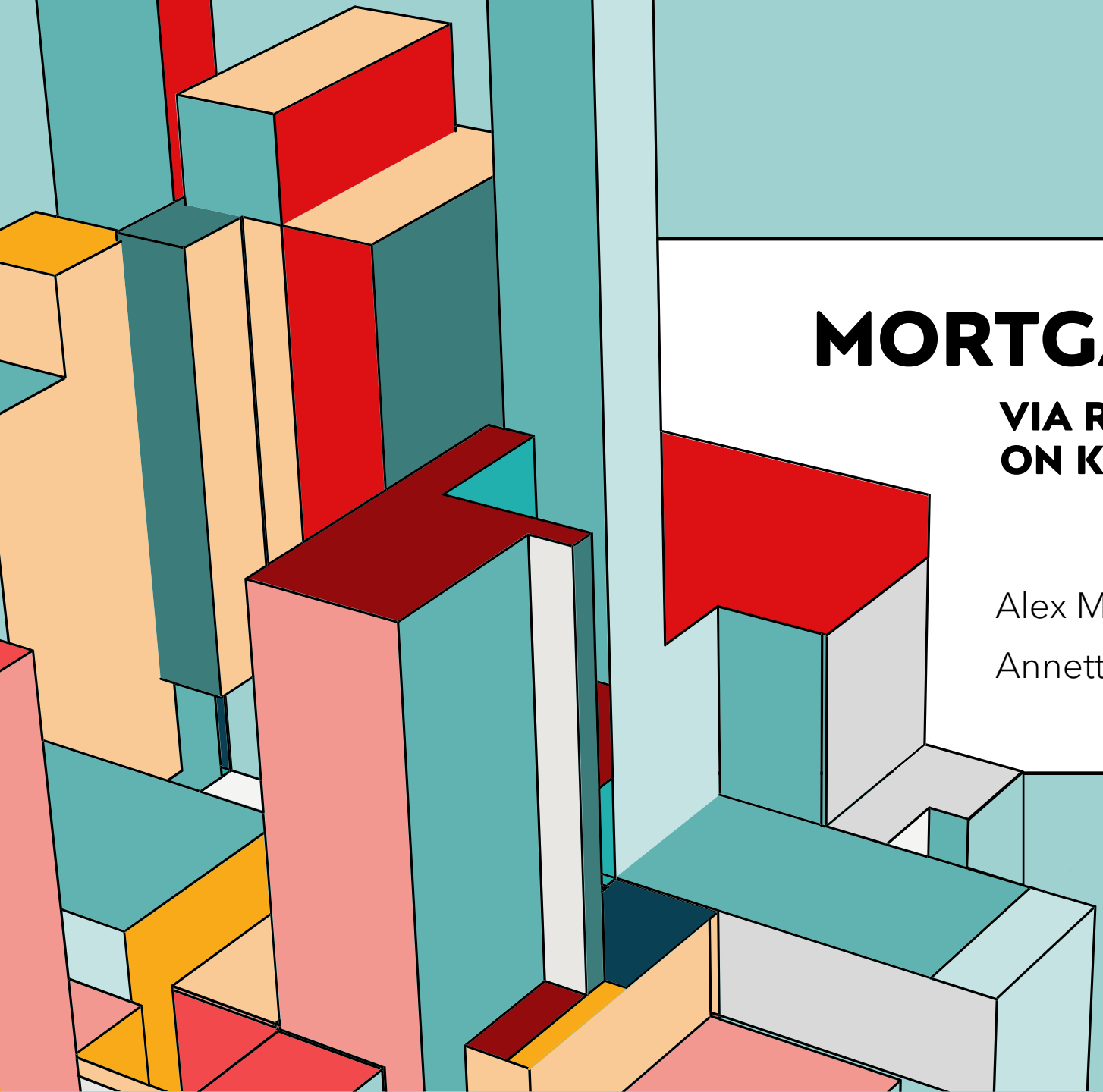
Public



MORTGAGE VALUATION

VIA REGRESSION MODELING ON KING COUNTY DATA SET

Group 19

Alex Muturi

Angela Nyaga

Annette Kalekye

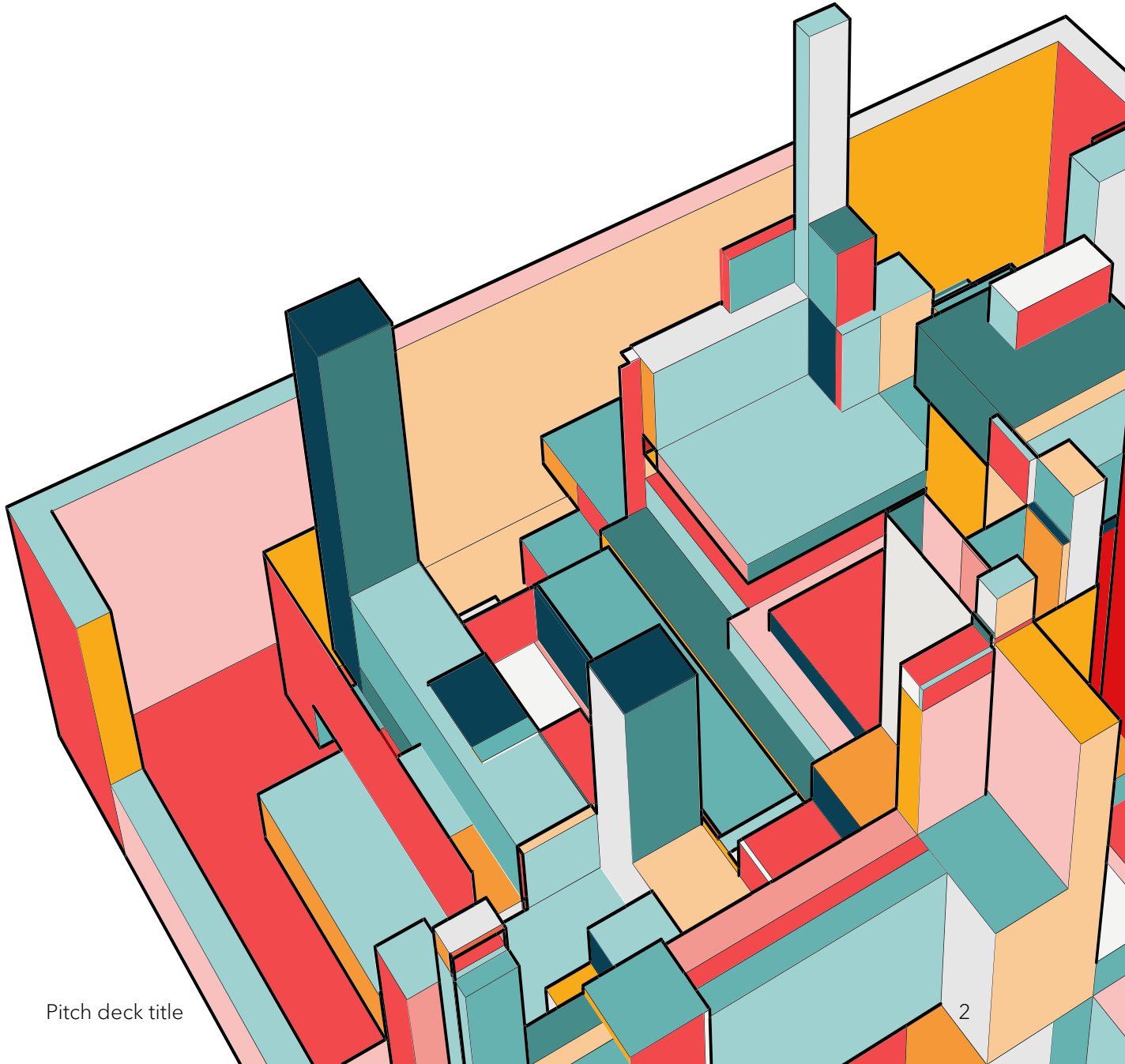
Chris Otieno

INTRODUCTION

We have analyzed King County House sales data to advise pricing decisions to be made by Mipaka, a real estate company.

Exploratory data analysis gave us insights into the data from which we derived price determinants using various modeling techniques.

From this we are able to get three main business recommendations.



Public

BUSINESS PROBLEM

PRICING CHALLENGES

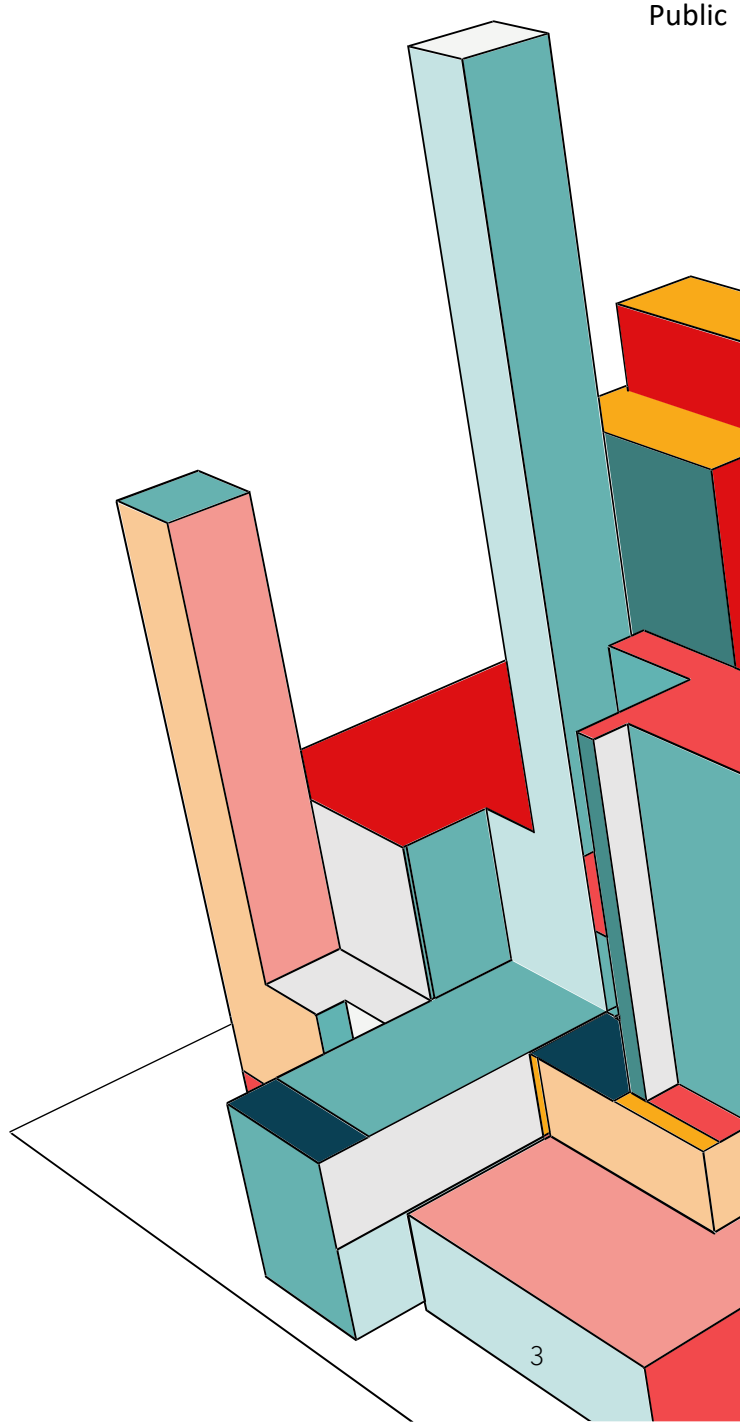
Mortgage valuation as is currently practiced doesn't take into advantage the wealth of data that currently exists

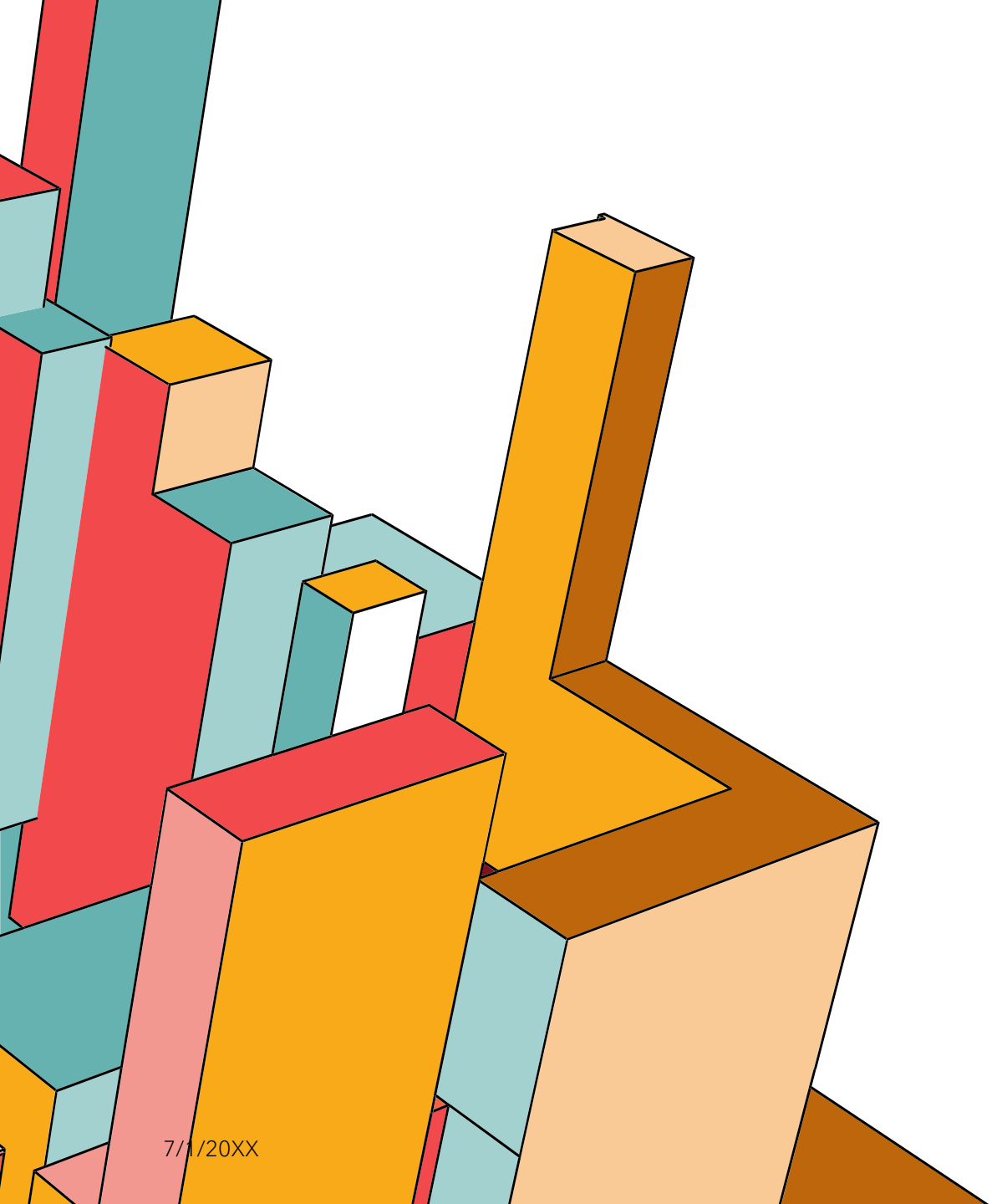
CUSTOMERS

There is a need to attract the best buyers through competitive pricing that focuses on the factors that a customer is most interested in

FINANCIALS

Regression models accurately predict house values to optimize Mipaka's pricing strategies to maximize profits





DATA UNDERSTANDING

King County House Sales dataset

The datasets contains valuable variables such as

- price,
- square feet covered by living room,
- year of construction,
- number of bathrooms,
- conditions of the house

TARGET VARIABLE

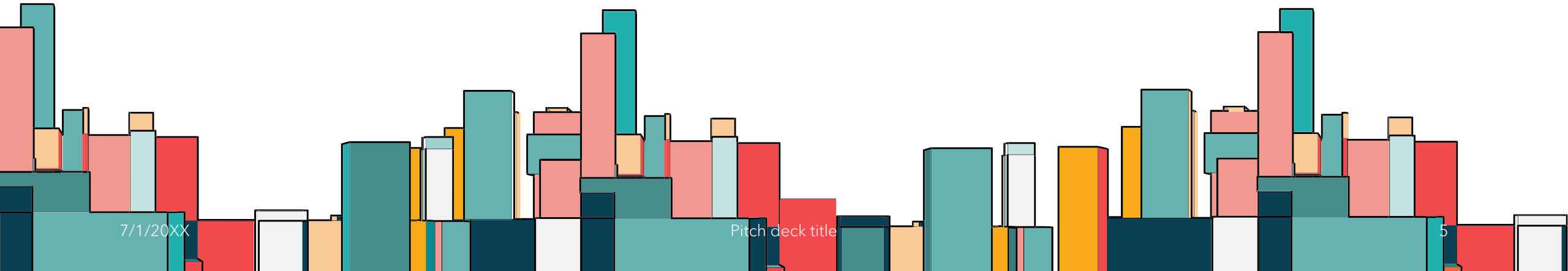
the **price**, it serves as the independent variable in our modeling, we determine what influences house pricing

BUSINESS STAKEHOLDER

MIPAKA

is a fictional company in the real estate field.

They specialize in mortgage evaluation they also assess property values to determine risk for market lenders.



MODELS

Sq. ft of the **living** room explains 49% of variance in price

Model I

Considered **age** of building, **living** room and number of **bathrooms**

Model II

Grade of house as categorical variable

Model III

MODELS

Model IV

- **Condition** of the house explains 64.0% of variance in price
- RMSE is 220,298.52, slightly lower than Model III
- **p-value** is below 0.05

Insights:

- cost of a house increases by $124.90 for 1 square foot increase in living space
- the cost of a house increases by $56,230 for each extra bathroom

Public

MODELS

Model V

- Sale **price** and **location** of the house (based on latitude and longitude)
  - **Location**s at centre of country have different prices

Prediction:

- Built two KRegressor models to predict houses based on location
  - k=25 and k = 1000

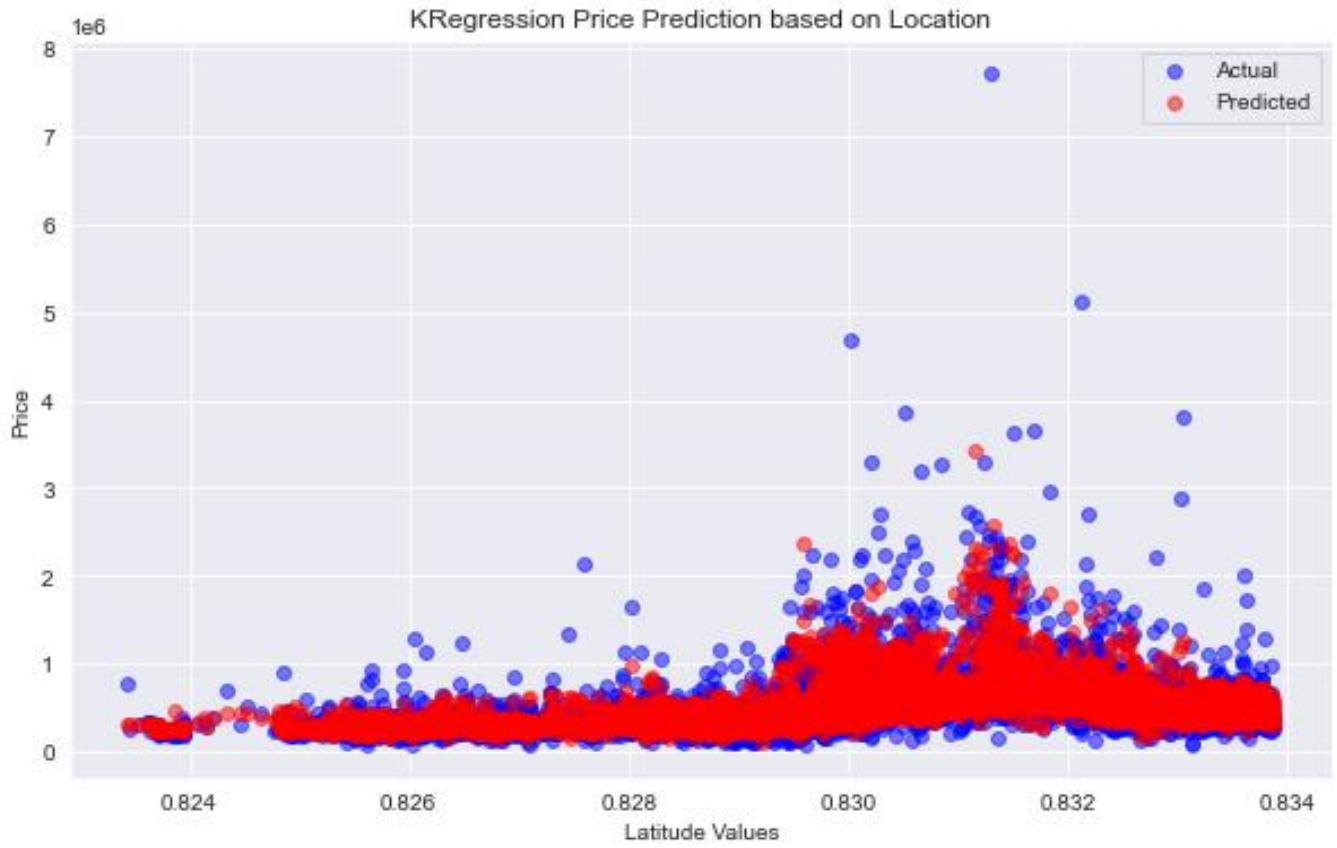
PREDICTION MODEL 1

Kregression Price Prediction based on Location

OBSERVATIONS

- Kregression model k = 25
- RMSE is 236992.59

PLOT



PREDICTION MODEL 2

Kregression Price Prediction based on Location

OBSERVATIONS

- Kregression model k = 100
- RMSE is 259889.70
- Prediction price on first model is off by $ 236992.59
- First model is predicting better than second model (RMSE is increasing with an increase in k-value)

PLOT

REGRESSION RESULTS

IMPROVE ACCURACY

- Mipaka should consider the **sq. ft** of the **living** rooms, **age**, number of **bathrooms**, **grade** and **condition** to improve price prediction
- These factors accounted for the most variance in change in prices

MAXIMUM RETURNS

- Mipaka should look to develop new housing developments on the Western Side of King County
- Prices towards the East show a slight decrease in price, therefore being less likely to bring in profits
- **Note:** Price prediction based on location is very sensitive to local variations

Public

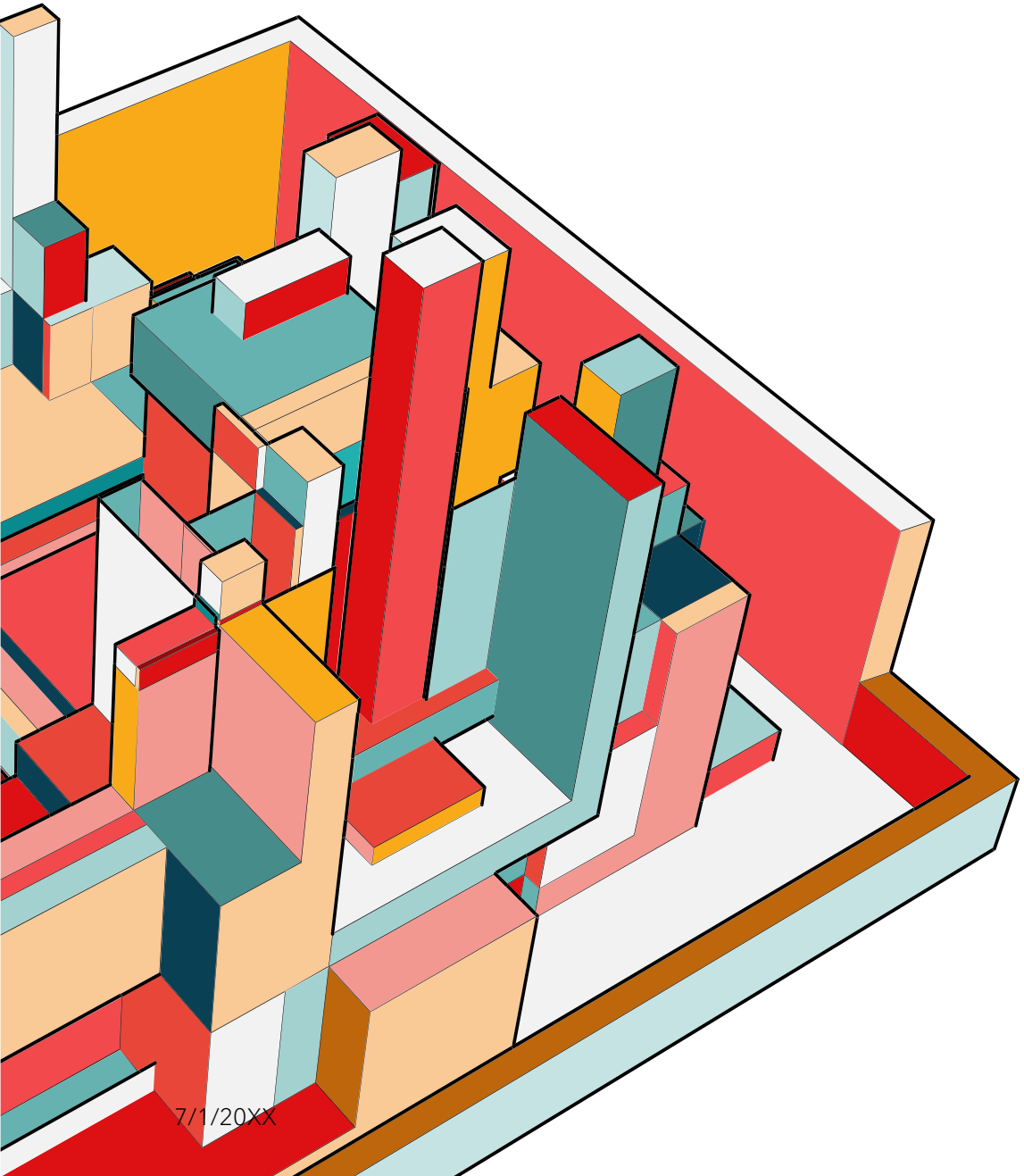
NEXT STEPS

EXPLORATORY DATA ANALYSIS

- More EDA on factors that might affect house prices
- Our best performing model accounted for less than 70% of the variance in sales
- Further exploration should be done on factors not in the dataset that might affect house prices

DIRECT MARKET RESEARCH

- Direct market research from real estate stakeholders
- Direct feedback will enhance the models by providing key dimensions missing in the datasets



CONCULSION

This analysis equips **Mipaka** with data-driven insights and predictive modeling strategies to make strategic decisions in the competitive real estate industry. By focusing on popular ways to determine fair pricing using factors like sqft living, age, grade, condition to improve its profit maximization