

# Intergenerational Families in Kings Counties

Beau Morton



# Overview

Client: National Association for Real Estate Brokers (NAREB) A black led organization that focuses on bringing together the minority professionals in the real estate industry to promote the meaningful exchange of ideas about our business and how best to serve their clientele. For this case the clientele will focus on black intergenerational families.



# Overview

**How can home renovations increase the estimated value of homes for black families in King County, WA? (specifically focusing on intergenerational homes)**

# Data Understanding

We begin by importing pandas, numpy, ols, plt, sns, durbin\_watson, and statsmodels to analyze our data and perform our regressions. We also import our data as df.

```
In [92]: import pandas as pd
import numpy as np
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt
from statsmodels.stats.stattools import durbin_watson
import seaborn as sns
import statsmodels.api as sm
%matplotlib inline

df = pd.read_csv('data/kc_house_data.csv')
df
```

```
Out[92]:
```

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft
0	7129300520	10/13/2014	221900.0	3	1.00	1180	5650	1.0	NaN	NONE	...	7 Average	
1	6414100192	12/9/2014	538000.0	3	2.25	2570	7242	2.0	NO	NONE	...	7 Average	
2	5631500400	2/25/2015	180000.0	2	1.00	770	10000	1.0	NO	NONE	...	6 Low Average	
3	2487200875	12/9/2014	604000.0	4	3.00	1960	5000	1.0	NO	NONE	...	7 Average	
4	1954400510	2/18/2015	510000.0	3	2.00	1680	8080	1.0	NO	NONE	...	8 Good	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
21592	263000018	5/21/2014	360000.0	3	2.50	1530	1131	3.0	NO	NONE	...	8 Good	
21593	6600060120	2/23/2015	400000.0	4	2.50	2310	5813	2.0	NO	NONE	...	8 Good	
21594	1523300141	6/23/2014	402101.0	2	0.75	1020	1350	2.0	NO	NONE	...	7 Average	

# Data Preparation

The first model iteration for our case will focus on square foot living, square foot above, and bathrooms. Before we can get to the analysis, we will drop 'zipcode' as a variable, because the numbers in zip code are not useful to us in this analysis without context of what each zip code means.

# Data Preparation

The first model iteration for our case will focus on square foot living, square foot above, and bathrooms. Before we can get to the analysis, we will drop 'zipcode' as a variable, because the numbers in zip code are not useful to us in this analysis without context of what each zip code means.

We also converted our categorical data into something more useful for us in this analysis.

# Data Preparation

The first model iteration for our case will focus on square foot living, square foot above, and bathrooms. Before we can get to the analysis, we will drop 'zipcode' as a variable, because the numbers in zip code are not useful to us in this analysis without context of what each zip code means.

# Model 1

```
=====
                        OLS Regression Results
=====
Dep. Variable:          price    R-squared:                0.493
Model:                  OLS      Adj. R-squared:            0.493
Method:                 Least Squares    F-statistic:          7003.
Date:                  Sun, 17 Apr 2022    Prob (F-statistic):    0.00
Time:                  20:58:24    Log-Likelihood:        -3.0005e+05
No. Observations:      21597    AIC:                   6.001e+05
Df Residuals:          21593    BIC:                   6.001e+05
Df Model:               3
Covariance Type:       nonrobust
=====
                        coef    std err          t      P>|t|      [0.025      0.975]
-----
Intercept    -3.815e+04    5252.893      -7.262     0.000    -4.84e+04    -2.79e+04
sqft_living   297.9354        4.483      66.458     0.000     289.148     306.723
sqft_above   -18.4232         4.479      -4.113     0.000     -27.202     -9.644
bathrooms    -3972.1769    3545.059     -1.120     0.263    -1.09e+04     2976.400
=====
Omnibus:            14748.993    Durbin-Watson:           1.982
Prob(Omnibus):      0.000    Jarque-Bera (JB):        538245.600
Skew:               2.806    Prob(JB):                 0.00
Kurtosis:           26.804    Cond. No.                 9.59e+03
=====
```

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 9.59e+03. This might indicate that there are strong multicollinearity or other numerical problems.



# Model 2

```
=====
                        OLS Regression Results
=====
Dep. Variable:          price    R-squared:                0.507
Model:                  OLS      Adj. R-squared:            0.507
Method:                 Least Squares    F-statistic:          7398.
Date:                   Sun, 17 Apr 2022    Prob (F-statistic):      0.00
Time:                   21:06:31    Log-Likelihood:         -2.9976e+05
No. Observations:       21597    AIC:                    5.995e+05
Df Residuals:           21593    BIC:                    5.996e+05
Df Model:                3
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept      7.704e+04    7684.108      10.026      0.000      6.2e+04    9.21e+04
bedrooms     -5.775e+04    2324.412     -24.846      0.000     -6.23e+04    -5.32e+04
sqft_living    313.8304      2.467     127.187      0.000      308.994    318.667
floors         3476.4322    3480.340      0.999      0.318     -3345.290    1.03e+04
=====
Omnibus:            14415.709    Durbin-Watson:          1.985
Prob(Omnibus):      0.000    Jarque-Bera (JB):       491526.502
Skew:               2.733    Prob(JB):               0.00
Kurtosis:           25.723    Cond. No.               1.04e+04
=====
```

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.04e+04. This might indicate that there are strong multicollinearity or other numerical problems.

# Model 3

```
=====
                        OLS Regression Results
=====
Dep. Variable:          price      R-squared:                0.509
Model:                  OLS        Adj. R-squared:            0.509
Method:                 Least Squares    F-statistic:          7451.
Date:                  Sun, 17 Apr 2022    Prob (F-statistic):    0.00
Time:                  22:15:14      Log-Likelihood:        -2.9972e+05
No. Observations:      21597         AIC:                  5.994e+05
Df Residuals:          21593         BIC:                  5.995e+05
Df Model:              3
Covariance Type:       nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept      8.424e+04    6637.291      12.692      0.000      7.12e+04    9.72e+04
bedrooms      -5.959e+04    2327.075     -25.605      0.000     -6.41e+04   -5.5e+04
sqft_living    318.6333      2.383     133.727      0.000      313.963    323.304
sqft_lot       -0.3847      0.043      -8.922      0.000      -0.469     -0.300
=====
Omnibus:          14308.034    Durbin-Watson:          1.985
Prob(Omnibus):    0.000      Jarque-Bera (JB):      477721.465
Skew:            2.710      Prob(JB):              0.00
Kurtosis:        25.394      Cond. No.              1.73e+05
=====
```

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.73e+05. This might indicate that there are strong multicollinearity or other numerical problems.

# Conclusion