



# KING COUNTY HOUSE SALES ANALYSIS REPORT

**Sneha Bhaskar**



# TABLE OF CONTENT

- BUSINESS PROBLEM
- KEY FINDINGS
- RECOMMENDATION

©Glasbergen / glasbergen.com



**"I used to sell mud pies, but I can charge more if I call it real estate."**







# BUSINESS PROBLEM

---

The Pro Home Builder Company aims to construct residences in King County Washington and wants to identify which housing features to focus on for maximum profitability. By understanding which amenities significantly influence home prices, the company can make data-driven choices on the types of houses to build in order to optimize profits





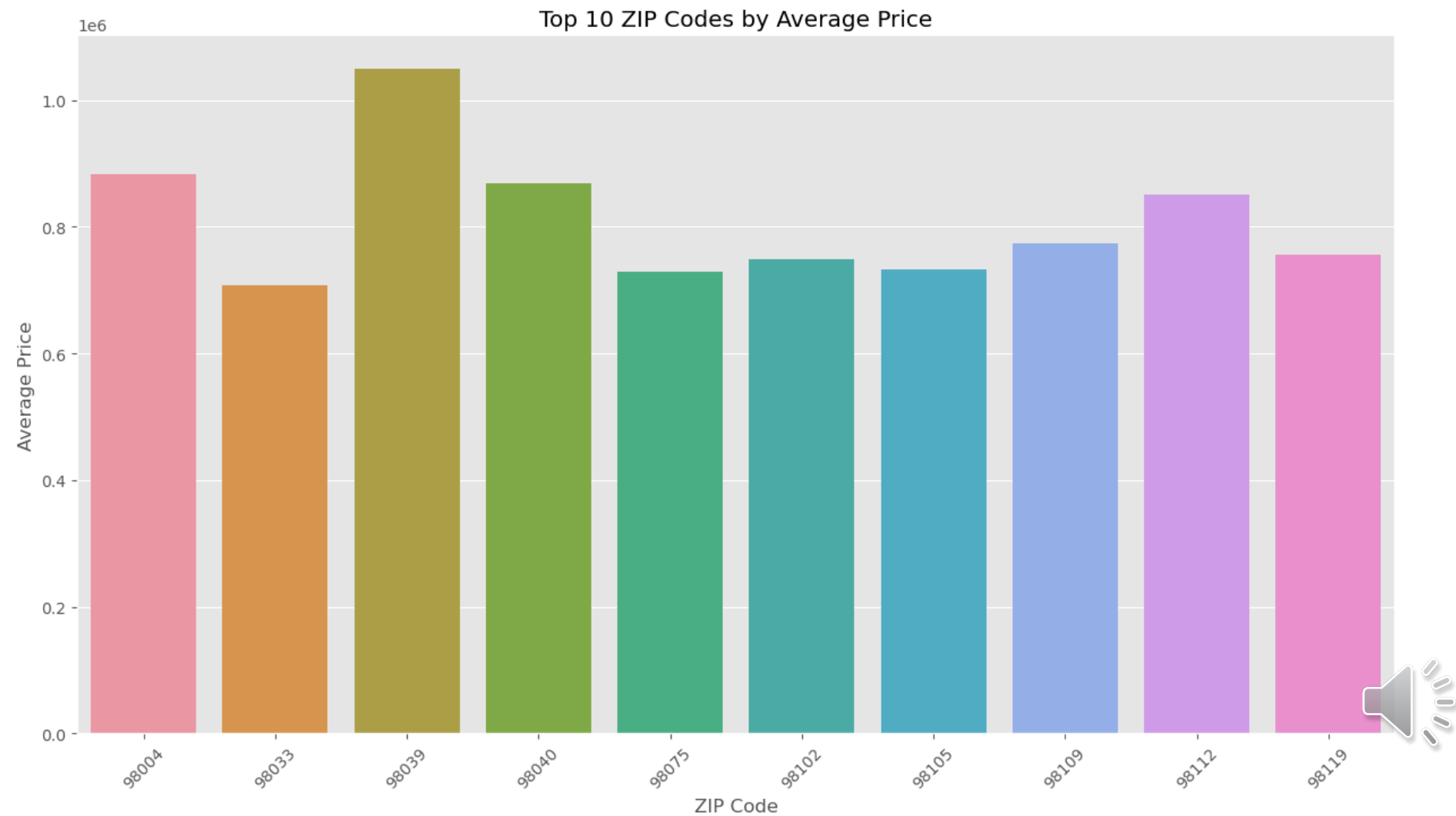


## KEY FINDINGS





# ZIP Codes Vs Price



# TOP SUBURBS IN KING COUNTY

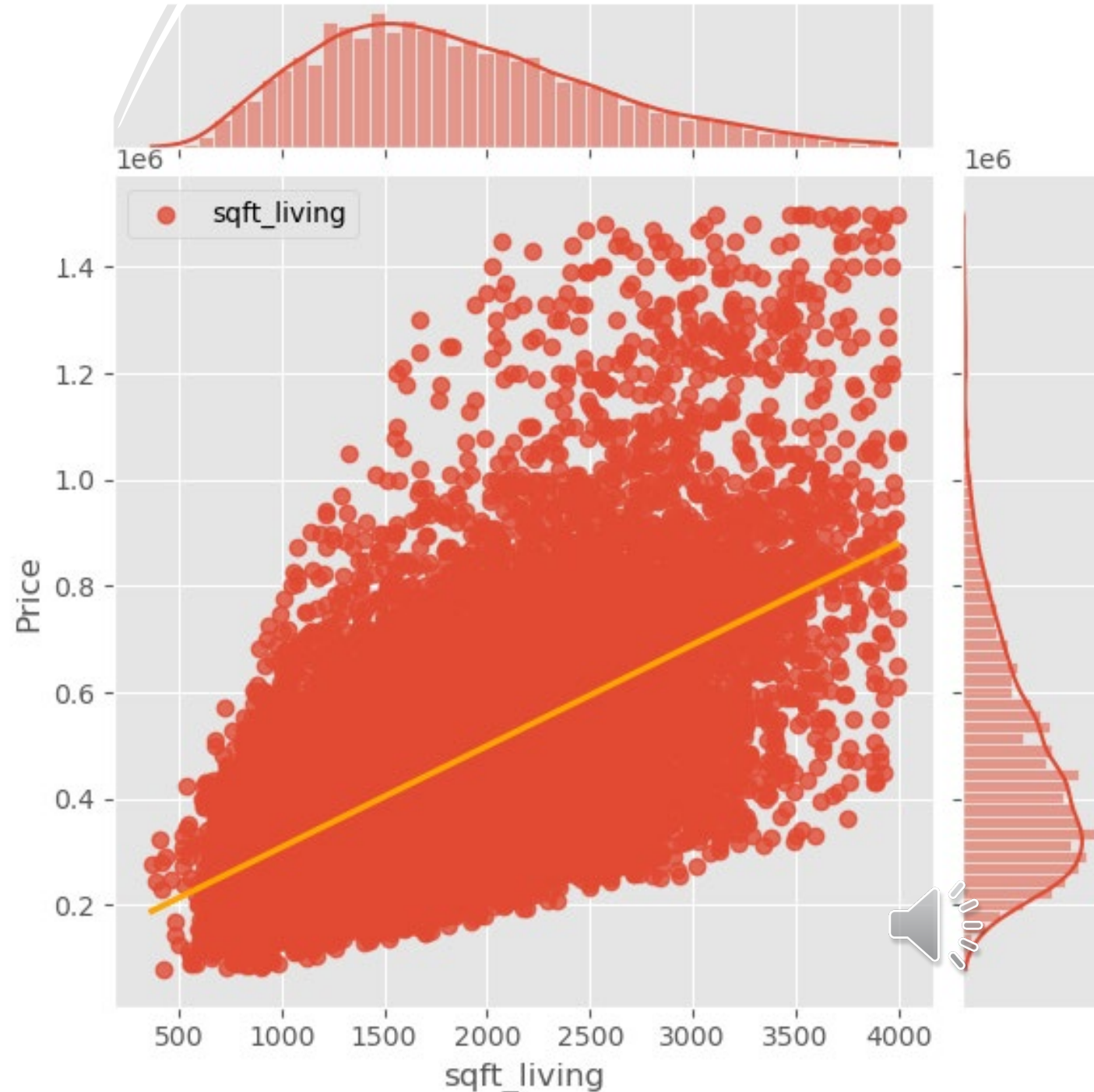
- West Bellevue
- Kirkland
- Medina
- Mercer Island
- Sammamish
- Seattle
- Overlake
- Education Hill



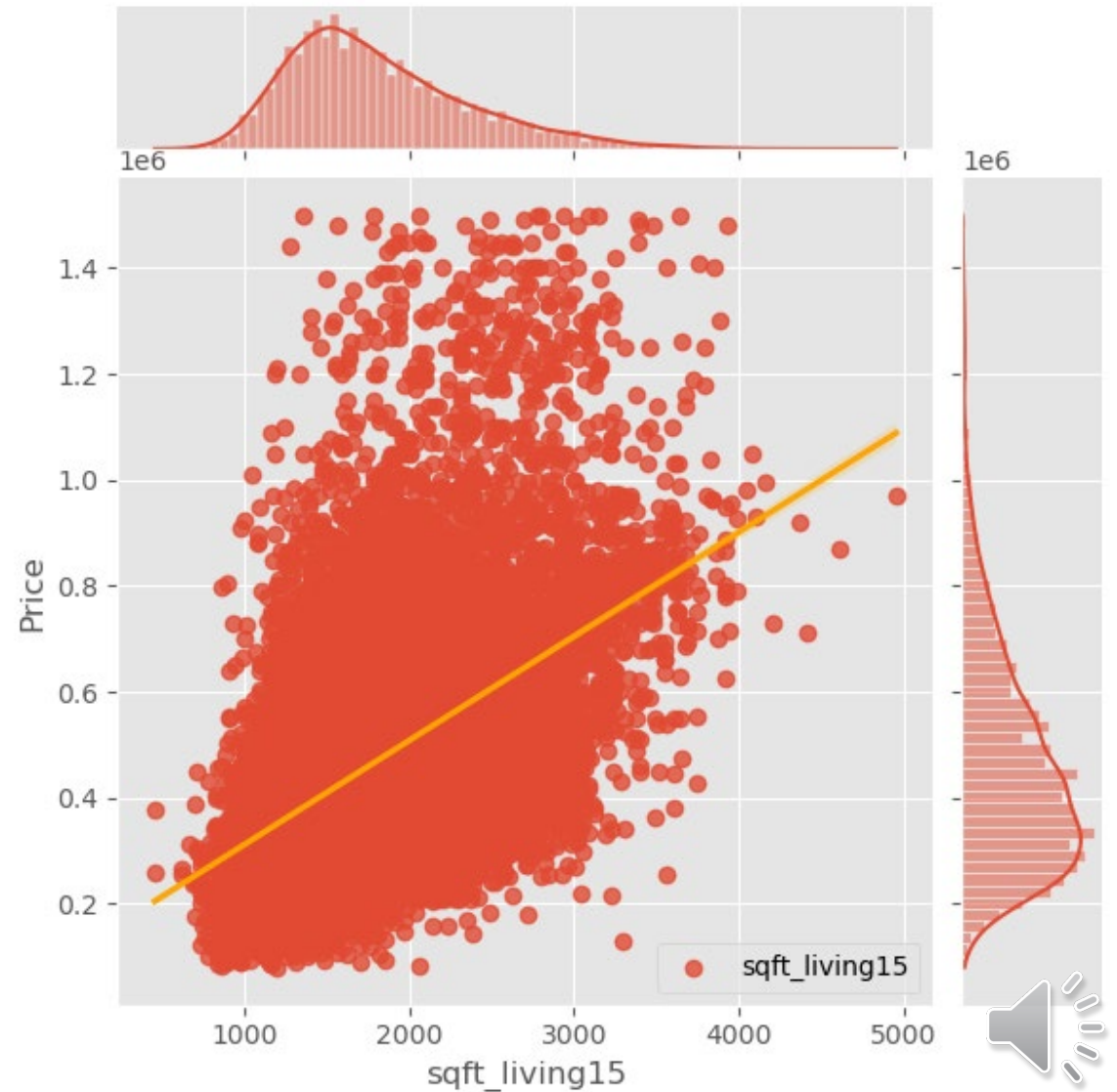


# Square Foot Living Vs Price

- This visualization shows there seems to be a relatively strong linear relationship between square feet of living space and the price of a house.
- 1500 - 2500 sqft\_living is the most observed footage of the house in this dataset

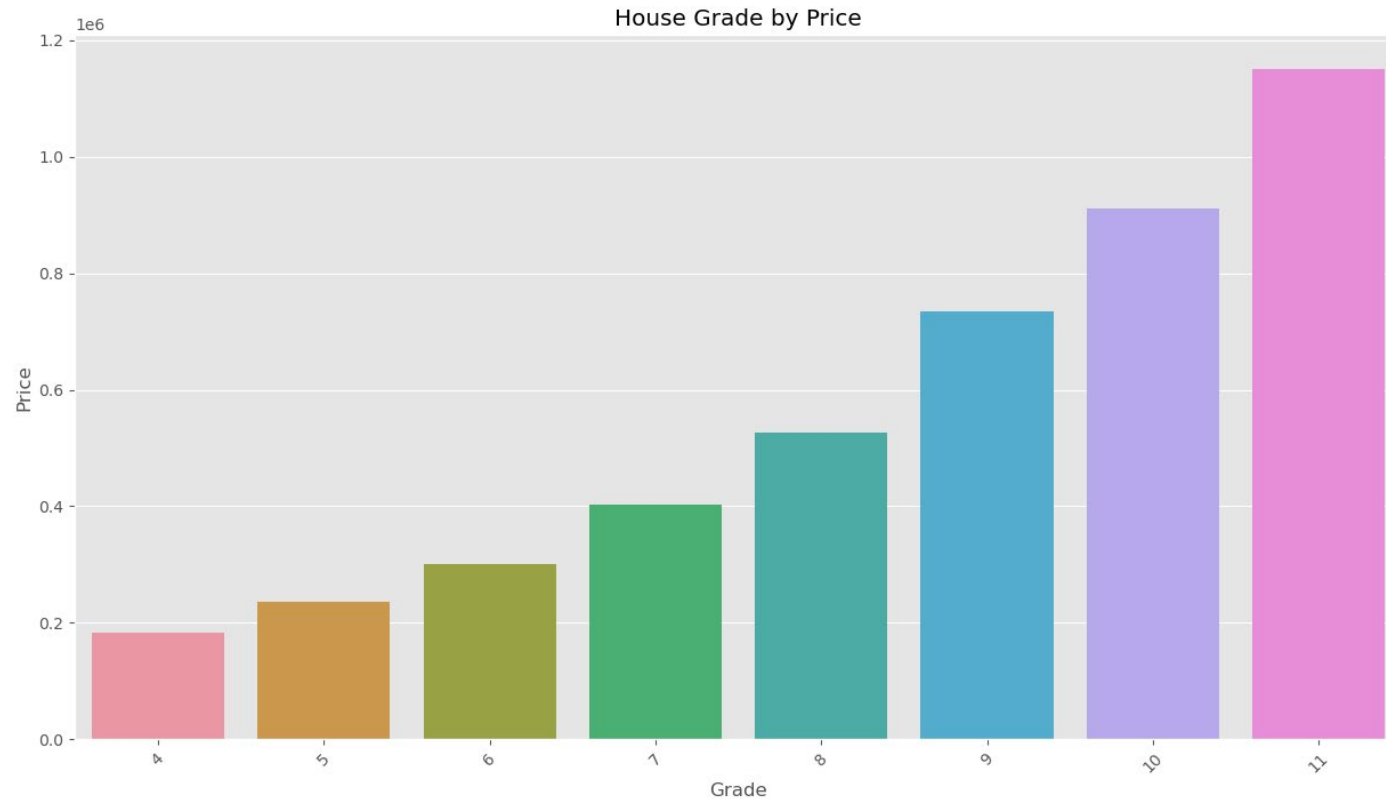


The square footage of interior housing living space for the nearest 15 neighbors Vs Price





# House Grade by Price



The visualization above shows that when the building grade improves, the house price rises as well. We can see in the bar plot that the mean house price for a home with a grade of 11 is far above other grades. While the building grade of 4 falls short of minimum building standards based on the King County grading system.



# Regression analysis

## The results of the analysis were as follows:

• The model above with an R-squared value of 0.775, telling us that the model fits the data with an accuracy of 78%, We can confidently say that zip code is one of the strongest influencers on the value of homes in King County.

## Other features that were positively correlated with price included:

- Square footage of living space
- Building grade
- Interior housing living space for the nearest 15 neighbors

```
In [107]: #include the zipcode column to the model
#zipcodes have no numerical significance in our data, this variable is treated as categorical
f = 'price~grade_sc+sqft_living_sc+sqft_living15_sc+C(zipcode)'
model = ols(formula=f, data=df).fit()
model.summary()
```

Out[107]: OLS Regression Results

Dep. Variable:	price	R-squared:	0.775			
Model:	OLS	Adj. R-squared:	0.774			
Method:	Least Squares	F-statistic:	730.7			
Date:	Wed, 20 Sep 2023	Prob (F-statistic):	0.00			
Time:	19:54:32	Log-Likelihood:	-1.9922e+05			
No. Observations:	15334	AIC:	3.986e+05			
Df Residuals:	15261	BIC:	3.991e+05			
Df Model:	72					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.625e+05	7103.962	36.951	0.000	2.49e+05	2.76e+05
C(zipcode)[T.98002]	5.112e+04	1.09e+04	4.684	0.000	2.97e+04	7.25e+04
C(zipcode)[T.98003]	1.568e+04	1.05e+04	1.496	0.135	-4859.404	3.62e+04
C(zipcode)[T.98004]	5.824e+05	1.3e+04	44.957	0.000	5.57e+05	6.08e+05
C(zipcode)[T.98005]	3.531e+05	1.68e+04	21.022	0.000	3.2e+05	3.86e+05
C(zipcode)[T.98006]	3.047e+05	1.01e+04	30.143	0.000	2.85e+05	3.25e+05
C(zipcode)[T.98007]	2.585e+05	1.31e+04	19.750	0.000	2.33e+05	2.84e+05
C(zipcode)[T.98008]	2.603e+05	1.03e+04	25.320	0.000	2.4e+05	2.8e+05
C(zipcode)[T.98010]	1.201e+05	2.1e+04	5.713	0.000	7.89e+04	1.61e+05
C(zipcode)[T.98011]	1.485e+05	1.16e+04	12.824	0.000	1.26e+05	1.71e+05
C(zipcode)[T.98014]	1.754e+05	2.61e+04	6.716	0.000	1.24e+05	2.27e+05
C(zipcode)[T.98019]	8.871e+04	1.32e+04	6.704	0.000	6.28e+04	1.15e+05





# Strategic Recommendations for a Pro Home Builder Company

- Strategic Location Selection
- Optimize Living Space
- Quality of Construction
- Neighborhood Composition







# FUTURE WORK





# Thank You

- Email: [Sneha.bhaskar1995@gmail.com](mailto:Sneha.bhaskar1995@gmail.com)
- GitHub: @Sdec30
- LinkedIn: <https://www.linkedin.com/in/sneha-bhaskar-8aa158186/>

