



KINGS COUNTY HOUSING PROJECT

DATA ANALYSIS PROJECT



CONTENT



01

INTRODUCTION

02

GOALS AND OBJECTIVES

03

BUSINESS UNDERSTANDING

04

OVERVIEW REPORT

05

CONCLUSION

06

RECOMMENDATION

GOALS AND OBJECTIVES

Objective n° 1

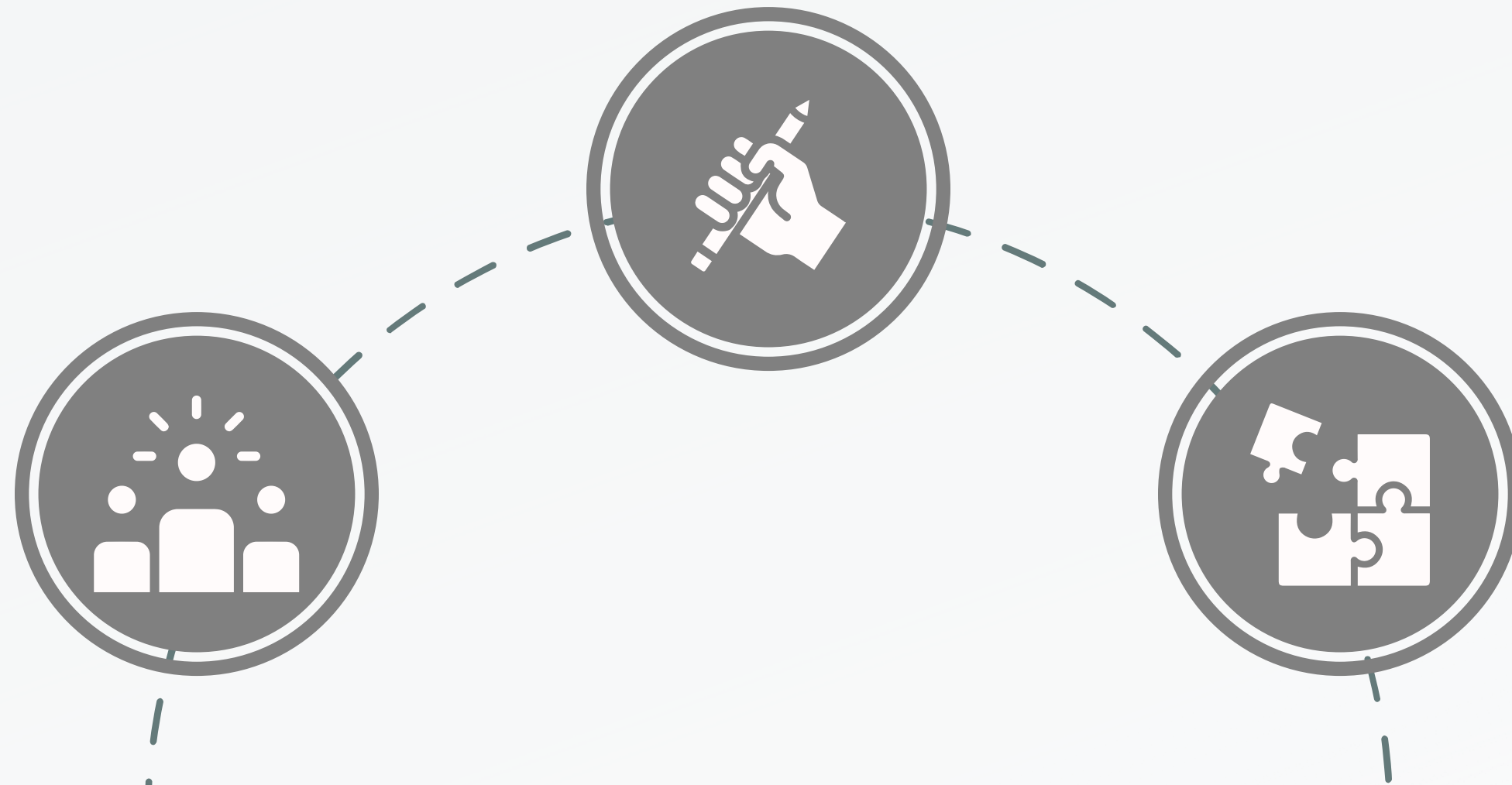
Develop a pricing
model.

Objective n° 2

Analyze seasonal trends.

Objective n° 3

Optimize home
renovations.





BUSINESS UNDERSTANDING

The real estate industry is a crucial sector that plays a significant role in the economy. The success of a real estate transaction depends on several factors, including the location, the size of the property, the condition of the property, the amenities, and the current market conditions. The housing market is subject to various external factors such as interest rates, economic conditions, and government policies that can impact the demand and supply of properties.

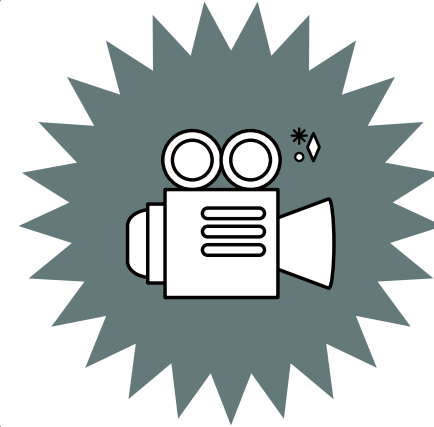
The background features abstract line art in a light teal color. On the left, there are several curved lines that sweep upwards and outwards. On the right, there are more curved lines, some of which are more densely packed and form a sort of 'combed' or 'feathered' shape. The overall effect is a modern, minimalist aesthetic.

BUSINESS PROBLEM

Some home owners are eager to sell their houses, and I have been approached by a real estate company about how to accurately appraise homes in King county so that they can give their customers accurate recommendations when it comes to buying and selling homes.

By using the dataset provided, I am able to build models that can actually predict sale price of houses. Real estate agents can better advise their clients on pricing strategies, investors can identify potentially undervalued properties, and homeowners can better estimate the value of their own properties. This can ultimately lead to more efficient and profitable real estate transactions in King County.

Tools for analysis



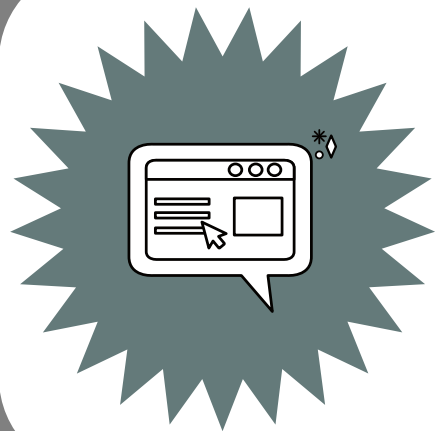
kc_house csv files

This was data that was collected from the various movies websites that are popular and they are used for our analysis.



Python

The programming language that is required to analyze the data.
This is used in regression, classification and anomaly detection of the csv data collected.

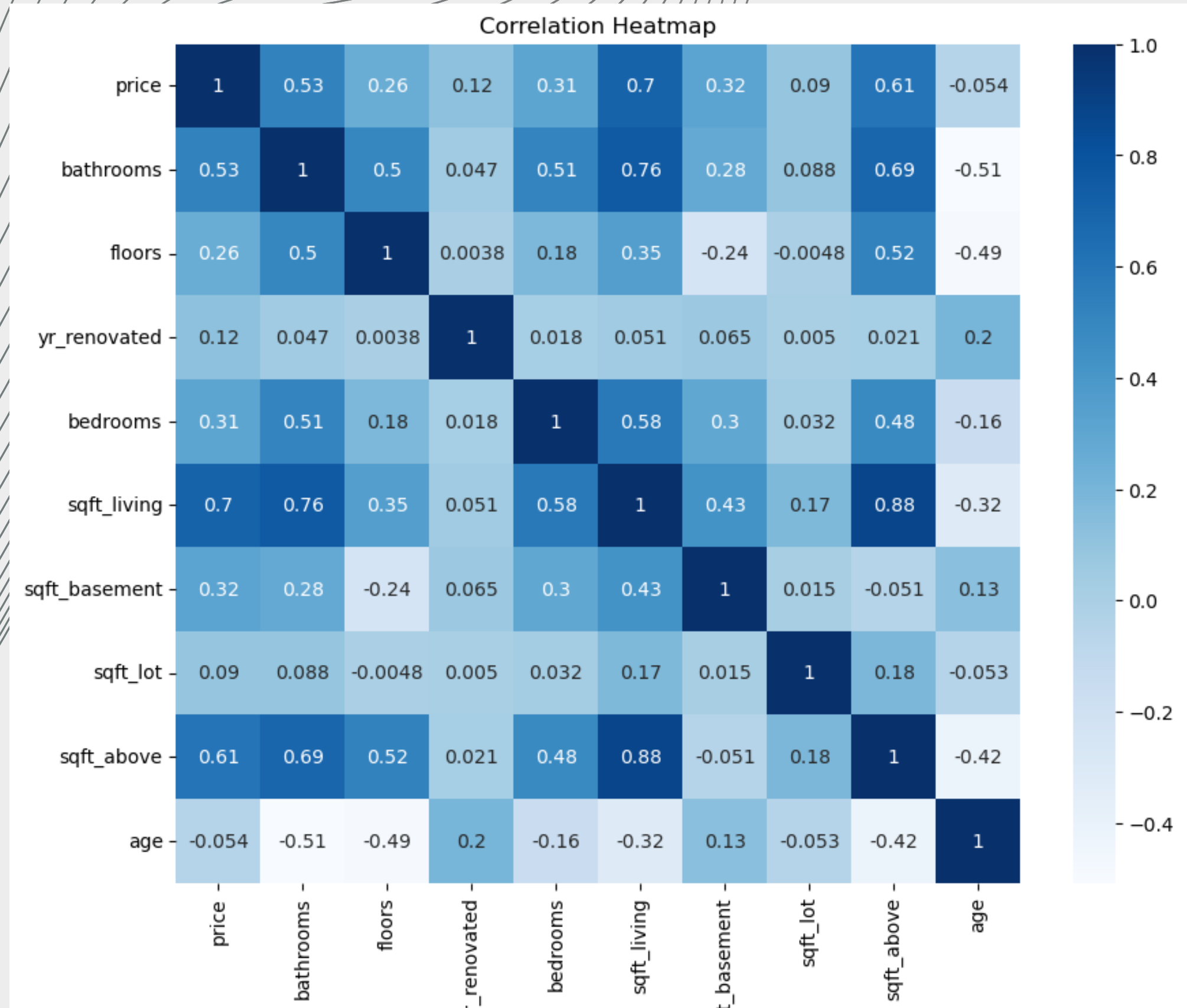


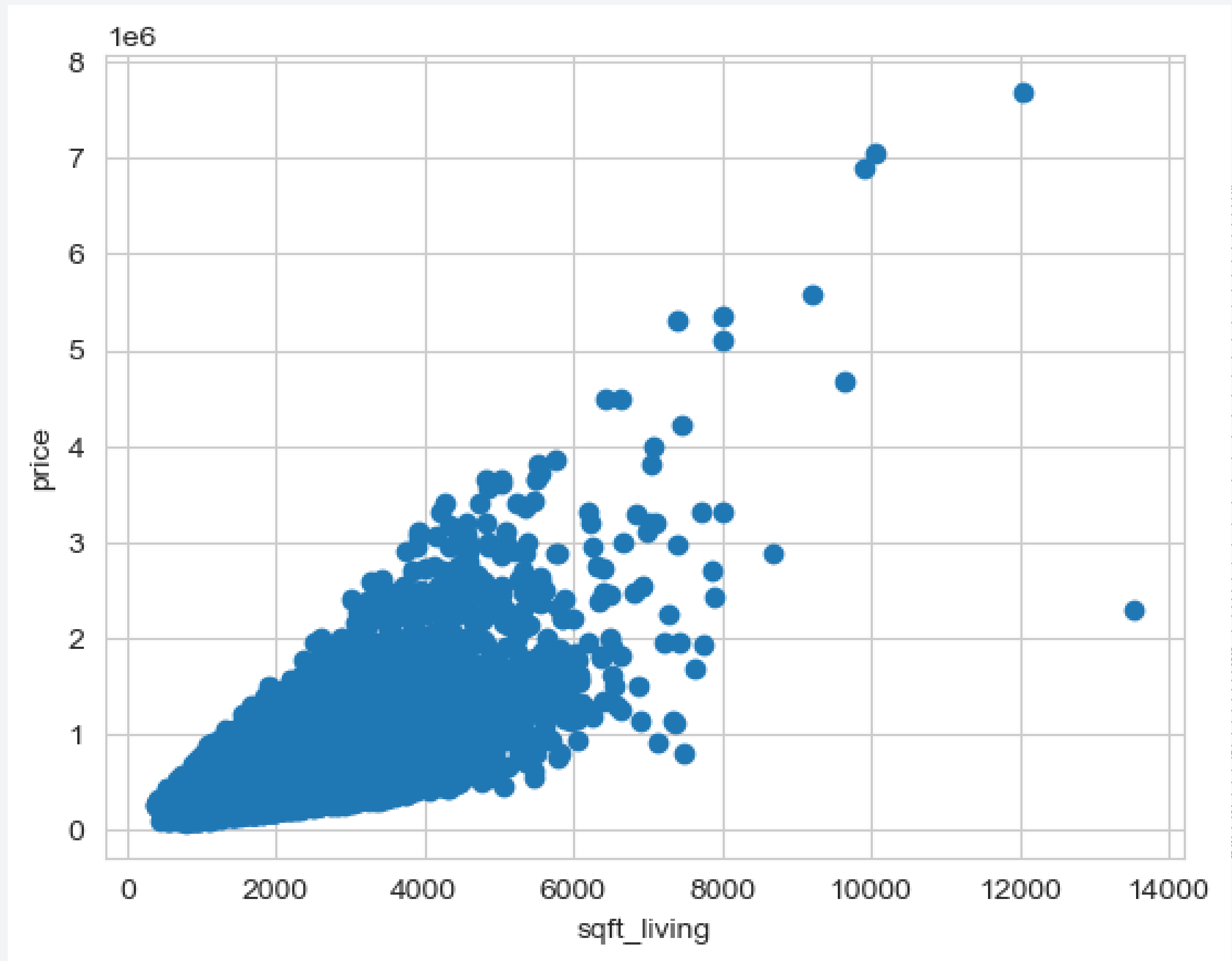
Modelling

After the analysis models were built in order to understand the relationships with our data

Overview Report

- A correlation heatmap showing how price is related to different continuous variables.





- Graph that shows the most correlated product.

Dep. Variable:	price		R-squared:	0.683		
Model:	OLS		Adj. R-squared:	0.683		
Method:	Least Squares		F-statistic:	1551.		
Date:	Fri, 07 Jul 2023		Prob (F-statistic):	0.00		
Time:	16:03:20		Log-Likelihood:	-2.9498e+05		
No. Observations:	21597		AIC:	5.900e+05		
Df Residuals:	21566		BIC:	5.903e+05		
Df Model:	30					
Covariance Type:	nonrobust					

const	3.222e+05	1.15e+04	27.941	0.000	3e+05	3.45e+05
bedrooms	-2.439e+04	1989.878	-12.259	0.000	-2.83e+04	-2.05e+04
bathrooms	4.714e+04	3369.685	13.991	0.000	4.05e+04	5.37e+04
sqft_living	127.9725	3.529	36.264	0.000	121.056	134.889
sqft_lot	-0.2457	0.035	-6.993	0.000	-0.315	-0.177
floors	4.157e+04	3624.566	11.469	0.000	3.45e+04	4.87e+04
sqft_basement	27.2393	4.319	6.307	0.000	18.773	35.705
yr_renovated	21.6365	4.098	5.280	0.000	13.604	29.669
age	3148.3857	68.844	45.732	0.000	3013.445	3283.326
waterfront_0	-1.018e+05	1.01e+04	-10.103	0.000	-1.22e+05	-8.21e+04
waterfront_1	4.24e+05	1.36e+04	31.167	0.000	3.97e+05	4.51e+05
view_0	1.164e+04	7247.563	1.606	0.108	-2564.098	2.58e+04
view_1	2.077e+05	1.17e+04	17.785	0.000	1.85e+05	2.31e+05
view_2	7.733e+04	1.03e+04	7.476	0.000	5.71e+04	9.76e+04
view_3	6.668e+04	8746.838	7.624	0.000	4.95e+04	8.38e+04
view_4	-4.122e+04	5295.848	-7.784	0.000	-5.16e+04	-3.08e+04

- This OLS regression model aims to predict the housing prices (dependent variable) based on several independent variables


season	mean	std	min	max	mean	std
season_0	7.31e+04	3834.198	19.066	0.000	6.56e+04	8.06e+04
season_1	1.001e+05	3696.936	27.071	0.000	9.28e+04	1.07e+05
season_2	7.77e+04	3721.393	20.881	0.000	7.04e+04	8.5e+04
season_3	7.128e+04	4015.187	17.751	0.000	6.34e+04	7.91e+04
=====						

- Continuation of the model that focuses mainly on the seasons.

Autumn - Season 0
Winter - Season 3
Spring - Season 1
Summer - Season 2

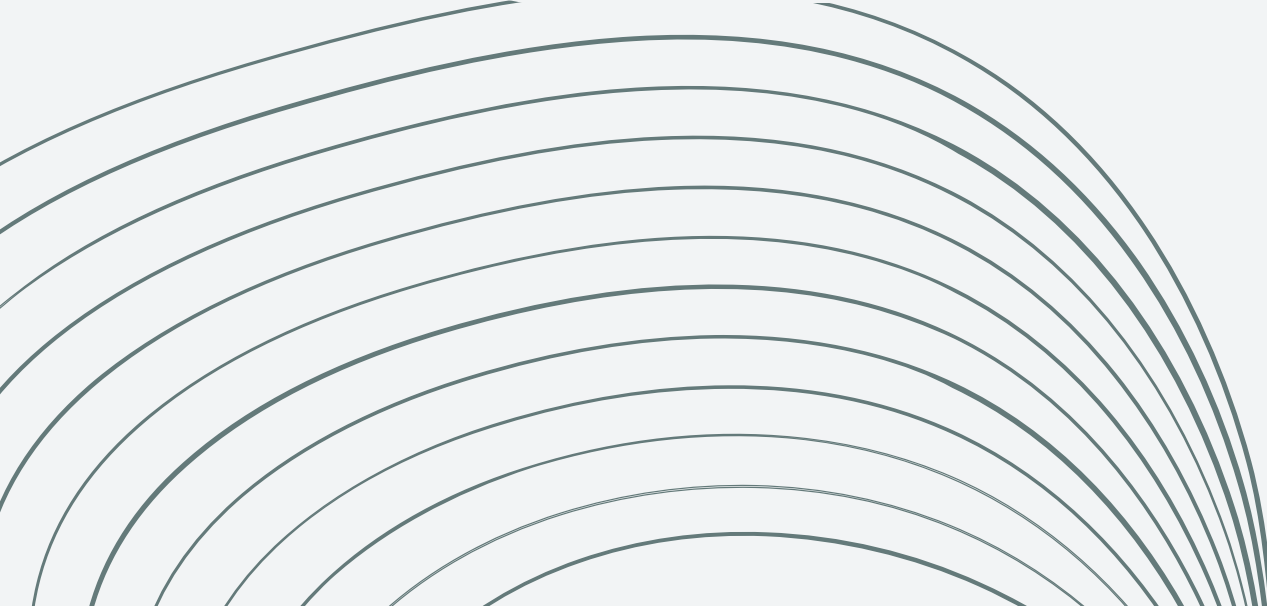
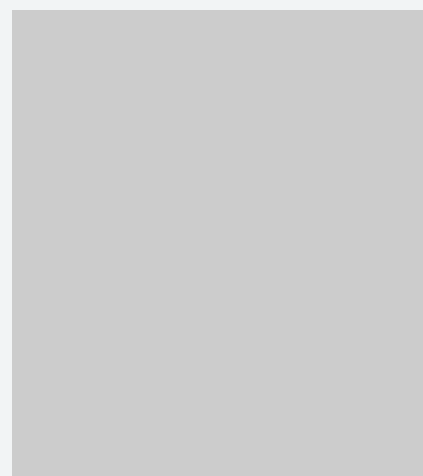


CONCLUSION

- The living room size has the highest correlation with price.
 - **Seasons** have the major influence with our price, as houses are sold the **highest in price in the spring** and the **lowest in the winter**.
 - Houses stated to be in very good condition and have excellent views get to have the highest price.
 - The number of **bathroom** and **floors** considered a major part of the house as an addtional of each of them amounts to an average of \$45,000.
 - The variables that have the least influence on the price of the house are: year of renovation, size of the basement and the lot.
- 



RECOMMENDATION

- Put up houses for sale in peak season spring.
 - Ensure that the houses are in good condition before putting it into the market for sale.
 - Revonate their house since this increases the value of the house.
 - Increase square footage of living space.
- 
- 



THANK YOU