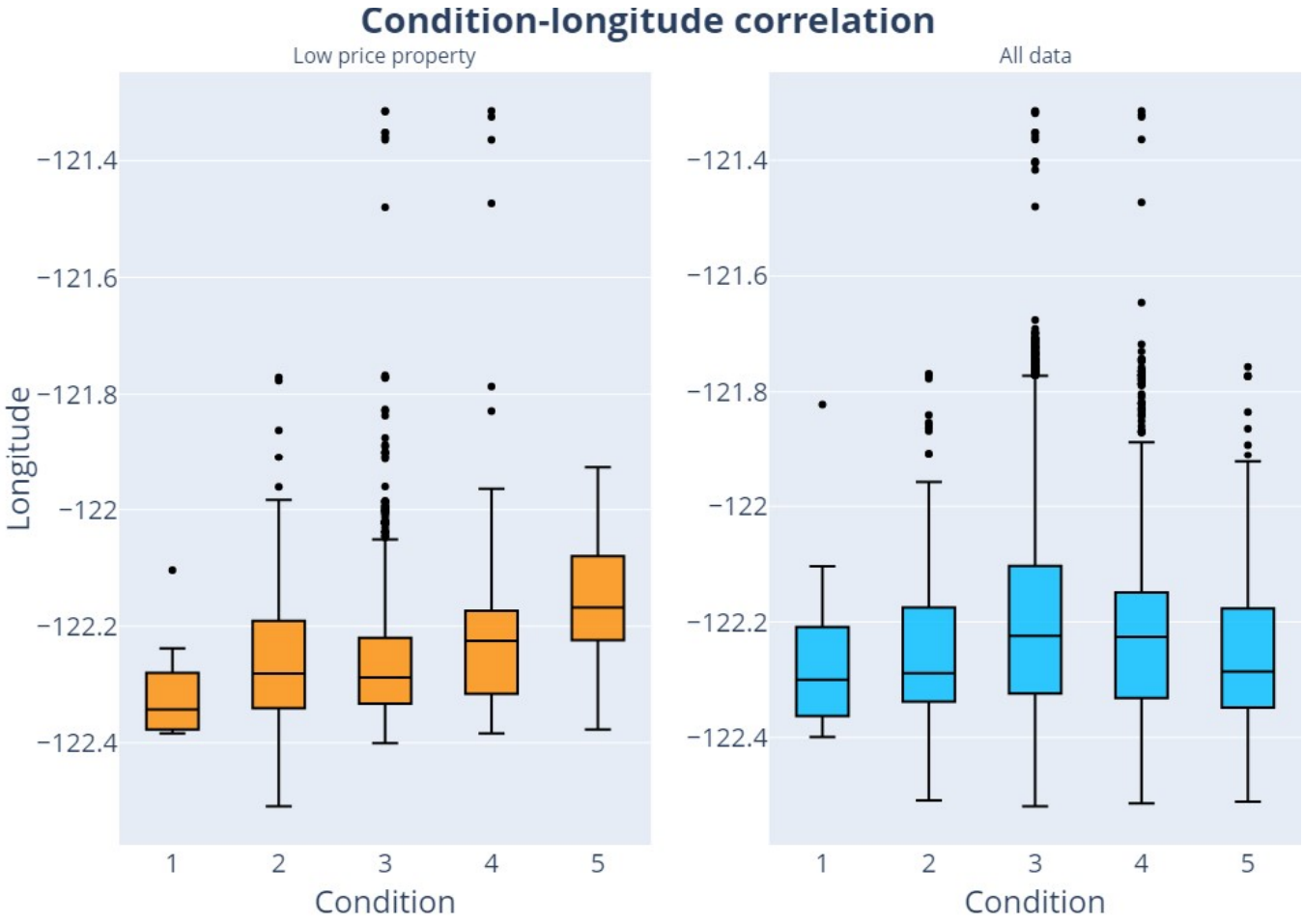
Living area **correlates** with price for both all households (table), and for poor neighborhoods (plots).



	price_t
date	-0.010353
price_t	1.000000
house_id	0.004267
id	0.040184
bedrooms	0.344245
bathrooms	0.497298
sqft_living	0.643993
sqft_lot	0.075071
floors	0.322482
waterfront	0.030893
view	0.287173
condition	0.017995
grade	0.658152
sqft_above	0.541563
sqft_basement	0.246800
yr_built	0.102058
yr_renovated	0.095829
zipcode	-0.009084
lat	0.456126
long	0.064058
sqft_living15	0.572263
sqft_lot15	0.063076
sqft_living_per_usd	-0.596537
sqft_lot_per_usd	-0.505835

Observation 1:

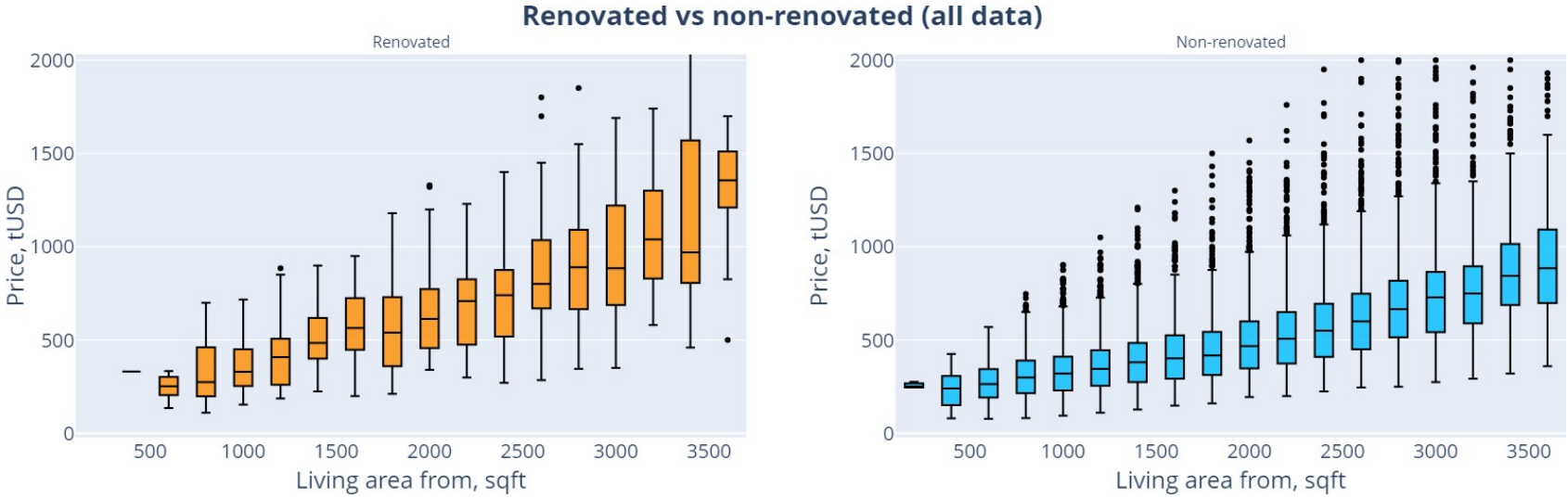
The condition of the houses increases alongside the longitude (i.e. the more far the property from the shore, the better the condition):



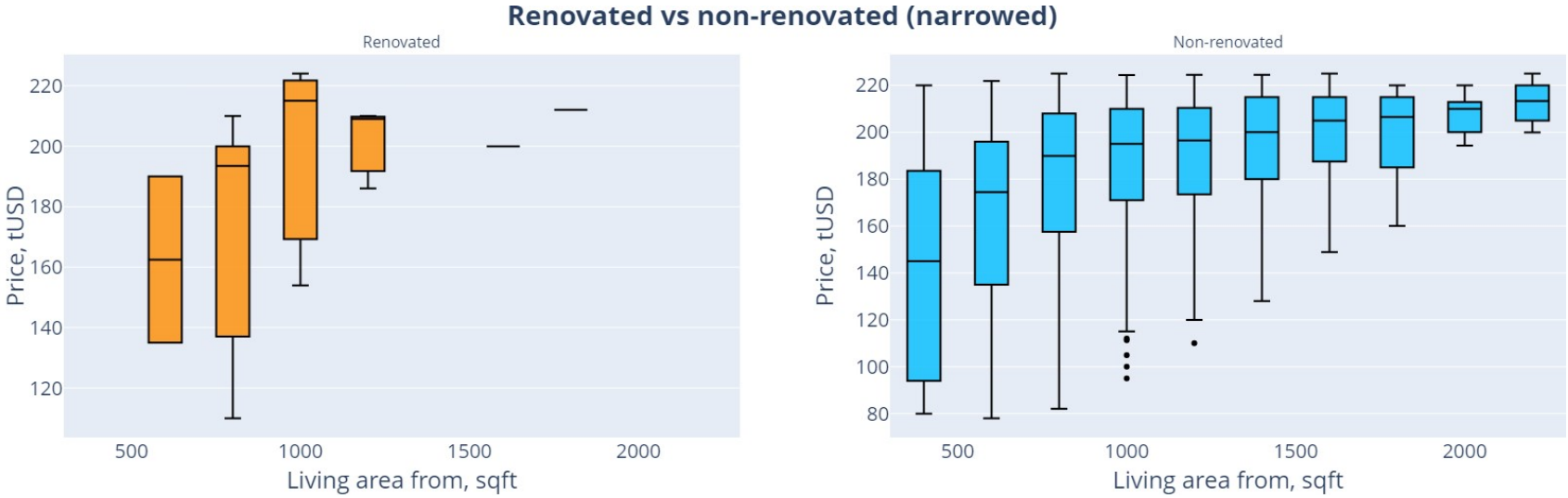
Nevertheless, I'd use a condition as a determining filtering factor.

Observation 2:

The influence of the renovation on the price within the same living area bracket (all data):



Unfortunately, data within lower price range is not conclusive. While it is possible to extrapolate the positive trend onto the narrowed data, only 14 out of 1415 have renovation done within last 30 years (assumming missing value = no renovation done).



Observation 3:

Total count of offers within lower price range varies with seasons: