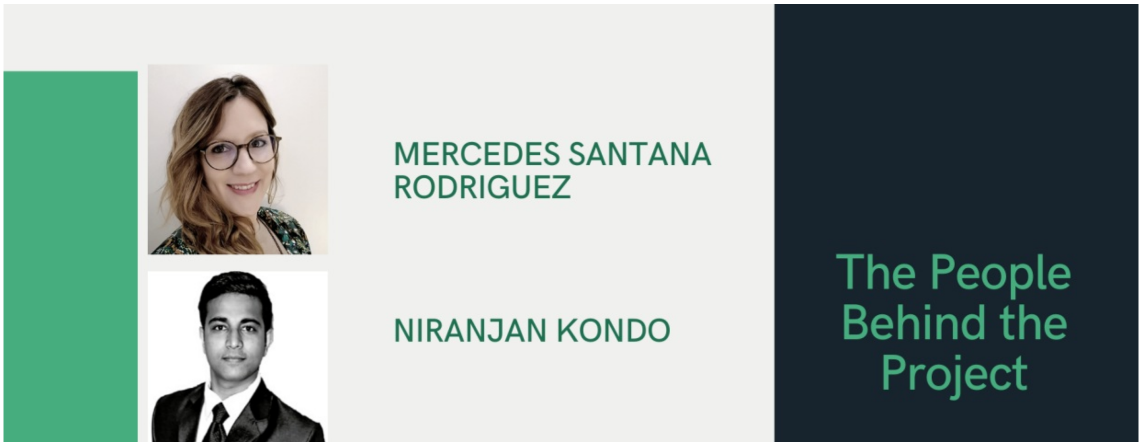
The housing story

Introduction | Team | Problem | Assumptions | Price analysis | Map1 | Map2 | Regression | outcomes | Conclusion | Tha nk..

House Price Prediction Model

The housing story



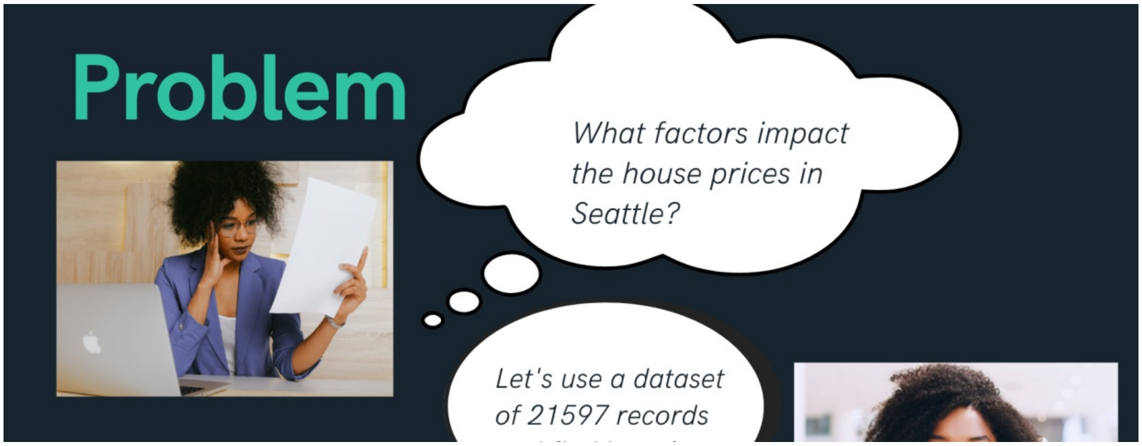
The People Behind the Project

MERCEDES SANTANA RODRIGUEZ

NIRANJAN KONDO

The housing story

Introduction | Team | Problem | Assumptions | Price analysis | Map1 | Map2 | Regression | outcomes | Conclusion | Tha nk..

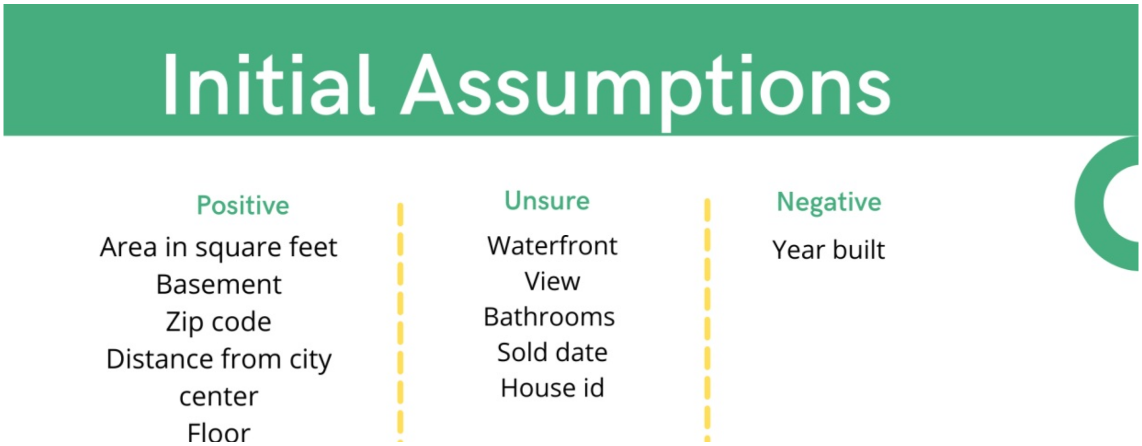


Problem

What factors impact the house prices in Seattle?

Let's use a dataset of 21597 records

The housing story



Initial Assumptions

Positive

Area in square feet
Basement
Zip code
Distance from city center
Floor

Unsure

Waterfront
View
Bathrooms
Sold date
House id

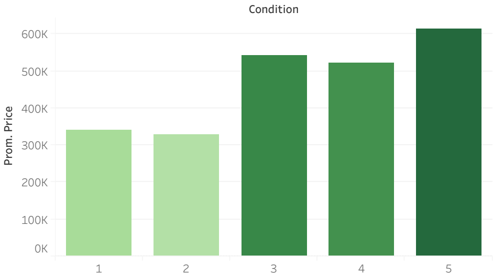
Negative

Year built

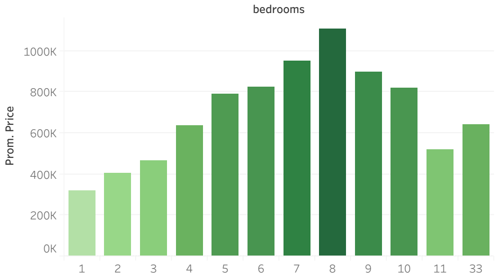
The housing story

Introduction | Team | Problem | Assumptions | Price analysis | Map1 | Map2 | Regression | outcomes | Conclusion | Tha nk..

Price - Condition



Price - Bedroom



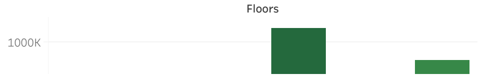
Year Built
Todo

Floors
Todo

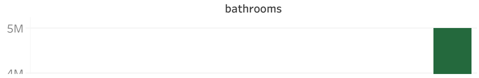
bathrooms
Todo

Prom. Price
291.867 4.990.000

Price - Floor



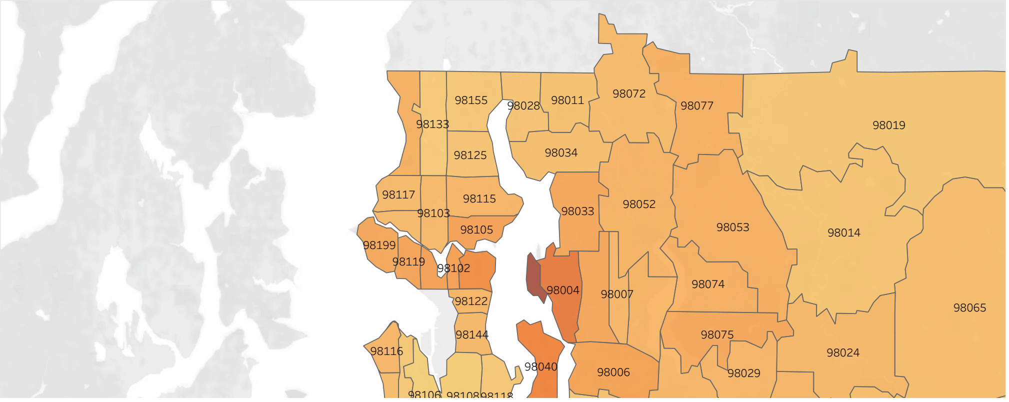
Price - Bathroom



The housing story

Introduction | Team | Problem | Assumptions | Price analysis | Map1 | Map2 | Regression | outcomes | Conclusion | Thank You

Mapping ZipCodes



Prom. Price
234.284,035175879 a 2...

Zipcode
Todo

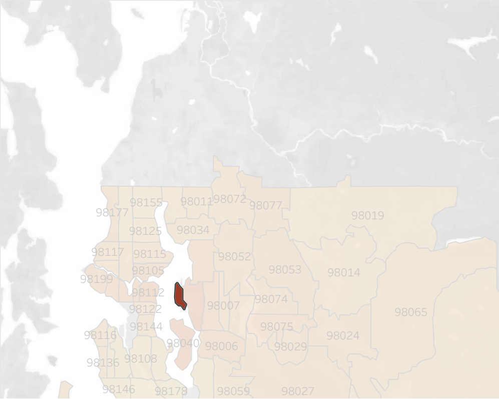
Prom. Price
234.284 2.161.300

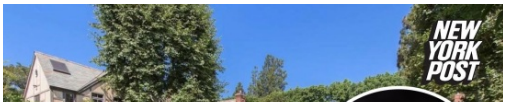
year built
Todo

The housing story

Intr od.. | Team | Problem | Assumptions | Price analysis | Map1 | **Map2** | Regression | outcomes | Conclusion | Thank You

Mapping ZipCodes



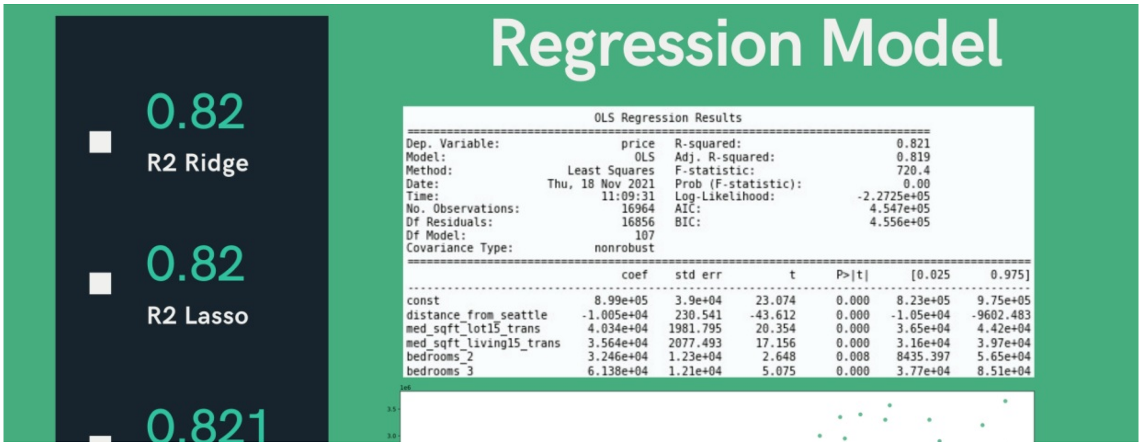


Prom. Price
234.284,035175879 a 2...

Zipcode
Todo

Prom. Price
234.284 2.161.300

year built
Todo

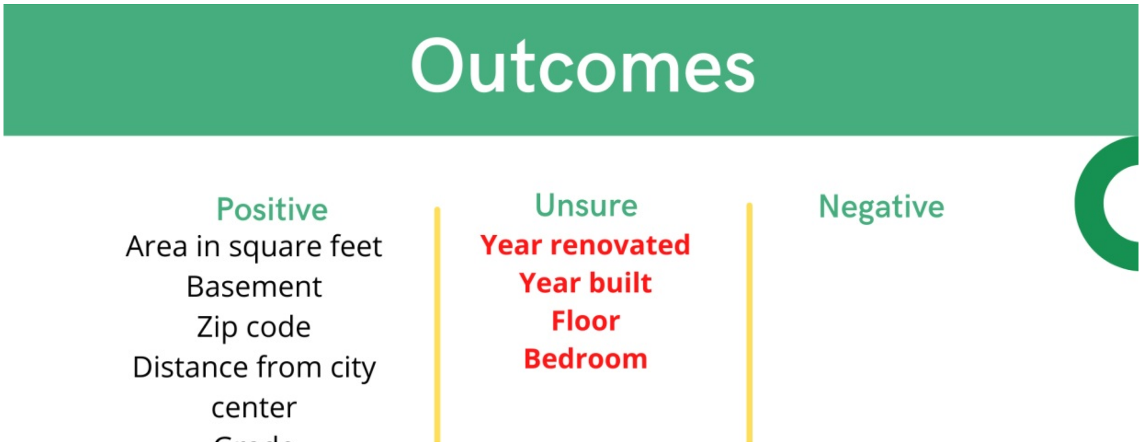


Regression Model

- □ 0.82 R2 Ridge
- □ 0.82 R2 Lasso
- □ 0.821

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.821
Model:                            OLS   Adj. R-squared:                  0.819
Method:                 Least Squares   F-statistic:                     720.4
Date:                Thu, 18 Nov 2021   Prob (F-statistic):               0.00
Time:                        11:09:31   Log-Likelihood:            -2.2725e+05
No. Observations:               16964   AIC:                         4.547e+05
Df Residuals:                   16856   BIC:                         4.556e+05
Df Model:                         107
Covariance Type:            nonrobust
================================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                     8.99e+05    3.9e+04     23.074      0.000    8.23e+05    9.75e+05
distance_from_seattle   -1.005e+04    230.541    -43.612      0.000   -1.05e+04   -9602.483
med_sqft_lot15_trans     4.034e+04   1981.795     20.354      0.000    3.65e+04    4.42e+04
med_sqft_living15_trans  3.564e+04   2077.493     17.156      0.000    3.16e+04    3.97e+04
bedrooms_2               3.246e+04   1.23e+04      2.648      0.008    8435.397    5.65e+04
bedrooms_3               6.138e+04   1.21e+04      5.075      0.000    3.77e+04    8.51e+04
```

The housing story



Outcomes

Positive

Area in square feet
Basement
Zip code
Distance from city center
Grade

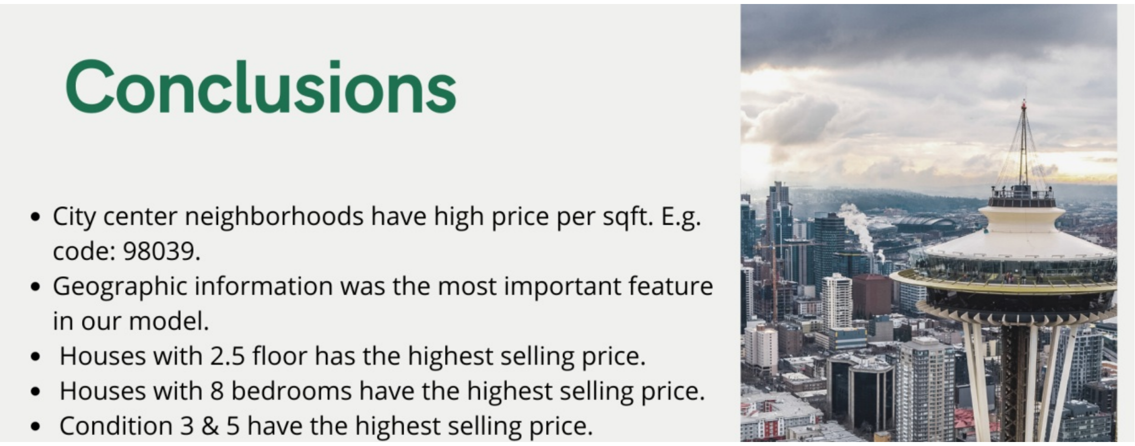
Unsure

Year renovated
Year built
Floor
Bedroom

Negative

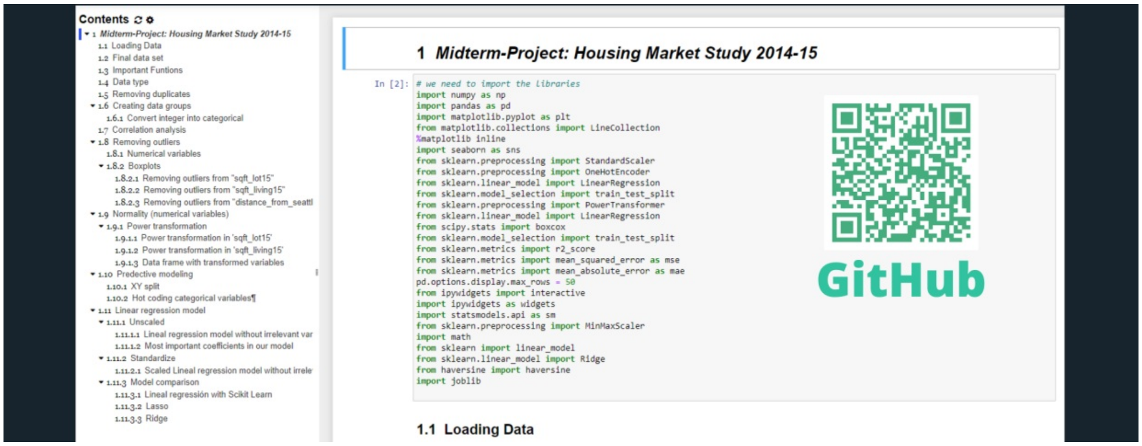
The housing story

Intr od.. | Team | Problem | Assumptions | Price analysis | Map1 | Map2 | Regression | outcomes | Conclusion | Thank You



Conclusions

- City center neighborhoods have high price per sqft. E.g. code: 98039.
- Geographic information was the most important feature in our model.
- Houses with 2.5 floor has the highest selling price.
- Houses with 8 bedrooms have the highest selling price.
- Condition 3 & 5 have the highest selling price.



Contents

1 *Midterm-Project: Housing Market Study 2014-15*

In [2]:

```
# we need to import the libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.collections import LineCollection
%matplotlib inline
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PowerTransformer
from sklearn.linear_model import LinearRegression
from scipy.stats import boxcox
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error as mae
pd.options.display.max_rows = 50
from ipywidgets import interactive
import ipywidgets as widgets
import statsmodels.api as sm
from sklearn.preprocessing import MinMaxScaler
import math
from sklearn import linear_model
from sklearn.linear_model import Ridge
from haversine import haversine
import joblib
```

1.1 Loading Data