Improving Real Estate Portfolio's Returns - Finding Investment Opportunities in Ames Housing Property Market

16th July 2021

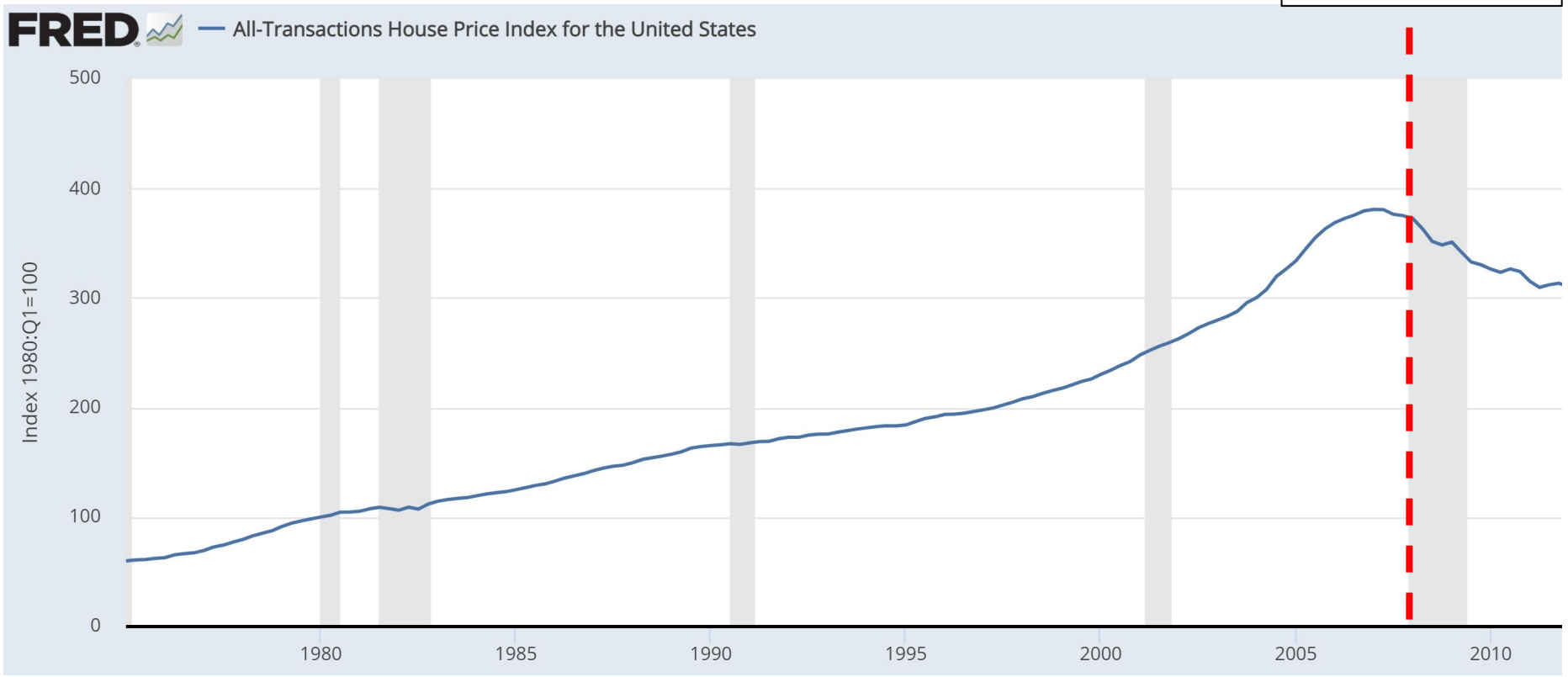
Valuation Team
Adriel Chen, Ash Ang, Estebelle Khong & Timothy Chan

Agenda

1. Introduction
2. Problem definition
3. Data preparation, engineering and analysis
4. Model selection and evaluation
5. Recommendation and conclusion

Dampening effect of 2008 Global Financial Crisis extends into recovery period

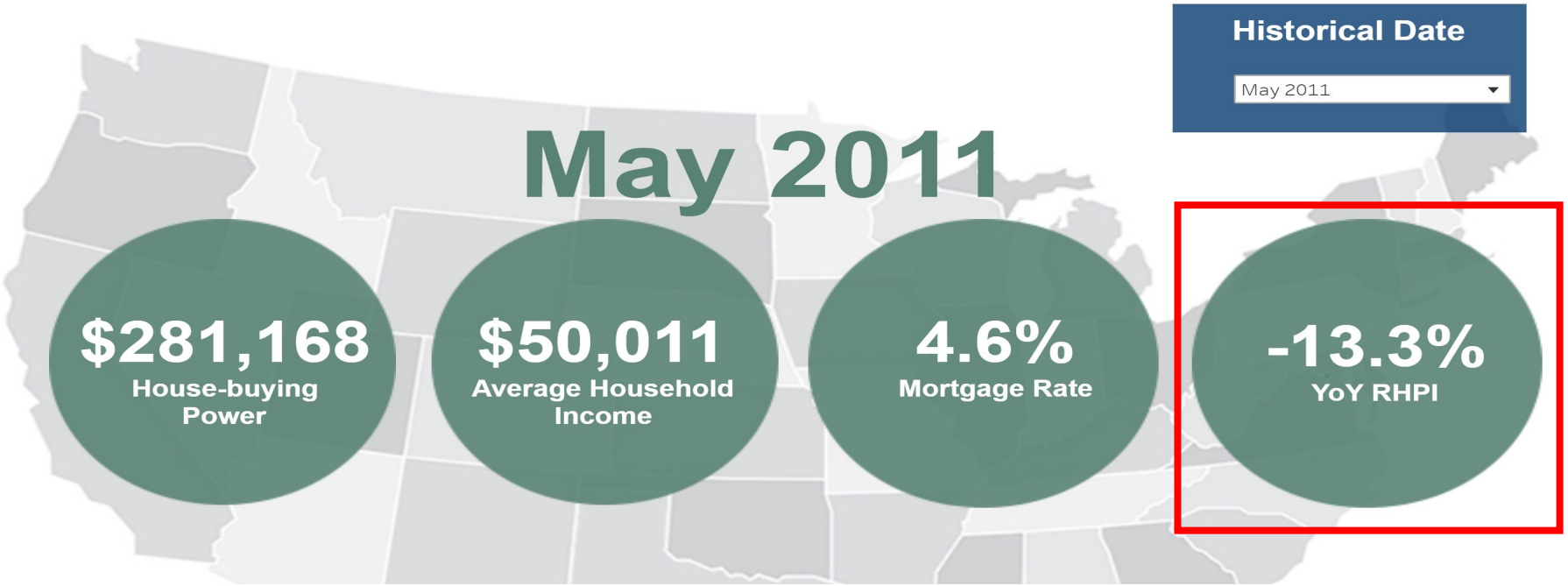
2008
Start of Global Financial Crisis



House Price Index for the U.S. is a weighted, repeat-sales index, meaning that it measures average price changes in repeat sales or refinancing on the same properties in 363 metropolises.
Source: https://fred.stlouisfed.org/series/USSTHPI

Snapshot of U.S. Housing Market

National



RHPI - Real House Price Index measures the price changes of single-family properties throughout the U.S. adjusted for the impact of income and interest rate changes on consumer house-buying power over time and across the United States at the national, state and metropolitan area level.
Source: https://www.firstam.com/economics/real-house-price-index/

Dismal performance of company's real estate portfolio

- 15.9 %

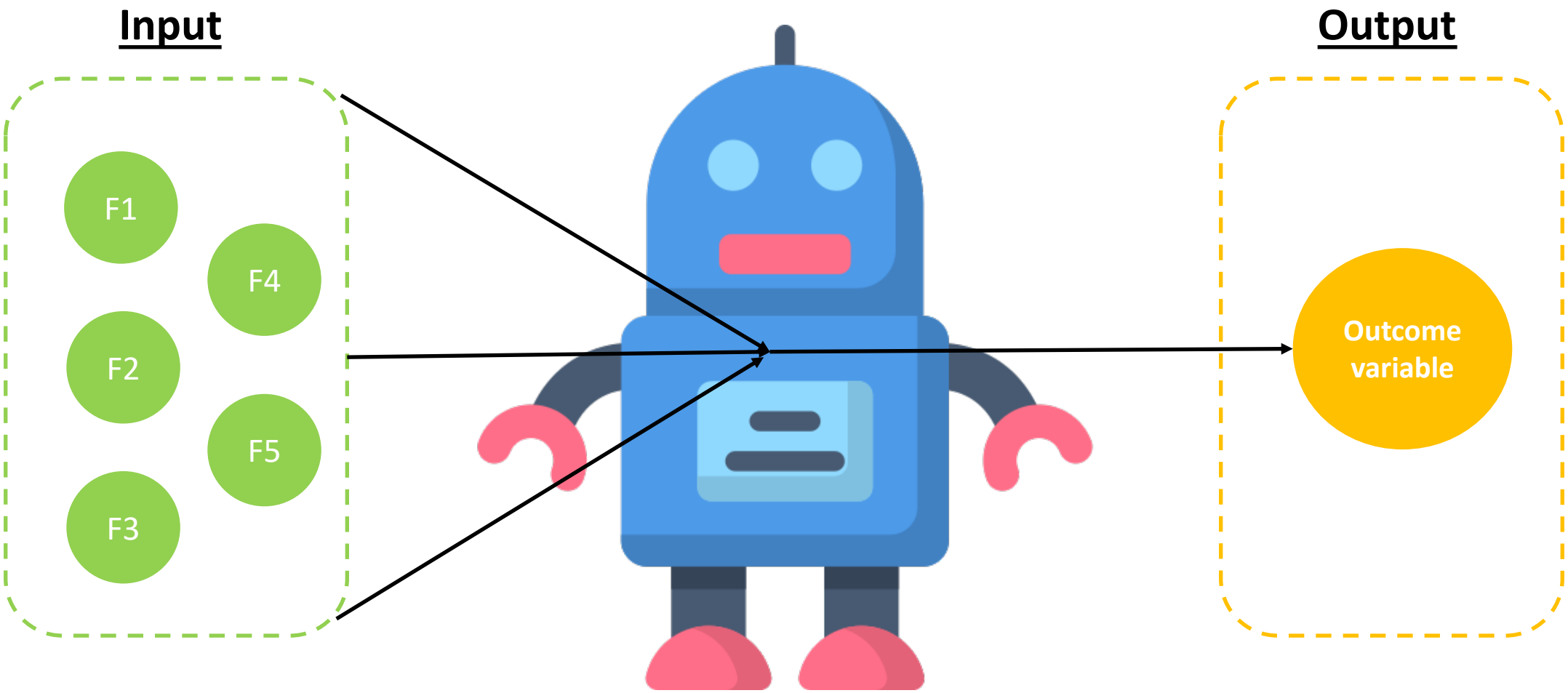
Company's adjusted YoY real estate portfolio returns

RHPI - Real House Price Index measures the price changes of single-family properties throughout the U.S. adjusted for the impact of income and interest rate changes on consumer house-buying power over time and across the United States at the national, state and metropolitan area level.

Objectives

1. Create a regression model to accurately predict Ames residential property sale price

What is Regression Analysis?



Motivations for using Regression Analysis

Traditional

- Common methods include, sales comparison approach, cost approach and income approach
- Useful to some extent
- Subjected to human biases such as confirmation and selection bias

Statistical

- Regression analysis or other more advanced techniques
- Flexible and works well with large dataset
- More objective and reduce human bias and errors
- Allows us to study relationships of independent variables and dependent variable on deeper level

Objectives

1. Create a regression model to accurately predict residential property sale price by reducing the incidences of human biases and errors,

2. Study how influential independent variables are in accounting for variation in sale price in the selected model,

3. Use it to guide buy/sell/hold recommendations and adjust real estate investment portfolio to maximize returns for shareholders and,

4. Make recommendations on how model can be fine-tuned with new data.

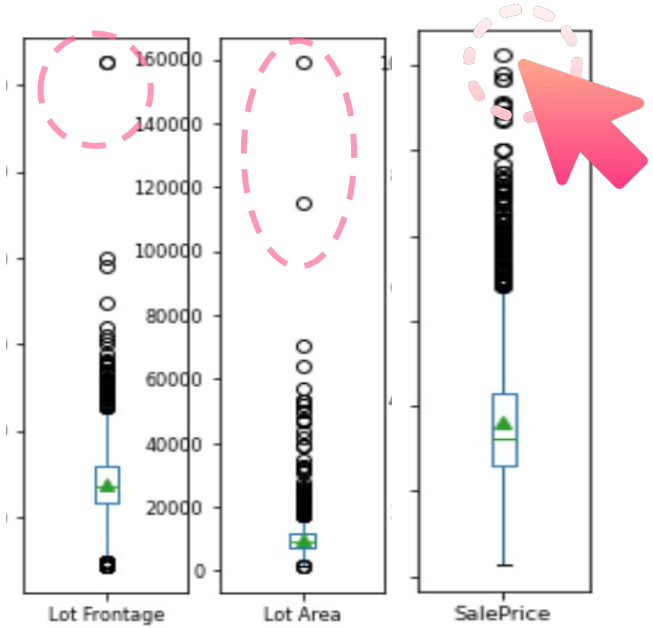
Data Preparation

2,390

Samples

Collected by Ames Assessor's Office, containing information of individual residential properties **sold in Ames**, IA.

Year 2006 ---------- > 2010



Extreme **outliers**, about 1% of the sample, are removed.

Non-residential properties (less than 1% records) are removed from the sample.

There are about **80 features** available for evaluation.

The features covered:

- Year built,
- Zoning, Street,
- Lot area/shape,
- Neighbourhood,
- Quality of house/building,
- Living area,
- Kitchen,
- Garage,
- Basement,
- Other additional features.

Correlation analysis was conducted among all the features, including our target Sale Price.

Any 2 independent features that are **highly correlated** (>70%), one of which is **removed** to minimize interference to the prediction model.

There are about 16 pairs.

Exploratory Data Analysis (EDA)

Features strongly correlate with Sale Price

Year Built

Overall Quality

Exterior Quality

Garage Quality

Sale Price

Basement Quality/ Size

Kitchen

Living Area

Legend:

80% correlated

70% correlated

60% correlated

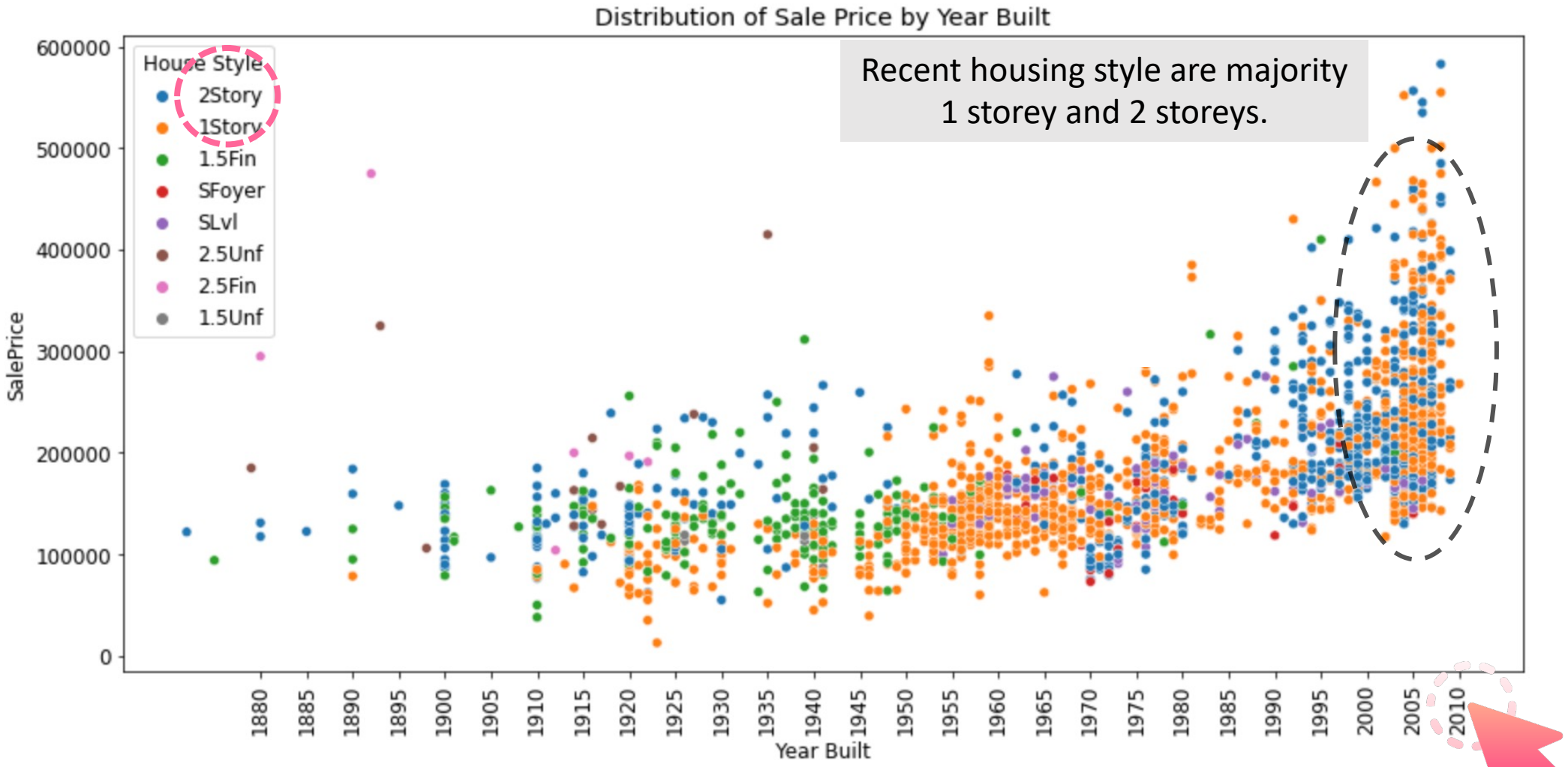
Introduction & Problem Definition

Data Preparation & Engineering

Modelling, Selection & Evaluation

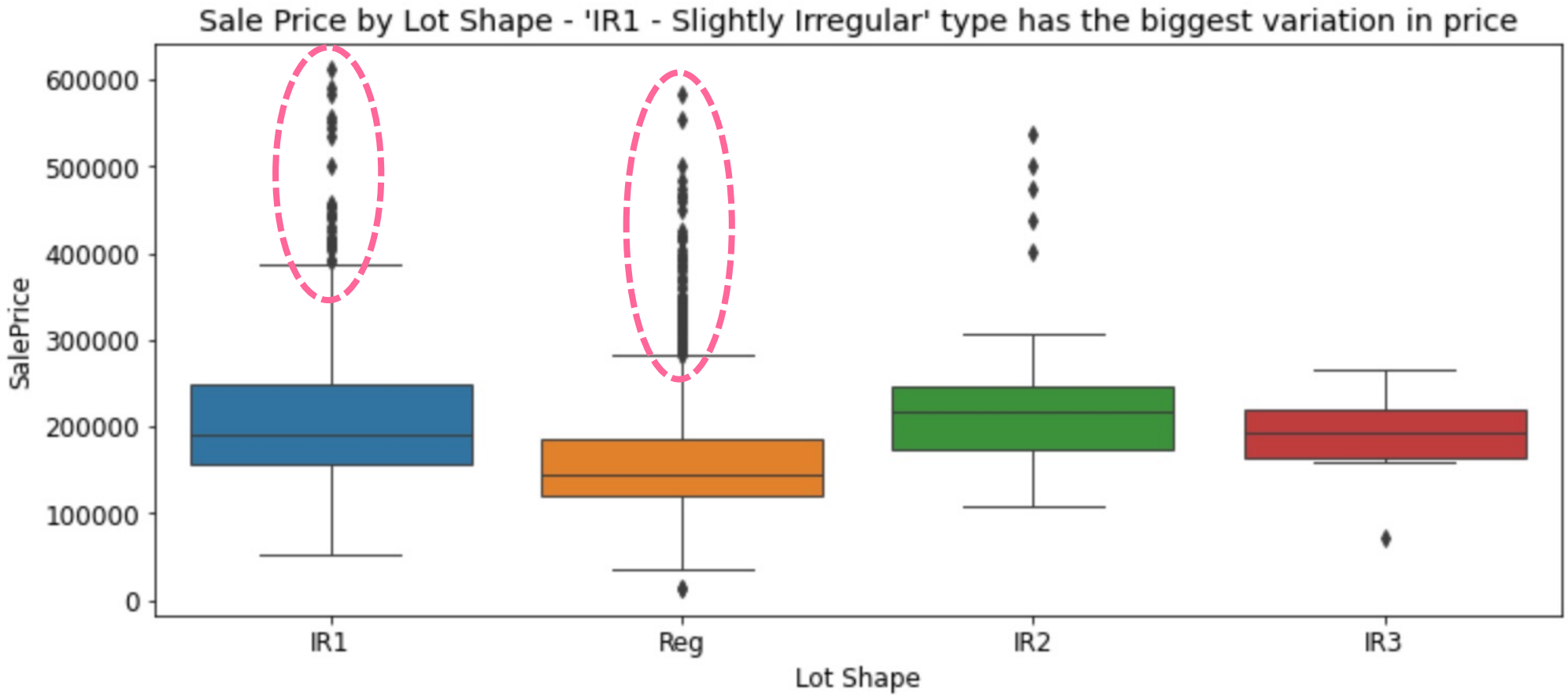
Recommendation & Conclusion

Sale volume is more promising if the built-year is more current.
Respectively, the sale price generally climbs if the age of unit is younger.

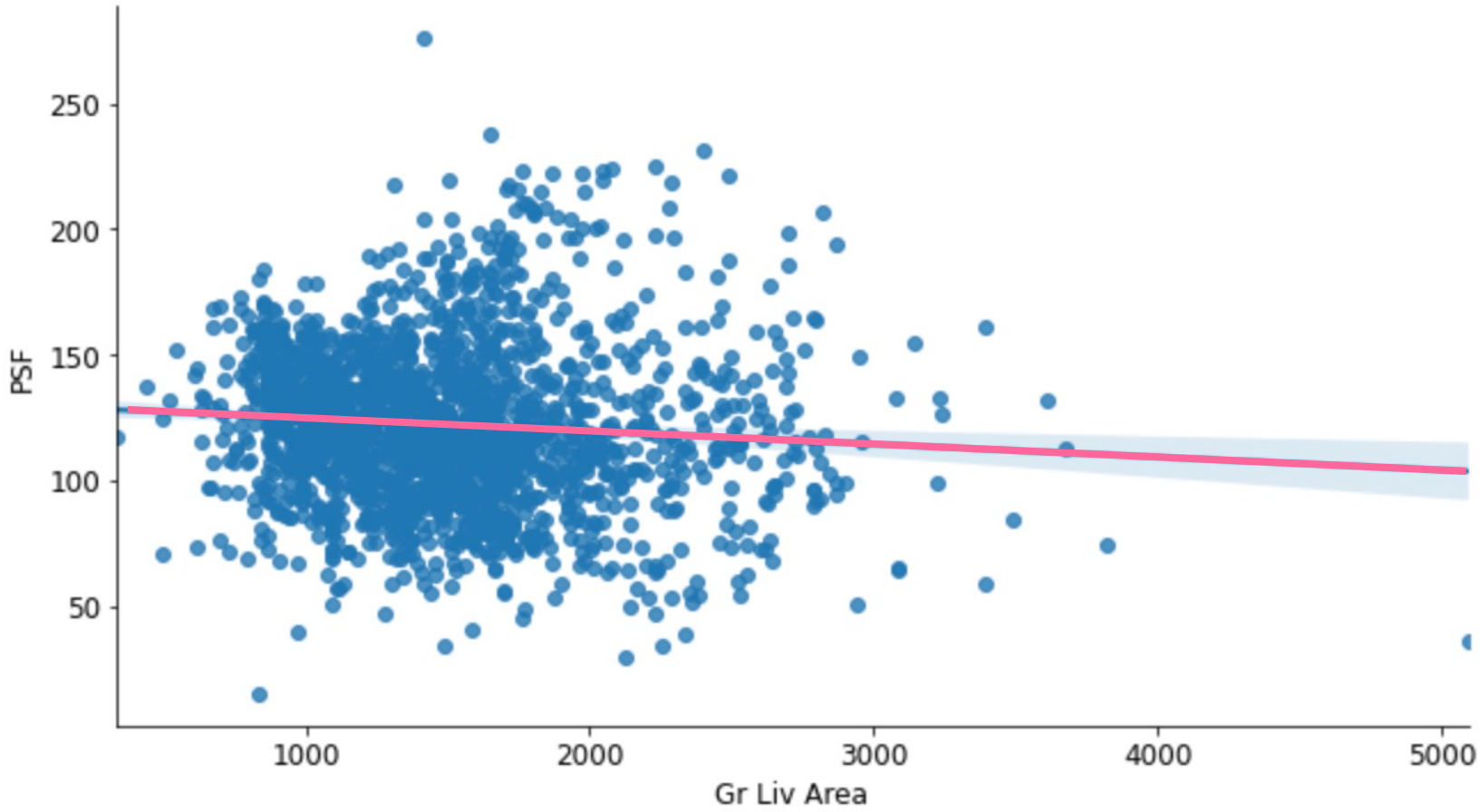


‘Regular’ and ‘Slightly Irregular’ house type made up almost 97% of the lot shapes.

Both display a larger lower and upper range, with noticeable outlier cluster.



Price-per-square-foot (PSF) has a negative linear relationship with ground living area. This relationship is valid because the bigger the house plan, all other things being equal, the cost can be spread over the square-feet. This is an important feature to consider in the model as it has direct impact on price.



Model Workflow

- Gather, prepare and partition the data
  - Train/Test Split
    - On the dataset, we performed a 70:30 - Train/Test Split

- Machine learning models used
  - Linear Regression
  - Ridge Regression
  - LASSO Regression

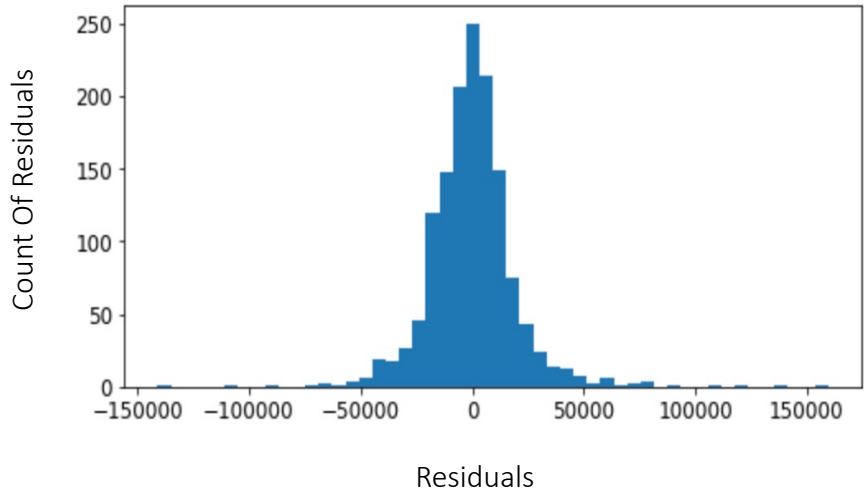
Evaluation and model selection

	Baseline Model (Null)	Linear Regression	Ridge Regression	LASSO Regression
R^2 Train Score	0	0.939	0.940	0.943
CV R^2 Train Score	NA	-7.753e+25	0.934	0.940
R^2 Test Score		-1.079e+23	0.853	0.822
Diff of CV Train Score & Test Score			0.081	0.118
RMSE	68,353		19,713	19,869

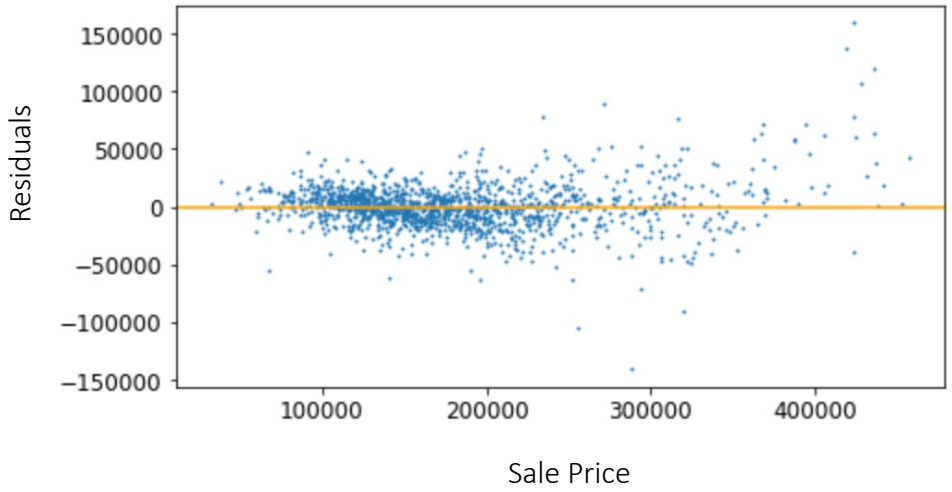
- Ridge Regression is the best performing model
- Ridge Regression Model has a smaller difference than Lasso Regression for the R^2 CV Train and Test score
- With lower RMSE, Ridge Regression will provide better predictions of Sale Price than the Null Model

Evaluation of Ridge Regression

Distribution of Residuals



Scatter Plot of Sale Price & Residuals



From the above plots, the model satisfies the 4 assumptions of **LINE**.

L - There is a linear relationship.

I - The errors are independent.

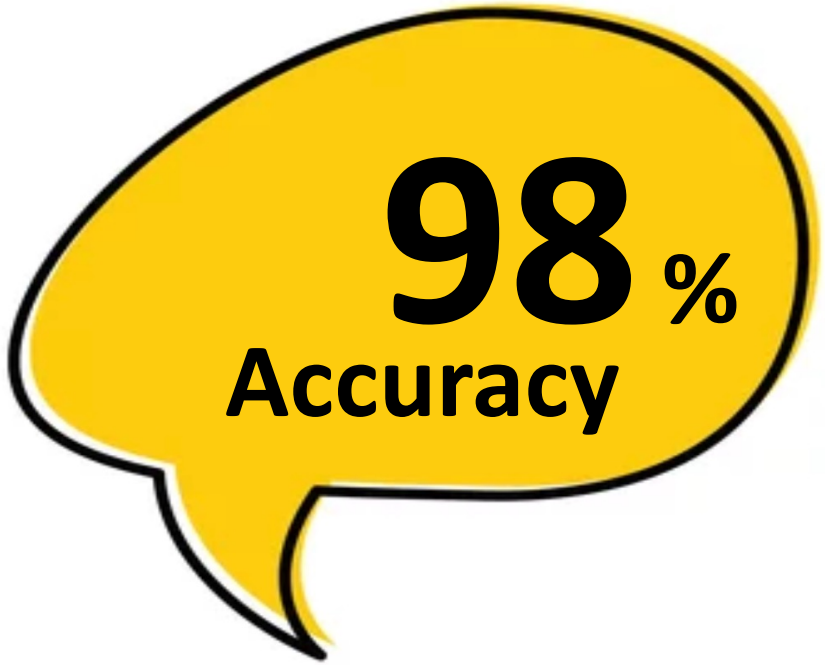
N - The errors are normally distributed.

E - There is an equal variance of errors.

Model Workflow

- Get predictions using the Ridge Regression Model
  - R^2 score for the dataset and the predicted sale price is **0.934.**
  - The score means that **93.4%** of the variability in the predicted price is explained by the features in our model.
  - This implies that the model would be very effective in predicting the sale price of a property.

Conclusion & Recommendations



- Production Ready

Conclusion & Recommendations

Above Ground (Grade) Living Area

Total Basement Area

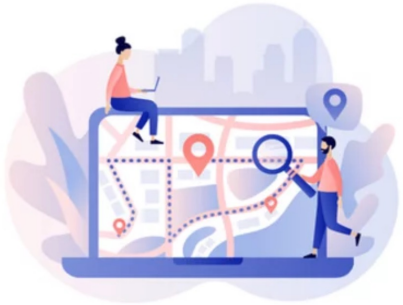
Overall Quality

Basement Height

Garage Size

Purchase Price < Predicted Price	**BUY**
Purchase Price > Predicted Price	**DON'T BUY**
Sell Price > Predicted Price	**SELL**
Sell Price < Predicted Price	**HOLD**

Going forward

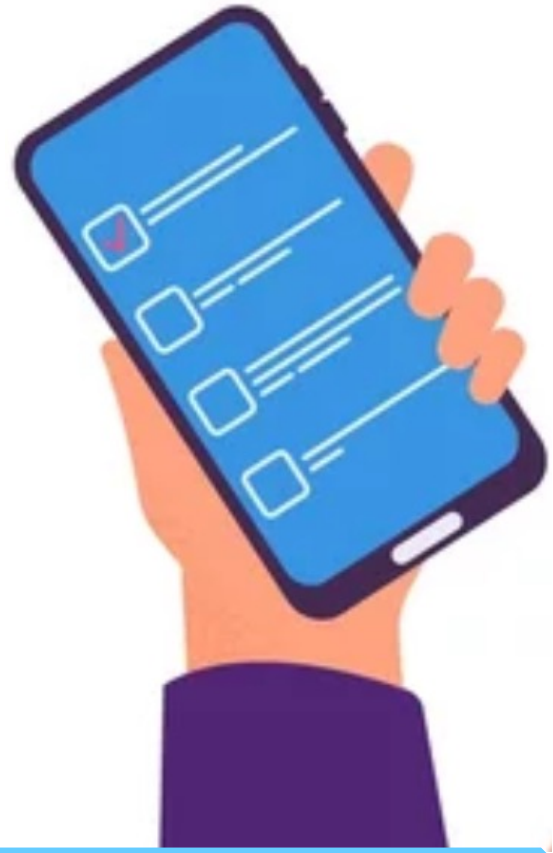


- Implementing this model in other towns
- Identify new features for finetuning
- Gather data about rental statistics

Going forward

Automation + Scraping

- MOBILE APP
- RECOMMENDATION
- GUIDE
- MONTHLY REPORT



Q&A