מבוא למערכות לומדות - תשפ"ד - תרגיל 1

תז : 324059856

7 ביוני 2024

תיאורטי:

אלגברה לינארית :

שאלה מספר 1 :

- נחשב את A^TA ו- AA^T :

1. נתחיל בלמצוא את A^T :

$$A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}$$

2. נחשב את A^TA :

$$A^TA = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

3. נחשב את AA^T :

$$AA^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

- נמצא את הערכים ווקטורים עצמיים של A^TA :
צריך לפתור $det\left(A^TA - \lambda I\right) = 0$:

$$A^TA - \lambda I = \begin{bmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{bmatrix}$$

הפולינום האופייני הוא:

$$det\left(A^TA - \lambda I\right) = (2-\lambda)\left((2-\lambda)(4-\lambda) - (-2)(2)\right) = \ldots$$

אני הולך לפתור את זה באמצעות $NumPy$ בצורך תשובות מדויקות :

```python
import numpy as np

# Define the matrix A
A = np.array([[1, 1, 0],
              [1, -1, 2]])

# Compute the SVD
U, Sigma, VT = np.linalg.svd(A, full_matrices=False)

# Print the results
print("Matrix U (left singular vectors):")
print(U)
print("\nDiagonal matrix Sigma (singular values):")
print(np.diag(Sigma))
print("\nMatrix V^T (right singular vectors):")
print(VT)
```

התוצאות :

```
Matrix U (left singular vectors):
[[ 2.35513869e-16 -1.00000000e+00]
 [ 1.00000000e+00  2.35513869e-16]]

Diagonal matrix Sigma (singular values):
[[2.44948974 0.        ]
 [0.         1.41421356]]

Matrix V^T (right singular vectors):
[[ 0.40824829 -0.40824829  0.81649658]
 [-0.70710678 -0.70710678  0.        ]]

Process finished with exit code 0
```

שאלה מספר 2 :

נשתמש בהגדרות ובתכונות של מכפלה חיצונית, דרגה, ותלות לינארית כדי לפתור את השאלה .

הגדרת מכפלה חיצונית המכפלה החיצונית A A של הווקטורים v v ו־ u u מוגדרת כך:

$$A = v \otimes u = vu^T$$

אם $v=\begin{bmatrix}v_1\\ .\\ .\\ .\\ v_n\end{bmatrix}$ ו־ $u=\begin{bmatrix}u_1\\ .\\ .\\ .\\ u_m\end{bmatrix}$ לכן המטריצה A אמורה להיות :

$$A=\begin{bmatrix}v_1u_1 & v_1u_2 & . & . & . & v_1u_m\\ v_2u_1 & v_2u_2 & . & . & . & v_2u_m\\ . & . & . & & & .\\ . & . & & . & & .\\ . & . & & & . & .\\ v_nu_1 & v_nu_2 & . & . & . & v_nu_m\end{bmatrix}$$

- דרגת המטריצה :
כל עמודה ב־A היא כפולה סקלרית של v . זה אומר שכל העמודות תלויות לינארית זו בזו.
הדרגה של מטריצה היא המספר המרבי של עמודות בלתי תלויות לינארית. לכן , $rank\,(A)=1$
- תלות לינארית של השורות והעמודות :
תלות לינארית של השורות:
כל שורה i ב־ A היא כפולה סקלרית של u^T אם $v_i\neq 0$ לכל שורה, אז כל השורות תלויות לינארית זו בזו.
תלות לינארית של העמודות:
כפי שהוזכר, כל העמודות הן כפולות סקלריות של v , ולכן תלויות לינארית.
- סיכום :
לכן, המכפלה החיצונית $A=vu^T$ נותנת מטריצה עם דרגה 1, כאשר כל השורות והעמודות בה תלויות לינארית. המסקנה הזו תקפה אלא אם v או u הם וקטור האפס, במקרה זה A עצמה היא מטריצת האפס עם דרגה 0.

שאלה מספר 3 :

ברצוני להראות כי לכל בסיס אורתונורמלי $(u_1,...,u_n)$ ולכל וקטור כלשהו $x\in\mathbb{R}^n$ שמוצג כך:

$$\sum_{i=1}^{n}u_i\cdot a_i=x$$

מתקיים כי $\langle x,u_i\rangle=a_i$ לכל $i\in[1,n]$ נשתמש בתכונות של בסיס אורתונורמלי ומכפלה פנימית כדי להוכיח זאת.

- בסיס אורתונורמלי :
בסיס $(u_1,...,u_n)$ ל־$\mathbb{R}^n$ וא אורתונורמלי אם כל אחד מוקטורי הבסיס הוא באורך יחידה והוקטורים הם מאונכים זה לזה. כלומר:

$$\langle u_i,u_j\rangle=\begin{cases}1 & if\ i=j\\ 0 & if\ i\neq j\end{cases}$$

- הצגת x בבסיס :
נתון ש־x ניתן להצגה כצירוף לינארי של וקטורי הבסיס:

$$u_1a_1+...+u_na_n=x$$

המטרה היא למצוא את המקדמים a_i
- חישוב של a_i :
כדי למצוא את המקדם a_i , נבצע מכפלה פנימית של x , עם כל אחד מוקטורי הבסיס u_i :

$$\langle x,u_i\rangle=\left\langle\sum_{j=1}^{n}a_ju_j,u_i\right\rangle$$

בהתבסס על תכונת הליניאריות של המכפלה הפנימית, הביטוי הזה נפתח ל:

$$\langle x,u_i\rangle=\sum_{j=1}^{n}a_j\left\langle u_j,u_i\right\rangle$$

בהתחשב בתכונה האורתונורמלית של הבסיס:

$$\langle u_j,u_i\rangle=\delta_{i,j}$$

בהחלפה במשוואה זו, נקבל:

$$\langle x,u_i\rangle=\sum_{j=1}^{n}a_j\delta_{i,j}=a_i$$

כיוון שכל האיברים שבהם $i\neq j$ מתאפסים בגלל ש־ $\delta_{i,j}=0$, וכאשר $i=j$, $\delta_{i,j}=1$, ומתקבל a_i .

שאלה מספר 4 :

1. צריך להראות ש־ P סימטרית :
מטריצת ההיטל האורתוגונלי P על V נתונה ע"י $P=\sum_{i=1}^{k}v_iv_i^T$ כאשר v_i הם וקטורים אורתונורמליים , כדי להראות ש־P סימטרית , צריך להראות ש־ $P=P^T$:
מכיוון שהשכפול של סכום הוא סכום השכפולים, והשכפול של מכפלה הוא מכפלת השכפולים בסדר הפוך, יש לנו:

$$P^T=\left(\sum_{i=1}^{k}v_iv_i^T\right)^T=\sum_{i=1}^{k}\left(v_iv_i^T\right)^T=\sum_{i=1}^{k}v_iv_i^T=P$$

לכן P סימטרית

2. צריך להוכיח ש־ $0,1$ הם ערכים עצמיים של P , ו־ $v_1,...,v_i$ הם הוקטורים העצמיים של הערך עצמי 1 :
ערך עצמי 1 : מכיוון ש־ v_i אורתונורמליים לכל v_i :

$$Pv_i=\sum_{j=1}^{k}v_jv_j^Tv_i=\sum_{j=1}^{k}v_i\delta_{ji}=v_i$$

כך ש־ δ_{ji} דלתא של קרונק , שהיא 1 כאשר $i=j$ אחרת 0 . זה מראה שכל v_i הוא וקטור עצמי של P עם ערך עצמי 1

3. צריך להראות שלכל $v\in V$, $Pv=v$
לכל $v\in V$, ניתן להביע את v כשילוב ליניארי של וקטורי הבסיס:

$$v=\sum_{i=1}^{k}c_iv_i$$

בהחלת P :

$$Pv=P\left(\sum_{i=1}^{k}c_iv_i\right)=\sum_{i=1}^{k}c_iPv_i=\sum_{i=1}^{k}c_iv_i=v$$

זה מראה ש־ $Pv=v$ לכל $v\in V$

4. צריך להוכיח ש־ $P^2=P$:

$$P^2=\left(\sum_{i=1}^{k}v_iv_i^T\right)\left(\sum_{j=1}^{k}v_jv_j^T\right)=\sum_{i=1}^{k}\sum_{j=1}^{k}v_i\left(v_i^Tv_j\right)v_j^T=$$

$$\sum_{i=1}^{k}v_iv_i^T=P$$

5. צריך להוכיח ש־ $(I-P)\,P=0$:

$$(I-P)\,P=P-P^2=0$$

$P^2=P$ הוכחנו את זה בסעיף שלעייל .

חשבון רב משתנים :

שאלה מספר 1 :

- ביטוי הפונקציה:
כדי לחשב את הגרדיאנט של הפונקציה $h(\sigma)$ שנתונה ע"י :

$$h(\sigma)=\frac{1}{2}\left\|h(\sigma)-y\right\|^2=\frac{1}{2}\sum_{i=1}^{d}\left(f_i(\sigma)-y_i\right)^2$$

כך ש־ $f_i(\sigma)$ ו־y_i הם הרכיב ה־i של $f(\sigma)$ ו־y , בהתאמה .

- החלת כלל השרשרת:
כדי למצוא את $\nabla h(\sigma)$, הגרדיאנט של h , לשים לב ש־ $h(\sigma)$ תלויה ב־σ דרך $f(\sigma)$. הגרדיאנט של פונקציה שמורכבת כך מחושבת על ידי שימוש בכלל השרשרת, שכולל גזירות חלקיות .
גזירת h לפי σ כוללת את גזירת הפונקציה הפנימית $f(\sigma)$ והפונקציה החיצונית (ריבוע הנורמה). במיוחד, משתמשים ב:

$$\frac{\partial h}{\partial \sigma_k} = \sum_{i=1}^{d} \frac{\partial}{\partial \sigma_k} \left(\frac{1}{2} (f_i(\sigma) - y_i)^2 \right)$$

- גזירת הריבועים:
לפי כלל השרשרת, עבור כל רכיב i

$$\frac{\partial}{\partial \sigma_k} \left(\frac{1}{2} (f_i(\sigma) - y_i)^2 \right) = (f_i(\sigma) - y_i) \frac{\partial f_i(\sigma)}{\partial \sigma_k}$$

כאן $\frac{\partial f_i(\sigma)}{\partial \sigma_k}$ היא הנגזרת החלקית של הרכיב ה־i של f לפי הרכיב ה־k של σ .

- סכימה של התרומות:
נסכם את התרומות של כל הרכיבים i :

$$\nabla h(\sigma) = \sum_{i=1}^{d} (f_i(\sigma) - y_i) \nabla f_i(\sigma)$$

כך ש־ $\nabla f_i(\sigma)$ מייצג את הגרדיאנט של f_i לפי σ .

שאלה מספר 2 :

1. מציאת הנגזרות החלקיות :
מקרה 1 : $i = j$:

$$\frac{\partial S(x)_i}{\partial x_i} = \frac{\partial}{\partial x_i} \left(\frac{e^{x_i}}{\sum_{l=1}^{k} e^{x_l}} \right)$$

נשתמש בכלל המנה:

$$\frac{\partial S(x)_i}{\partial x_i} = \frac{e^{x_i} \cdot \sum_{l=1}^{k} e^l - e^{x_i} \cdot e^{x_i}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} = \frac{e^{x_i} \left(\sum_{l=1}^{k} e^{x_l} - e^{x_i}\right)}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}$$

נבצע פישוט:

$$\frac{\partial S(x)_i}{\partial x_i} = S(x)_i (1 - S(x)_i)$$

מקרה 2 : $i \neq j$:

$$\frac{\partial S(x)_i}{\partial x_j} = \frac{\partial}{\partial x_j} \left(\frac{e^{x_i}}{\sum_{l=1}^{k} e^{x_l}} \right)$$

שוב נשתמש בכלל המנה :

$$\frac{\partial S(x)_i}{\partial x_j} = \frac{0 \cdot \sum_{l=1}^{k} e^l - e^{x_i} \cdot e^{x_j}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} = \frac{-e^{x_i} e^{x_j}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}$$

נבצע פישוט:

$$\frac{\partial S(x)_i}{\partial x_i} = -S(x)_i S(x)_j$$

2. בניית מטריצת היעקוביאן :
באמצעות התוצאות מהמקרים הקודמים, נבנה את מטריצת היעקוביאן J :

$$J_{i,j} = \begin{cases} S(x)_i (1 - S(x)_i) & i = j \\ -S(x)_i S(x)_j & i \neq j \end{cases}$$

אז מטריצת היעקוביאן J של פונקציית הסופטמקס היא:

$$J_{i,j} = S(x)_i \delta_{i,j} - S(x)_i S(x)_j$$

כאשר $\delta_{i,j}$ הוא דלתא קרונקר, שהוא 1 אם $i = j$ ו־0 אחרת .

רגרסיה לינארית :

שאלה מספר 1 :

1. צריך להוכיח ש- $ker\left(X\right)=ker\left(X^{T}X\right)$
נוכיח ש- $ker\left(X\right)\subseteq ker\left(X^{T}X\right)$:
נניח ש- $v\in ker\left(X\right)$, לכן $Xv=0$ מכפלת שני הצדדים פ- X^{T} , נקבל $X^{T}\left(Xv\right)=X^{T}0=0$
לכן $X^{T}Xv=0$, שמקיים $v\in ker\left(X^{T}X\right)$
כעת נוכיח ש- $ker\left(X^{T}X\right)\subseteq ker\left(X\right)$:
נניח $v\in ker\left(X^{T}X\right)$. אז $X^{T}Xv=0$. נשקול את הצורה הריבועית :

$$v^{T}X^{T}Xv=0$$

אני יכולים לכתוב $\left(Xv\right)^{T}\left(Xv\right)=\|Xv\|^{2}=0$, שמקיים $Xv=0$. לכן $v\in ker\left(X\right)$
ואז קיבלנו ש- $ker\left(X\right)=ker\left(X^{T}X\right)$.

2. צריך להוכיח ש- $A:Im\left(A^{T}\right)=ker\left(A\right)^{\perp}$
נניח ש- $b\in Im\left(A^{T}\right)$ אז קיים $x\in\mathbb{R}^{m}$ עם $A^{T}x=b$ ..
נניח ש- $w\in ker\left(A^{T}\right)$:

$$A^{T}w=0$$

$$\Downarrow$$

$$0=\left\langle A^{T}W,X\right\rangle =\left\langle W,Ax\right\rangle =\left\langle W,b\right\rangle$$

לכן $Im\left(A\right)\subset ker\left(A^{T}\right)^{\perp}$. להראות ההכלה ההפוכה $ker\left(A^{T}\right)^{\perp}\subset Im\left(A\right)$ נניח ש- $b\neq Im\left(A\right)$ ונראה ש- $b\notin ker\left(A^{T}\right)^{\perp}$. מספיק למצוא וקטור $c\in ker\left(A^{T}\right)$ כך $\left\langle b,c\right\rangle \neq0$. אכן , לפי מההנחה ש- $b\notin Im\left(A\right)$, הוקטור b אמור להכיל רכיב ב- $Im\left(A\right)^{\perp}$.
נניח ש- $c\in Im\left(A\right)^{\perp}$ כך ש- $\left\langle b,c\right\rangle \neq0$
כעת , מכיוון ש- $c\in Im\left(A\right)^{\perp}$ אנחנו יודעים ש- c אורתוגונלי לכל וקטור ב- $Im\left(A\right)$ אז $\left\langle c,AA^{T}c\right\rangle =0$

$$\left\Vert A^{T}c\right\Vert =\left\langle A^{T}c,A^{T}c\right\rangle =\left\langle c,AA^{T}c\right\rangle =0$$

לכן $A^{T}c=0\leftarrow c\in ker\left(A^{T}\right)$.

3. מכיוןש- X אינה הפיכה , למערכת אין פתרונות או יש א"ס פתרונות .
אנחנו יודעים שלמערכת יש לפחות פתרון יחיד אם"ם $y\in Im\left(X\right)$, וגם למערכת יש א"ס פתרונות אם"ם :

$$y\in Im\left(X\right)\overset{*}{\Longleftrightarrow}y\in ker\left(X\right)^{\perp}\Longleftrightarrow y\perp ker\left(X\right)$$

* : משאלה 2

4. פתרון יחיד : אם $X^{T}X$ הפיכה אז ל- X יש את כל העמודות ב- $rank$ ללא אפס ע"ע . ובכך המערכת $X^{T}Xw=X^{T}y$ יש לה פתרון יחיד שניתן ע"י $w=\left(X^{T}X\right)^{-1}X^{T}y$.
א"ס פתרונות : אם $X^{T}X$ אינה הפיכה אז קיים $kernel$ שאינו טריוויאלי . כל וקטור v ב- $kernel$ מקיים $X^{T}Xw=0$. לכן, לכל פתרון w של- $X^{T}Xw=X^{T}y$, $w+v$ הוא גם פתרון לכל וקטור ב- $ker\left(X^{T}X\right)$.
זה מרמז על אינסוף פתרונות מכיוון שיש אינסוף אפשרויות עבור v .

הריבועים הפחותים :

שאלה מספר 1 :

משימה : בהנחה ש- $X^{T}X$ הפיכה, הראה כי הפתרון הכללי שהושג בשיעור 3 (באמצעות הפסאודו הופכי X^{+}) שווה לפתרון שנראה בהרצאה 1 $\left(X^{T}y^{1-\left(X^{T}X\right)}\right)$.
פתרון :

1. נזכור את הנוסחה לפסאודו הופכי X^{+}:

$$X^{+}=C\overset{+}{\sum}U^{T}$$

2. נוסחה לפתרון כללי באמצעות פסאודו הופכי:

$$w = X^+ y = V \sum^{+} U^T y$$

3. מטריצה X^TX וההפיכות שלה:

$$X^TX = \left(U \sum V^T\right)^T U \sum V^T = V \sum^{T} U^T U \sum V^T =$$

$$V \sum^{2} V^T$$

מכיוון ש- $U^TU = I$ ו- $V^TV = I$ (בגלל ש-U ו-V מטריצות אורתוגונולי) , והנחת X^TX הפיכה מקיים ש- ל- $\sum$ יש $rank$ מלא (ללא ע"ע) .

4. נוסחה באמצעות שיטת משוואת הנורמל מהרצאה 1:

$$\left(X^TX\right)^{-1} X^T y = \left(V \sum^{2} V^T\right)^{-1} \left(V \sum U^T\right) y$$

נפשט באמצעות $V^{-1} = V^T$ (מכיוון ש- V אורתוגונולי):

$$= V \sum^{-2} V^T V \sum U^T y = V \sum^{-1} U^T y$$

זה נובע מ-$\sum^2$ היא מטריצה אלכסונית של σ_i^2 , לכן ההופכי שלה הוא σ_i^{-2}

5. השוואה בין שני התוצאות:

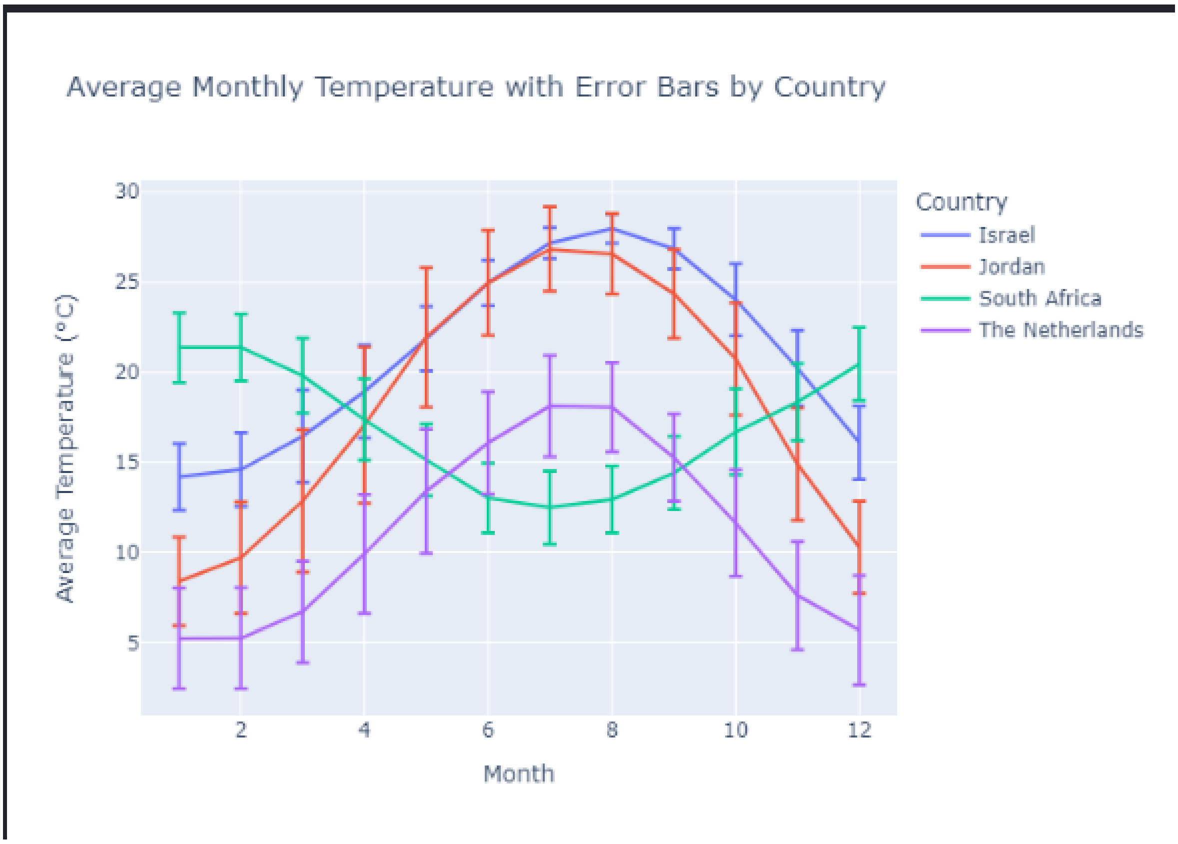
$$V \sum^{+} U^T y = V \sum^{-1} U^T y$$

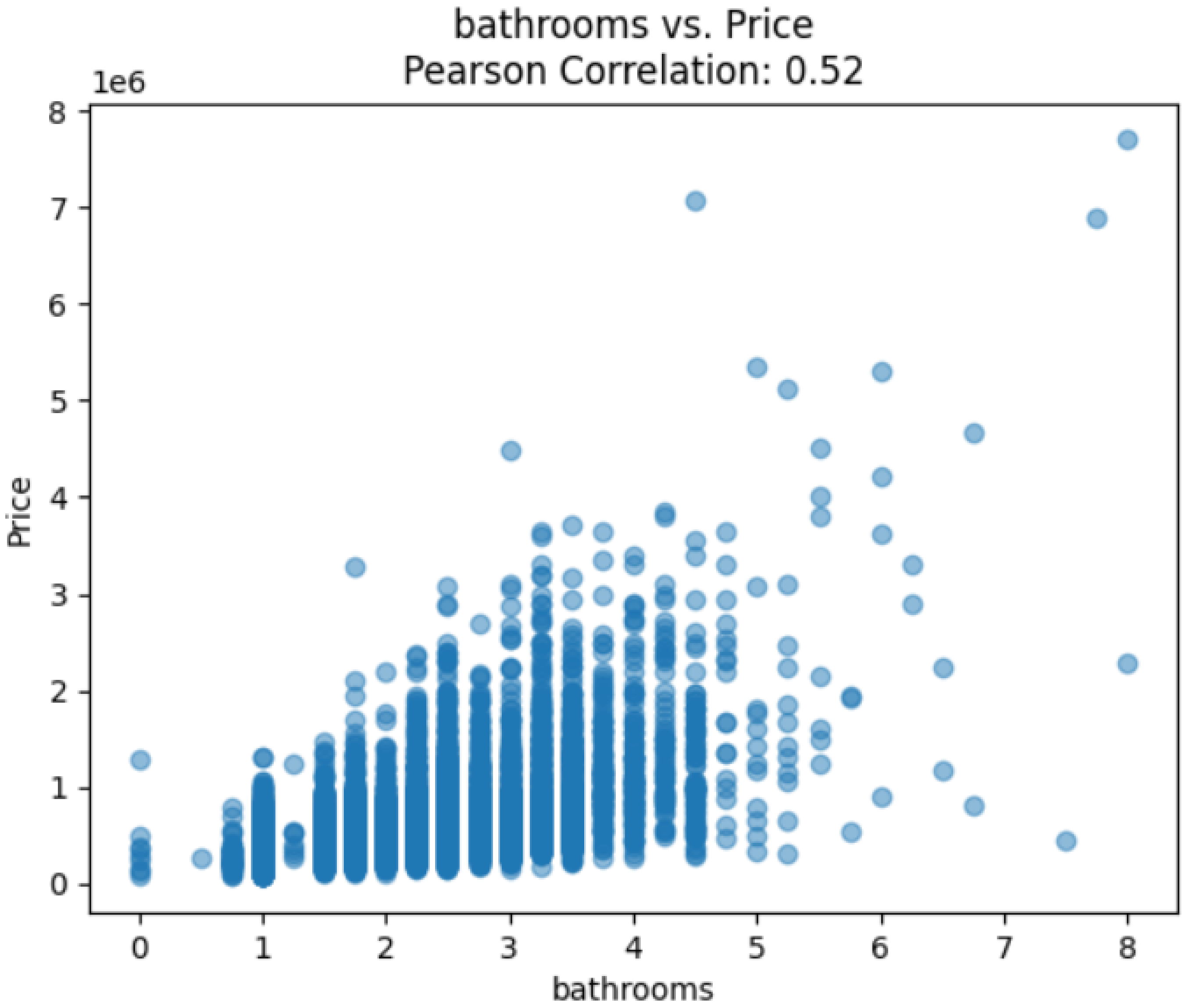
מכיוון $\sum^+ = \sum^{-1}$ (כאשר $\sum^+$מחליפה σ_i שאינו אפס עם σ_i^{-1}) תי השיטות נותנות את אותה תוצאה:

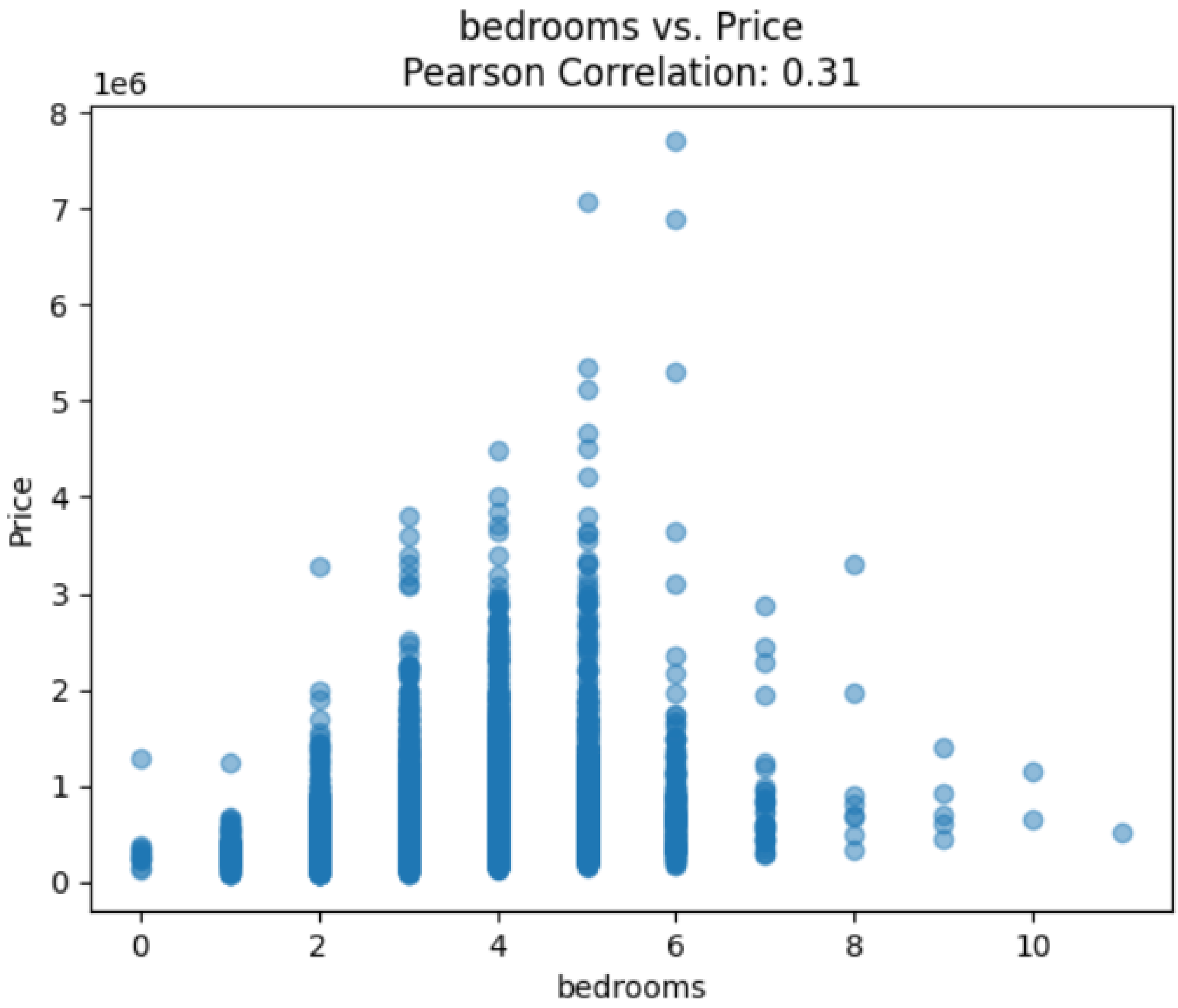
$$X^+ y = \left(X^TX\right)^{-1} X^T y$$

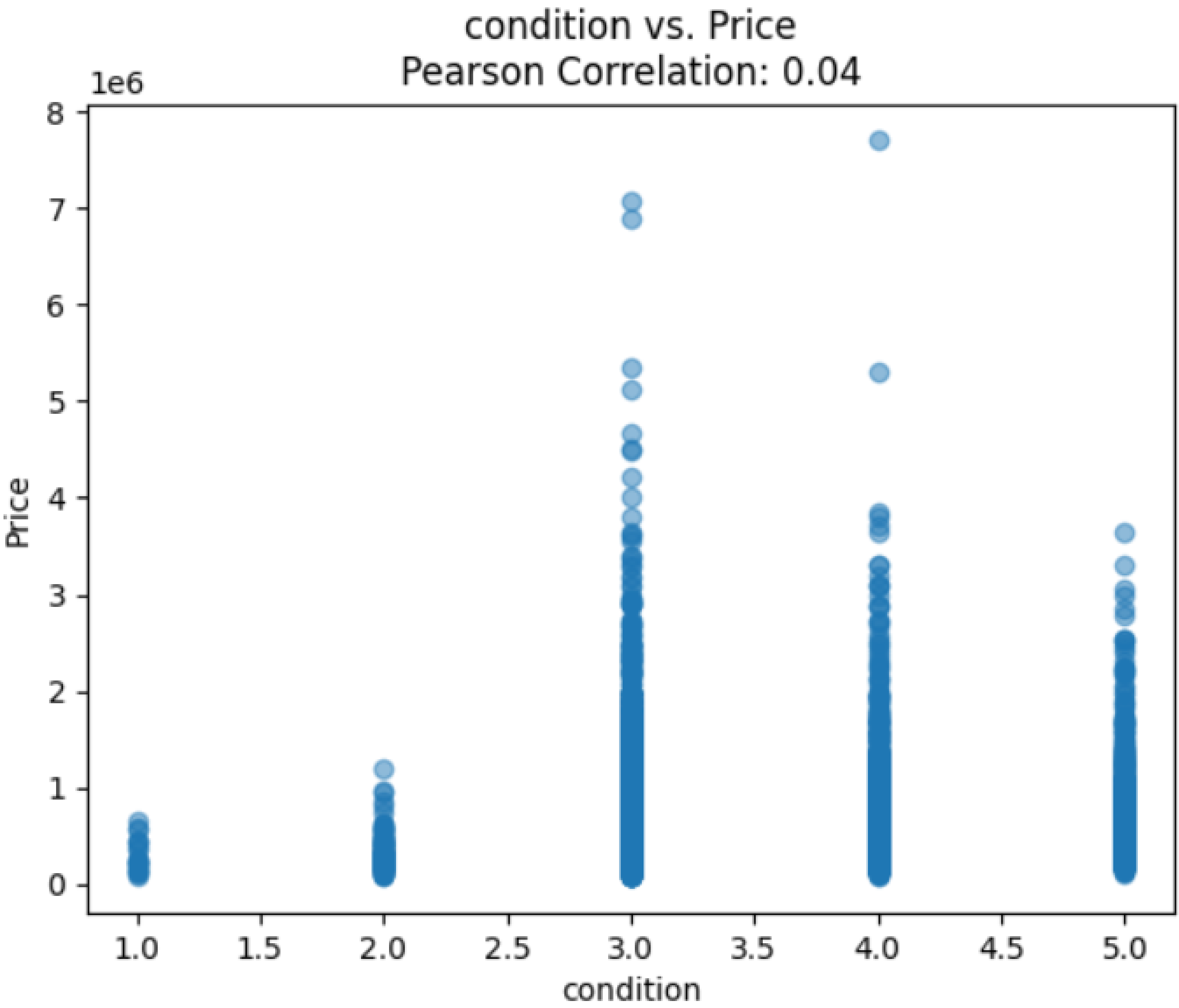
ה מראה שכאשר X^TX הפיכה, הפתרונות באמצעות הפסאודו הופכי ושיטת משוואת הנורמל שקולים, תומכים בתוצאות ההרצאה והתרגול .

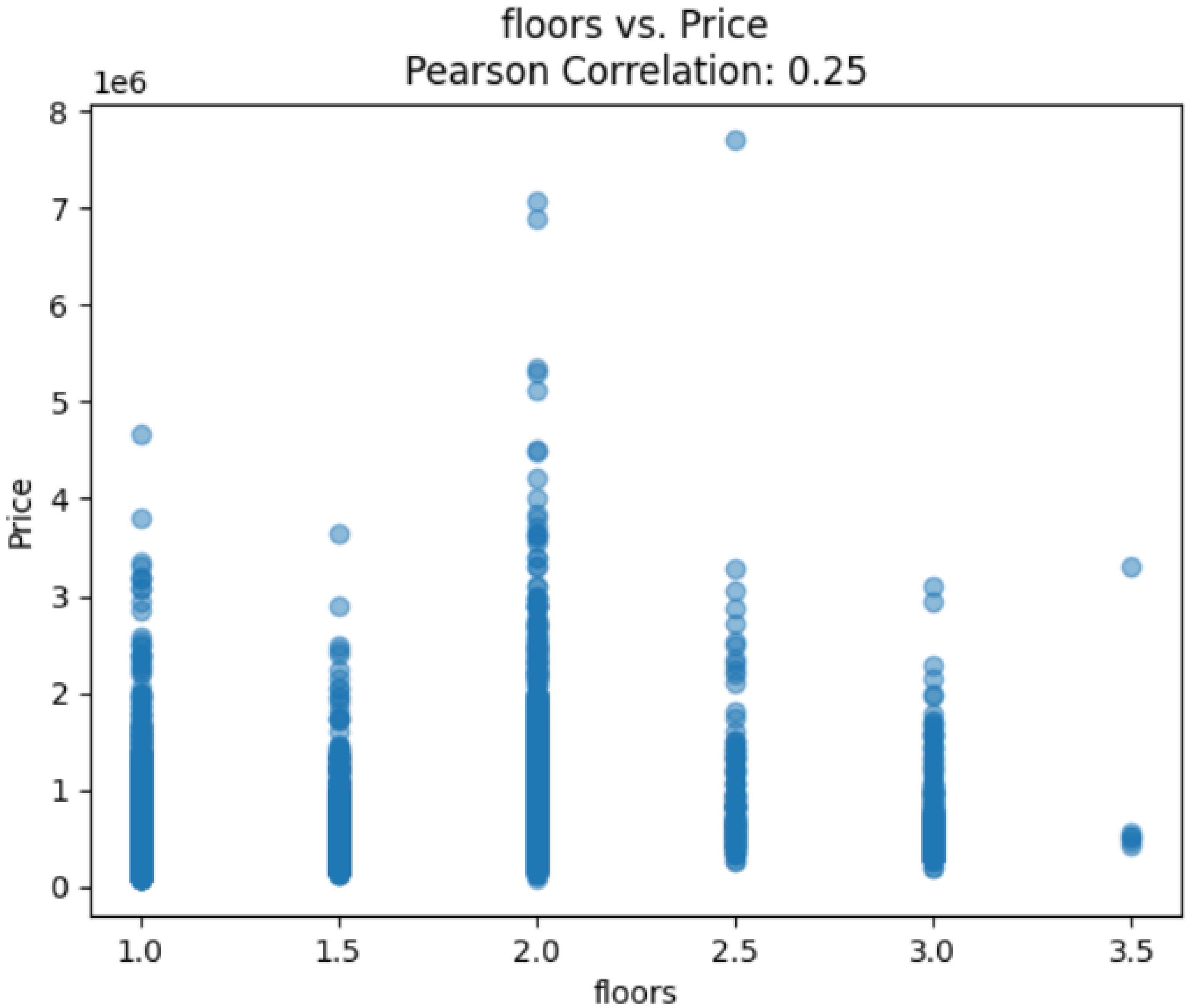
מעשי :

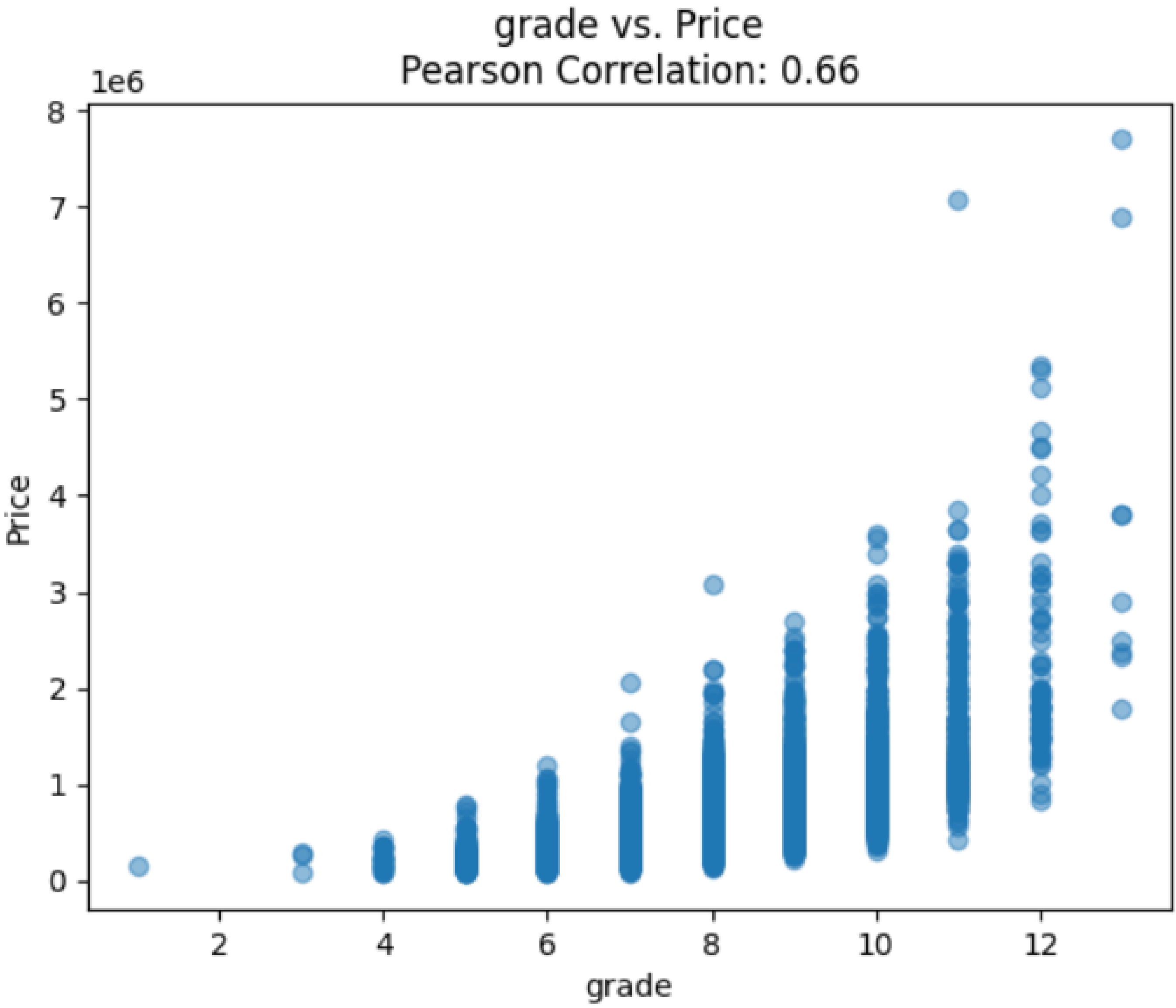


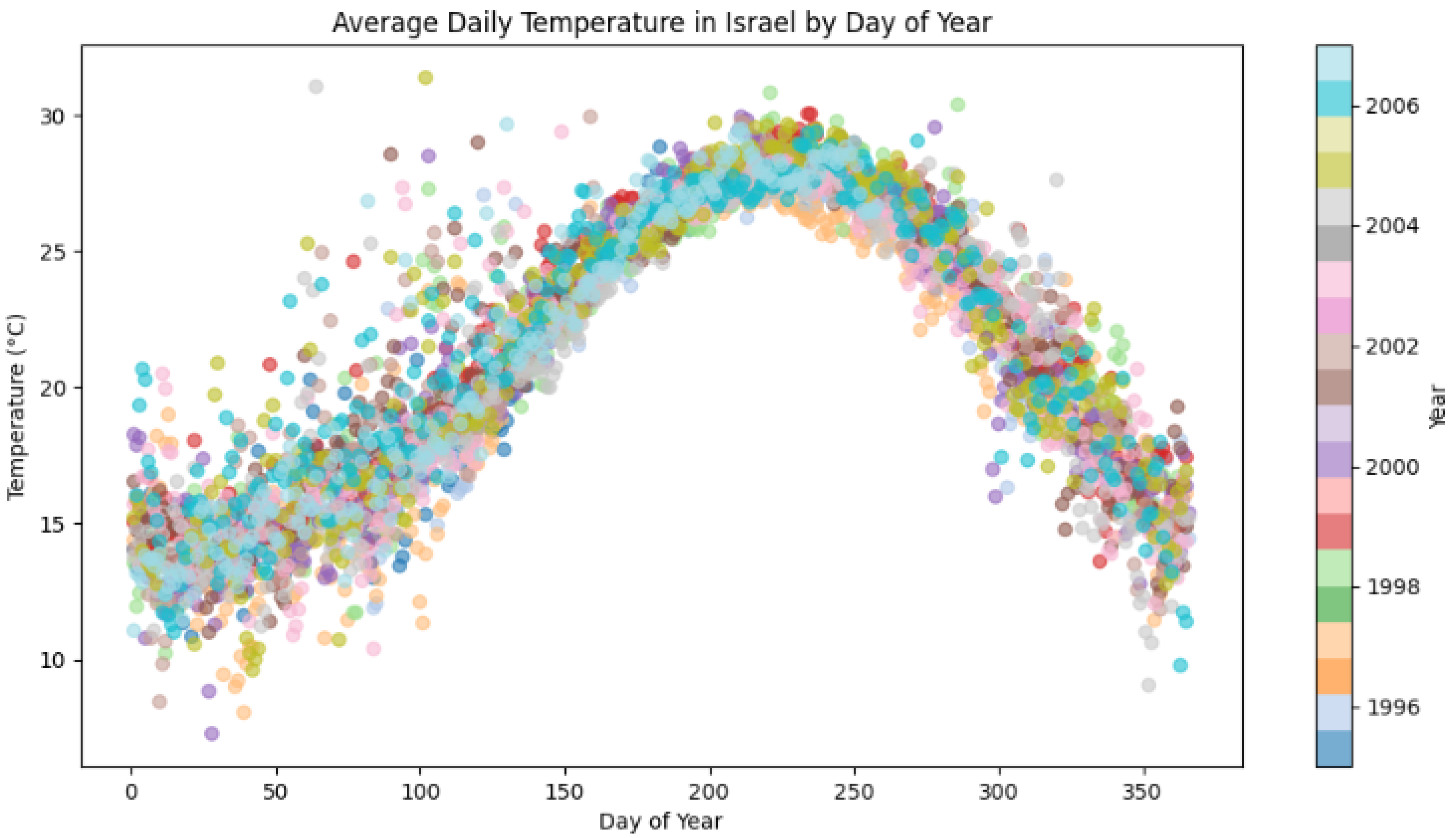

bathrooms vs. Price
Pearson Correlation: 0.52
1e6
Price
bathrooms
0
1
2
3
4
5
6
7
8

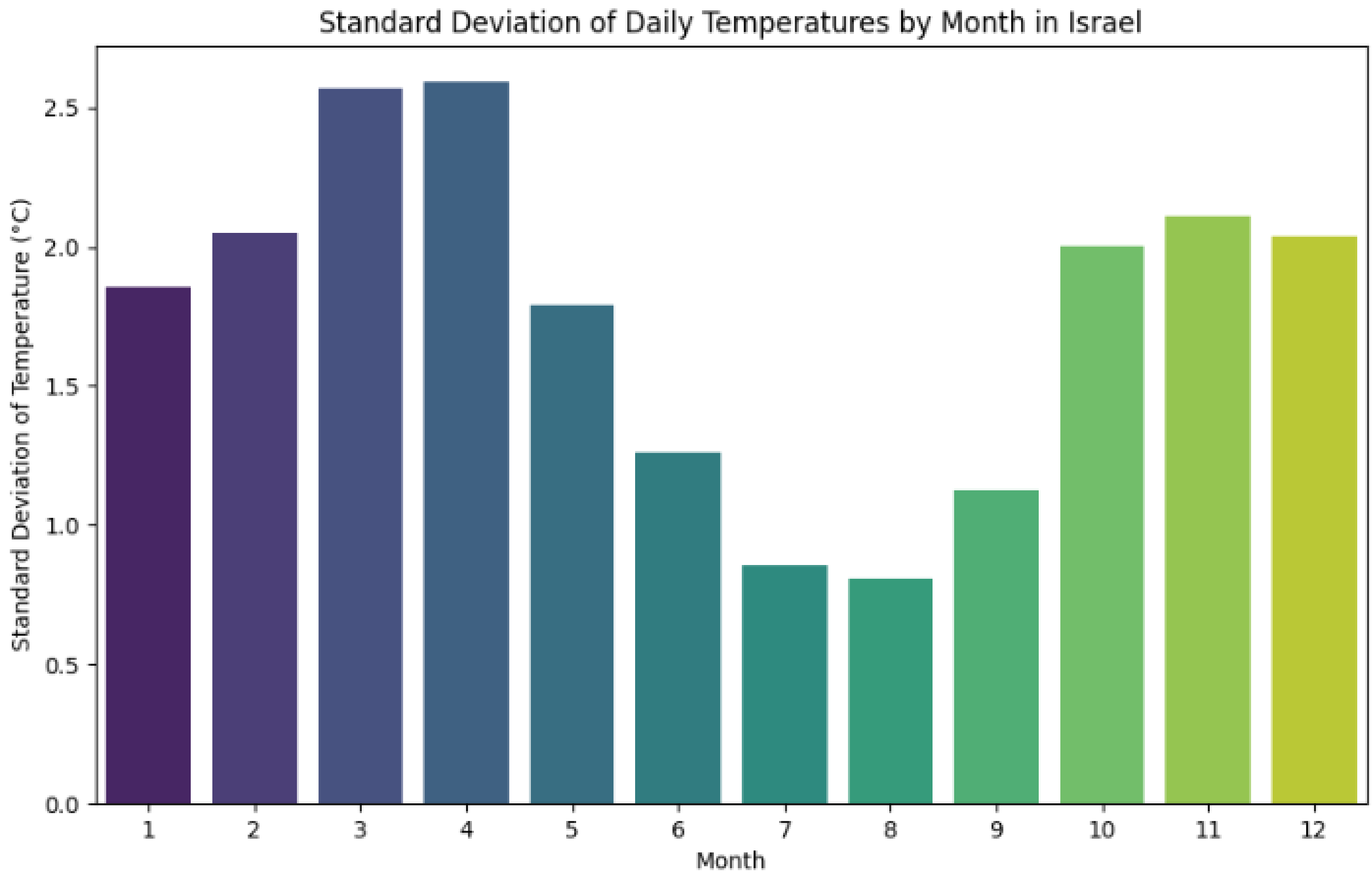

bedrooms vs. Price
Pearson Correlation: 0.31
1e6
8
7
6
5
4
3
2
1
0
Price
0
2
4
6
8
10
bedrooms

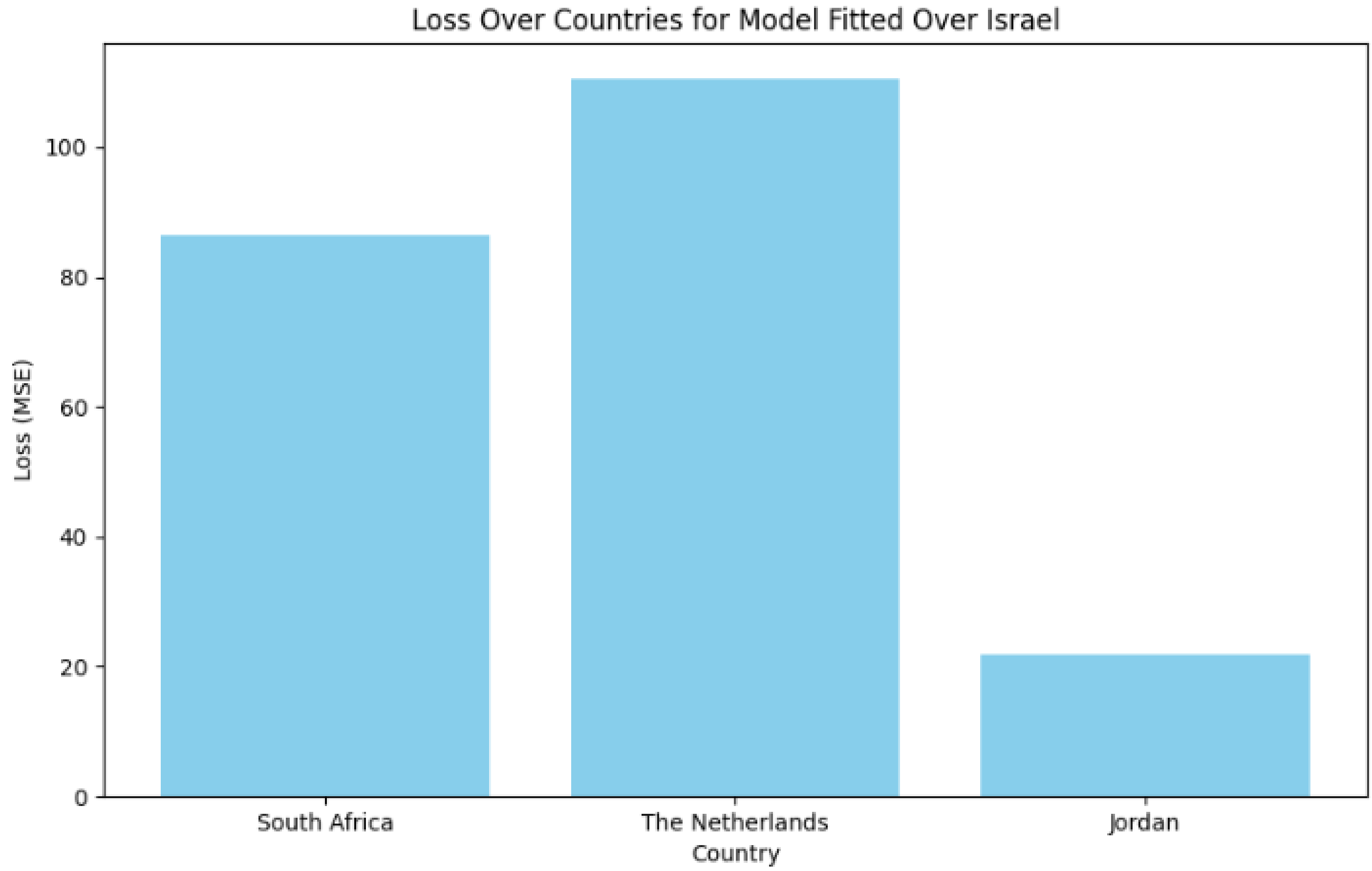

condition vs. Price
Pearson Correlation: 0.04
1e6
8
7
6
5
4
3
2
1
0
Price
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
condition

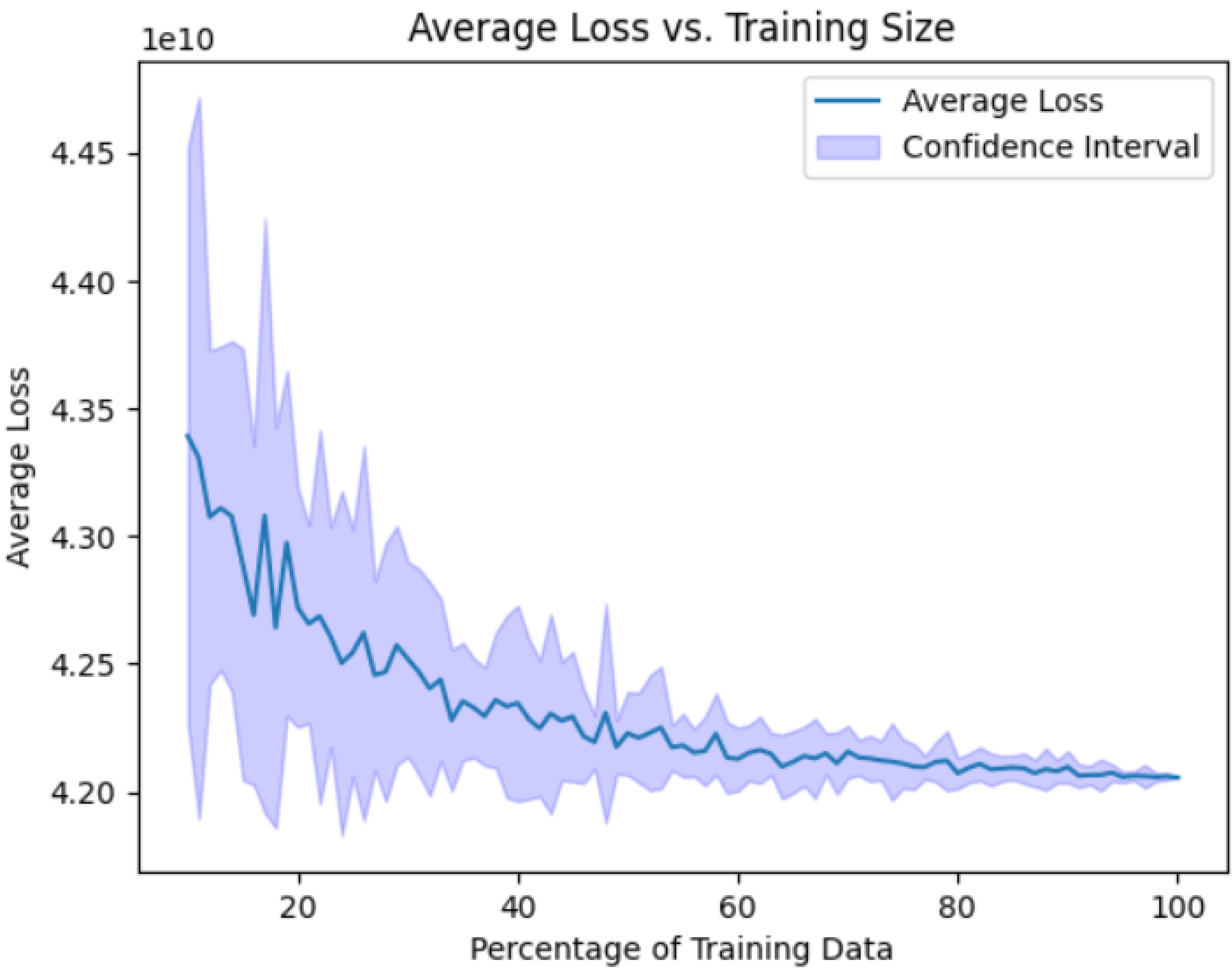

floors vs. Price
Pearson Correlation: 0.25
1e6
8
7
6
5
4
3
2
1
0
Price
1.0
1.5
2.0
2.5
3.0
3.5
floors

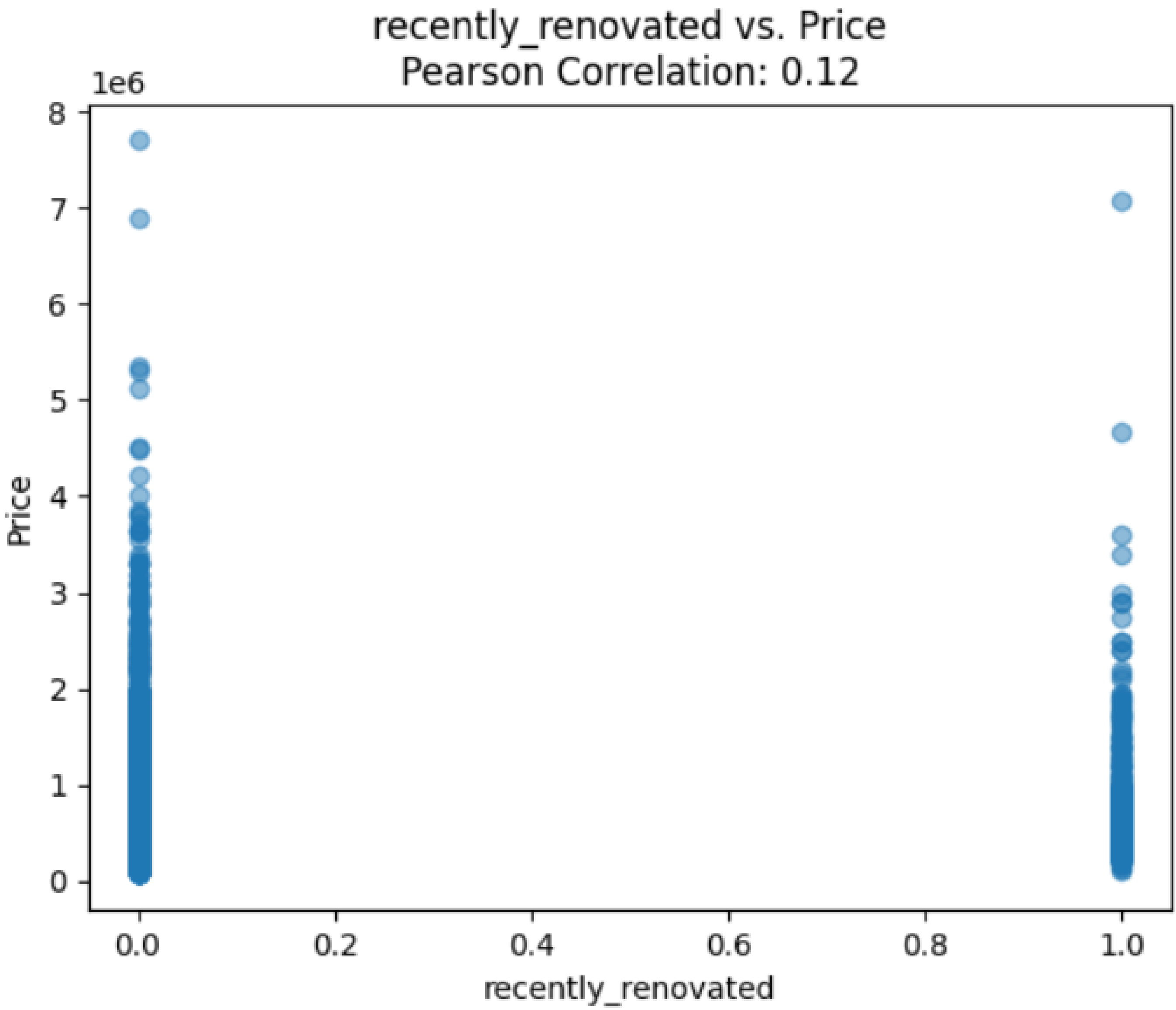

grade vs. Price
Pearson Correlation: 0.66
1e6
8
7
6
5
4
3
2
1
0
Price
2
4
6
8
10
12
grade

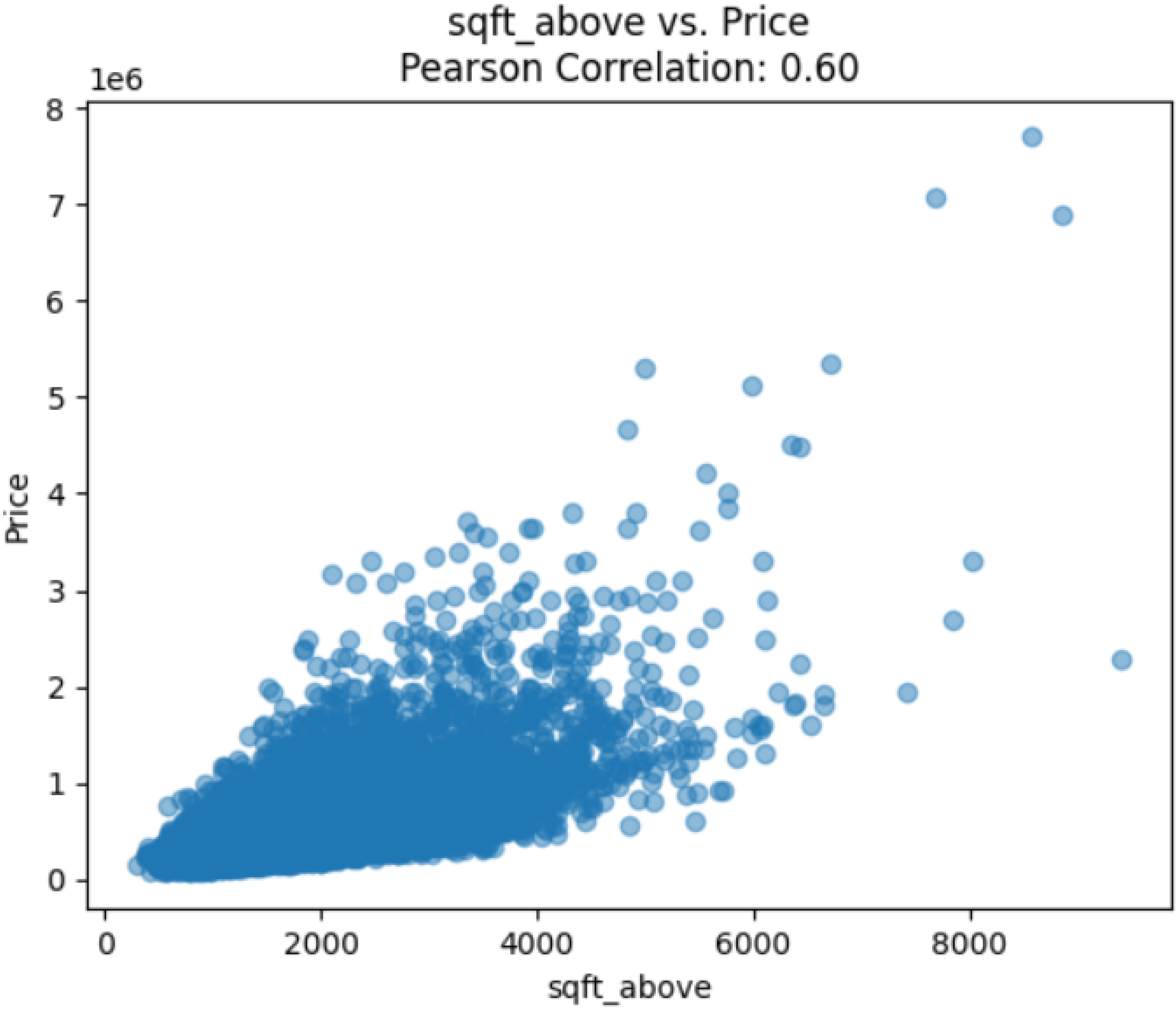

Average Daily Temperature in Israel by Day of Year
Temperature (°C)
Day of Year
Year
30
25
20
15
10
0
50
100
150
200
250
300
350
2006
2004
2002
2000
1998
1996

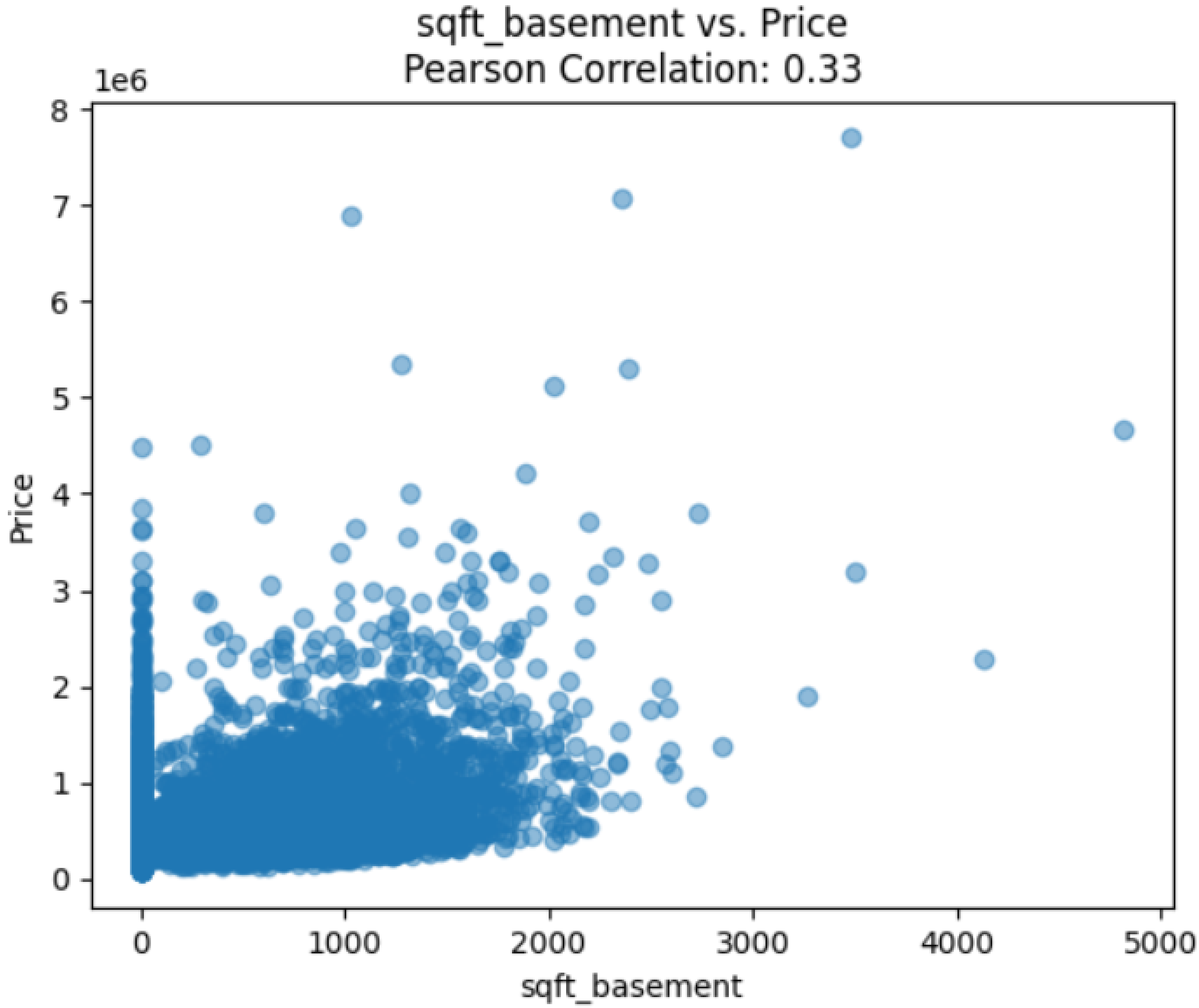

Standard Deviation of Daily Temperatures by Month in Israel
Standard Deviation of Temperature (°C)
Month
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12

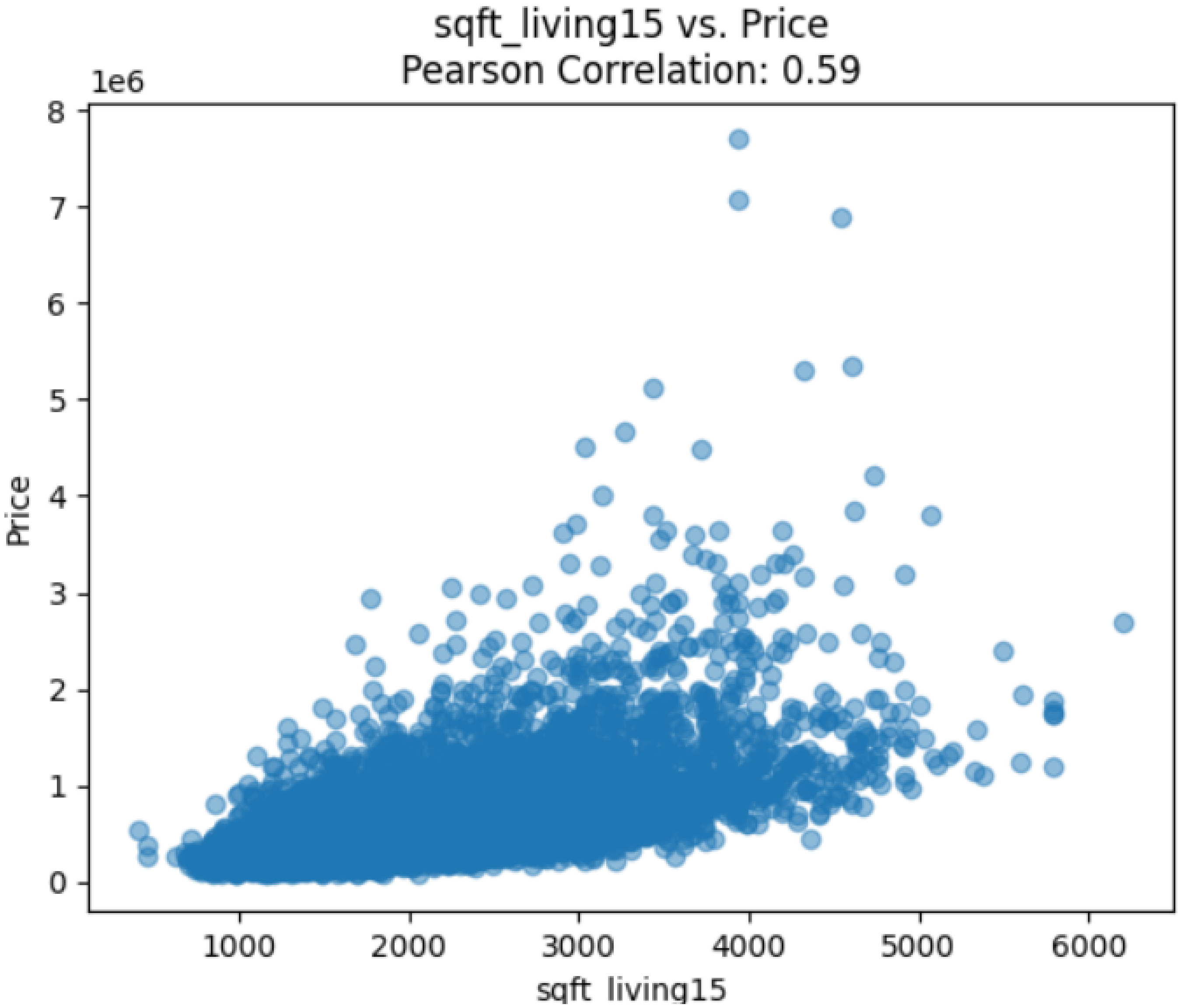

Loss Over Countries for Model Fitted Over Israel
Loss (MSE)
100
80
60
40
20
0
South Africa
The Netherlands
Jordan
Country


Average Loss vs. Training Size
1e10
Average Loss
Confidence Interval
4.45
4.40
4.35
4.30
4.25
4.20
Average Loss
20
40
60
80
100
Percentage of Training Data

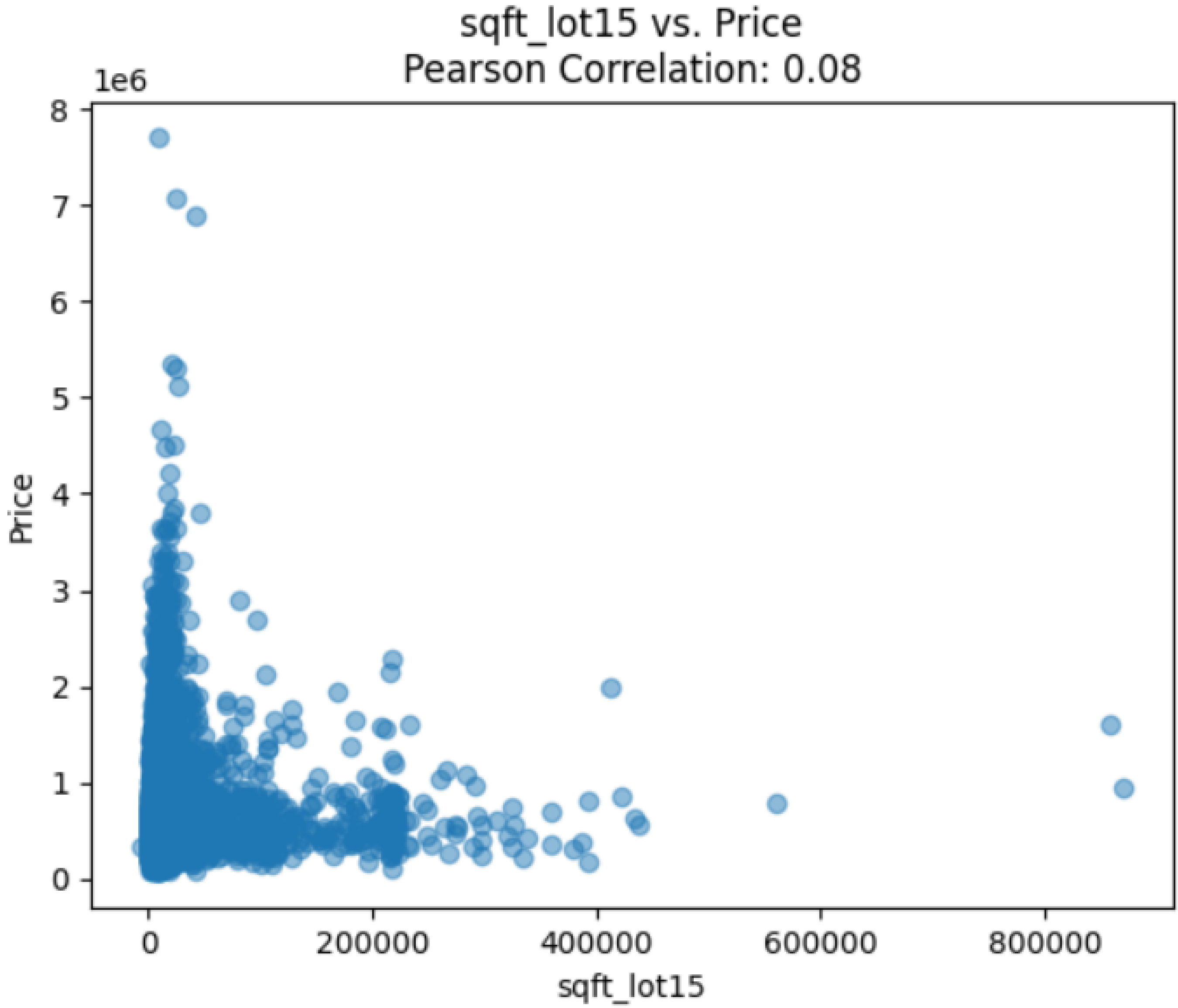

recently_renovated vs. Price
Pearson Correlation: 0.12
1e6
8
7
6
5
4
3
2
1
0
Price
0.0
0.2
0.4
0.6
0.8
1.0
recently_renovated

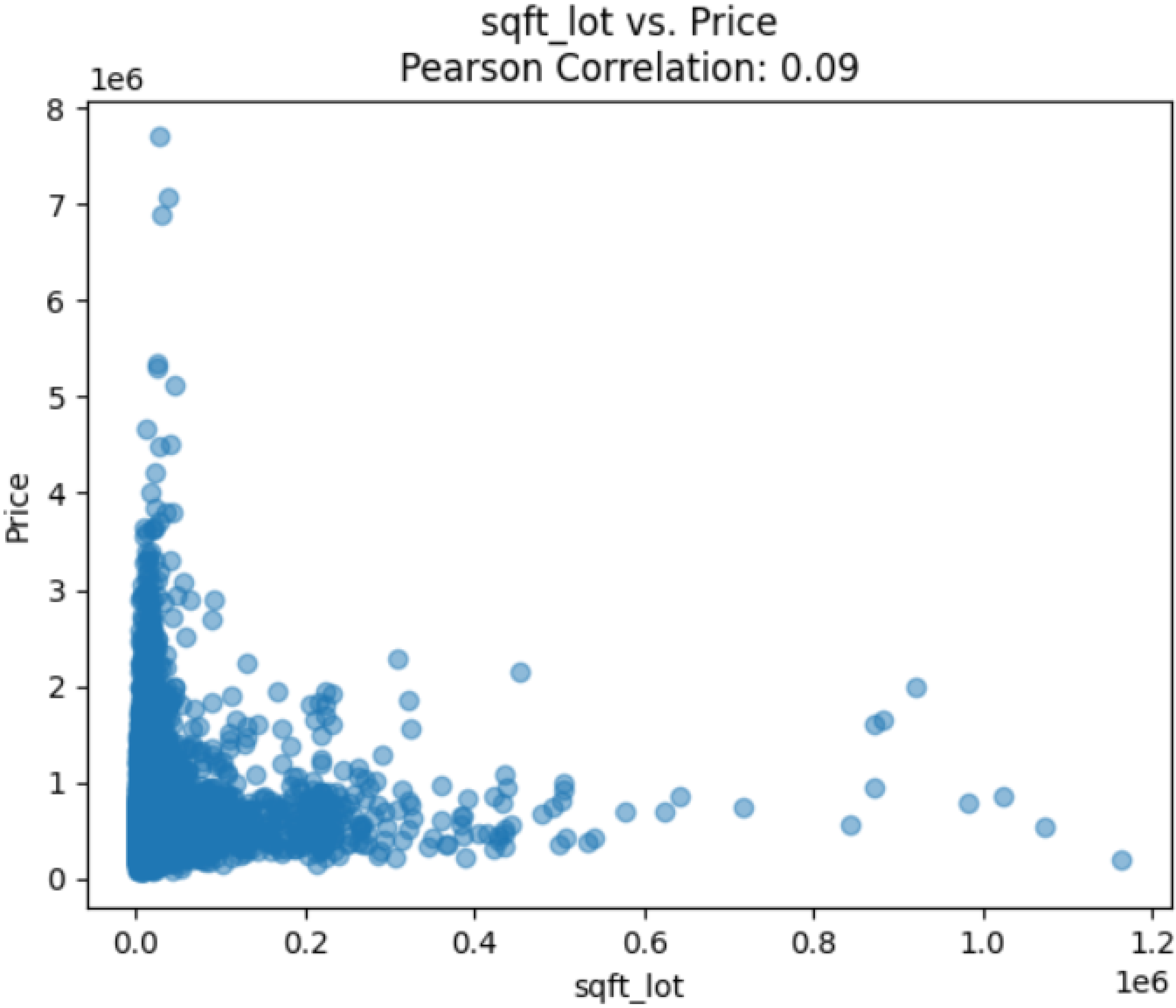

sqft_above vs. Price
Pearson Correlation: 0.60
1e6
8
7
6
5
4
3
2
1
0
Price
0
2000
4000
6000
8000
sqft_above

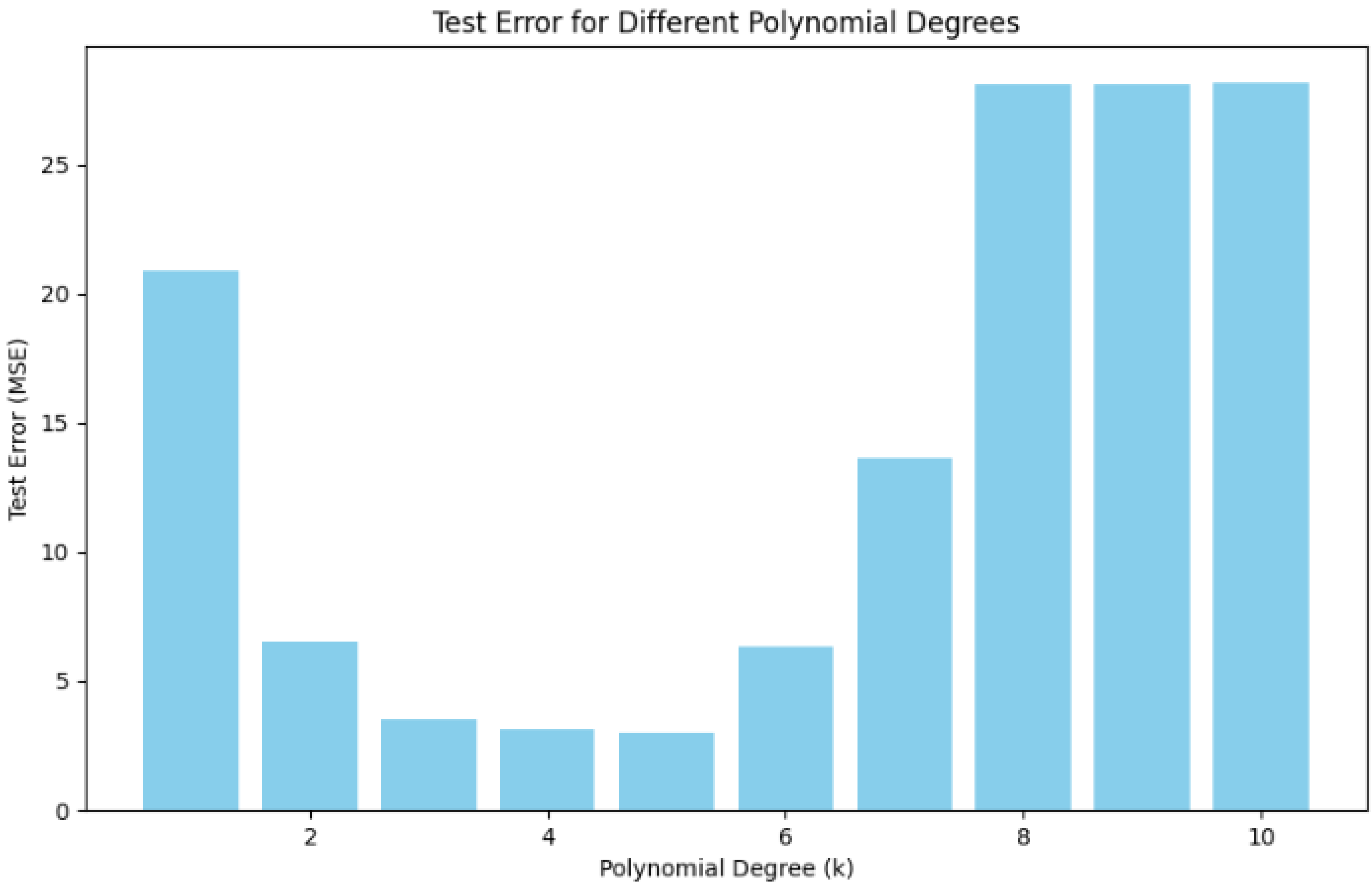

sqft_basement vs. Price
Pearson Correlation: 0.33
1e6
Price
sqft_basement
0
1
2
3
4
5
6
7
8
0
1000
2000
3000
4000
5000

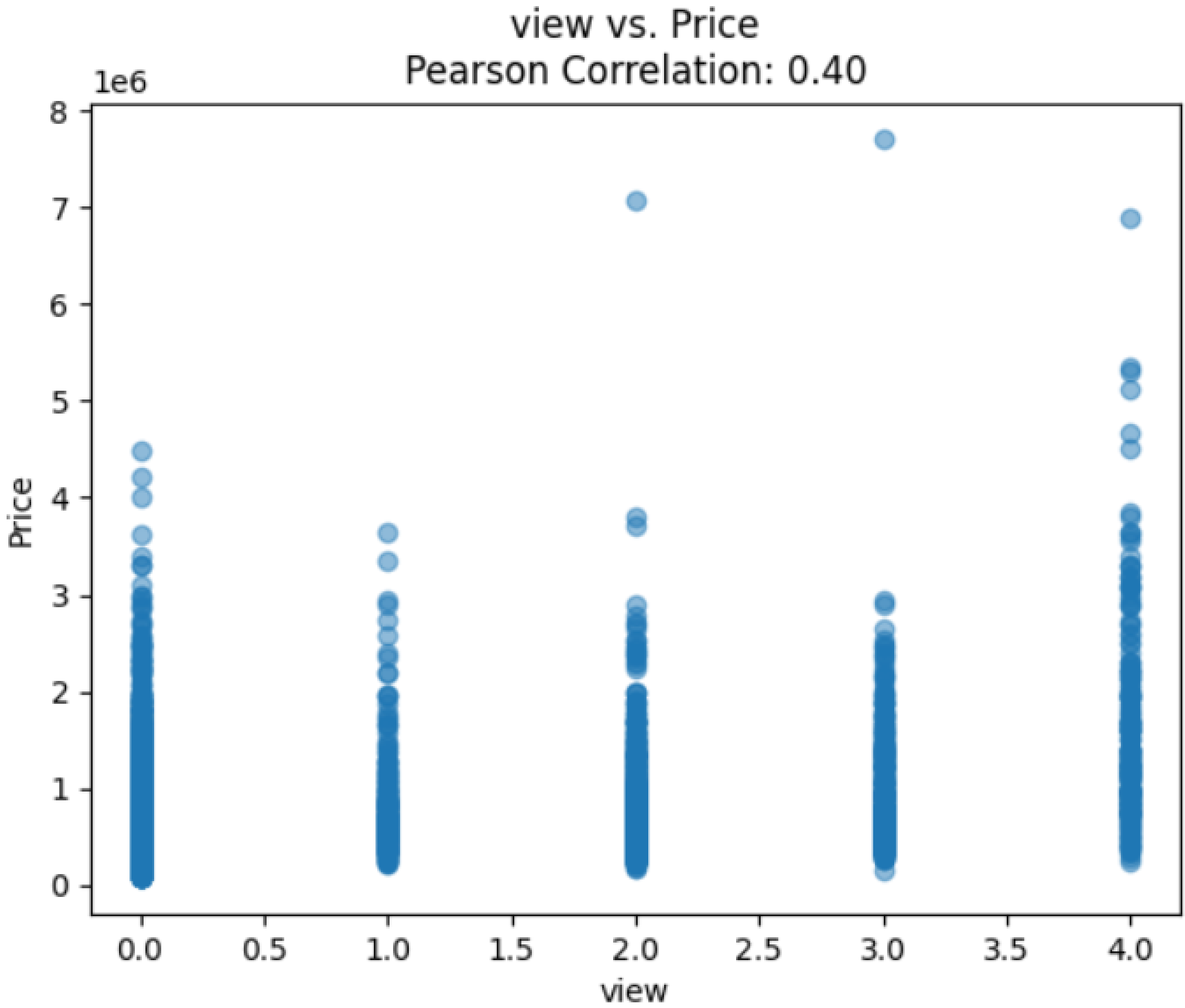

sqft_living15 vs. Price
Pearson Correlation: 0.59
1e6
8
7
6
5
4
3
2
1
0
Price
1000
2000
3000
4000
5000
6000
sqft_living15

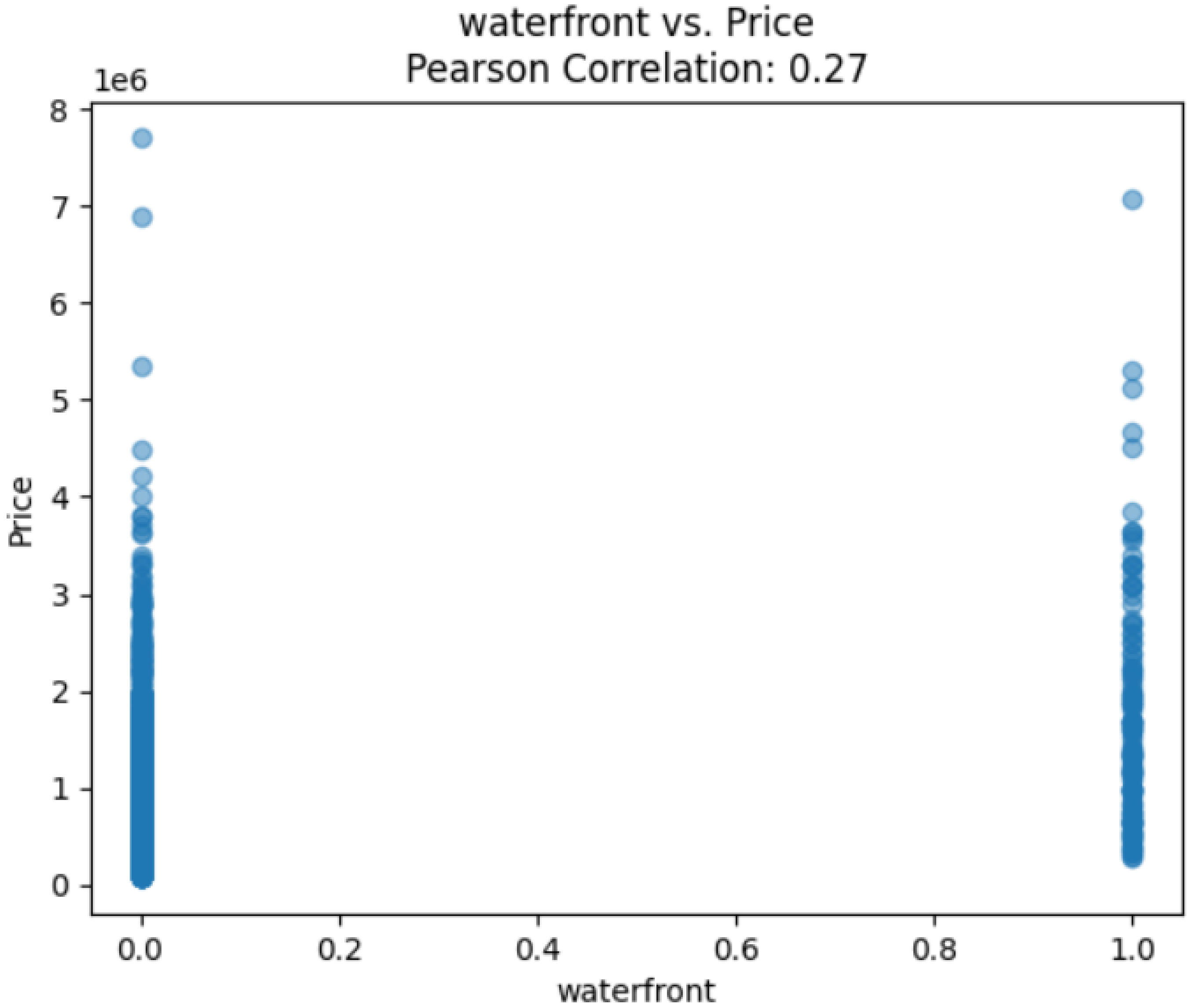

sqft_living vs. Price
Pearson Correlation: 0.70
1e6
8
7
6
5
4
3
2
1
0
Price
0
2000
4000
6000
8000
10000
12000
14000
sqft_living


sqft_lot15 vs. Price
Pearson Correlation: 0.08
1e6
8
7
6
5
4
3
2
1
0
Price
0
200000
400000
600000
800000
sqft_lot15


sqft_lot vs. Price
Pearson Correlation: 0.09
1e6
8
7
6
5
4
3
2
1
0
Price
0.0
0.2
0.4
0.6
0.8
1.0
1.2
sqft_lot
1e6


Test Error for Different Polynomial Degrees
Test Error (MSE)
0
5
10
15
20
25
2
4
6
8
10
Polynomial Degree (k)


view vs. Price
Pearson Correlation: 0.40
1e6
8
7
6
5
4
3
2
1
0
Price
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
view


waterfront vs. Price
Pearson Correlation: 0.27
1e6
8
7
6
5
4
3
2
1
0
Price
0.0
0.2
0.4
0.6
0.8
1.0
waterfront