Exercise 2 IML
Eyal Perets 209541903

1. Prove that: $Ker(\mathbf{X}) = Ker(\mathbf{X}^\top\mathbf{X})$

1.

וקטור v שייך לגרעין של X אזי $X\cdot v=0$.

יהי: v וקטור שונה מאפס השייך לגרעין X, לכן:

$$X\cdot v=0 \Longrightarrow X^TX\cdot v=0 \Longrightarrow v\in ker(X^TX)$$

כלומר אם $v\in ker(X)$ בהכרח גם $v\in ker(X^TX)$

כעת נראה את הכיוון השני: יהי $u\in ker(X^TX)$, לכן:

$$X^TXu=0 \Rightarrow u\cdot X^TX\cdot u=0 \Rightarrow u^T\|X\|^2\cdot u=0$$

$$\Rightarrow \|Xu\|^2=0 \Rightarrow Xu=0$$

כלומר אם $u\in ker(X)$ אז $u\in ker(X^TX)$,

לכן $ker(X)=ker(X^TX)$

2. Prove that for a square matrix A: $Im(A^\top) = Ker(A)^\perp$

2.

יהי: $v\in\mathbb{R}^m$, נניח $b\in Im(A^T)$ כך ש
$A^Tv=b$

נוכיח הכלה דו כיוונית,

$:Im(A^T)\subset ker(A)^\perp$

יהי: $u\in ker(A) \Rightarrow A\cdot u=0 \Rightarrow \langle Au,v\rangle=0$

$$\langle u, A^Tv\rangle=0 \Rightarrow \langle u,b\rangle=0$$

כלומר אם $b\in Im(A^T)$ נקבל $b\in ker(A)^\perp$

$Im(A^T) \supset ker(A)^\perp$:

נניח ש $b \notin Im(A^T)$ אזי $b \in Im(A^T)^\perp$

יהי $x \in Im(A^T)^\perp$ כך ש $\langle b, x\rangle \neq 0$, ומכאן ש x אורתוגונלי:

לכל וקטור b כלשהו ב $Im(A^T)$ אזי מתקיים

$$\|A\cdot c\|^2 = \langle Ac, Ac\rangle = \langle c, A^TAc\rangle = 0$$

לפי הנתון מתקיים, כלומר $\|Ac\|^2 = 0$ ולכן $Ac = 0$, $c \in ker(A)$

וקיבלנו שאם $b \notin Im(A^T)$ אזי $b \notin ker(A)^\perp$, כנדרש, ולכן

$$ker(A)^\perp \subset Im(A^T)$$

ולכן $Im(A^T) = ker(A)^\perp$

3. Let $\mathbf{y} = \mathbf{X}\mathbf{w}$ be a non-homogeneous system of linear equations. Assume that $\mathbf{X}$ is square and not invertible. Show that the system has ∞ solutions $\Leftrightarrow y \perp Ker(\mathbf{X}^\top)$.

3. X לא הפיך, כלומר $det(X)=0$ ולמערכת משוואות

יש אינסוף פתרונות, או שאין פתרון כלל.

למערכת משוואות הזו יש אינסוף פתרונות אמ"מ

$y \in Im(X)$, ולפי סעיף 2 בקודם אמ"מ $y \in ker(X^T)^\perp$

כלומר $y \perp ker(X^T)$.

4. Consider the (normal) linear system $\mathbf{X}^\top\mathbf{X}\mathbf{w} = \mathbf{X}^\top\mathbf{y}$. Using what you have proved above prove that the normal equations can only have a unique solution (if $\mathbf{X}^\top\mathbf{X}$ is invertible) or infinitely many solutions (otherwise).

4.

לכן: הפתרון היחידי הינו כאשר X^TX לא הפיך

אחרת: מערכת אינסוף משוואות אמ״מ $X^+y \perp \ker(X^TX)$

$\Updownarrow$

לפי סעיף 1 $\ker(X^TX)=\ker(X)$ ולכן $X^Ty \perp \ker(X)$.

כי: $v \in \ker(X)$ אזי $\langle v, X^+y\rangle = \langle Xv, y\rangle = 0$

$y \in Im(X)$

ולכן $X^Ty \perp \ker(X)$ וקיבלנו שמשפט אינסוף משוואות אם X^TX

לא הפיך, ופתרון יחיד אם X^TX הפיך.

5. Based on Recitation 1 In this question you will prove some properties of orthogonal projection matrices seen in recitation 1. Let $V \subseteq \mathbb{R}^d$, $dim(V) = k$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be an orthonormal basis of V. Define the orthogonal projection matrix $P = \sum_{i=1}^k \mathbf{v}_i\mathbf{v}_i^\top$ (notice this is an outer product).

Prove the following properties in any order you wish:

(a) Show that P is symmetric.

(b) Prove that the eigenvalues of P are 0 or 1 and that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are the eigenvectors corresponding the eigenvalue 1.

(c) Show that $\forall \mathbf{v} \in V$ $P\mathbf{v} = \mathbf{v}$.

(d) Prove that $P^2 = P$.

5. a.

(a) Show that P is symmetric.

$V \subseteq R^d$

$B = (v_1, \ldots, v_k)$

בסיס אורתונורמלי של V

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

והיות והמכפלה החיצונית הינה פעולה סימטרית

כלומר $v \otimes u = u \otimes v$ כאשר v ו- u בעלי אותו המימד,

ומכאן P הינה סימטרית

(b) Prove that the eigenvalues of P are 0 or 1 and that $\mathbf{v}_1,\dots,\mathbf{v}_k$ are the eigenvectors corresponding the eigenvalue 1.

ב.

יהי: v_j ו"ע שבשאלתנו ע"ע λ_j שלו

$$\lambda_j v_j = P\cdot v_j = \sum_{i=1}^{k} v_i v_i^T \cdot v_j \overset{\circledast}{=} \sum_{i=1}^{k} v_i \cdot a_{ij} = v_j$$

כמו שאמרנו הנ"ל $\circledast$

$v_i^T \cdot v_j = a_{ij}$ $\circledast$

רק הרכיב ה- $i=j$ שווה

ולכן $\lambda = 1$ או $\lambda = 0$

וכמו כן ו"ע המתאימים הם הבסיס $v_1,\dots,v_k$

(c) Show that $\forall \mathbf{v} \in V \ P\mathbf{v} = \mathbf{v}$.

ג.

אמרנו כי נבחר ו"ע שיוך מהבסיס הקודם, נניח שהם $v_1,\dots,v_k$

ו"ע נניח אינם 0 אלא הם מתאימים לע"ע 1,

ולכן $Pv = \lambda v = 1\cdot v = v$.

(d) Prove that $P^2 = P$.

ד.

P מטריצה סימטרית ולכן ניתן ללכסן כך:

$P = UDU^T$ כאשר $UU^T = I$

$$P^2 = UDU^TUDU^T = UD^2U^T = UDU^T = P$$

וגם: D היא מטריצה אלכסונית שבאלכסונה ע"ע של P -

כל אחד 0 או 1, ולכן $D^2 = D$

(e) Prove that $(I-P)P=0$.

ה.

$$(I-P)\cdot P = P - P^2 \overset{\text{מהסעיף הקודם}}{=} 0$$

6. Show that if $\mathbf{X}^\top\mathbf{X}$ is invertible, the general solution we derived in recitation equals to the solution you have seen in class. For this part, assume that $\mathbf{X}^\top\mathbf{X}$ is invertible.

.6

מתרגול ראינו שהפתרון כאשר X^TX הפיך הינו $\hat{w}=X^+y$

בכיתה ראינו שהפתרון הוא $w=[X^TX]^{-1}X^Ty$

אז צריך להראות כי $X^+y=[X^TX]^{-1}X^Ty$

יהי $X=U\Sigma V^T$ ה-SVD של X כמשמעו כפשוטו:

אזי:

$$[X^TX]^{-1}X^Ty=\left[(U\Sigma V^T)^T\cdot U\Sigma V^T\right]^{-1}(U\Sigma V^T)^T\cdot y$$

$$=\left[V\Sigma^T \underbrace{U^TU}_{I}\Sigma V^T\right]^{-1}\cdot V\Sigma^TU^Ty=\left[V\Sigma\Sigma V^T\right]^{-1}V\Sigma^TU^Ty$$

כי V אורתוגונלית: $V^{-1}=V^T$

$$\overset{\Downarrow}{=} V(\Sigma^T\Sigma)^{-1}\underbrace{V^TV}_{=I}\Sigma^TU^Ty = V(\Sigma^T\Sigma)^{-1}\Sigma^TU^Ty = V(\Sigma^2)^{-1}\Sigma^TU^Ty$$

כלומר Σ^2 הינה מטריצה שבאלכסונה הערכים הסינגולריים בריבוע, אזי $(\Sigma^2)^{-1}$ הינה מטריצה אלכסונית שבאלכסונה: $\frac{1}{\sigma^2}$, ו- $(\Sigma^2)^{-1}\Sigma$ הינה מטריצה אלכסונית, נקרא לה Σ^+ שבאלכסונה: $\frac{1}{\sigma^2}\cdot\sigma=\frac{1}{\sigma}$

ולכן

$$[X^TX]^{-1}X^T=V\Sigma^+U^T=X^+$$

וקיבלנו $[X^TX]^{-1}X^Ty=X^+y$ והתוצאות הינן שוות.

7. Show that $\mathbf{X}^\top\mathbf{X}$ is invertible if and only if $\operatorname{span}\{\mathbf{x}_1,\ldots,\mathbf{x}_m\} = \mathbb{R}^d$.

.7

X^TX מטריצה בגודל $d\times d$ והפיכה אמ"מ

דרגתה היא d, כלומר בעלת פתרון יחיד.

דרגת X היא כדרגת המרחב הנפרש ע"י $\{x_1,\dots,x_m\}$

לפי שאלה 1: דרגת X ו X^TX זהה, ולכן

X^TX הפיכה אמ"מ דרגתה היא d, כלומר נפרש ע"י $\operatorname{span}\{x_1,\dots,x_m\}=\mathbb{R}^d$

8. Recall that if $\mathbf{X}^\top\mathbf{X}$ is not invertible then there are many solutions. Show that $\hat{\mathbf{w}} = \mathbf{X}^\dagger\mathbf{y}$ is the solution whose L_2 norm is minimal. That is, show that for any other solution $\overline{\mathbf{w}}$, $||\hat{\mathbf{w}}|| \leq ||\overline{\mathbf{w}}||$.

.8

$\hat{w}=X^+y$ ה-SVD של X היא $X=U\Sigma V^T$

$\hat{w} = V\Sigma^+U^Ty$

אנו צריכים להוכיח שהפתרון לפי נורמה 2 הינו מינימלי.

אנו מנסים למזער את המינימום של $\|Xw-y\|_2$

$$\|Xw-y\| = \|U\Sigma V^Tw-y\| \overset{*}{=} \|\Sigma V^Tw-U^Ty\|$$

$*$ כפלנו ב U^T, והרי $UU^T=I$ נכון כי U אורתוגונלית

נגדיר $Z=V^Tw$, והרי $\|w\|=\|V^Tw\|=\|Z\|$ כי V אורתוגונלית.

ולכן אנו צריכים למזער את המינימום של הנורמה הבאה: $\|\Sigma Z-U^Ty\|$

מכיוון ש Σ תלויה למטריצה האלכסונית, המינימום בהכרח מתקבל עבור (הנורמה תלויה רק בשאר ערכים) $Z=\Sigma^+U^Ty$!

כעת הפתרון היחידי יתקבל עבור $w^+ = V\cdot z^+ = V\Sigma^+ U^T y$

וזה בדיוק הפתרון w^+ הרצוי.

חלק הפרקטי

2.

1. אילו פיצ'רים להשאיר ואילו למחוק?

מחקתי פיצ'רים שרואים שאין בהם קשר למחיר, כמו האיקוד, קווי אורך ורוחב, והשארתי פיצ'רים שנראה כי יכול להיות ביניהם לבין המחיר קשר.

2. אילו פיצ'רים הם קטגוריאליים?

הפיצ'רים הקטגוריאליים הם ה Lat ו Long אשר חשבתי תחילה לבדוק אילו אזורים היקרים יותר, ואולי לאנדקס ולסמנם במספרים, כאשר 1 הוא האיזור הזול יותר ו 5 הוא היקר. אך הסתבר לאחר בדיקה (גרפית) כי אין כל קשר בין האזורים ב dataset הנבחר.

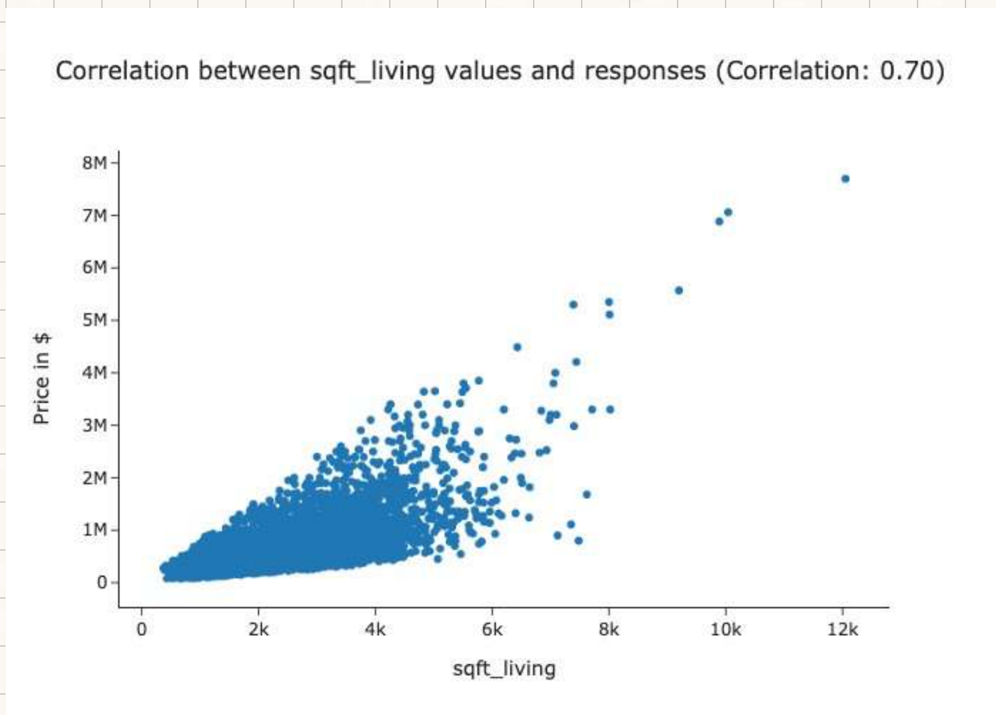
כמו כן שנת השיפוץ היא גם פיצ'ר קטגוריאלי, ועבורו התייחסתי לבתים שעברו שיפוץ ונתתי להם ערך 1, ולבתים שלא עברו שיפוץ ונתתי להם ערך 0.

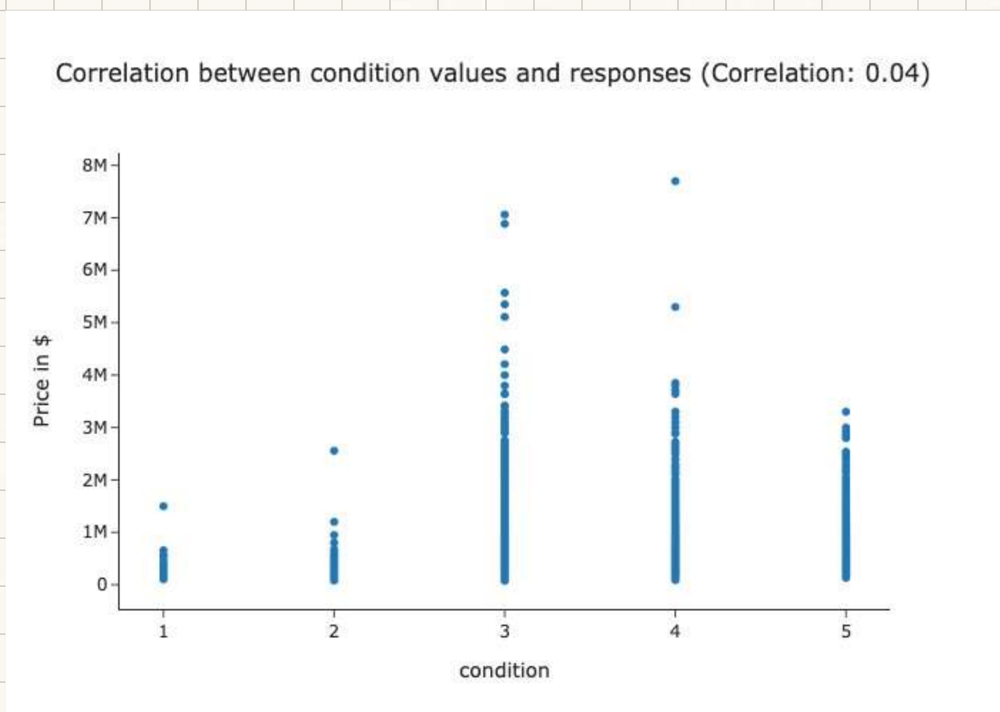
3. יצרתי פיצ'ר חדש בשם decade ושמרתי בו את שנת הבנייה והמצמצם אותה לפי מנת שנים יותר ונתתי לה ערך וכך יכולתי לבנות Linear regression טובה יותר.

4. מחקתי אינדקסים שגויים - מחיר שלילי, חדרים וכו'

5. ראינו שב dataset אין מספיק מידע עבור מיקוד בקוד יותר
מ- 13 בתים ולכן הורדתי אותו מה dataset א.
כמו כן עשיתי פעולה דומה עבור sqft_lot שגדול מ- 120,000

3.

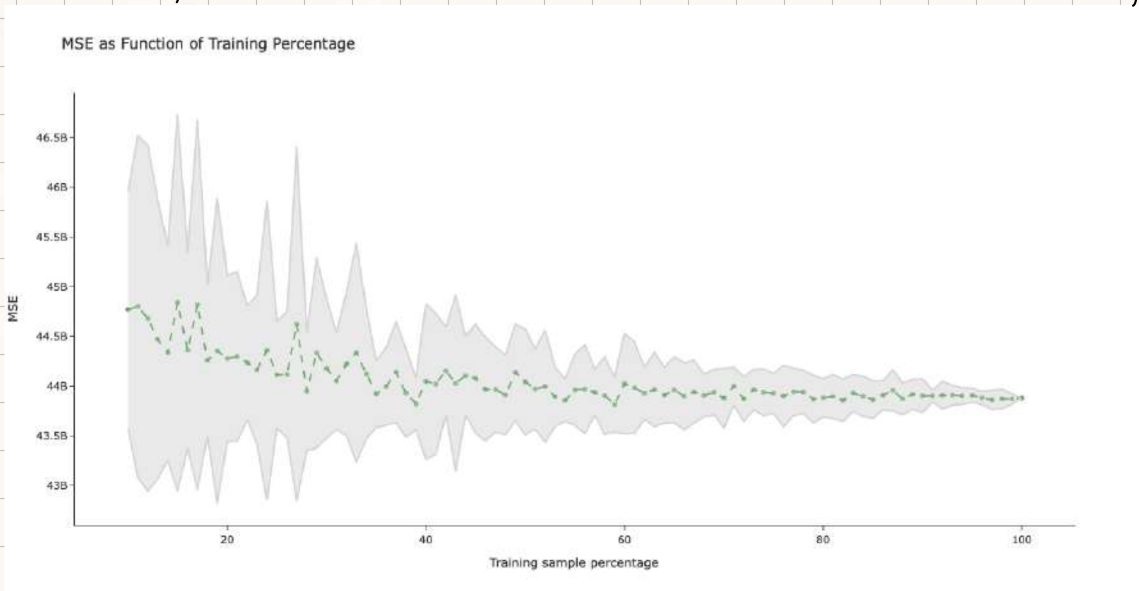




אך מספיק ראות הגרפים הללו קשורים כי לא לפי הקורלציה לבחור,
ניתן לראות שב sqft_living ישנה קורלציה גבוהה ועל כן ניתן להסיק כי הוא

משפיע עיקרי את המחיר, ואילו עבור condition ההשפעה מאוד נמוכה.

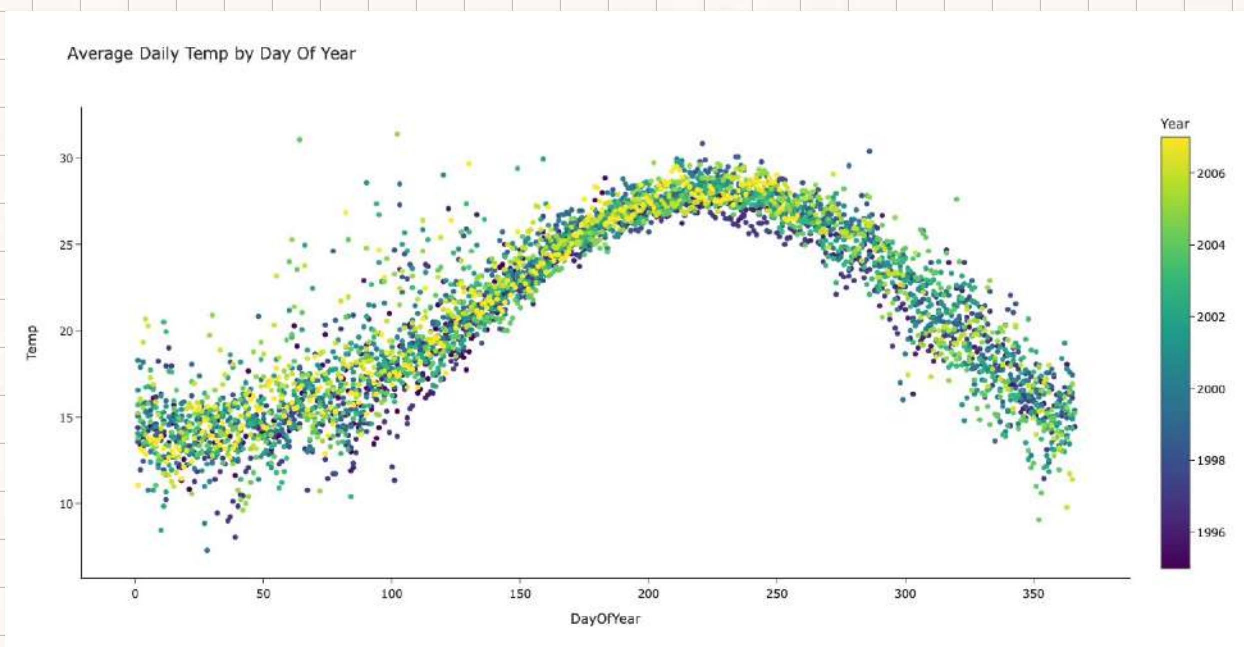
4.



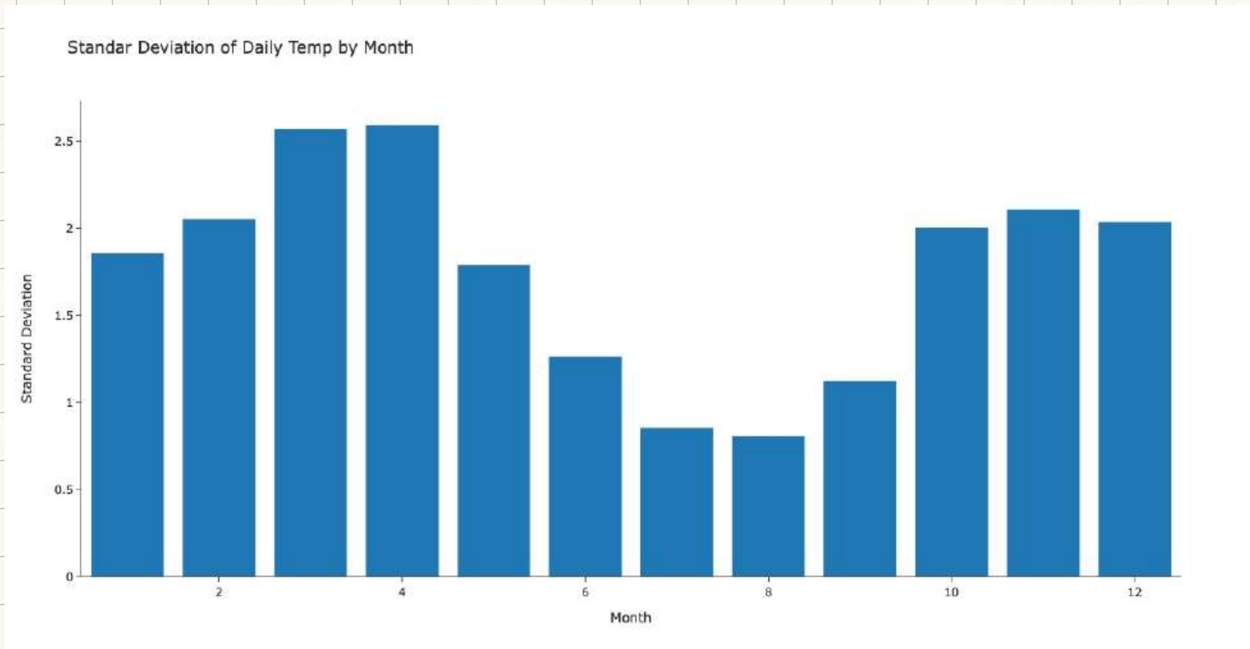
ניתן לראות מן הגרף שככל שאחוז הדגימות גדל יותר, כך ה confidence interval יורד, כמו גם השגיאה שלנו
שהשונות קטנה.

Polynomial Fitting

2.

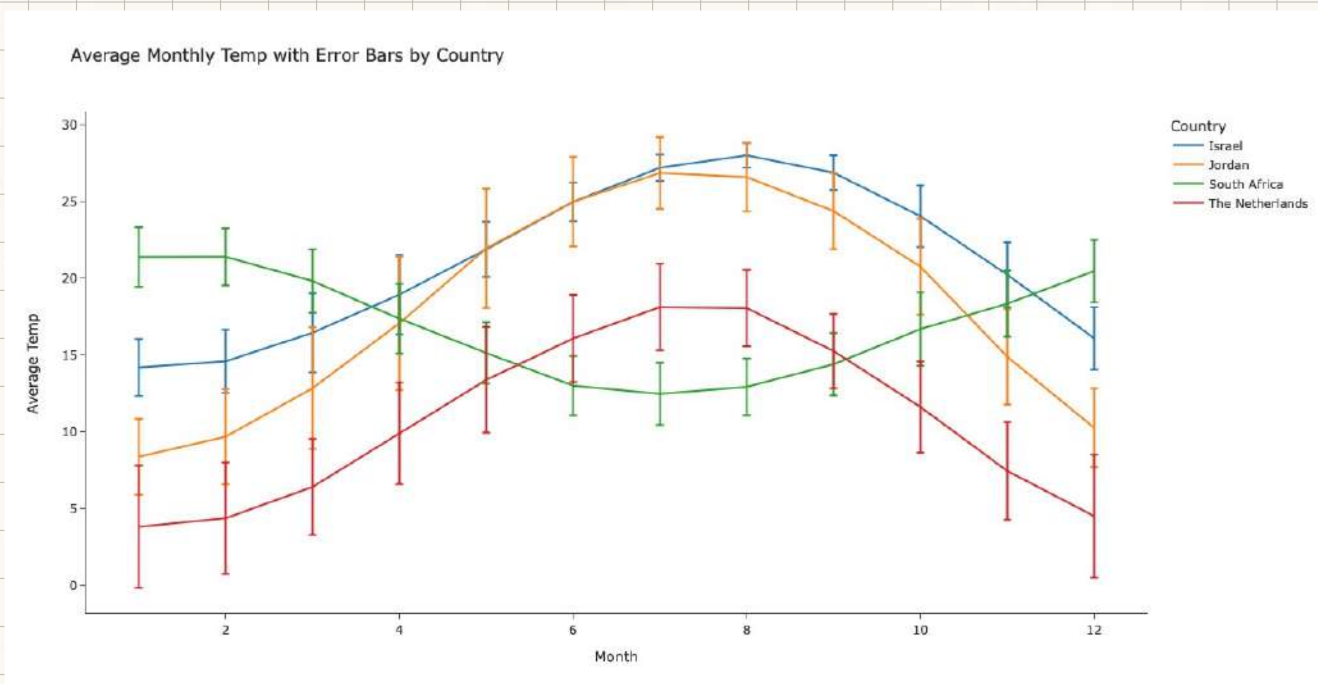


ניתן לראות כי הגרף אינו דומה לפולינום מעלה 2 שכן באמצע-סוף הגרף נראה איטי, ולכן הייתי אומר שמעלה הקטנה ביותר המתאימה היא 3.



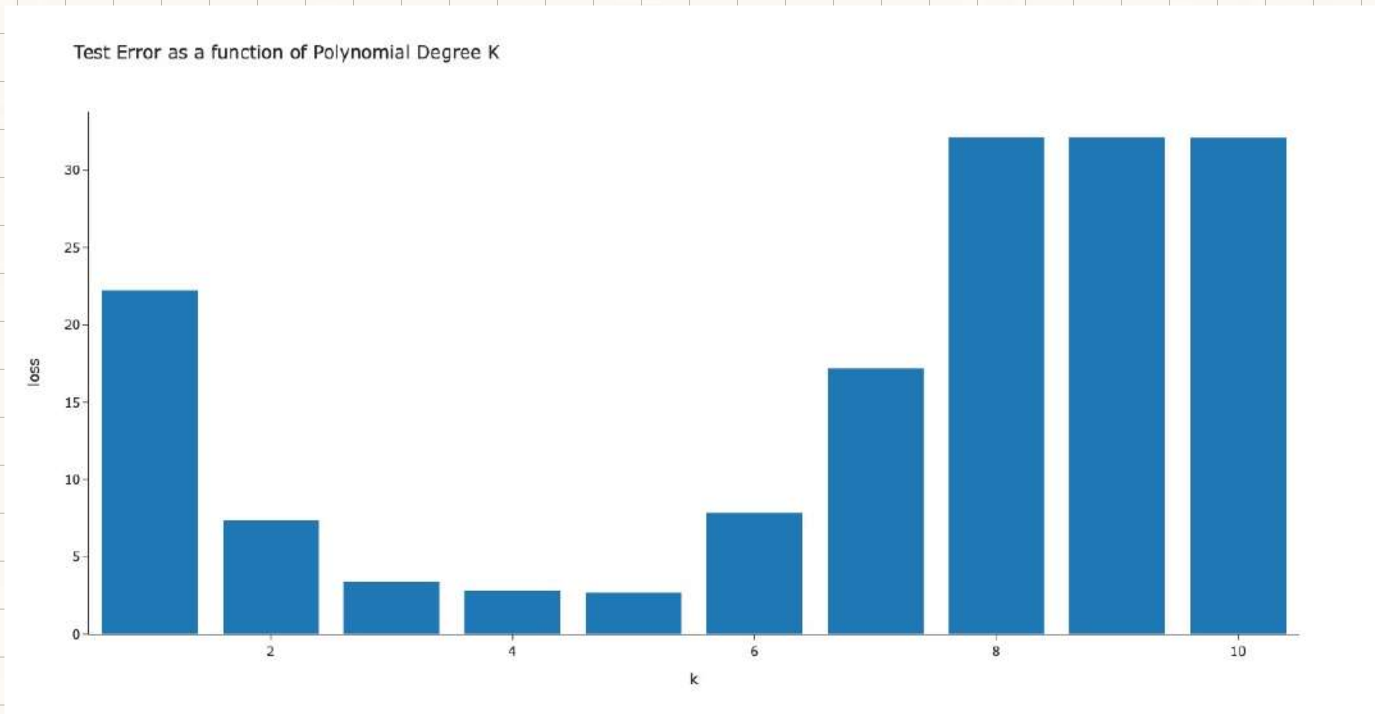
לפי הגרף אני מצפה שהמזג יציב יותר בחודשי יוני עד ספטמבר,
ומשאר החודשים פחות, שכן השונות גדולה יותר.

3.



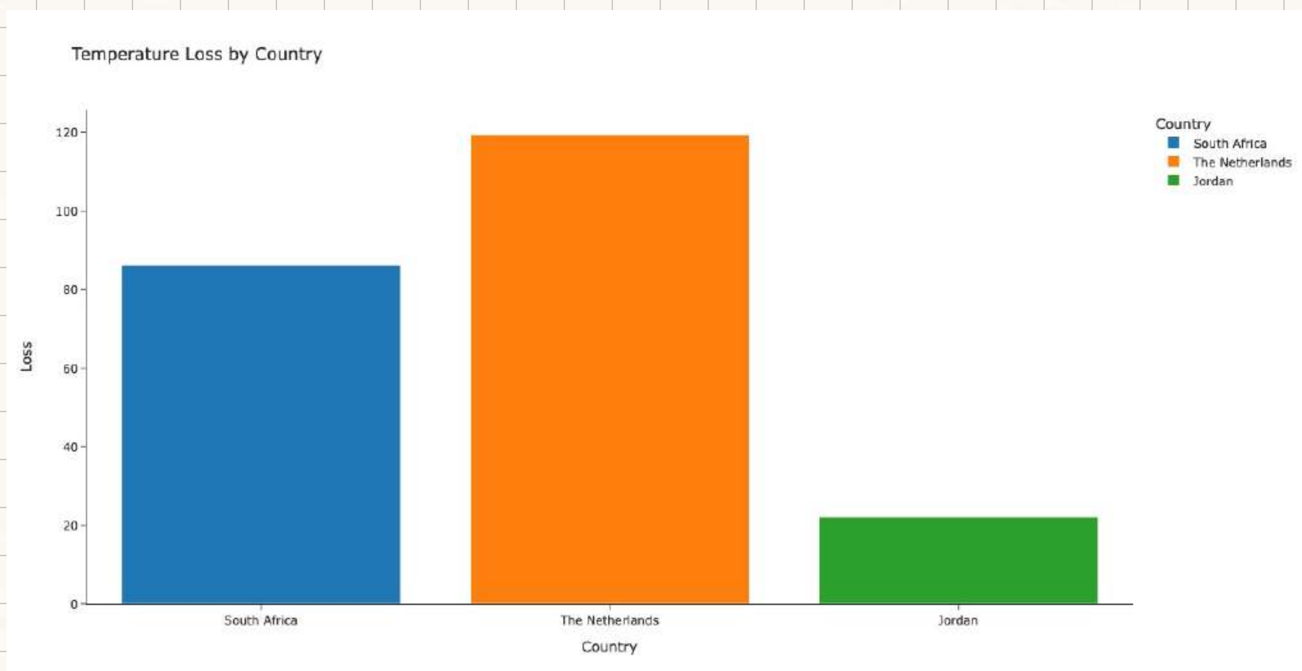
על סמך הגרף ניתן לראות כי לירדן המזג יציב טוב, בשונה להולנד
לדרום אפריקה פחות, וזאת מפני שירדן וישראל בעלי מזג קרובים,
ואילו דרום אפריקה ומצויה בצד הדרומי של כדור הארץ ולכן
העונות שונות בה הפוכות.

4.



ערך ה-k שעבורו המודל ישקף את הדאטה הכי טוב הוא $k=5$,
וערכי $k=3,4$ מקבלים loss נמוך כמעט באותה המידה ולכן ניתן לבחור
גם אותם.

5.



התוצאות הגיוניות הן כמו שציפיתי בסעיף 3, שעבור ירדן נקבל מודל כמעט
זהה לישראל ולכן ה-loss יהיה נמוך, ואילו עבור דרום אפריקה (הולנד
ו-loss יהיה גבוה בהרבה (ציפיתי שעבור דרום אפריקה יהיה loss
גבוה יותר משל הולנד)