PROPERTY ANALYSIS IN KINGS COUNTY

Presented by Yessy Rayner

CONTENT

- Business Overview
- Data Understanding & Preparation
- Data Modelling
- Regression Results
- Recommendation/Next Steps

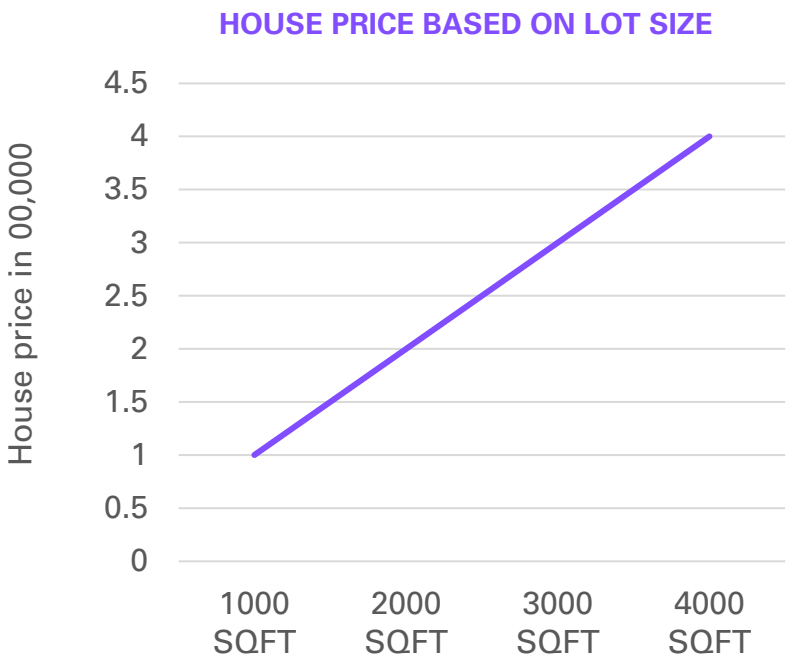
BUSINESS OVERVIEW

- KC Financial Investment would like to expand their business portfolio to include property buy and sell
- Area of interest: Kings County, CA
- Goal: Maximise profits through buy & sell by purchasing properties under predicted price
- Lucrative market due to short turn around process (~50 days in the market) and low outgoing fees (if they don't engage a real estate agent)
- Analysis: Build a model to predict property prices

DATA UNDERSTANDING & PREPATION

- Using Kings County's property sales datas in the past 2 years
- Over 21,000 sales
- Cleaning and reviewing all the datas available for analysis
- Break datas into 3 variables:
  1. Continuous
  2. Categorical and
  3. Discrete

DATA MODELLING



- Using **linear regression model** to predict house price based on numerous features
- Simple statistical way to measure the relationship between variables in order predict a value or a future outcome
- **Group the datas into 3 different variables:**
- **Continuous** - can be any value in a range, for example: sqft living and lot sizes
- **Discrete** - certain number of particular values that can be counted, for example: number of bedrooms, bathrooms
- **Categorical** - descriptive categories instead of numerical measured, for example: zipcodes, waterfront property (Yes or No value)

REGRESSION RESULTS

BASE MODEL:

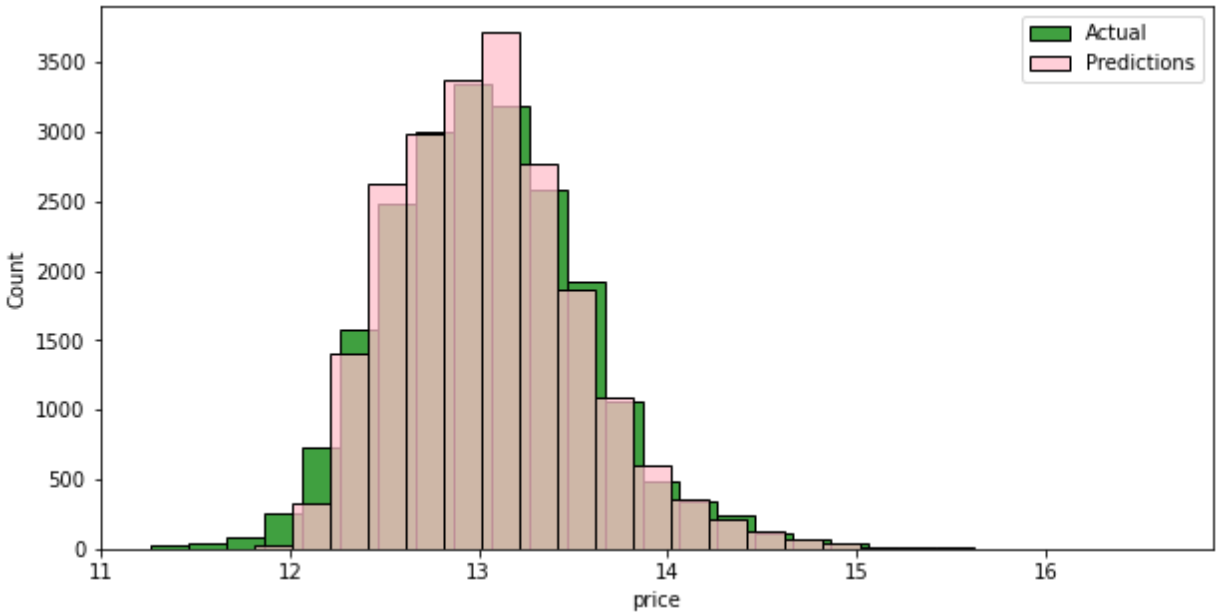
- Using 3 apparent base features (with the highest correlation/relevancy score)
- Sqft_living, grade, sqft_above
- R2 score at 0.564 (1 indicate perfect score)

FINAL MODEL:

- Using 12 features carefully selected
- Numerous testing, combining variables
- R2 score at 0.876 (87.6% accuracy in predicting house price)

REGRESSION RESULTS

ACTUAL VS. PREDICTIONS PRICE



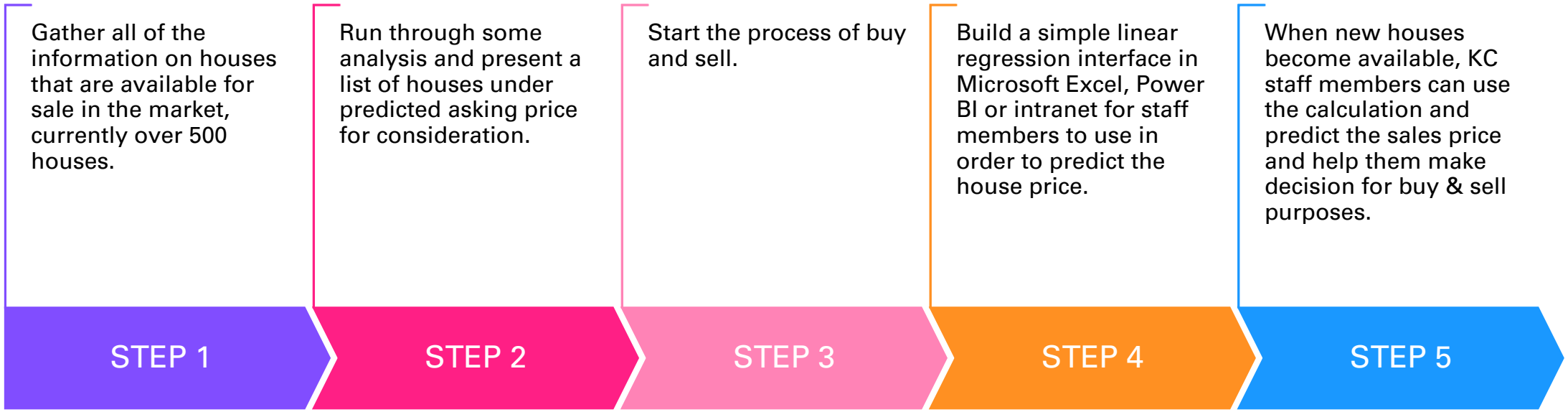
FINAL MODEL: Using 12 features

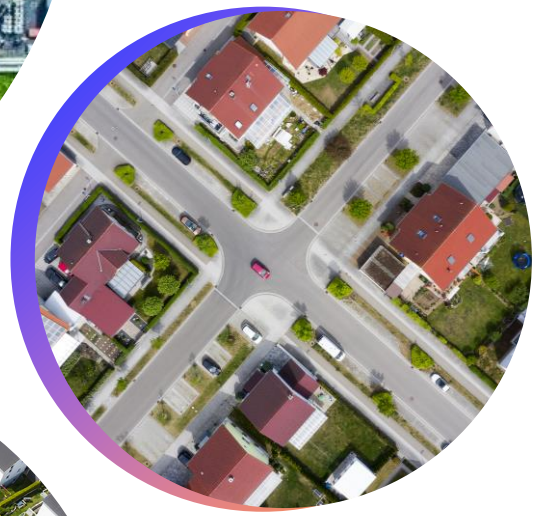
- 10 continuous & discrete features: yr_built, lat, long, sqft_living, grade, condition, bedrooms, bathrooms, floors, view.
  R2 score increased from 0.56 to 0.77

- ADD 2 categorical feature: zipcode & waterfront (Y/N feature)
  R2 score increased to 0.876
  (10% increased in accuracy)

LOCATION, LOCATION, LOCATION!!!

NEXT STEP/RECOMMENDATION





THANK YOU

Presenter name:

Yessy Rayner

Email:

yessy.rayner@gmail.com

LinkedIn:

https://www.linkedin.com/in/yessy-rayner-perth

GitHub:

https://github.com/YessyLee