**IML – ex2 יאיר שטרן ת.ז. 318442241**

**1.**

הפיצ'רים שחשבתי שאין להם השפעה על מחיר הדירה והורדתי:

id, lat, long, date

כמו כן, סיננתי ערכים לא הגיוניים (שקטנים או שווים לאפס) עבור:

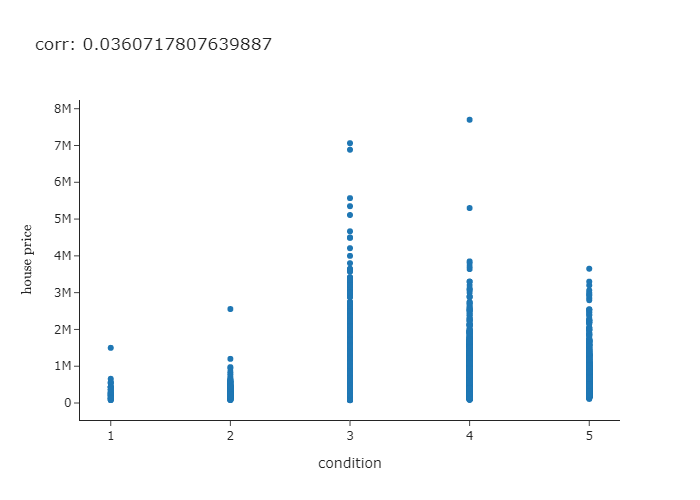
price, sqrt\_lot, sqrt\_living, bedrooms, floors

כמו כן הסרתי ערכי Nan, וחזרות.

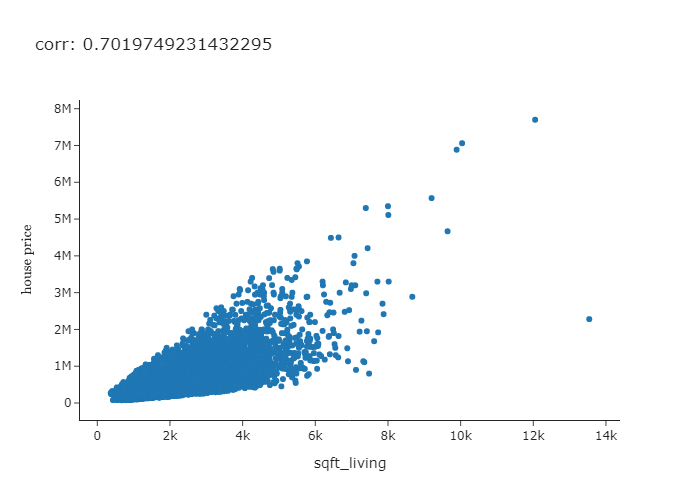
הוספתי עמודה חדשה "bedrooms\_for\_floor" המכילה את מספר החדרים בכל קומה, שחשבתי שהוא יכול להוסיף מידע על שווי הבית (שהרי בית עם הרבה קומות עם מעט חדרים בכל קומה לא בהכרח שווה יותר מבית עם פחות קומות והרבה חדרים בכל קומה).

הפיצ'ר "zipcode" הוא ערך קטגוריאלי ולכן יצרתי dummies שמחלק את הערכים לעמודות המכילות 0/1 אם הבית נמצא בזיפ קוד הספציפי.

**2.**

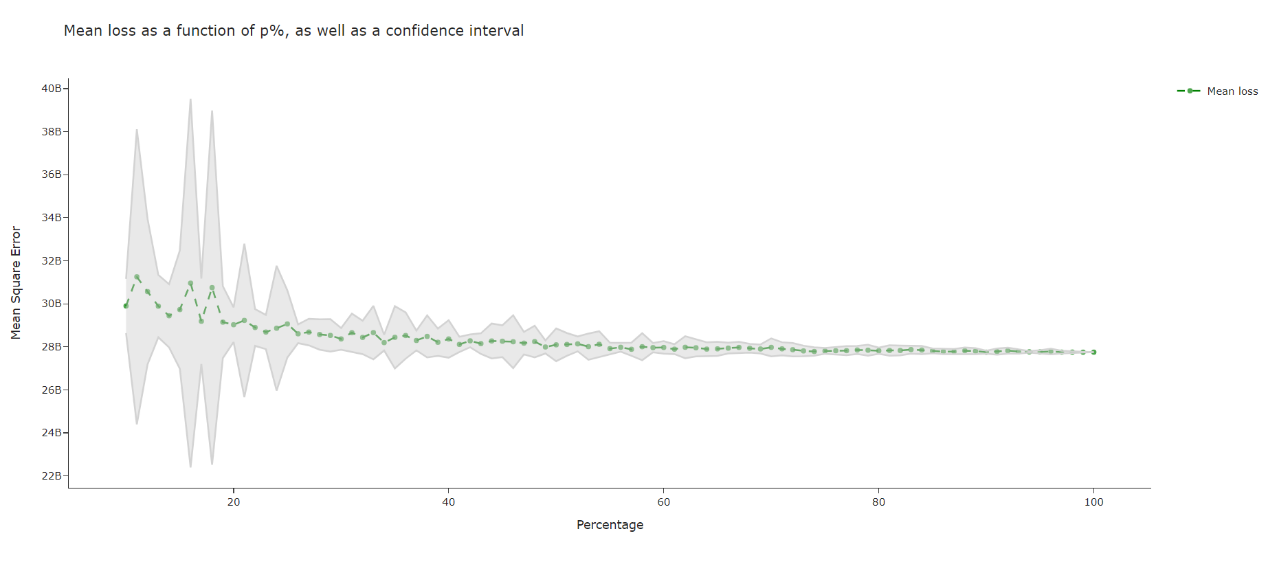
הקורולציה של הפיצ'ר conditions נמוכה וניתן לראות בגרף שאין קשר בין מצב הדירה למחיר. 

לעומת זאת בפיצ'ר sqrt\_living שהקורולציה שלו גבוהה ניתן לראות בגרף שיש קשר בין הפיצ'ר למחיר הבית.



**4.**

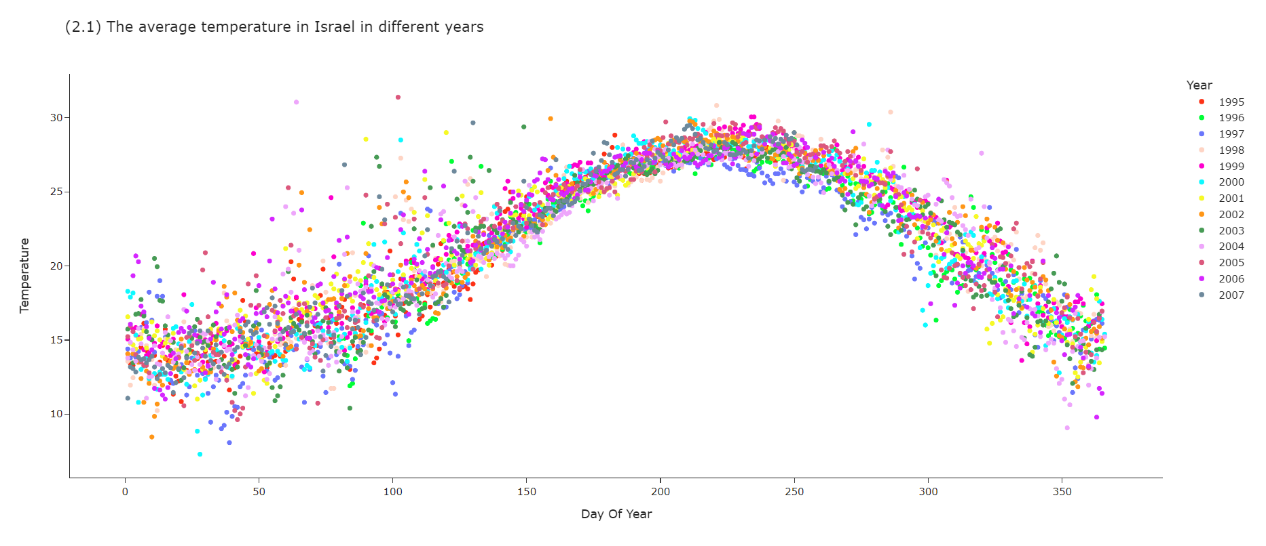
הקו הירוק מתאר את ה loss הממוצע, השטח האפור מתאר את הconfidence interval . ניתן לראות שככל שאנחנו לוקחים יותר אחוזים מהדגימות ה loss הממוצע מתכנס לערך כל שהוא כאשר גם ה confidence intervalמתכנס אל אותו הערך.



חלק 2:

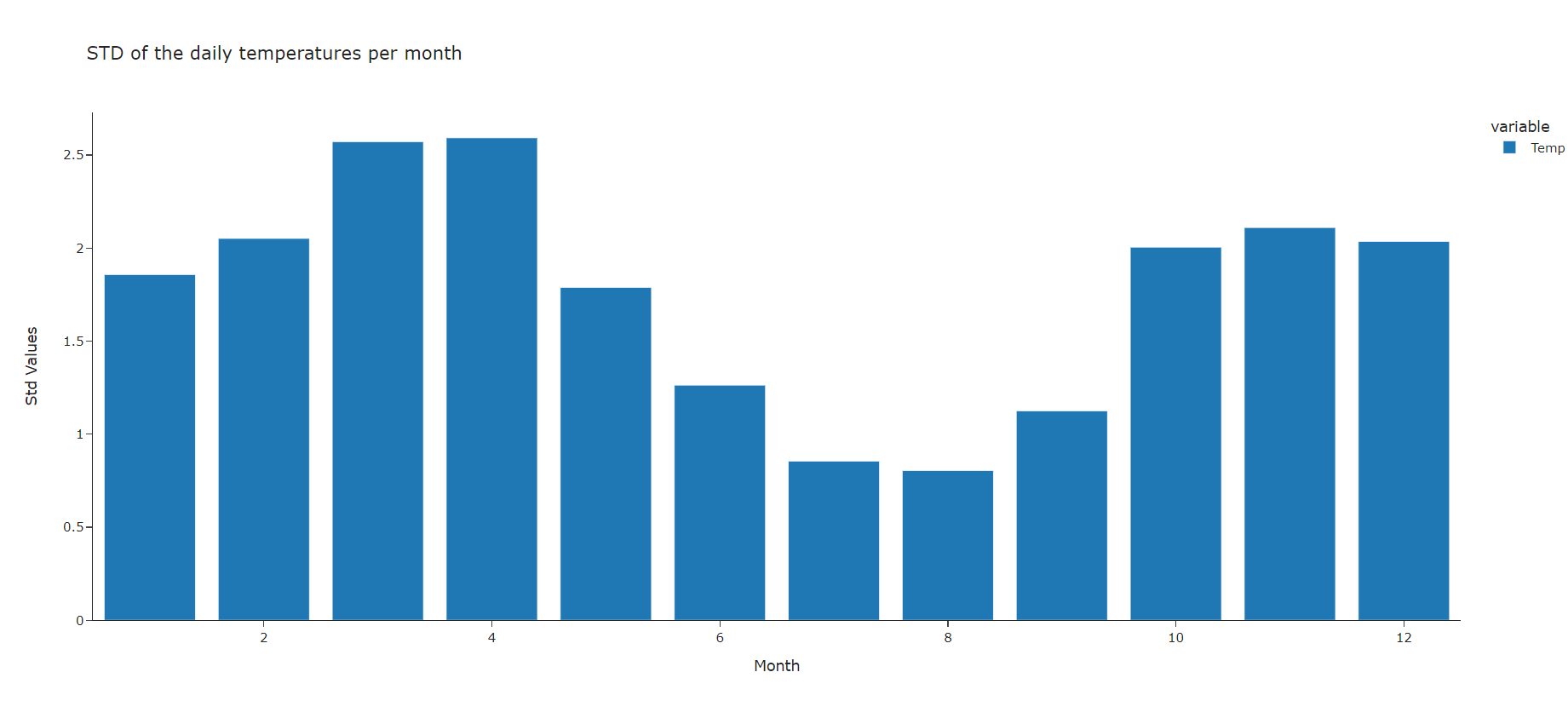
**1.2**

כפי שנראה הגרף מתאים לפולינום מדרגה הגבוהה מ-2



**2.2**

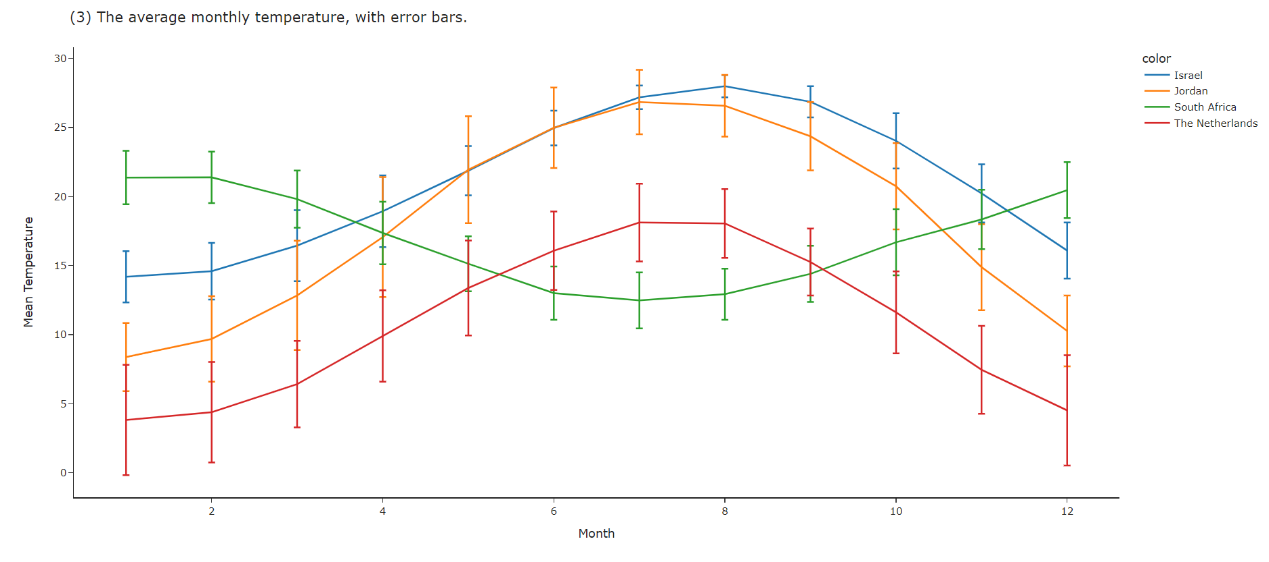
ניתן לראות שסטיית התקן בחודשים 7,8 היא נמוכה יותר (ניתן לשער שזה כיוון שבחודשים האלו התחזית יחסית קבועה בניגוד לשאר החודשים ולתקופות מעבר.) לכן ניתן לשער שהמודל יחזה טוב יותר בחודשים אלו.



**3.**

ראשית, ניתן לראות מהגרף שהדפוס של ירדן דומה לשל ישראל, כמו כן הדפוס של הולנד דומה לשל ישראל אך נמוך יותר (קר שם יותר...). לעומת זאת הדפוס של דרום אפריקה ממש שונה משניהם כאשר בחודשים הקרים בישראל חם באפריקה ולהפך.

ניתן להסיק שהמודל של ישראל יכול לעבוד טוב (יחסית) עבור ירדן אך לא עבור שאר המדינות.

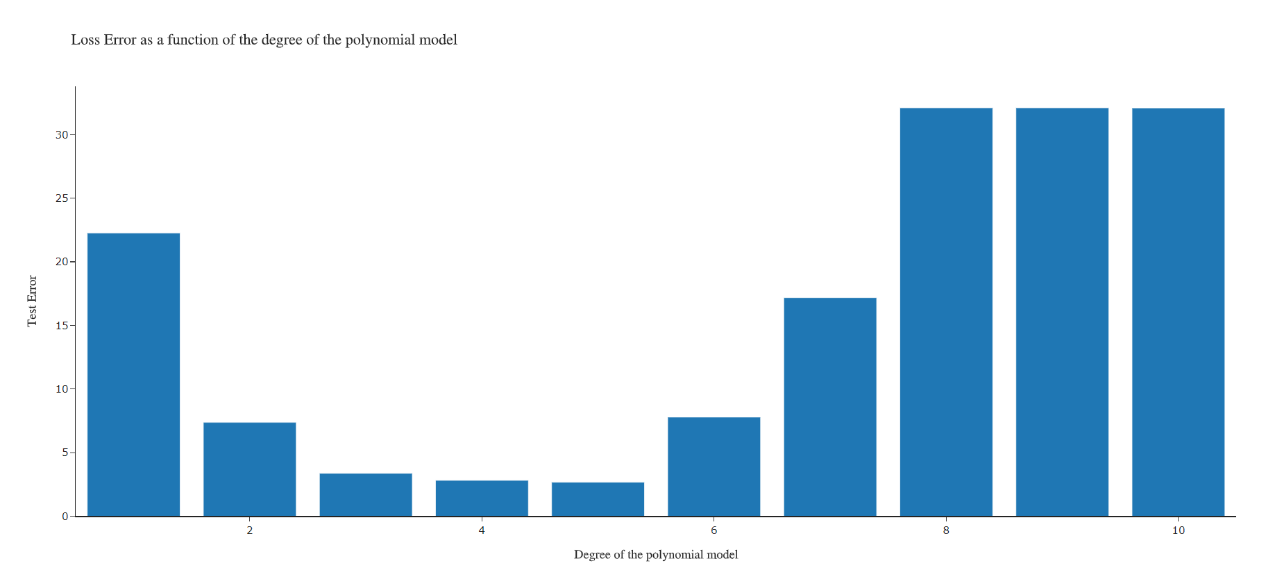


**.4**

**תמונה שמכילה טקסט, אלקטרוניקה, לוח מקשים

התיאור נוצר באופן אוטומטי**

ניתן לראות שהמעלה המביאה את ה loss הנמוך ביותר היא 5 לכן מודל ממעלה 5 הכי טוב. (3,4 נותנים גם כן loss נמוך, לכן הם גם אפשריים).



**.5**

כמו שראינו בגרף בשאלה 3, ניתן לראות שחיזוי לפי ישראל עבור ירדן מביא לשגיאה נמוכה ביחס לשאר המדינות.

נשים לב, שהשגיאה עבור הולנד גדולה יותר מהשגיאה עבור דרום אפריקה (למרות שראינו שהדפוס דומה) כי אמנם הדפוס דומה לשל ישראל אך ברוב השנה הערכים של דרום אפריקה קרובים יותר לשל ישראל מאשר ערכי הולנד.

