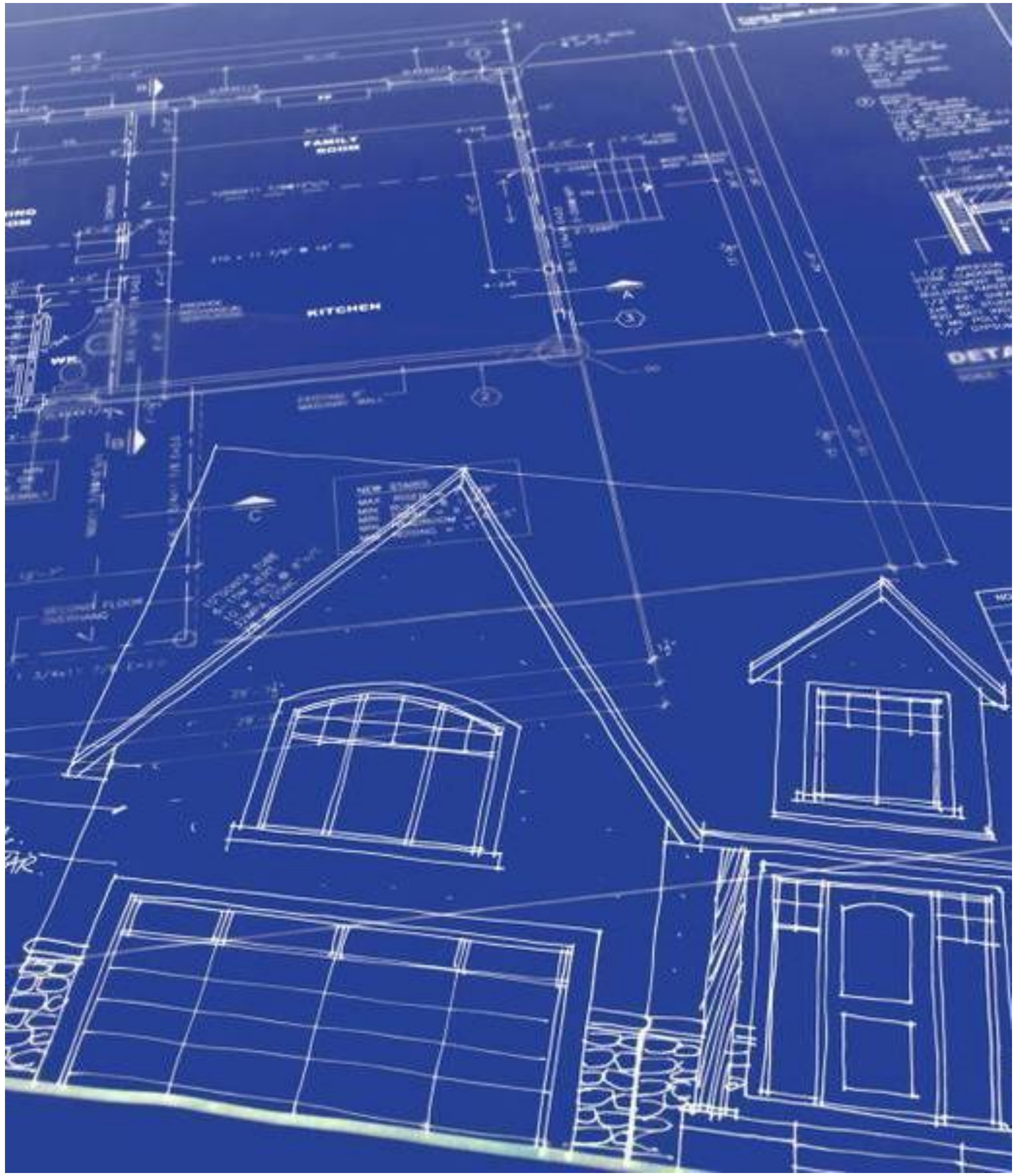
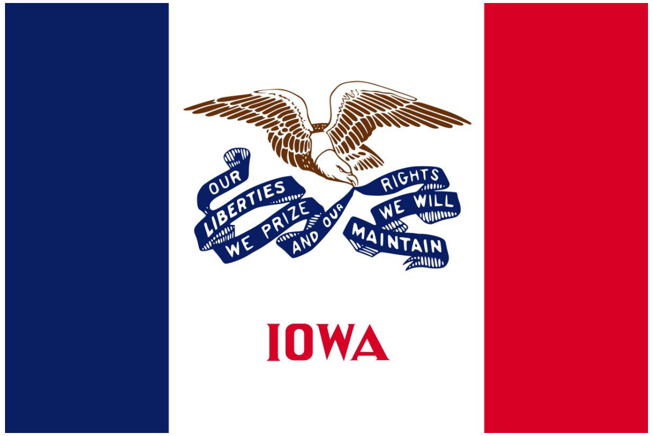
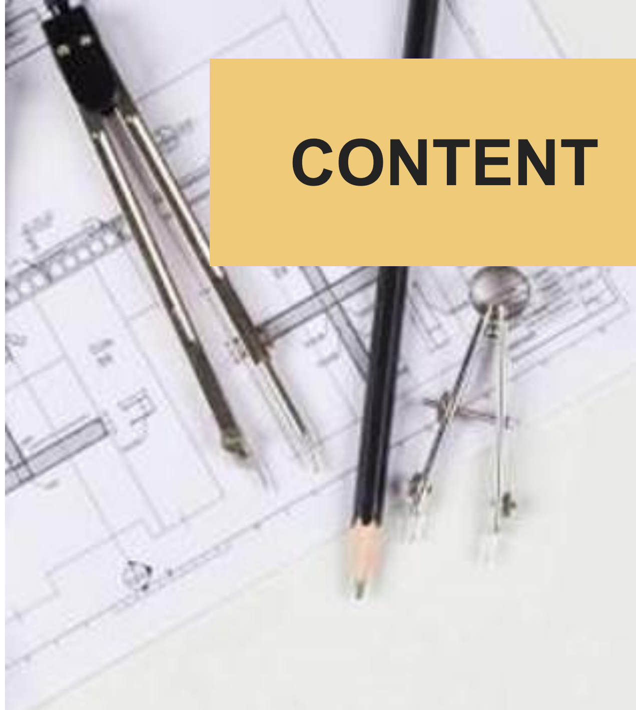




Project 2: Ames Housing Price Prediction

Team Members:

Cheryl Leong, Pan Kah Fei, Ong Song Yi(Group 6)



CONTENT

- **Problem Statement**
- **Background**
- **Prediction Model**
- **Methodology**
- **Analytical Result**
- **Summary & Recommendations**

Problem Statement & Background

OUR GOALS AND TARGET AUDIENCE

Problem Statement

We are a team of real estate consultants providing advice to property developers as clients for asset appreciation

Aims: Identify features with a strong positive correlation to the sale price of a home and generate business insights to maximize the ROI

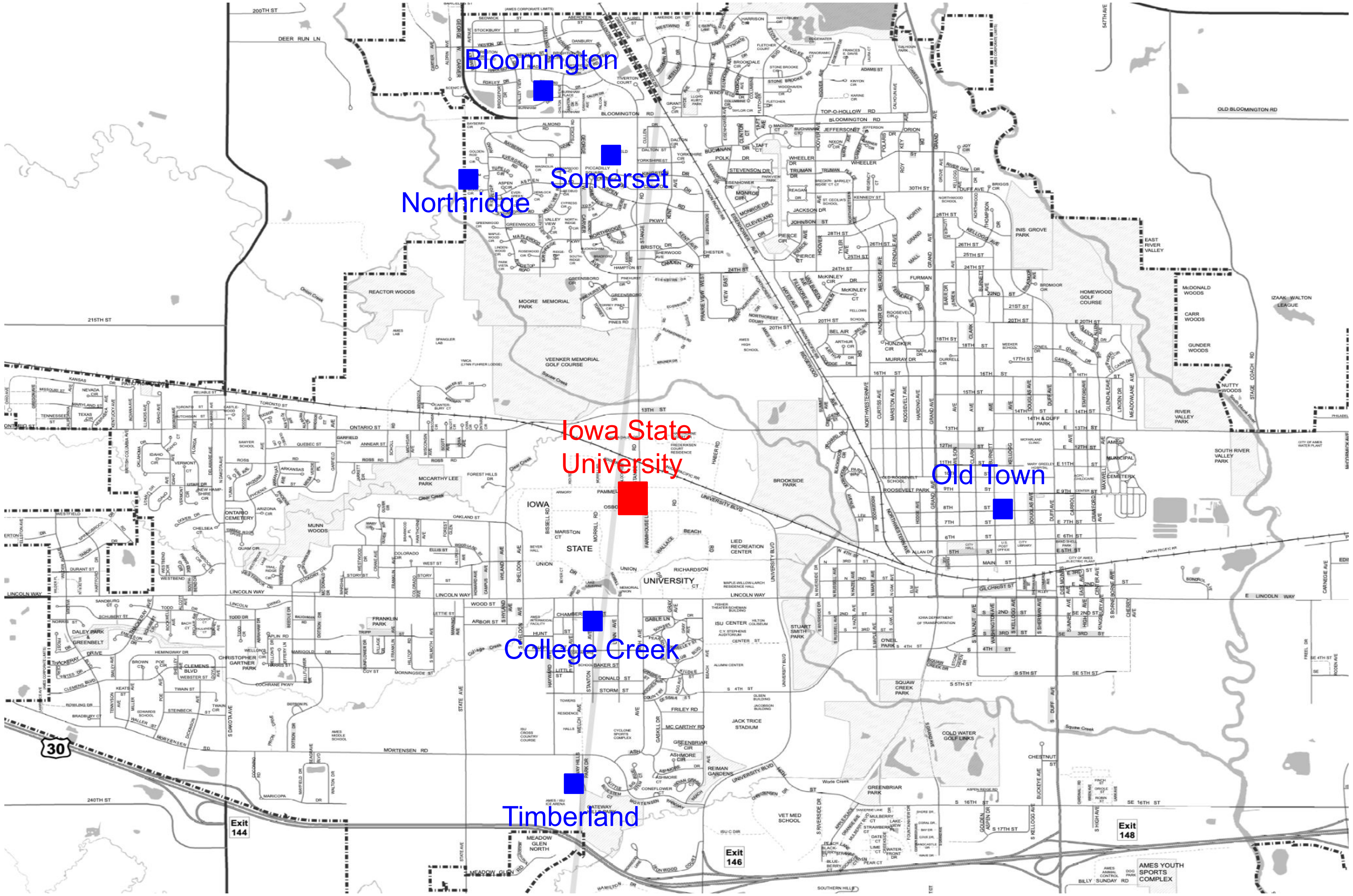
We will focus on the neighbourhood(s) as well as the features that can fetch the highest sale price

Background

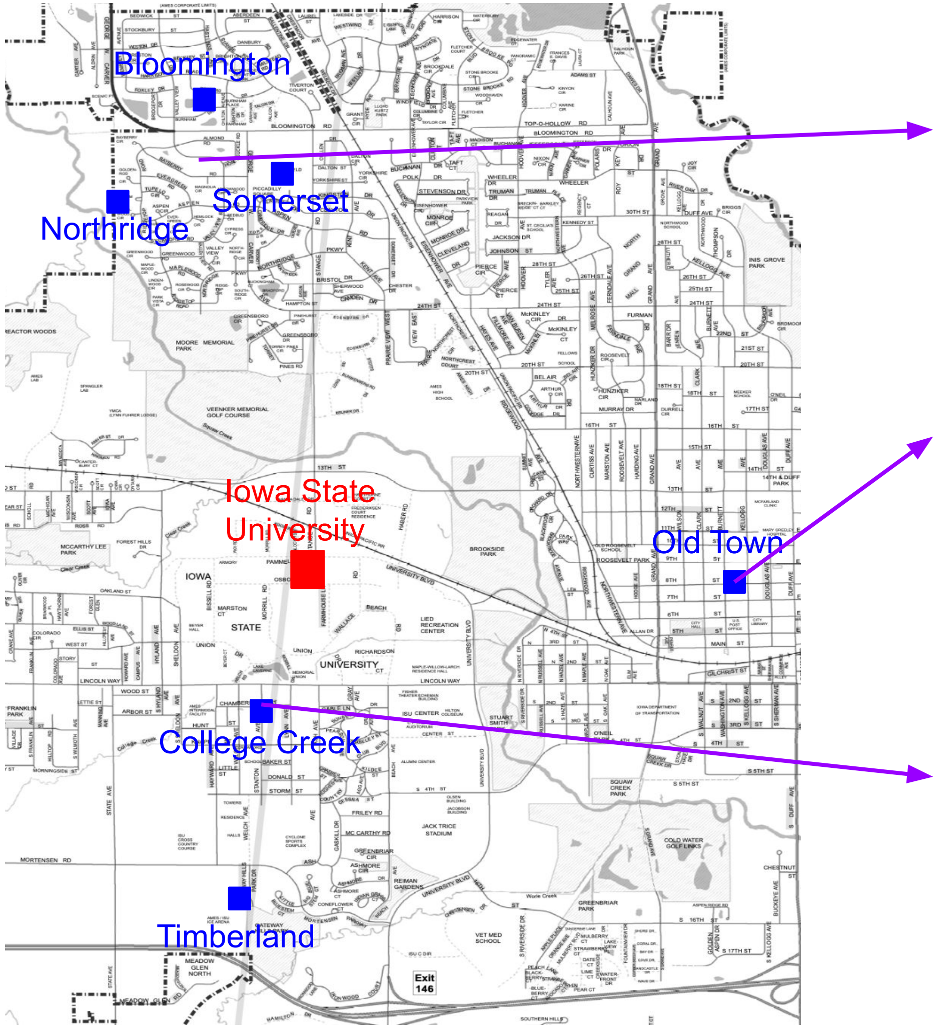
Ames is a city in the state of Iowa, USA. It is best known as the home of Iowa State University, with leading agriculture, design, engineering, and veterinary medicine colleges. It is the ninth largest city, with a population of about 67,000 people.

Ames is also known as a college town, where the students makes up about half of its population. This also means that property rental is a huge market in this city.

Background


Bloomington
Somerset
Northridge
Iowa State University
Old Town
College Creek
Timberland

Background



This is a cluster of new neighbourhoods. Amenities in the area are attractive to the college crowd, ranging from restaurants to cafes, as well as gyms and more importantly, department stores like Kohl’s, T.J. Maxx and Walmart.

Old Town is located north of Ames’ CBD. Old Town is identified as a historic district, and consist of properties that are ‘contributing’ or ‘non-contributing’. A property can change from “contributing” to “non-contributing” and vice versa if significant alterations take place.

Arguably the closest neighbourhoods to Iowa State University. However, from 2008 to 2010, there were studies and restoration work done to the area as it was facing soil erosion issues, which affected water quality and stability around the area.

source: [*https://www.cityofames.org/home/showpublisheddocument/6565/635809338107530000*](https://www.cityofames.org/home/showpublisheddocument/6565/635809338107530000)

Prediction Model

PROPERTY SALE PRICE PREDICTION

Model Comparison:

Rank	Model	Hyper Parameter	Train MSE	Test MSE	Generalisation (<5%)	Kaggle Score(Public)	Kaggle Score(Private)
1st	**Ridge**	**Alpha = 40**	**420,092,468**	**419,973,040**	**0.0284%**	**22,578**	**19,456**
2nd	Ridge	Alpha = 100	1,071,613,309	1,031,319,693	-3.76%	36,335	27,848
3rd	Linear Regression	(Polynomial) n = 2	681,045,924	683,769,285	0.3998%	196,362	197,842

- Model: Prediction of AMES Housing Sale Price
- The Ames Housing data examines the houses sold between 2006 - 2010.
- The Data contains 81 features and 1 output variable, the Sale Price.

Data Source: https://www.kaggle.com/c/dsi-us-11-project-2-regression-challenge

Modelling flow

EDA

Data Cleaning

- Remove features which have more than 3% Null values →
- Drop outliers and features not linear with Sale Price →
- Identify Correlation

Preprocessing

Feature Engineering

- Train test split (Test Size = 25%)
- Reduce Multicollinearity by utilising VIF
- Correct skewness of feature
- Standard Scaler plus 1 Hot Encoder

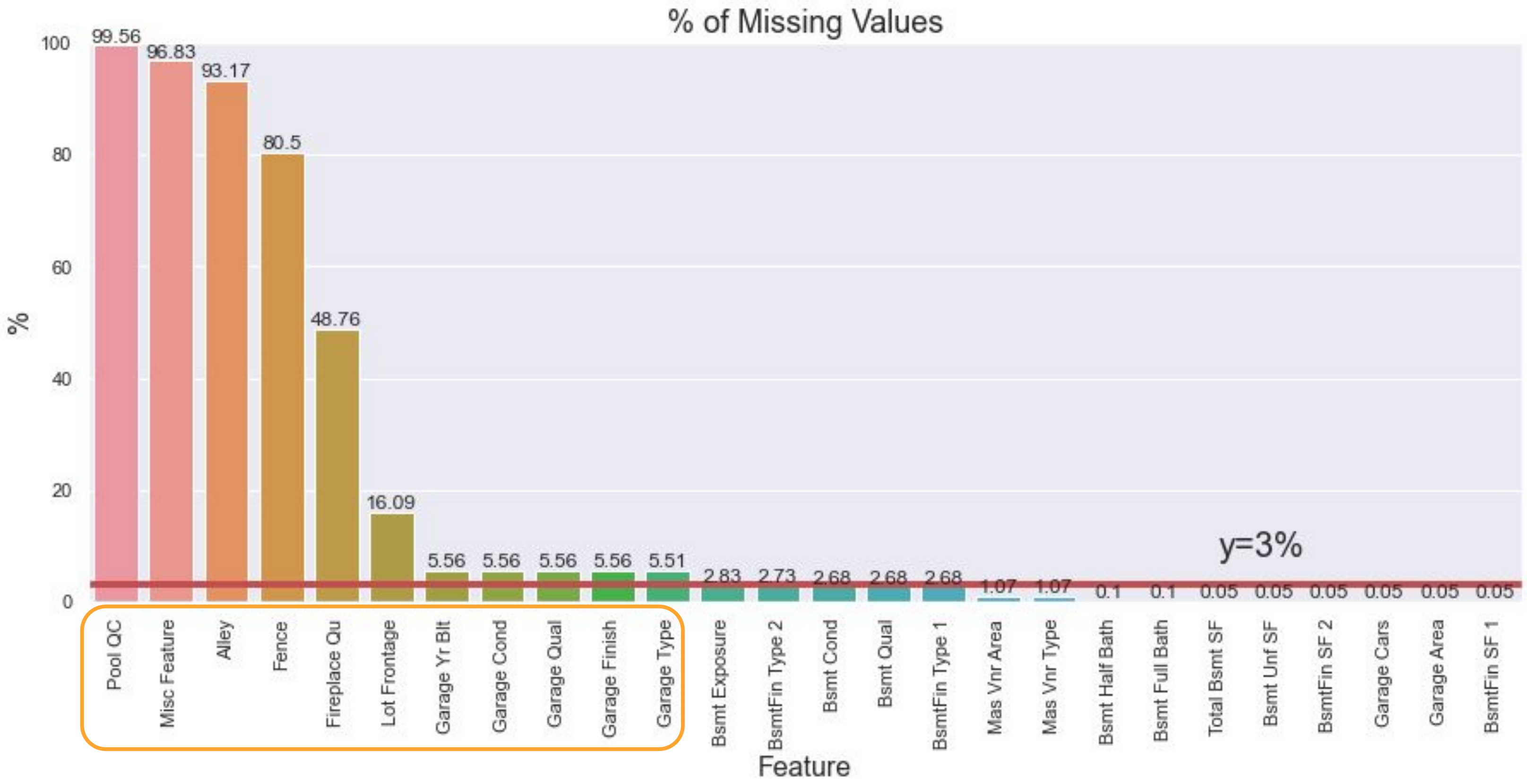
Model Selection

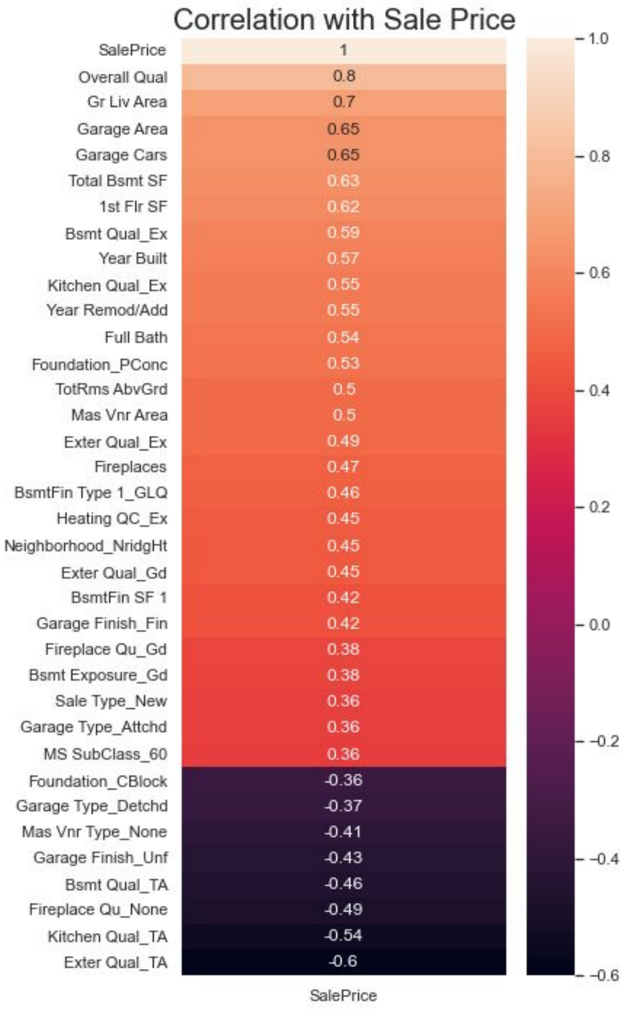
- Grid Search for Hyperparameter tuning
- Optimize Generalisation (≤ 5%)

Final Model Evaluation

- Distribution of Standard Error
- Equal Variance Error →
- MSE Check

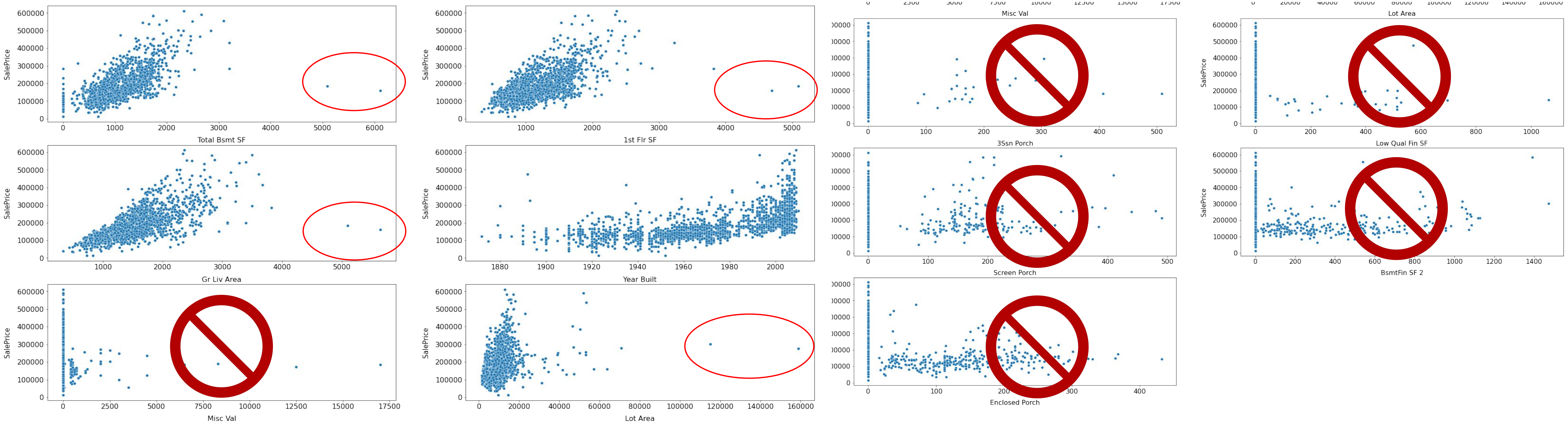
Null Values and Correlation





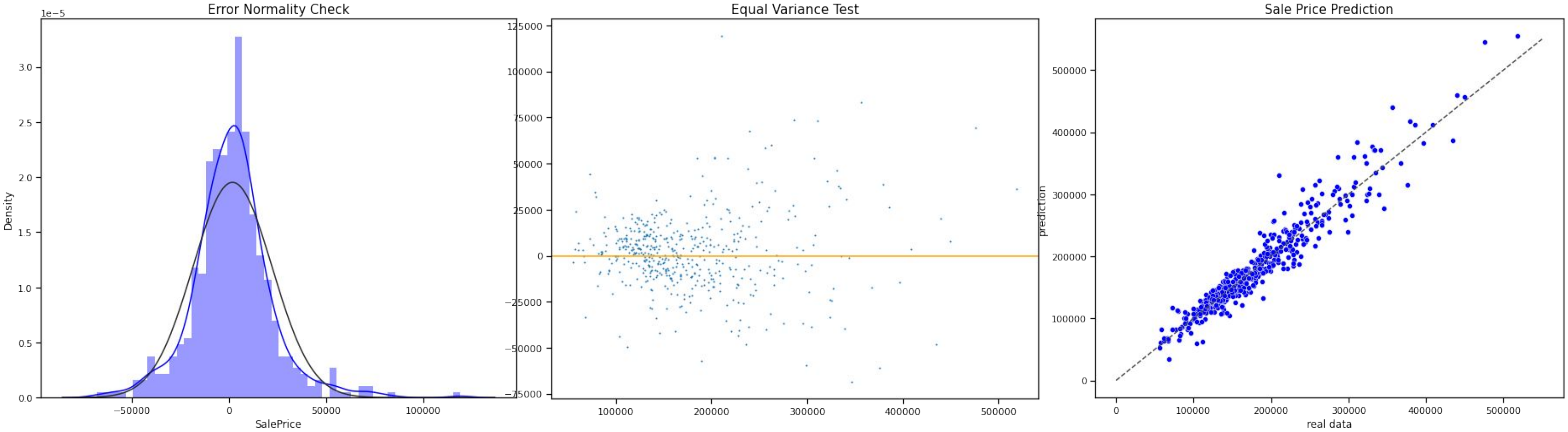
- Features in the box were dropped
- Dropped features account for about 3% of the total dataset

Outliers and Linearity



- Outliers in the dataset are removed.
- Features that do not have a linear relationship with sale price are dropped

Final Model Evaluation



Rows removed: 3.22%

Columns removed: 35 out of 80

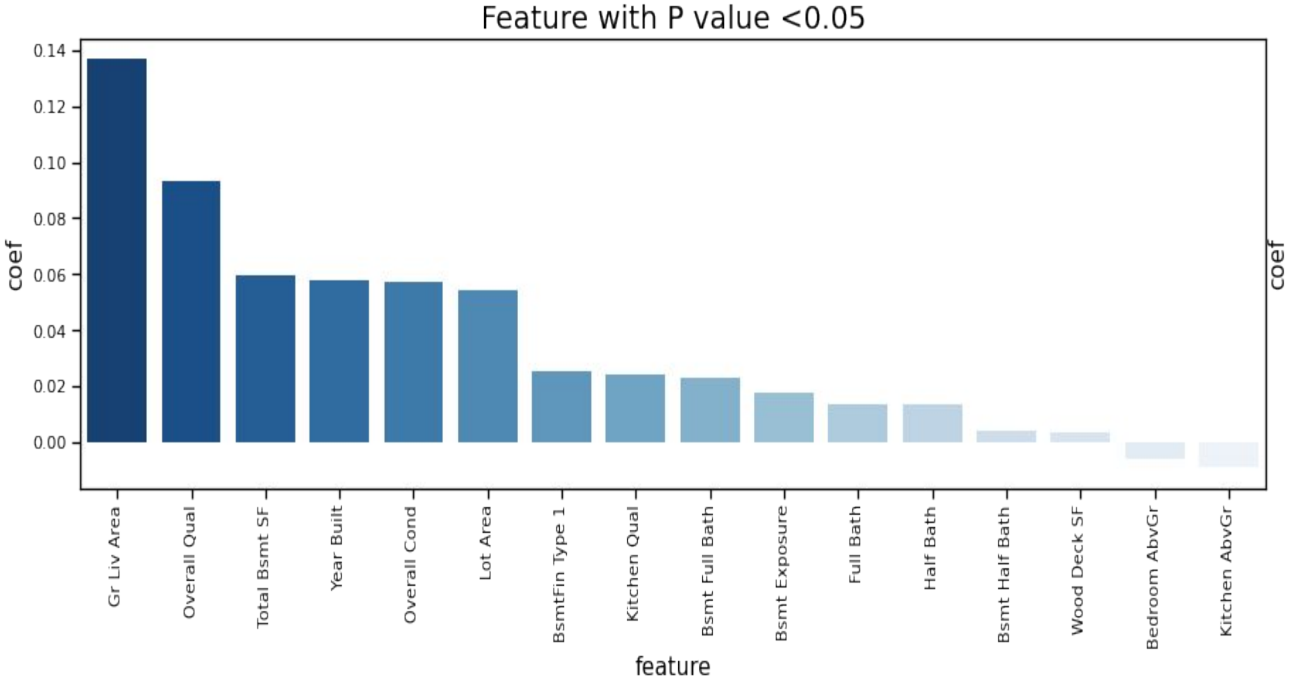
Prediction Model: Ridge (Alpha=40)

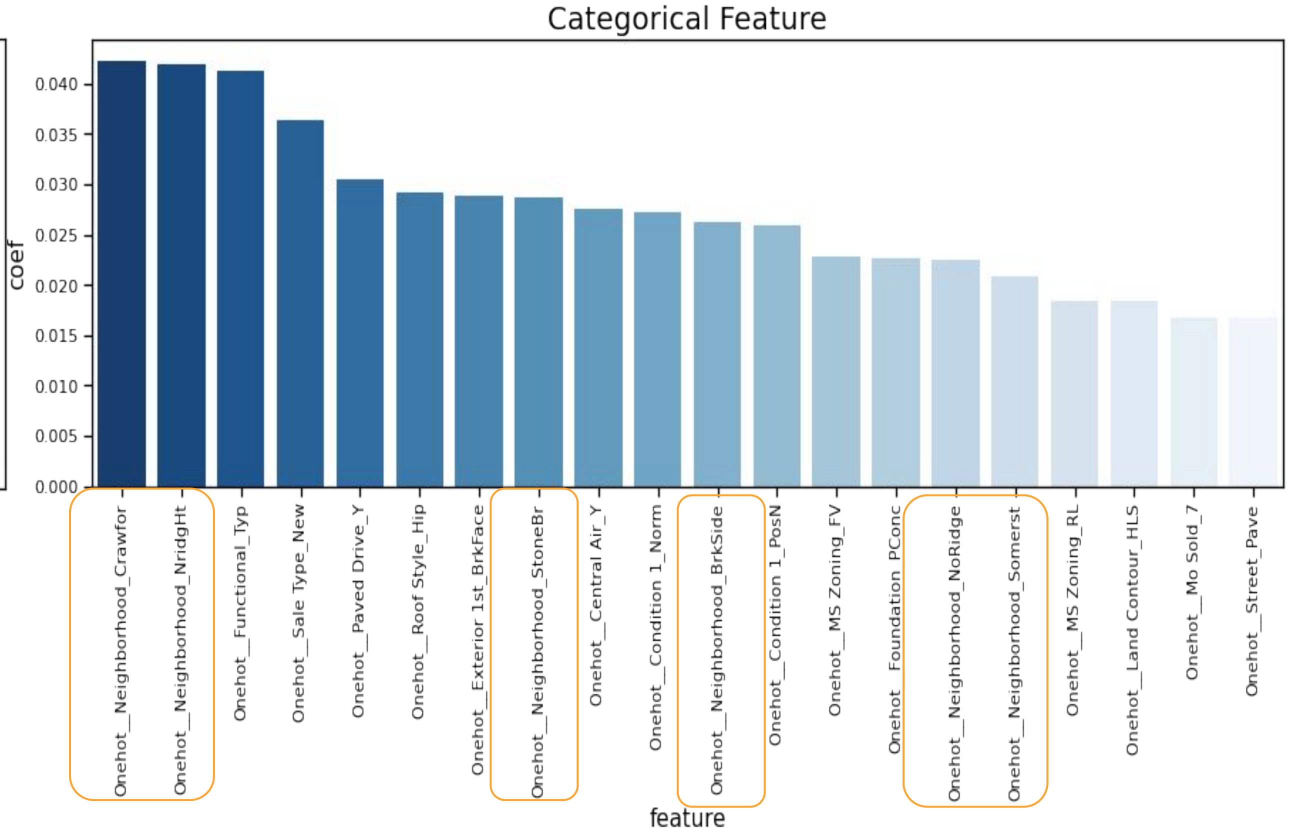
Model generalisation : 0.02 % MSE different between test/train

Methodology

UTILIZE MODEL FOR ANALYSIS

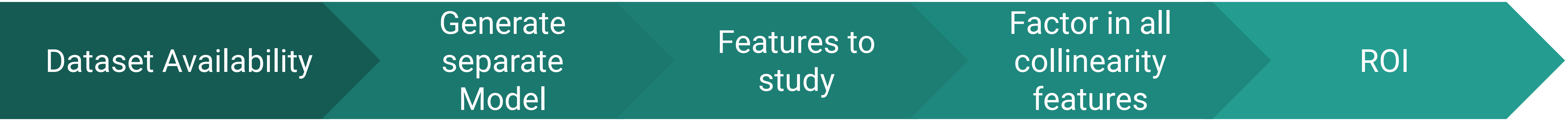
Coefficient from the General Model:





- Most of the numerical features have 95% confidence interval, but none of the categorical feature is true in this case
- Numerical features has relatively higher coefficient
- Neighbourhood has higher association with sales price compare to other one hot encoded features

Flow to Generate Investment Idea by Selecting Neighborhood


Dataset Availability

Generate separate Model

Features to study

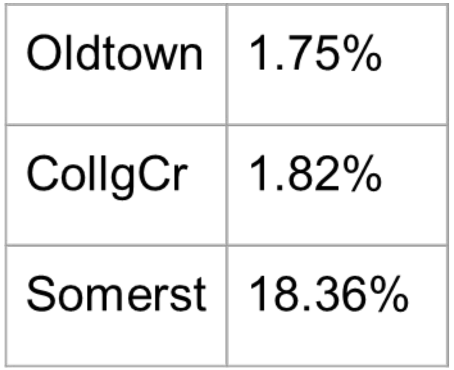
Factor in all collinearity features

ROI

To generate a good model, the top neighbourhoods with most number of data available is considered:

1. NAmes
2. CollgCr
3. Oldtown
4. Somerst

Fit in general model and get the MSE score to compare with training set.
Top performers:



Oldtown	1.75%
CollgCr	1.82%
Somerst	18.36%

1. **Bedroom**
2. **Full Bathroom**
3. **Half Bathroom**

From 3 separate models, all correlation within the numerical features are factored in to generate sales price prediction

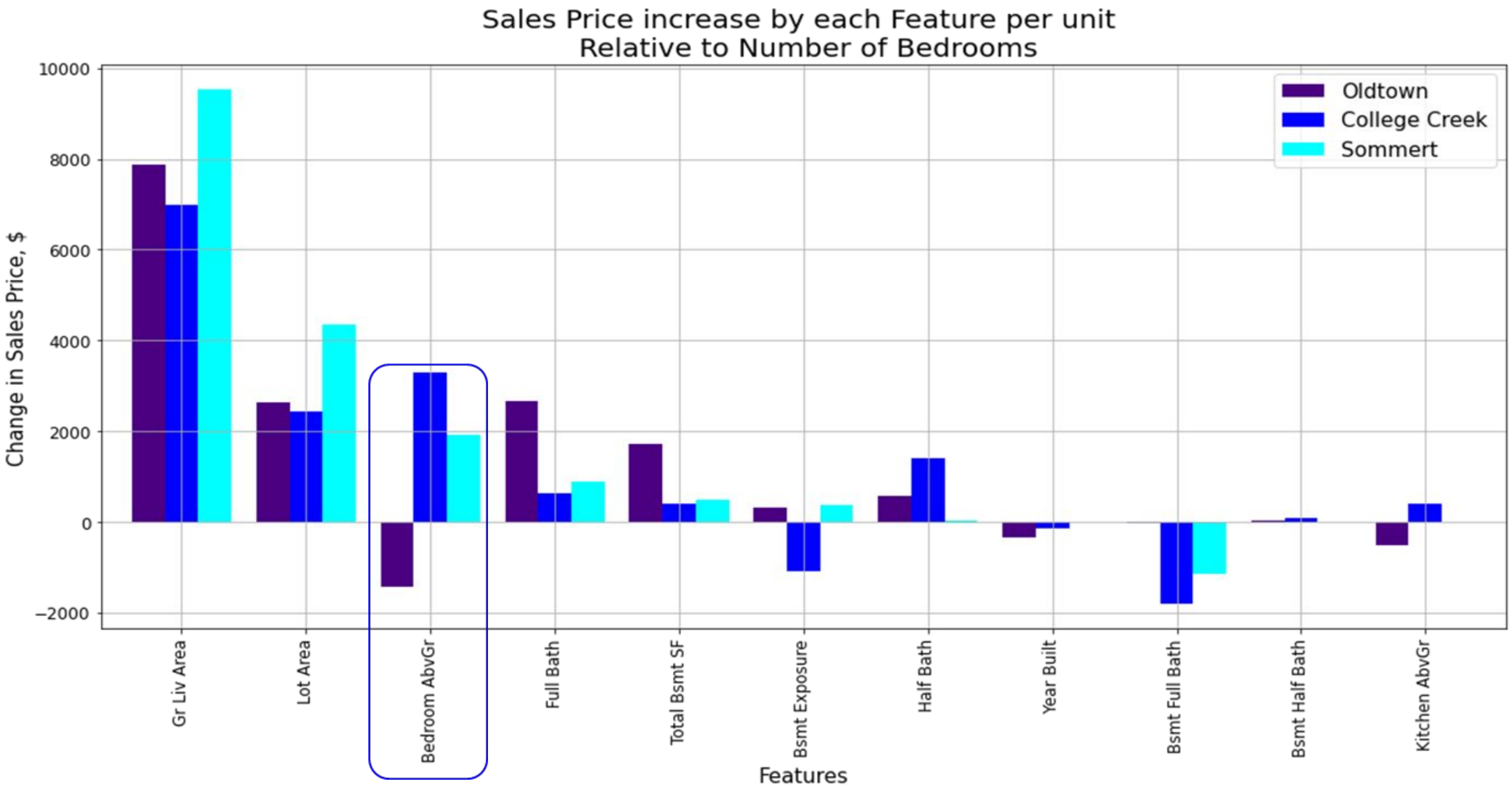
Generate recommendations to invest on the features in different neighbourhoods

Average building cost for each features are taken from: https://homeguide.com/costs/

RESULT

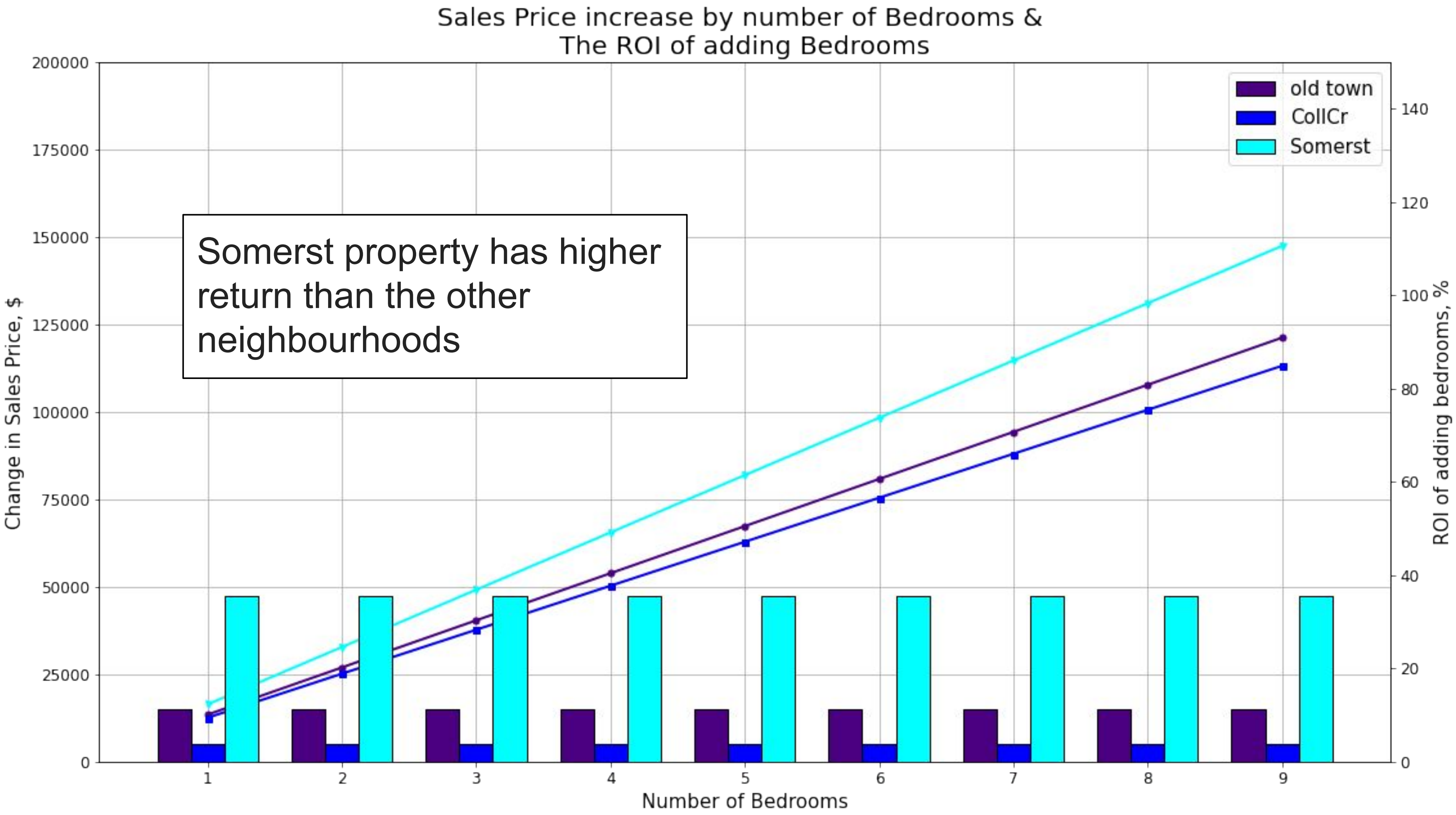
ANALYSIS

Factor in correlation with other features relative to Bedroom



1st Feature: Bedrooms

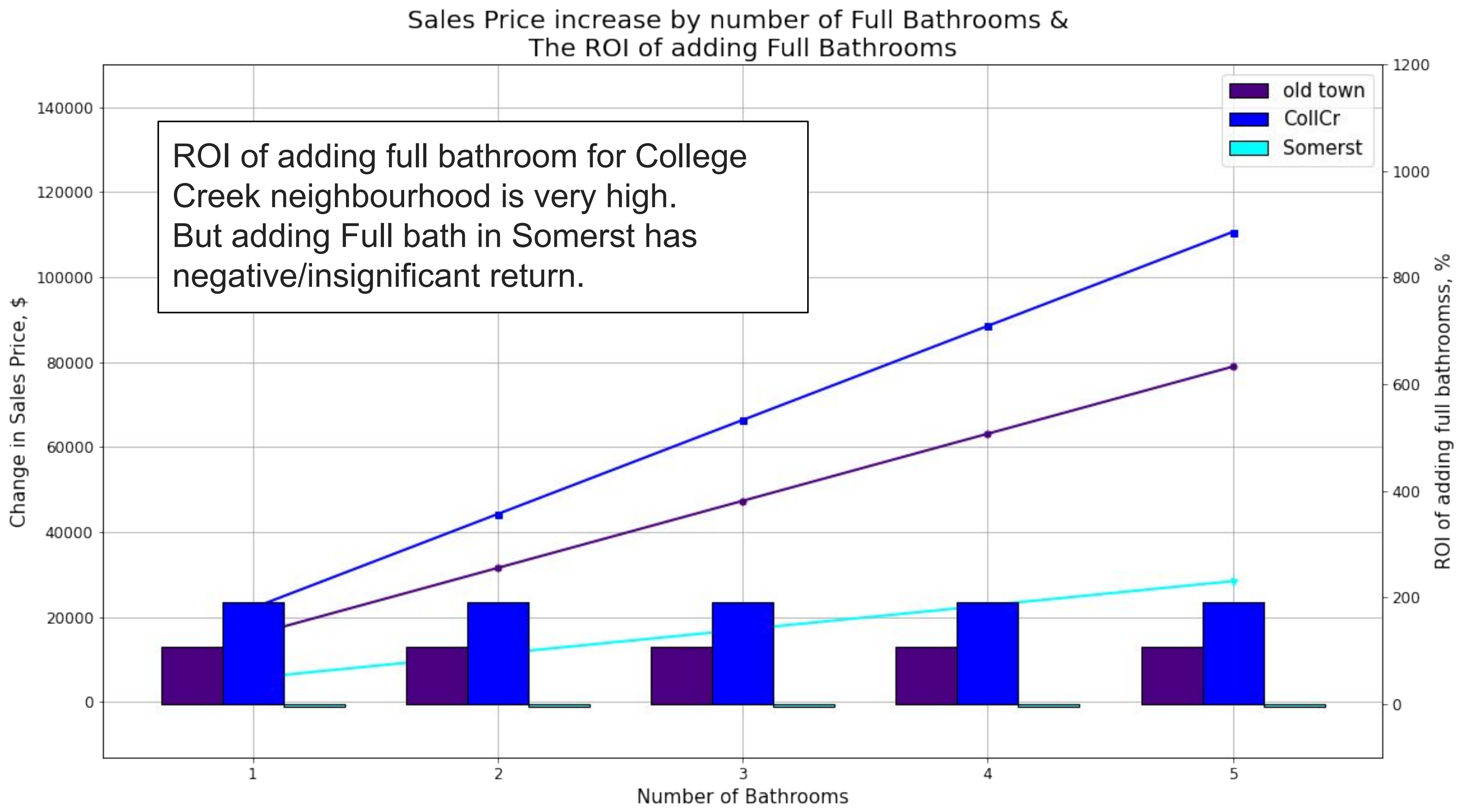
Increase number of bedrooms increase the property sale price with positive ROI for all neighbourhoods



Somerst property has higher return than the other neighbourhoods

2nd Feature: Full Bathrooms

Increase number of Full Bathroom in property increase the sale price, with positive ROI only in Old Town and College Creek

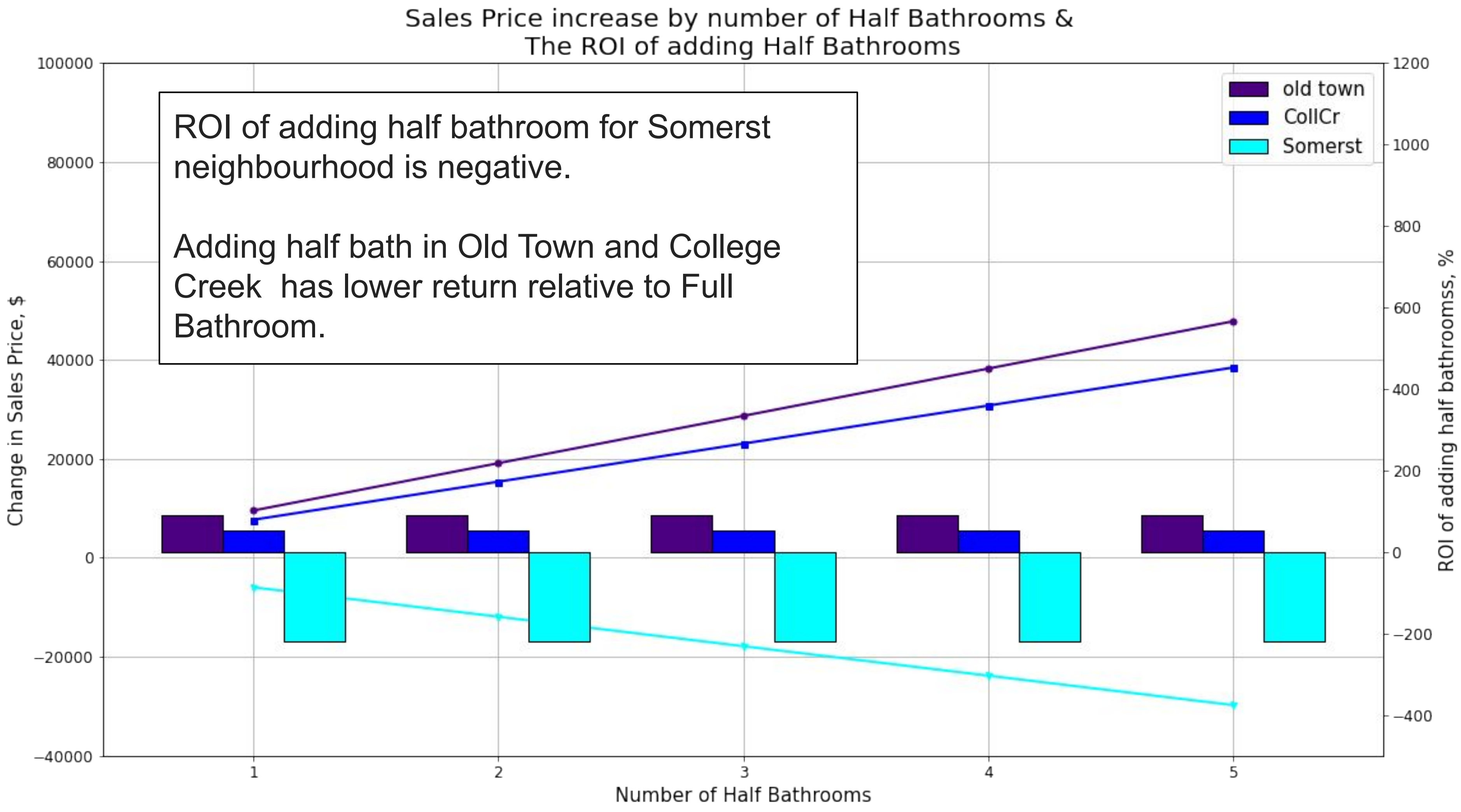


Sales Price increase by number of Full Bathrooms &
The ROI of adding Full Bathrooms

ROI of adding full bathroom for College Creek neighbourhood is very high.
But adding Full bath in Somerst has negative/insignificant return.

3rd Feature: Half Bathrooms

Increase number of Half Bathroom in property increase the sale price for only Old town and College Creek



ROI of adding half bathroom for Somerst neighbourhood is negative.

Adding half bath in Old Town and College Creek has lower return relative to Full Bathroom.

Summary

- We recommend developers who want to build property in the three neighbourhoods, to optimize their investments by increasing the number of different feature type by referring to the analytical result.

Neighbourhood	Best Feature to invest	Average ROI
Somerst	Bedroom	35%
College Creek	Full Bathroom	200%
Old Town	Full Bathroom	107%

- There is an increased demand for properties with multiple bedrooms in **Somerst**, this could be due to to the higher proportion of students. However a word of caution to investors would be not to invest in bathrooms as this feature is not profitable.
- The number of bedrooms has higher limitations in gains compared to the other features given a limited lot area.
- Increasing bedrooms has larger impact on the absolute value of property.

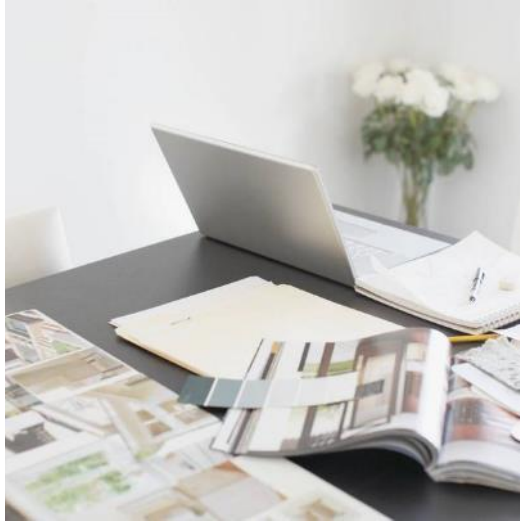
Limitation and Potential Study

Limitations:

1. **Multicollinearity still persists despite actions taken to reduce it.**
2. **Cost in calculating ROI is a rough estimation, might be different depend on location and season.- But its good enough for comparison between the neighbourhoods.**
3. **Insufficient data to analyse other neighbourhoods.**

Potential Study:

1. **Use models besides Linear regression.**
2. **Analyse other features requested and generate recommendation for property developers in AMES city.**
3. **Deploy property price predictor for property investor.**



Thank you
Q&A