WHPO PLANIFICATION FOR HOUSE PURCHASE PORTFOLIO

New Insights

Ariadna Recasens

OVERVIEW

- In this project, we help the non-for-profit Washington Help People organization to find the most affordable houses.

OUTLINE

- Business problem
- Data
- Methods
- Results
- Conclusions

BUSINESS PROBLEM

- The non-for-profit Washington Help People Organization (WHPO) has recently received an anonymous donation to build the maximum number of houses for homeless people.
- WHPO wants to get some insights to find the cheapest houses so they can buy more units to help more homeless people.
- In this project, we will use regression models to identify the factors that have a significant impact on house price.

DATA

King County House Sales dataset

The dataset contains house sale prices for King County. It includes homes sold between May 2014 and May 2015.

It contains 21597 entries and 21 columns. This project will focus on how the following variables influence house price:

- Zipcode
- Waterfront
- Grade
- Condition
- bedrooms number
- sqft_livingsquare
- sqft_lotsquare
- yr_built

METHODS

Filtering process to create a subset of interesting cheap houses

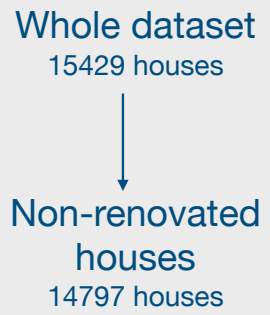
1. T-test analysis to study the effect of renovation and water views on house prices.

2. Linear regression models to study the individual effect of zip-code, grades, conditions and number of bedrooms on houses price

3. Linear regression model in our subset of cheap houses to understand the effect of the following continuous variables on price: sqft_livingsquare, sqft_lotsquare, yr_built

Results

EFFECT OF RENOVATION ON HOUSE PRICES



	Renovated houses	Non-renovated houses
Average price	759316	532194

Analysis: t-test, p_value < 0.05
Meaning: there is a significant difference between renovated and non-renovated houses

Results

EFFECT OF WATER-VIEW ON HOUSE PRICES

	Houses with water-view	Houses without water-view
Average price	1725217	524975

Analysis: t-test, p_value < 0.05
Meaning: there is a significant difference between houses with and without water-views

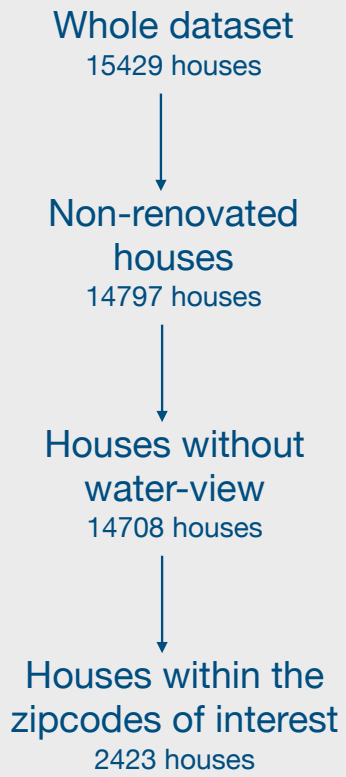
Results

EFFECT OF ZIPCODE ON HOUSE PRICES

Analysis: linear regression
Results:

Zipcodes that significantly increase house prices: 98004, 98005, 98006, 98007, 98008, 98010, 98011, 98014, 98019, 98024, 98027, 98028, 98029, 98033, 98034, 98038, 98039, 98040, 98042, 98045, 98052, 98053, 98056, 98058, 98059, 98065, 98070, 98072, 98074, 98075, 98077, 98092, 98102, 98103, 98105, 98107, 98108, 98109, 98112, 98115, 98116, 98117, 98118, 98119, '98122', 98125, 98126, 98133, 98136, 98144, 98146, 98155, 98166, 98177, 98199

Zipcodes that do not significantly increase house prices: 98002, 98003, 98022, 98023, 98030, 98031, 98032, 98055, 98106, 98148, 98168, 98178, 98188, 98198.



Results

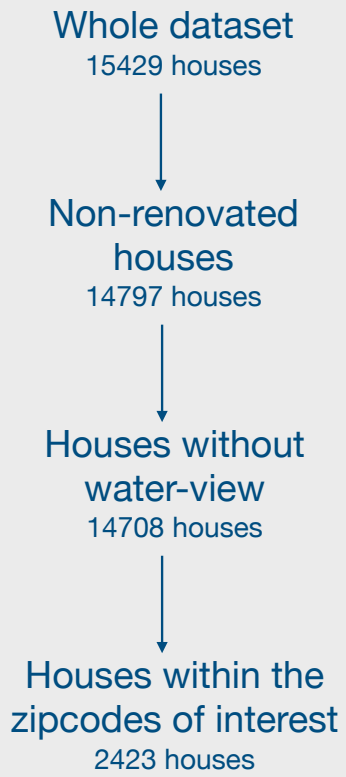
EFFECT OF CONDITION ON HOUSE PRICES

Analysis: linear regression

Results:

Conditions 3, 4 and 5 significantly and positively contribute to the price of houses (coef. Positive and p_values < 0.05)

Note: there are only 27 houses on the condition below 3, so we can't discard them or we would lose too many data points.



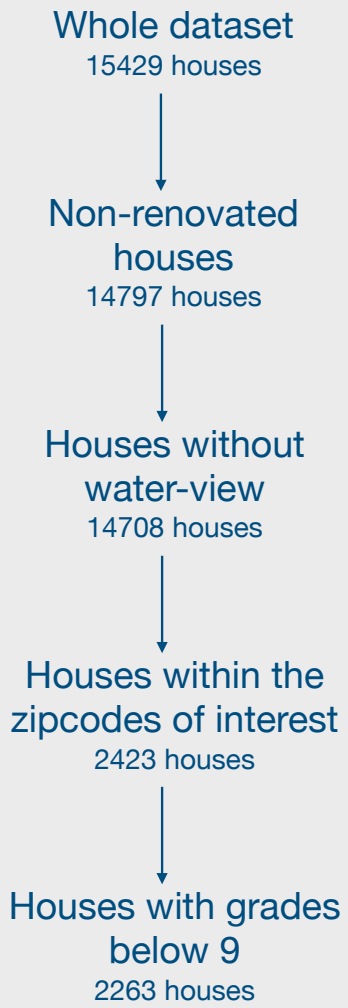
Results

EFFECT OF GRADE ON HOUSE PRICES

Analysis: linear regression

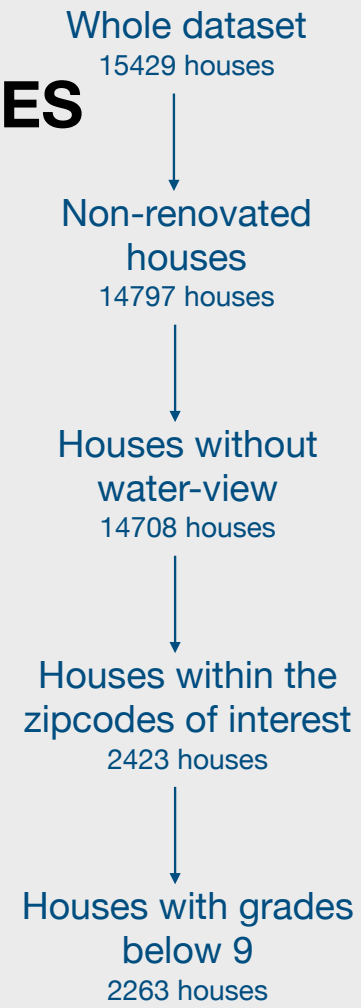
Results:

Grades 9, 10 and 11 significantly and positively contribute to the price of houses



Results

EFFECT OF NUMBER OF BEDROOMS ON HOUSE PRICES



Analysis: linear regression

Results:

There is a correlation between the number of bedrooms and prices. The more number of bedrooms, the more number of houses.

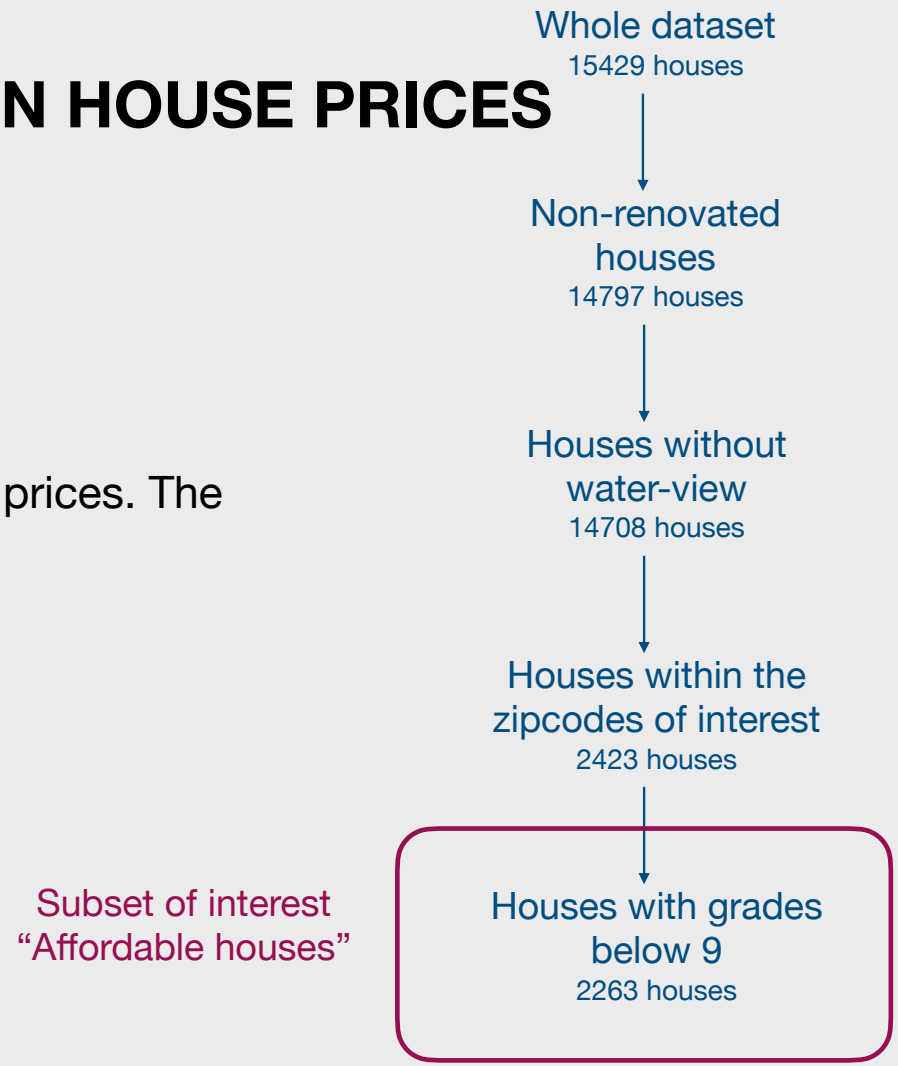
Results

EFFECT OF NUMBER OF BEDROOMS ON HOUSE PRICES

Analysis: linear regression

Results:

There is a correlation between the number of bedrooms and prices. The more number of bedrooms, the more number of houses.



Results

EFFECT OF YEARS SINCE BUILT, SQFT LIVING AREA AND SQFT LIVING LOT IN OUR SUBSTE OF INTEREST

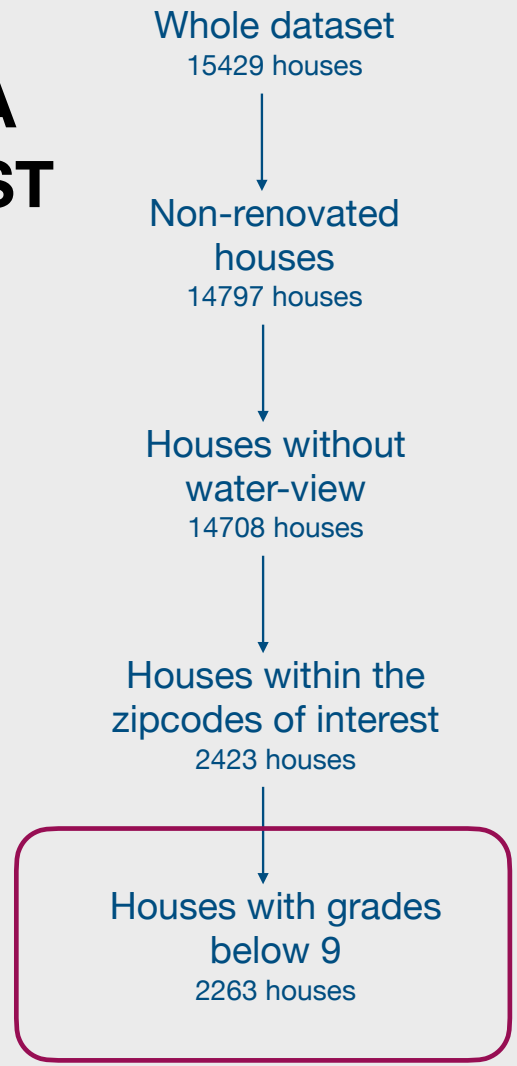
Analysis: linear regression

Results:

SQFT living area and lot have a significant positive contribution to the house price (the more area, the more expensive)

Years since built has a significant negative contribution to the price (the oldest, the cheapest)

sqft living area is the variable that affects the most to the price, as it has the highest coefficient value (coef = 0.56) and a p_value < 0.05.



CONCLUSIONS

Business recommendations

- Look for non-renovated houses.
- Look for houses that have no waterview.
- Look for houses in the following zipcodes: 98002, 98003, 98022, 98023, 98030, 98031, 98032, 98055, 98106, 98148, 98168, 98178, 98188, 98198.
- Look for houses below condition 3.
- Look for houses below grade 9.
- Centre your search on houses with the lowest sqft living. Older houses will be cheaper