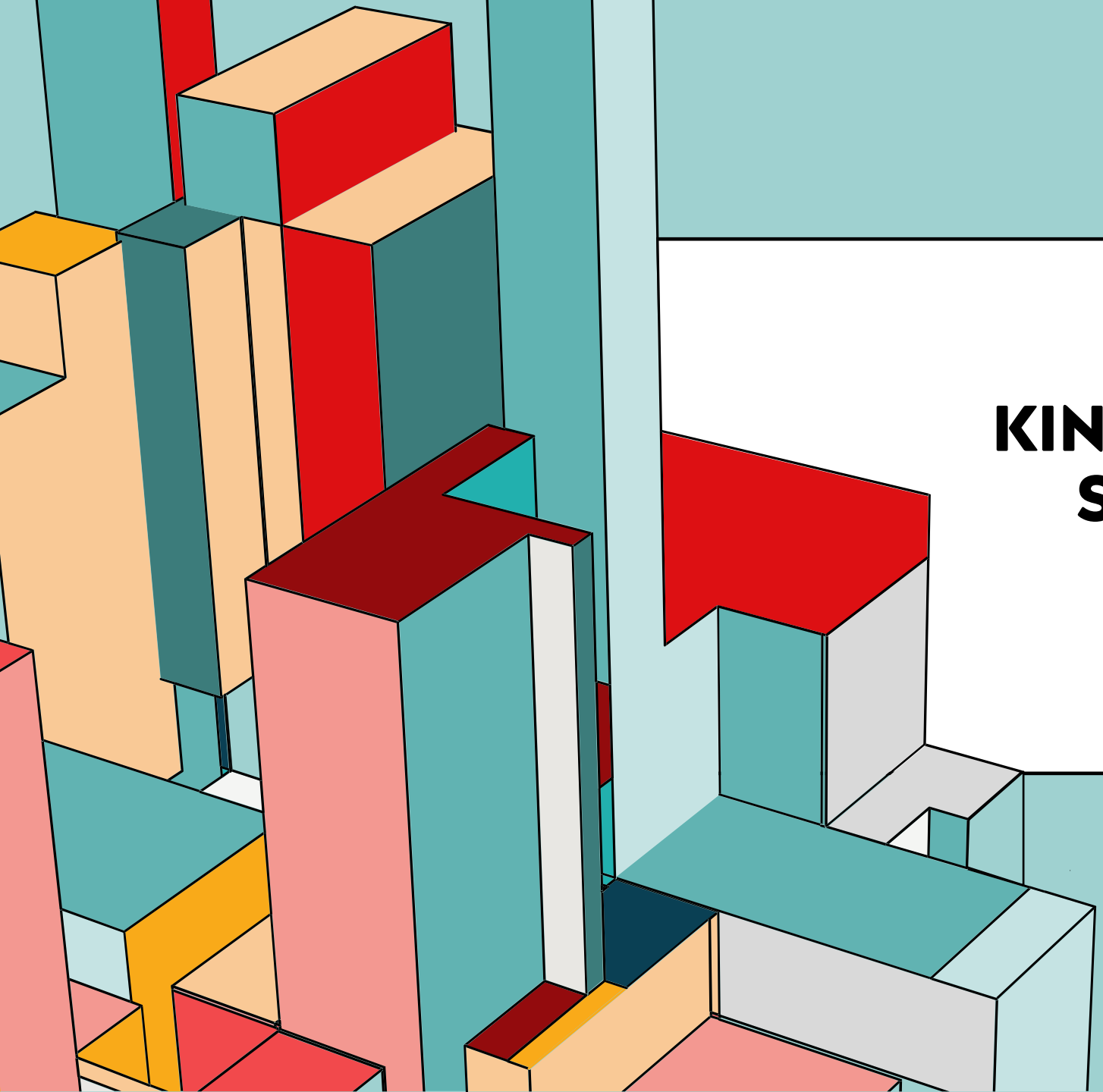
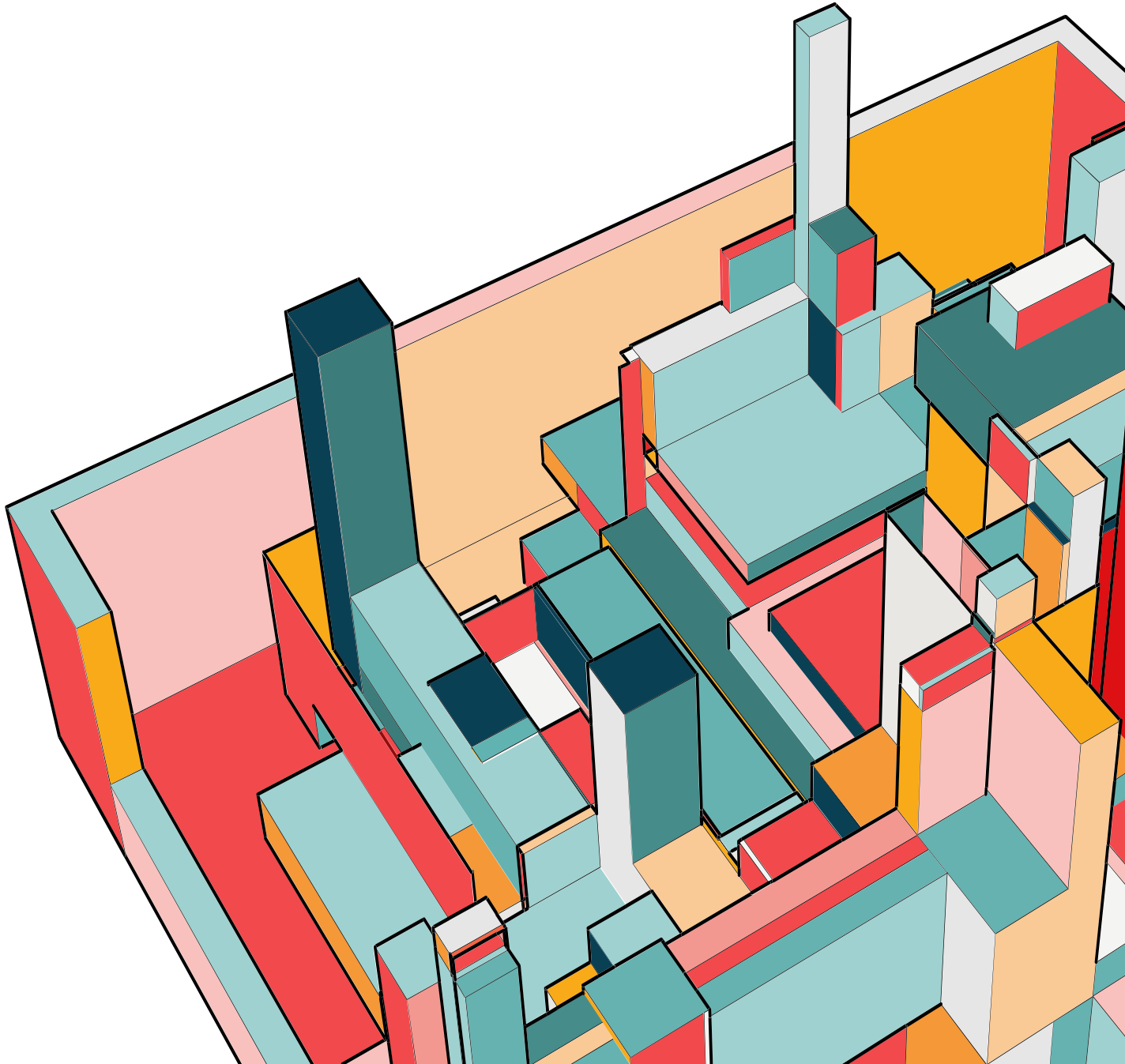


KING COUNTY HOUSE SALES ANALYSIS

Femi Kamau

CONTENTS



OVERVIEW

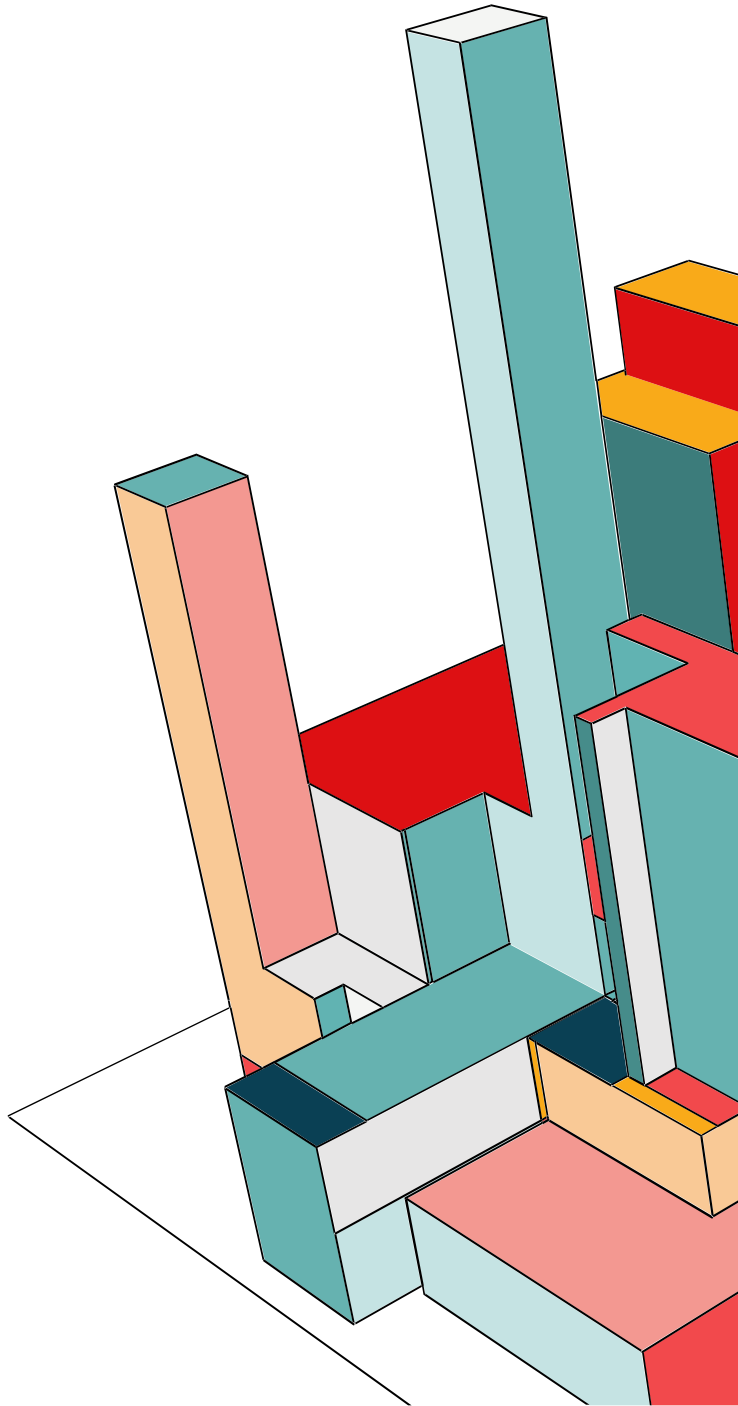
HOUSE BY THE WATER

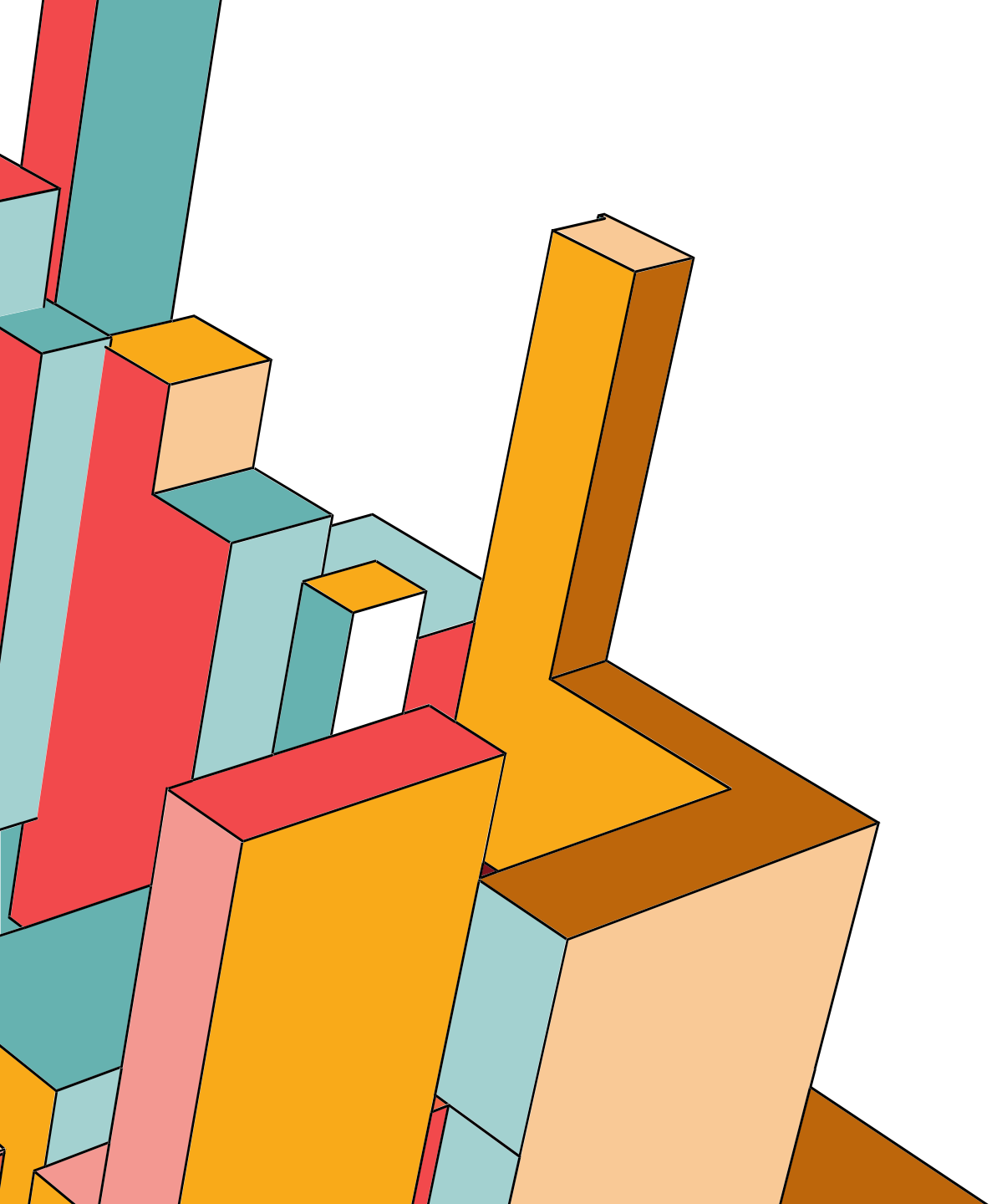
INCREASE THE BASEMENT SIZE

MORE BEDROOMS

HIGHER HOUSE GRADE

INCREASE THE FLOORS





BUSINESS UNDERSTANDING

A REAL ESTATE ANGENCY LOCATED IN KING COUNTY IS LOOKING TO ADVICE HOMEOWNERS ABOUT HOW HOME RENOVATIONS MIGHT INCREASE THE VALUE OF THEIR HOMES.

THE AGENCY IS LOOKING TO USE THE KING COUNTY HOUSE DATA PROVIDED BY THE STAKEHOLDER TO DETERMINE THE BEST RENOVATIONS TO MAKE TO INCREASE THE VALUE OF A HOME.

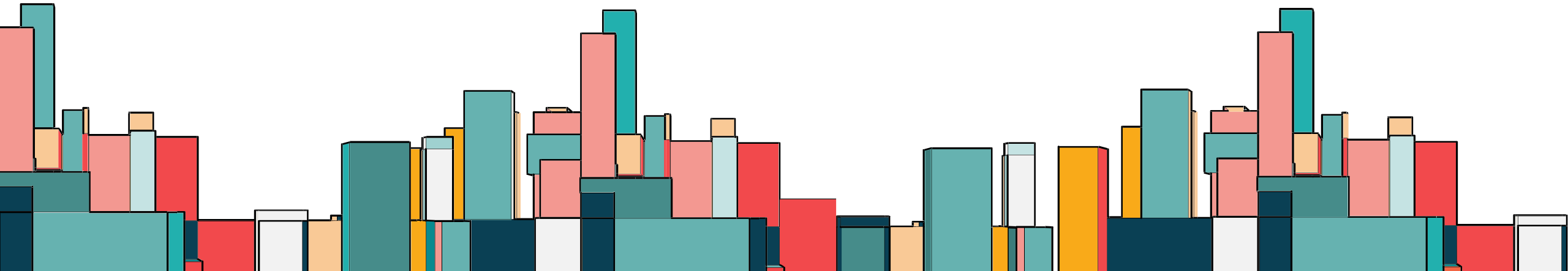
DATA UNDERSTANDING

NUMERICAL

DATE, PRICE, BEDROOMS, BATHROOMS
LIVING SPACE SIZE, LOT SIZE, FLOORS,
HOUSE SIZE (ABOVE GROUND),
BASEMENT SIZE, YEAR BUILT, YEAR
RENOVATED, LATITUDE, LONGITUDE
LIVING SPACE (NEAREST 15
NEIGHBORS),
LOT SIZE (NEAREST 15 NEIGHBORS)

CATEGRORICAL

ID, WATERFRONT, VIEW, CONDITION, GRADE,
ZIPCODE

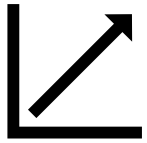


MODELING

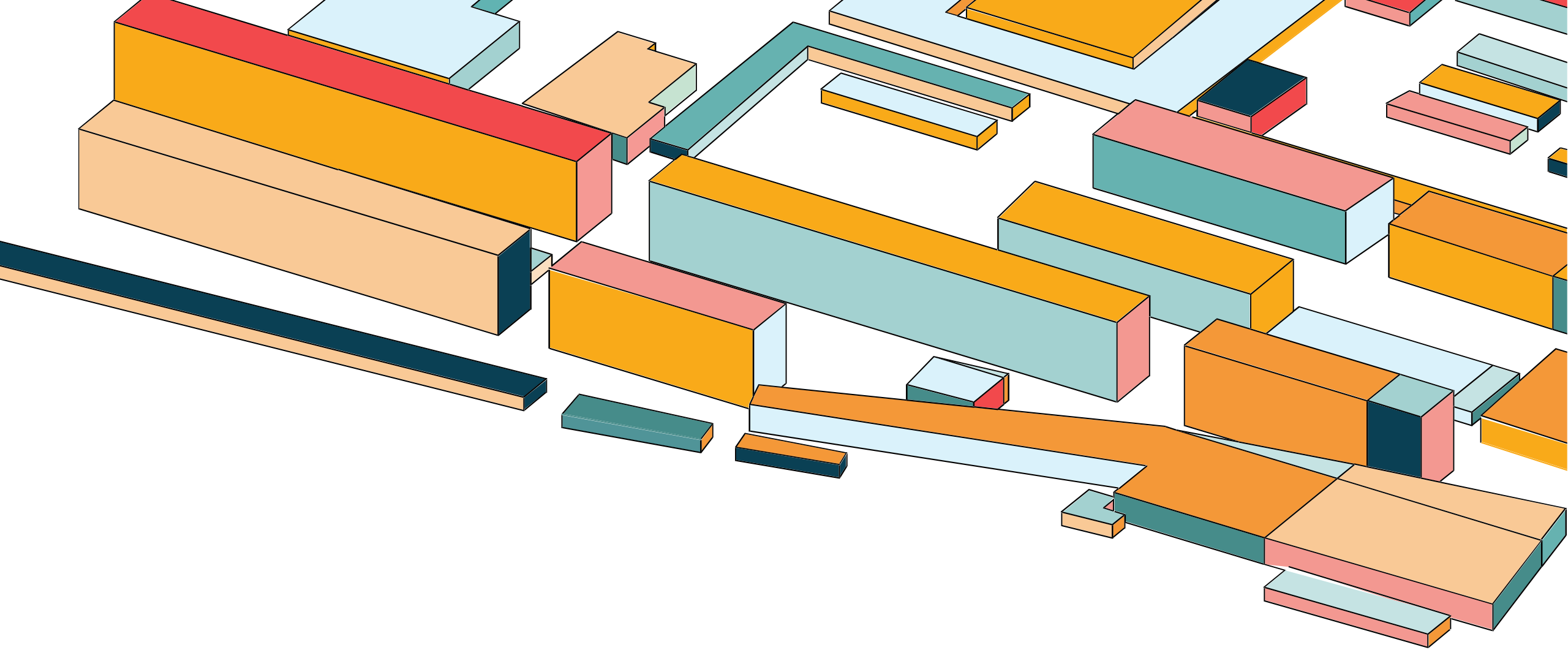
WE OPTED TO USE MULTIPLE LINEAR REGRESSION FOR MODELLING

MULTIPLE INDEPENDENT VARIABLES

ONE TARGET VARIABLE



OUR AIM IS TO IDENTIFY THE IMPACT



REGRESSION RESULTS

BASELINE MODEL

USING THE CORRELATION MATRIX, WE IDENTIFIED SQFT_LIVING AS THE MOST BEST VARIABLE TO USE FOR THE LINEAR REGRESSION

BASELINE MODEL (RESULTS)

FROM THE BASELINE MODEL, WE WERE ABLE TO ESTABLISH THE FOLLOWING RESULTS:

MEAN ABSOLUTE ERROR

$173,713.2

ADJUSTED R-SQUARED

49.3%

ITERATED MODEL (ENCODING)

IN ORDER TO BUILD OUR ITERATED MODEL, WE HAD TO FIRST ENCODE OUR CATEGORICAL COLUMNS:

ORDINAL ENCODING:

CONDITION

GRADE

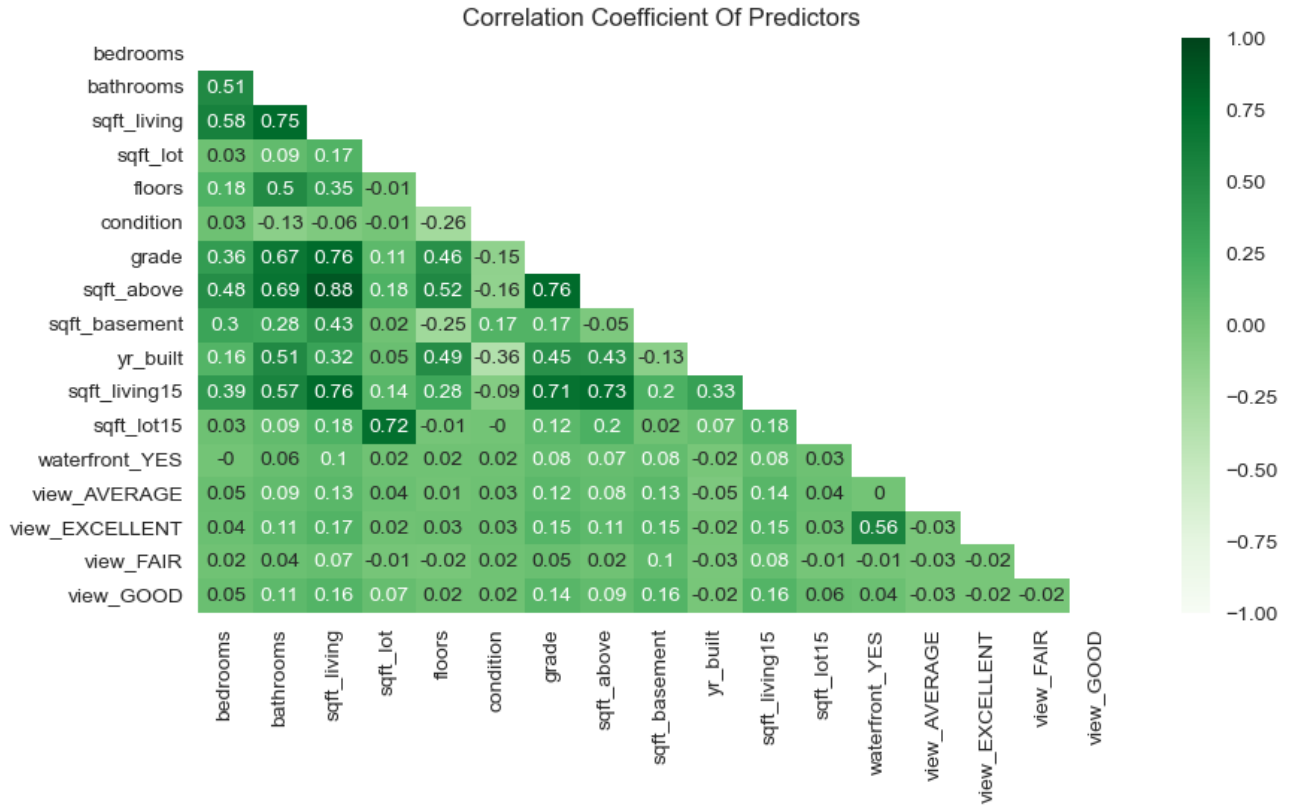
ONE-HOT ENCODING:

WATERFRONT

VIEW

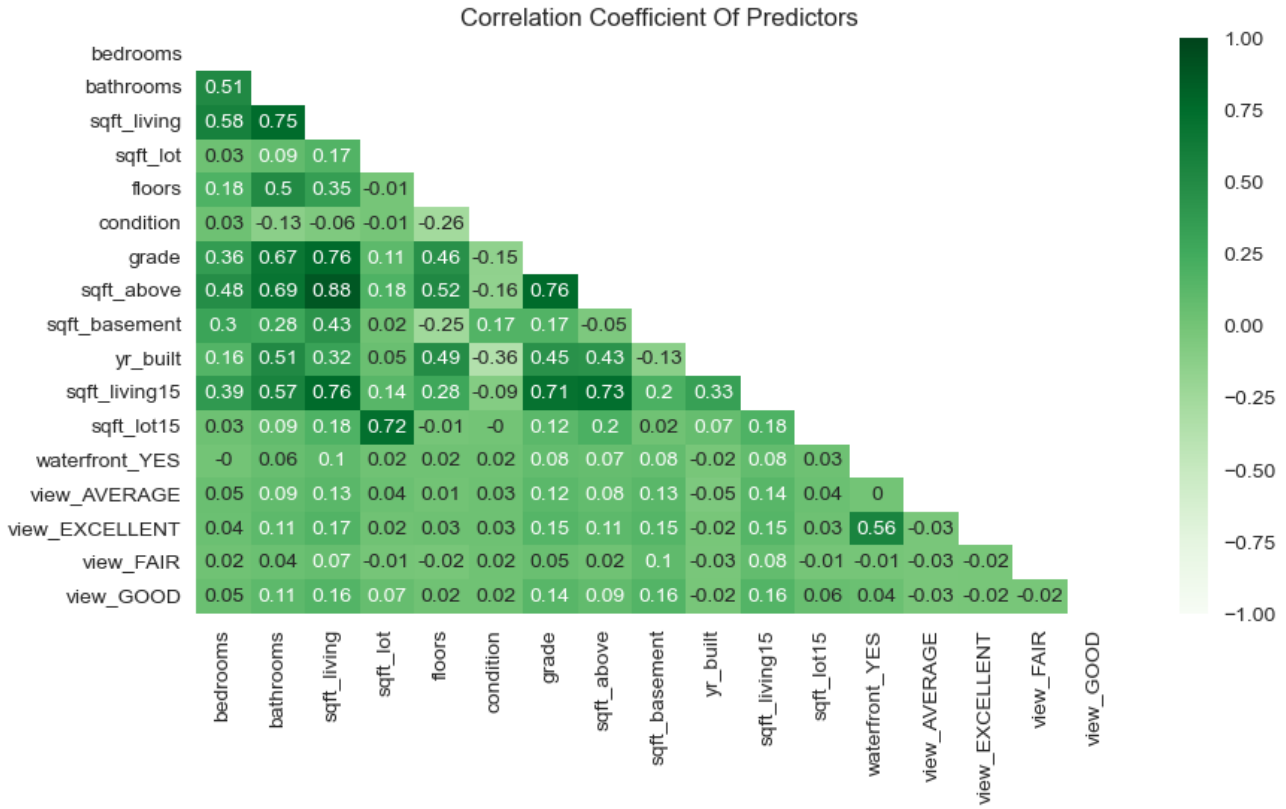
ITERATED MODEL (INITIAL CORRELATION MATRIX)

OUR ORIGINAL CORRELATION MATRIX LOOKED LIKE THIS:



ITERATED MODEL (FINAL CORRELATION MATRIX)

WITH A THRESHOLD OF 0.6 (TO REDUCE MULTICOLINEARITY), OUR FINAL MATRIX LOOKED LIKE THIS:



ITERATED MODEL (COEFFICIENTS)

AFTER OUR COEFFICIENTS ON THE REGRESSION MODEL THESE WERE OUR RESULTS:

	coefficient	p-value
const	5137454.742	0.000000
bedrooms	16239.726	0.000000
sqft_lot	0.137	0.000498
floors	78262.473	0.000000
condition	19442.583	0.000000
grade	205091.669	0.000000
sqft_basement	118.865	0.000000
yr_built	-3277.171	0.000000
waterfront_YES	540924.431	0.000000
view_AVERAGE	69940.897	0.000000
view_EXCELLENT	340899.246	0.000000
view_FAIR	131058.754	0.000000
view_GOOD	137050.033	0.000000

MEAN ABSOLUTE ERROR

$147,296.6

ADJUSTED R-SQUARED

60.1%

RECCOMENDATIONS

$540,924

INCREASE IN VALUE OF HAVING A HOUSE BY THE WATER

$205,092

VALUE INCREASE PER GRADE INCREMENT

$78,262

VALUE INCREASE PER NUMBER OF FLOORS

RECCOMENDATIONS

$16,240

INCREASE IN VALUE PER NUMBER OF BEDROOMS

$119

VALUE INCREASE PER SQUARE FOOT IN BASEMENT SIZE

NEXT STEPS

WHAT NEXT? HOW CAN WE IMPROVE THE ANALYSIS

1

WE WOULD NEED TO IMPROVE THIS ANALYSIS BY INCLUDING MORE CURRENT HOUSE SALES DATA

2

MARKET RESEARCH (SUCH AS: POPULATION, AVERAGE FAMILY SIZE, ETC.) WOULD IMPROVE THE RECOMMENDATION

3

LOWERING THE CORRELATION THRESHOLD (FROM 0.6) MAY REDUCE THE MULTICOLINEARLITY

THANK YOU

Femi Kamau

Email: femikkamau@gmail.com

GitHub: ctrl-Karugu