MURAT KURT

EDA-PROJECT SPICED

2025-07-11

To stay sharp-eyed, even when the data piles up!

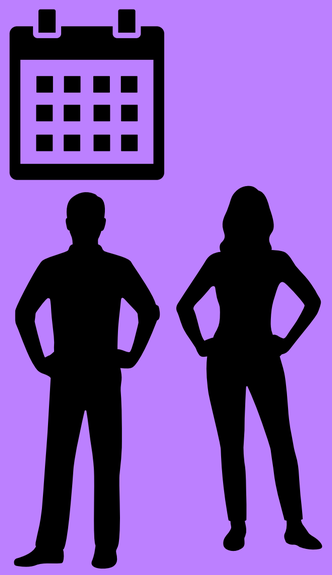
CLIENT-REQUIREMENTS

AMY WILLIAMS

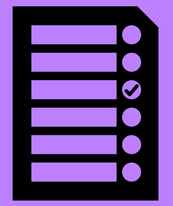
- INTERESTED IN PURCHASING PROPERTIES IN A PERIPHERAL LOCATION
- PREFERS TO AVOID ANY INTERACTION WITH FBI
- OWNS PRIME-LOCATION PROPERTIES

CUSTOMER REQUEST FROM AMY

AMY WILLIAMS IS VISITING MY OFFICE IN SEATTLE, DOWN-TOWN, BEST LOCATION OF COURSE...



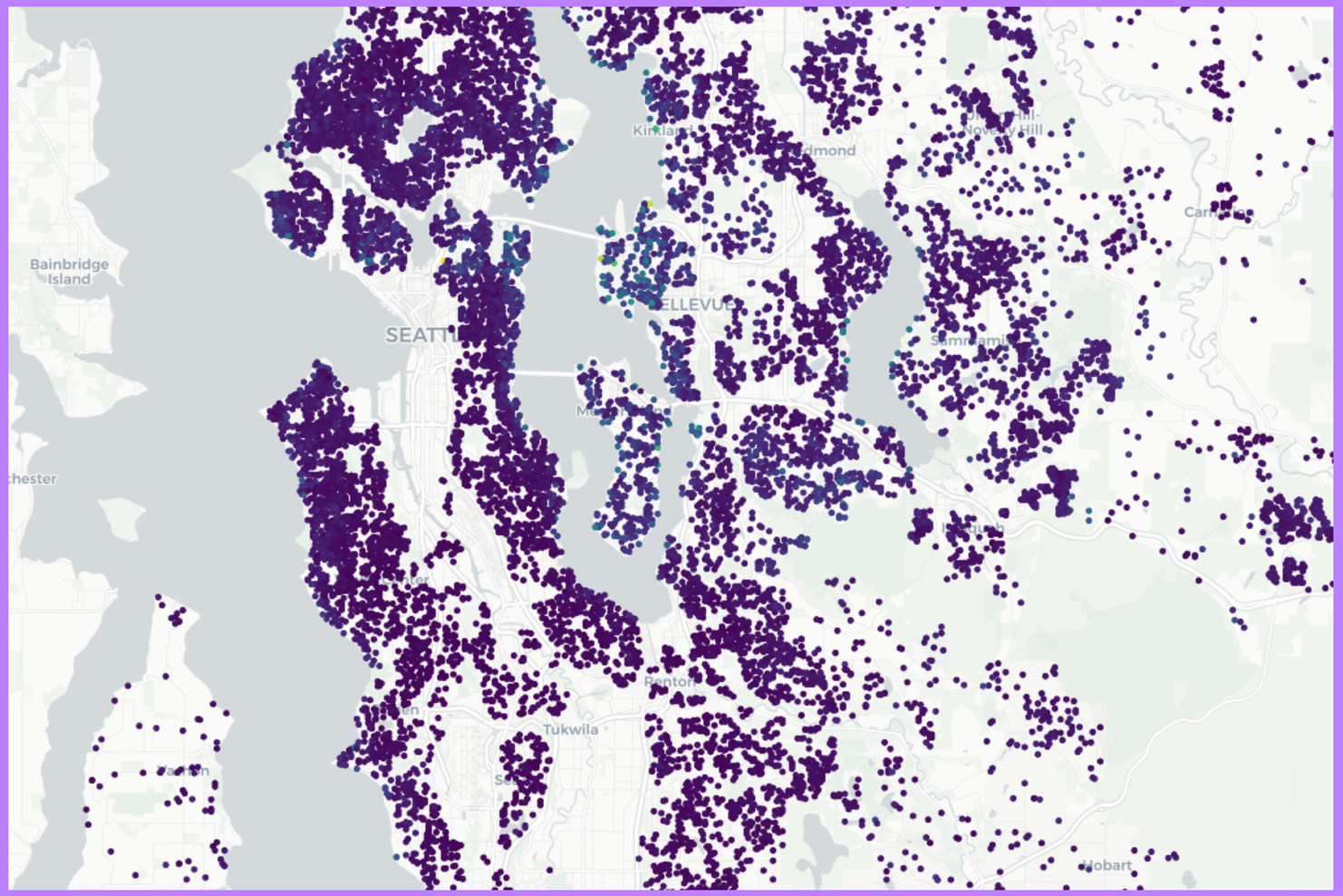
WHERE AND HOW TO HIDE
TIPS FOR PROPERTY SEARCH
CHECKING HER TOP-RATED PROPERTIES

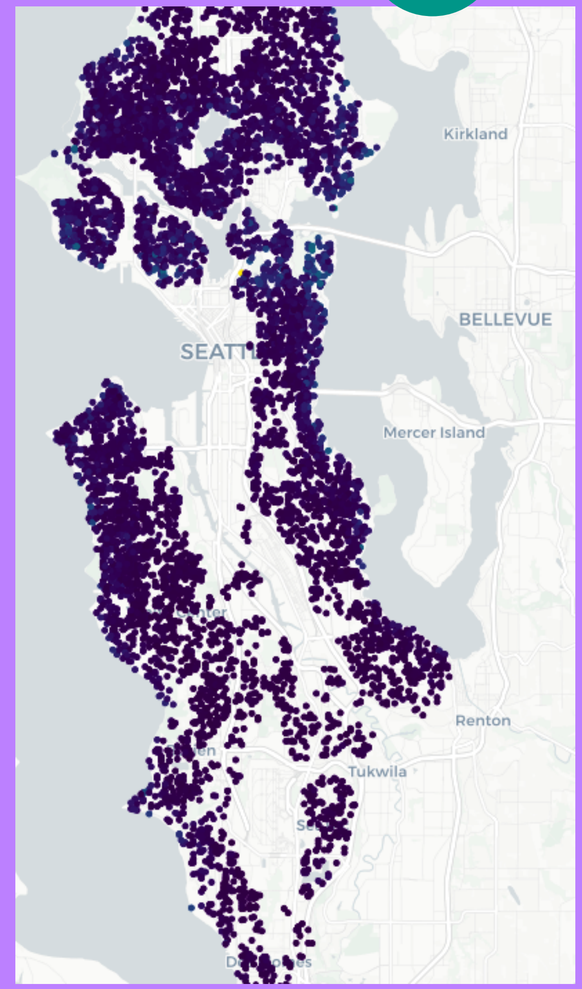


HYPOTHESIS (1)

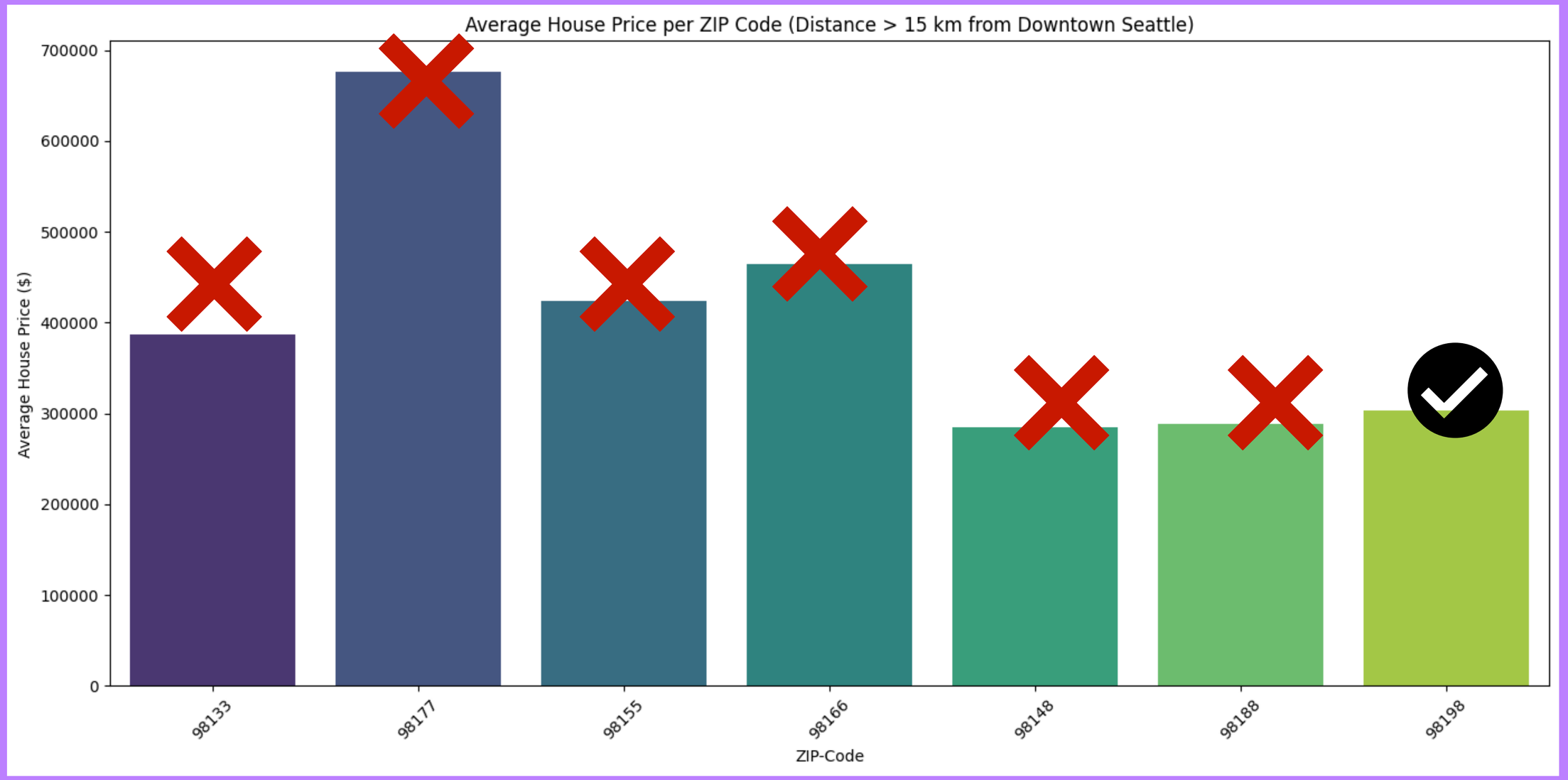
THE PRICE DEPENDS ON THE DISTANCE TO THE CENTER-COORDINATES OF SEATTLE, WASHINGTON (DOWN-TOWN)

PERIPHAL DEFINITION (OUTSIDE CITY YES/NO)



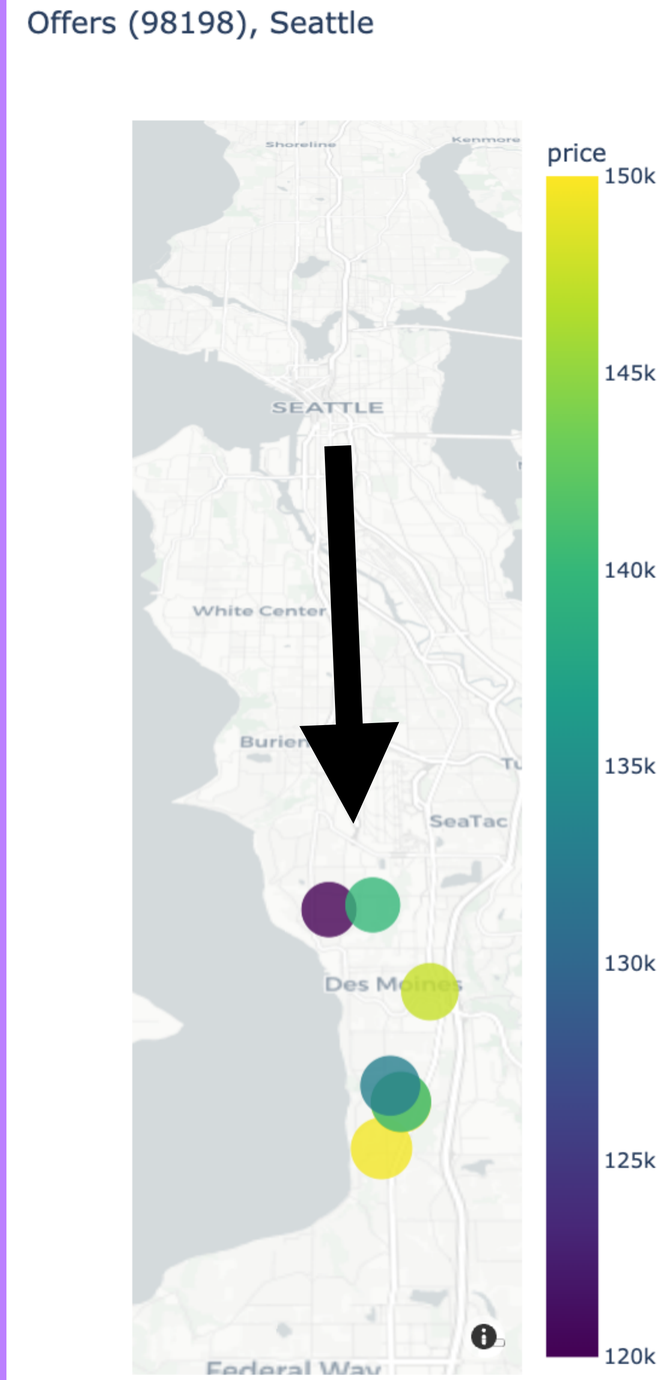


PERIPHAL DEFINITION > 15 KM (SEARCH FARTHEST LOCATION)

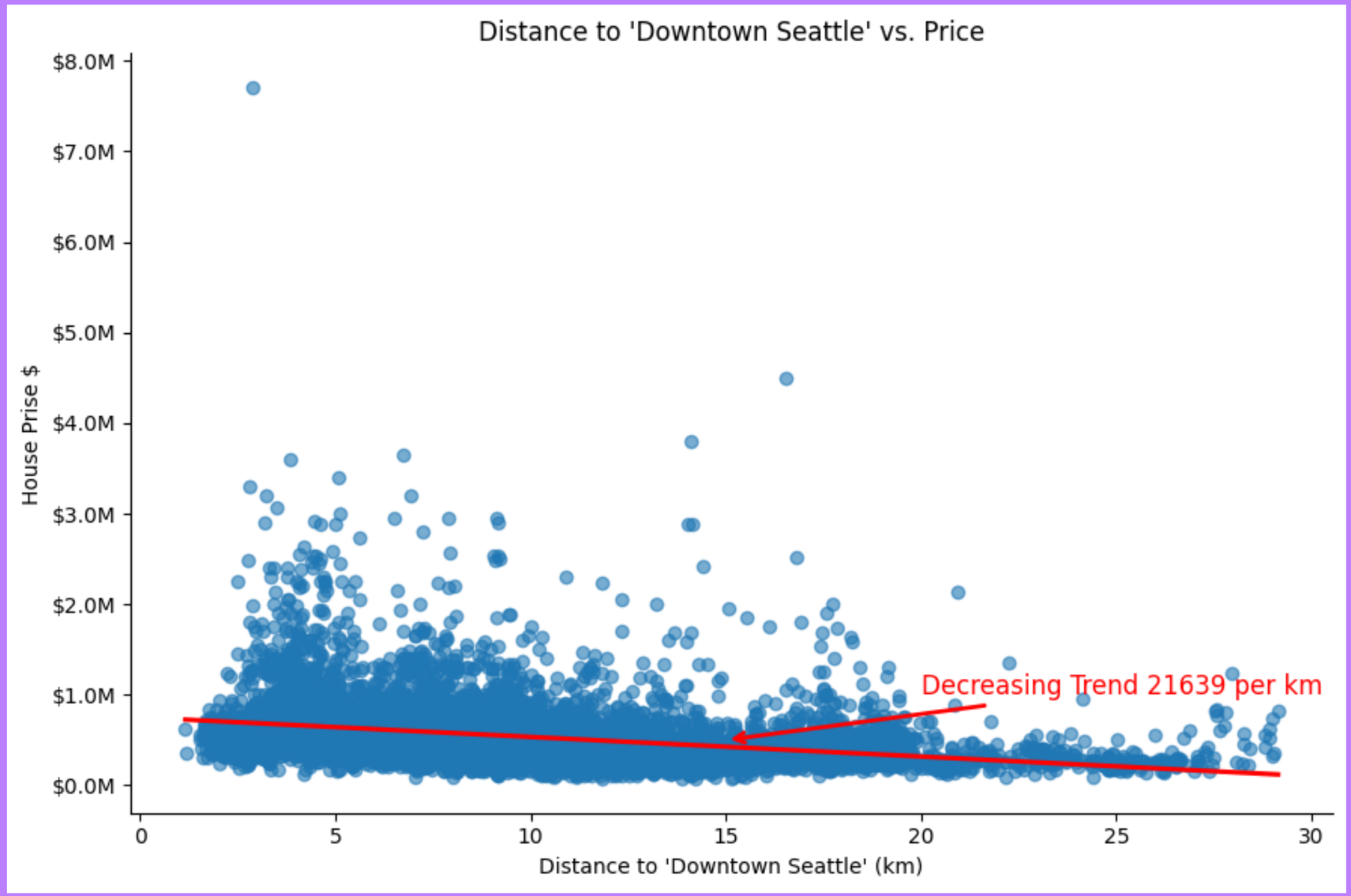


PERIPHAL PROPERTY OPPORTUNITY

- ADDITIONAL CONDITION ZIP=98198 **AND BEDROOM COUNT > 2**
- PROPERTY IS FAR AWAY FROM DOWN-TOWN
- PROPERTY IS WITHIN SEATTLE CITY



HOW BELIEVE VISUALLY



HOW BELIEVE ARGUMENTATIVE

NULL-HYPOTHESIS: THERE IS NO STATISTICALLY SIGNIFICANT BETWEEN PRICE AND DISTANCE
->SLOPE = 0

ALTERNATIVE -HYPOTHESIS: THERE IS A STATISTICALLY SIGNIFICANT BETWEEN PRICE AND DISTANCE)
->SLOPE NOT 0

SLOPE: -21639 $ FOR EVERY ADDITIONAL KM

P-VALUE (SLOPE < 0): 0.000

CHECK WITH ONE-TAIL METHOD:

SLOPE: -21639 $ FOR EVERY ADDITIONAL KM
P-VALUE (SLOPE < 0): 0.000

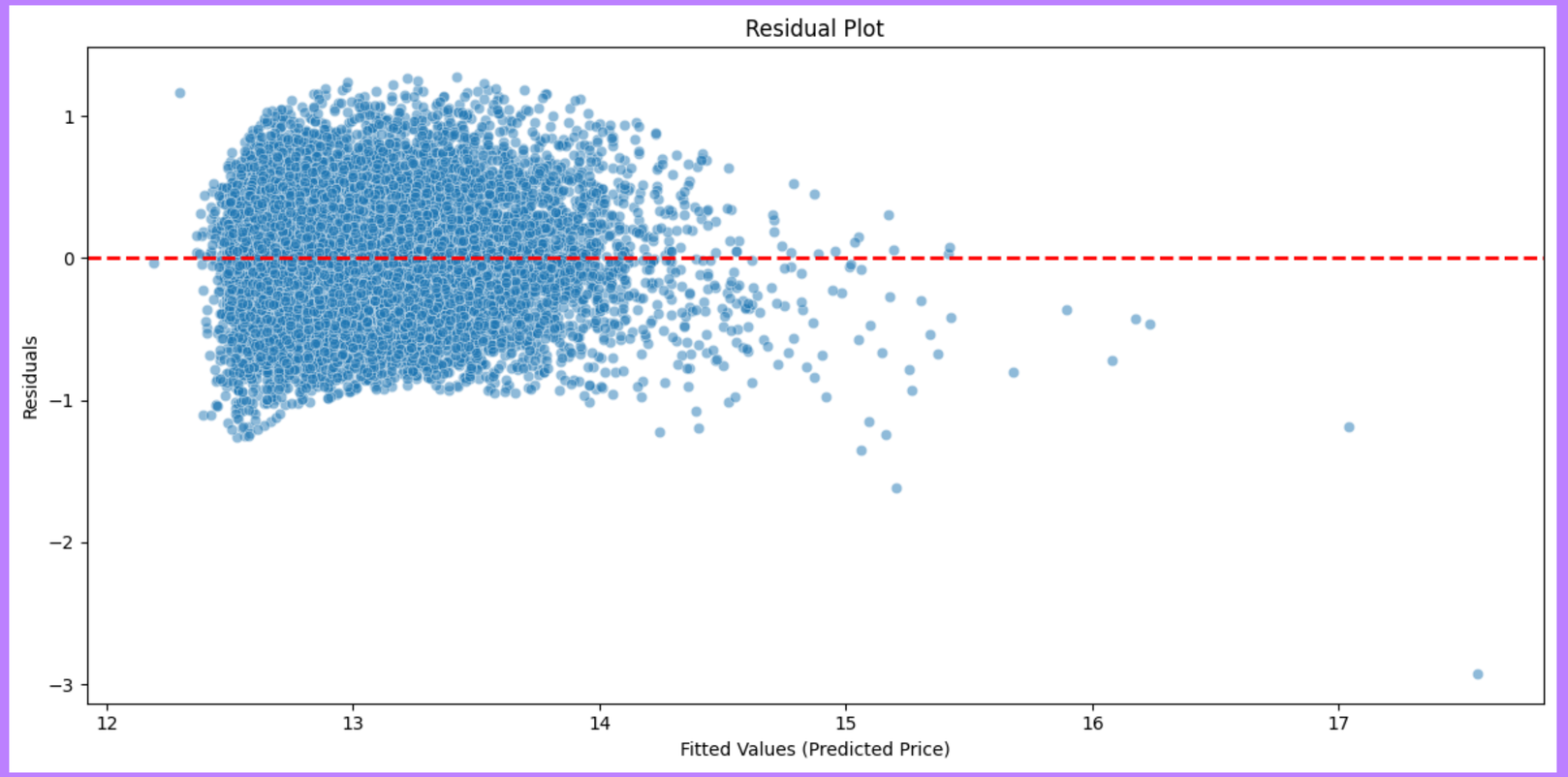
SINCE THE ESTIMATED SLOPE IS **NEGATIVE** AND THE P-VALUE IS PRACTICALLY ZERO, THE NULL-HT CAN REJECTED
—> ALTERNATIVE-HT IS VALID

HYPOTHESIS (2)

BOTH, LIVING SPACE AND LOT ARE SIGNIFICANT FOR PRICE

PRICE -> LOG(PRICE)

—> LOG($ 50000000) = 7.67
—> LOG($ 50000) = 4.67
—> AVOIDING EFFECT OF OUTLINERS



HYPOTHESIS (2) ANALYSIS

SQFT_LIVING:

SLOPE/COEF > 0 AND P = 0

—> A SIGNIFICANT POSITIV EFFECT CAN BE DETECTED

SQFT_LOT:

SLOP/COEF < 0 AND P = 0

—> A SMALL NEGATIVE EFFECT CAN BE DETECTED

OLS Regression Results

Dep. Variable:	log_price	R-squared:	0.484
Model:	OLS	Adj. R-squared:	0.484
Method:	Least Squares	F-statistic:	1.012e+04
Date:	Do, 10 Jul 2025	Prob (F-statistic):	0.00
Time:	17:33:44	Log-Likelihood:	-9653.1
No. Observations:	21597	AIC:	1.931e+04
Df Residuals:	21594	BIC:	1.934e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.2185	0.006	1915.967	0.000	12.206	12.231
sqft_living	0.0004	2.85e-06	140.760	0.000	0.000	0.000
sqft_lot	-2.695e-07	6.31e-08	-4.269	0.000	-3.93e-07	-1.46e-07

Omnibus:	3.789	Durbin-Watson:	1.978
Prob(Omnibus):	0.150	Jarque-Bera (JB):	3.795
Skew:	0.027	Prob(JB):	0.150
Kurtosis:	2.963	Cond. No.	1.09e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.09e+05. This might indicate that there are strong multicollinearity or other numerical problems.

HYPOTHESIS (2) CONCLUSION FOR AMY

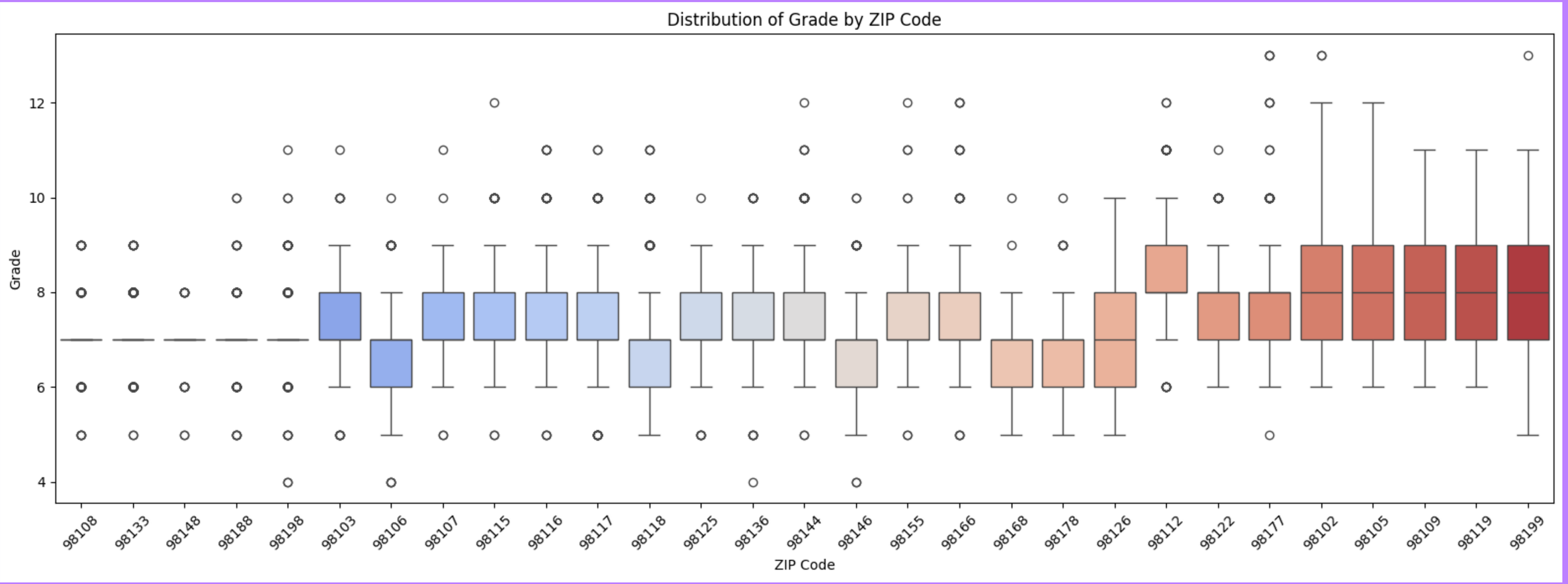
LOT SIZE ARE NOT SO SIGNIFICANT,
BECAUSE THE LOT SIZES ARE
GENERALLY MORE OR LESS SAME

AMY SHOULD CONCENTRATE MORE ON
THE SQUARE FOOTAGE DETAILS :-)

HYPOTHESIS (3)

THE DISTRIBUTION OF GRADE DEPENDS ON THE ZIP (981..)

HYPOTHESIS (3) ANALYSIS



HYPOTHESIS (3) ANALYIS

THE DISTRIBUTION OF GRADE VALUE DEPENDS ON ZIP, YES

GRADES ARE NOT EVENLY DISTRIBUTED, IT CLUSTERS DIFFERENTLY DEPENDING ON NEIGHBORHOOD

BUYERS AND SELLERS SHOULD CONSIDER LOCATION WHEN COMPARING GRADE AND PRICES

SIGNIFICANT NEGATIVE ON GRADE -> IN AVERAGE LESS GRADE THEN THE BASELINE (INTERCEPT)

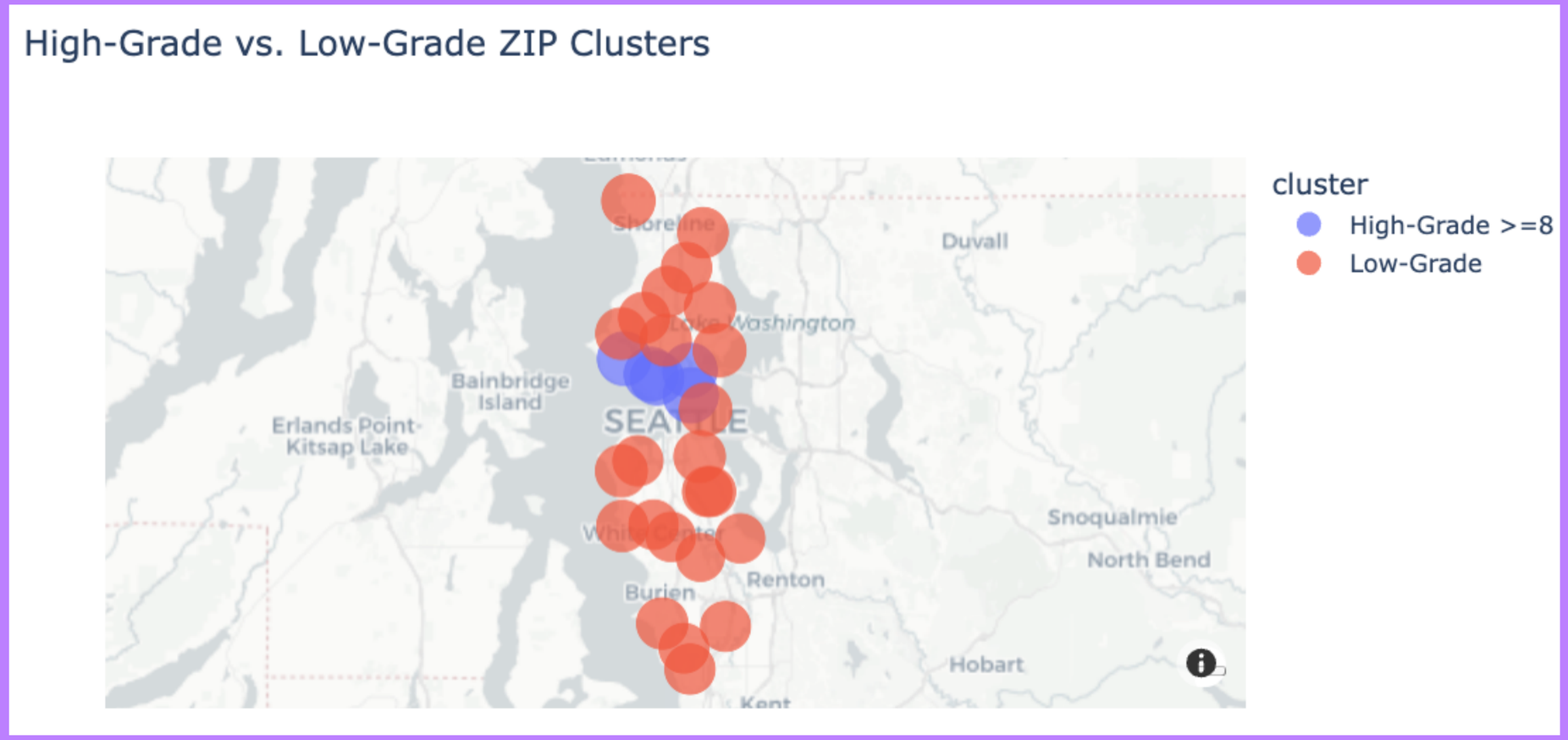
OLS Regression Results

Dep. Variable:	grade	R-squared:	0.169
Model:	OLS	Adj. R-squared:	0.166
Method:	Least Squares	F-statistic:	64.90
Date:	Do, 10 Jul 2025	Prob (F-statistic):	0.00
Time:	18:25:09	Log-Likelihood:	-12015.
No. Observations:	8973	AIC:	2.409e+04
Df Residuals:	8944	BIC:	2.429e+04
Df Model:	28		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.2981	0.091	91.518	0.000	8.120	8.476
C(zipcode)[T.98103]	-0.8878	0.098	-9.041	0.000	-1.080	-0.695
C(zipcode)[T.98105]	-0.4596	0.109	-4.204	0.000	-0.674	-0.245
C(zipcode)[T.98106]	-1.4324	0.104	-13.800	0.000	-1.636	-1.229
C(zipcode)[T.98107]	-0.8733	0.107	-8.166	0.000	-1.083	-0.664
C(zipcode)[T.98108]	-1.2819	0.113	-11.323	0.000	-1.504	-1.060
C(zipcode)[T.98109]	-0.2797	0.127	-2.207	0.027	-0.528	-0.031
C(zipcode)[T.98112]	0.1443	0.107	1.352	0.177	-0.065	0.354
C(zipcode)[T.98115]	-0.9362	0.098	-9.511	0.000	-1.129	-0.743
C(zipcode)[T.98116]	-0.7284	0.104	-7.005	0.000	-0.932	-0.525
C(zipcode)[T.98117]	-1.0160	0.099	-10.280	0.000	-1.210	-0.822
...						

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

HYPOTHESIS (3) CONCLUSION FOR AMY



98102	**High-Grade >=8**
98119	High-Grade >=8
98112	High-Grade >=8
98109	High-Grade >=8
98199	High-Grade >=8

AMY HAS PRIME-LOCATION PROPERTIES IN DOWN-TOWN SEATTLE, SHE TOLD ME :-)

TAKE-AWAY AMY

HIDE ON THE OUTSKIRTS OF THE CITY BUT NOT OUTSIDE SEATTLE

DO NOT OVERVALUE LOT-SIZE

THEIR PROPERTIES ARE LOCATED IN PRIME LOCATIONS