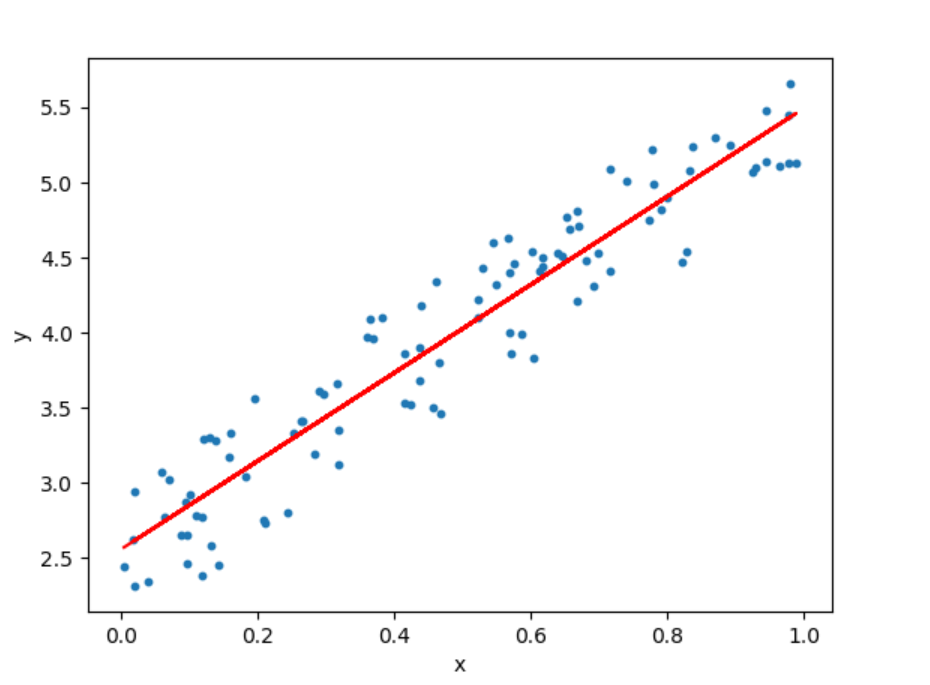
***פרויקט ברגרסיה***

**משתתפים:**

**בר גרינבלט 207333287  
איתן האלי 314090838**

****

**תוכן עניינים:**

1. **טבלת משתנים..............................................................................3**
2. **תיאור המשתנים............................................................................4**
3. **תיאור קשרים בין המשתנים............................................................5**
4. **ניתוח תיאורי של המשתנים............................................................5-7**
5. **ניתוח חריגים................................................................................8-10**
6. **פונקציית צפיפות ופונקציית התפלגות..............................................11-13**
7. **ייצוג קשרים בעזרת תרשימים.........................................................14-16**
8. **טבלת שכיחויות.............................................................................16-17**

**נספחים...............................................................................................18-27**

1. **מאגר נתונים**: המאגר נתונים שאנו שובצנו אליו מדבר על מחירי דירות (המשתנה המוסבר) בהתייחס לגורמים שונים שיכולים להסביר את מחיר הדירה.
2. **המשתנים במודל:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| סוג המשתנה -מוסבר/מסביר | סימון במודל | יחידת מידה | סוג המשתנה – רציף / קטגוריאלי | הסבר קצר על המשתנה |
| משתנה מוסבר  **מחיר** | Y | שקל חדש | רציף | מחיר הנכס בשקלים חדשים |
| משתנה מסביר **ID** | X1 | חסר יחידות | קטגוריאלי | מזהה ייחודי של הנכס |
| משתנה מסביר **מספר חדרי שינה** | X2 | חסר יחידות | קטגוריאלי | מספר חדרי שינה בנכס, המשתנה הוא קטגוריאלי כיוון שהטווח של המשתנה מצומצם בין ערכים תחומים. |
| משתנה מסביר **גדול הבית** | X3 | מ"ר | רציף | גודל השטח הבנוי של הנכס |
| משתנה מסביר **גודל המגרש** | X4 | מ"ר | רציף | גודל השטח הכולל של הנכס |
| משתנה מסביר **גודל המרתף** | X5 | מ"ר | רציף | גודל המרתף בנכס,0 אם אין |
| משתנה מסביר **מספר קומות בנכס** | X6 | חסר יחידות | קטגוריאלי | מציין את מספר הקומות שהנכס מכיל (מתחיל מ1) |
| משתנה מסביר **מצב הנכס** | X7 | [1-5] | קטגוריאלי | מצב הנכס 1 – לא טוב 5- טוב מאוד. המשתנה הוא קטגוריאלי כיוון שהטווח של המשתנה מצומצם בין ערכים תחומים. |
| משתנה מסביר **השנה בה הנכס נבנה** | X8 | שנים | רציף | מציין את השנה שבה נבנה הנכס, המשתנה הוא רציף כיוון שיש מספר רב של תצפיות והטווח של השנים יכול להכיל מספר רב של ערכים |
| משתנה מסביר **האם הנכס שופץ** | X9 | [0-1] | קטגוריאלי | האם הנכס עבר שיפוץ 0 אם לא 1 אם כן |

1. **תיאור המשתנים**

|  |  |  |
| --- | --- | --- |
| **המשתנה** | **סוג** | **מדוע מוסיף מידע לגבי מגמות השינוי במשתנה המוסבר** |
| מחיר | משתנה מוסבר | נרצה להסביר כיצד המחיר של נכס מושפע מכל אחד מהמשתנים המסבירים שמוצגים בהמשך |
| ID | משתנה מסביר | משתנה זה מאפשר בצורה הטובה ביותר לזהות את הדירה כיוון שמרכיב זה מציין את כל שאר המשתנים וכל ערך בו ייחודי |
| מספר חדרי שינה | משתנה מסביר | מסביר ומתאר את מספר החדרים בנכס וכך את החלוקה הפנימית של הנכס ואת יכולת האכלוס שלו לפי החלוקה. משום כך נכס עם מספר חדרי שינה גדול מצביע על יכולת אכלוס הנכס לפי מספר החדרים ומשמש כמדד ישיר שמציע על מגמות השינוי של המחיר |
| גדול הבית | משתנה מסביר | מצביע על הכמות של משאבים הושקעו על מנת לבנות את הנכס וכך גם על זמן בניית הנכס משום כך אנו צופים שבין מדד זה לבין המחיר יהיה קשר ישיר |
| גודל מגרש | משתנה מסביר | מצביע על כמות השטח של הנכס (גם בנוי וגם שטח נוסף במידה ויש) אחד הפרמטרים שקובעים את הערך הכספי של השטח הוא הגודל שלו ולכן גודל השטח צפוי להסביר על מחיר הנכס. |
| גודל המרתף | משתנה מסביר | מרתף בנכס מצביע על יכולת נוספת לאחסן צויד של המשתמשים בנכס וכך יכול להשפיע על הערך שלו. |
| מספר קומות בנכס | משתנה מסביר | מספר הקומות בנכס מצביע על החלוקה של החלל בנכס ועל היכולת לחלק את הנכס וליצור הפרדה בין חללים(בדרך כלל בין החדרים לחללים המשותפים כגון סלון ומטבח) דבר שצפוי להעלות את ערך הנכס בהתאם למספר הקומות בו. |
| מצב הנכס | משתנה מסביר | המצב של הנכס יכול להשפיע על מחיר הנכס לפי העובדה שאנשים ירצו לרכוש נכס במצב טוב וכך הביקוש שלו יעלה וכך גם ערכו של הנכס צפוי לגדול. |
| השנה בה הנכס נבנה | משתנה מסביר | השנה שבה נבנה הנכס יכול להעיד על ערכו משום שכלל שהנכס חדש יותר צפוי כי יהיה בו פחות ליקויים וכך גם ייתכן כי התקציב של האחזקה השוטפת צפוי לרדת, משום כך אנו צופים שנכס שנבנה לאחרונה יהיה שווה יותר מנכס ישן יותר וכך תהיה השפעה כלשהי על המחיר של הנכס, אולם עוד גורמים מעורבים במצב של הנכס כגון שיפוצים שנעשו ותחזוקה שוטפת ולכן צפוי קשר כלשהו. |
| האם הנכס שופץ | משתנה מסביר | נכס ששופץ לאחרונה לרוב יהיה במצב טוב יותר וכך גם מחירו יעלה, בנוסף שיפוץ נחשב השקעה בנכס וטיפול שותף ועל כן שיפוץ יכול להשפיע בצורה ישירה על מחיר הנכס |

1. **קשרים סיבתיים:**
2. **בין האם הנכס שופץ 9 Xלשנה שבה הנכס נבנה 8X-** להערכתנו נכס שנבנה לפני זמן רב יצטרך שיפוץ לשם תחזוקת הנכס ומשום כך אנו צופים קשר הפוך בין השנים- ככל שהבית נבנה מוקדם יותר כך יהיו יותר נכסים ששופצו.  
   מתוך נתוני הקורלציה ניתן לראות כי אכן קיים קשר הפוך בין השניים -0.31410348
3. **בין מספר חדרי שינה 2 Xלגודל הבית** 3X – להערכתנו כלל שהבית יהיה גדול יותר כך יהיו בו יותר חדרי שינה כיוון שיהיה יותר מקום שכך ישמש לעוד חדרי שינה, משום כך אנו מצפים לקשר ישיר בין השניים. הקורלציה היא: 0.46961853 ולכן לפי הממצאים הערכתינו אכן נכונה.
4. **בין מספר הקומות 6 Xלגודל הבית 3X** – להערכתנו ככל שהבית גדול כך יהיה מקום לעוד קומות בבית ולפיכך אנו מצפים שיהיה קשר ישיר בין שני המשתנים, לפי נתוני הקורלציה ניתן לראות כי: 0.1853100 ישנו קשר חיובי
5. **בין גודל המגרש 4 Xלגודל הבית 3X** – להערכתנו כלל שהבית יותר גדול כך אנו צופים כי השטח של המגרש יהיה גם בהתאם גדול, לפי הציפייה שלנו אנשים שיכולים להרשות לעצמם בית גדול יוכלו גם לרכוש להם מגרש גדול עם גינה ואז יהיה קשר ישיר בין השניים והקורלציה שלהם תהיה חזקה יחסית, לפי נתוני הקורלציה ניתן לראות שהיא: 0.4276966 ומשום כך ההערכה שלנו הייתה נכונה.
6. **בין גודל המרתף 5 Xלגודל הבית 3X** – להערכתנו יהיה קשר ישיר בין השניים כיוון שבית גדול זה מותרות וכך גם מרתף גדול, להערכתנו לאנשים שיש כסף לבית גדול יהיה להם מרתף וגם יחסית גדול כי הוא גם נכלל בגודל של הבית. לפני נתוני הקורלציה ניתן לראות כי: 0.5857622 משום כך השערתנו נכונה.

**קשרים מדגמיים:**

1. **בין האם הנכס שופץ 9 Xלמצב הנכס 7X** – להערכתנו יהיה קשר ישיר בין השניים כיוון שאנו צופים שנכס ששופץ יהיה במצב טוב יותר עקב השיפוץ. לפי נתוני הקורלציה הקשר הוא: -0.20801426 ולכן אנו רואים כי קיים קשר הפוך דווקא בין השניים דבר שלא ציפינו אותו ונוגד את ההסבר ולכן נסיק כי קיימים משתנים נוספים על הקשר בצורה עקיפה במדגם.
2. **ניתוח תיאורי של המשתנים:**

|  |  |
| --- | --- |
| 1. תוצאה | נתון סטטיסטי |
| 58.5 | ממוצע |
| 58.5 | חציון |
| 33.6303 | סטיית תקן |
| 29.75 | רבעון ראשון |
| 87.25 | רבעון שלישי |
| 0 | א-סימטריה |

**:X1 – ID**

המשתנה הוא קטגוריאלי, ולכן אי אפשר להסיק על פי density כי ההתפלגות סימטרית או לא סימטרית.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 3.4396 | ממוצע |
| 3 | חציון |
| 0.8052 | סטיית תקן |
| 3 | רבעון ראשון |
| 4 | רבעון שלישי |
| 0.541 | א-סימטריה |

**X2-מספר חדרי שינה**

המשתנה הוא קטגוריאלי, ולכן אי אפשר להסיק על פי density כי ההתפלגות סימטרית או לא סימטרית.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 2136.3103 | ממוצע |
| 2045 | חציון |
| 898.077 | סטיית תקן |
| 1445 | רבעון ראשון |
| 2570 | רבעון שלישי |
| 1.1184 | א-סימטריה |

**3X – גודל אזור מחייה**

ניתן לראות כי ערך הממוצע וערך החציון יחסית קרובים זה לזה, אך הממוצע מעט גדול יותר. עובדה זאת יחד עם ערך א-סימטריה חיובי מביאים לכך שאנחנו נצפה לקבל התפלגות א-סימטרית חיובית עם זנב ימני. בשל שהפרש בין הרבעונים גדול יחסית מצביע על כך שהפיזור יהיה רחב יותר.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 13037.1206 | ממוצע |
| 7680 | חציון |
| 18490.0431 | סטיית תקן |
| 5000 | רבעון ראשון |
| 10449.75 | רבעון שלישי |
| 3.7481 | א-סימטריה |

**4X – גודל שטח הנכס**

ניתן לראות כי ערך הממוצע גדול בהרבה יחסית מערך החציון. עובדה זאת יחד עם ערך

א-סימטריה חיובי מביאים לכך שאנחנו נצפה לקבל התפלגות א-סימטרית

חיובית עם זנב ימני ארוך. בשל שהפרש בין הרבעונים גדול יחסית וסטיית התקן גדולה יחסית ביחס לנתונים, מה שמצביע על כך שהפיזור רחב מאוד.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 362.3103 | ממוצע |
| 0 | חציון |
| 484.7658 | סטיית תקן |
| 0 | רבעון ראשון |
| 702.5 | רבעון שלישי |
| 1.1956 | א-סימטריה |

**5X – דוגל המרתף**

ניתן לראות כי ערך הממוצע גדול בהרבה יחסית מערך החציון. עובדה זאת יחד עם ערך

א-סימטריה חיובי מביאים לכך שאנחנו נצפה לקבל התפלגות א-סימטרית

חיובית עם זנב ימני ארוך. בשל שהפרש בין הרבעונים גדול יחסית וסטיית התקן גדולה יחסית ביחס לנתונים, מה שמצביע על כך שהפיזור רחב מאוד.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 1.5603 | ממוצע |
| 1 | חציון |
| 0.6499 | סטיית תקן |
| 1 | רבעון ראשון |
| 2 | רבעון שלישי |
| 0.72 | א-סימטריה |

**6X-מספר קומות**

המשתנה הוא קטגוריאלי, ולכן אי אפשר להסיק על פי density כי ההתפלגות סימטרית או לא סימטרית.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 3.4655 | ממוצע |
| 3 | חציון |
| 0.7274 | סטיית תקן |
| 3 | רבעון ראשון |
| 4 | רבעון שלישי |
| 0.5217 | א-סימטריה |

**7X-מצב הדירה**

המשתנה הוא קטגוריאלי, ולכן אי אפשר להסיק על פי density כי ההתפלגות סימטרית או לא סימטרית.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 1962.8448 | ממוצע |
| 1967 | חציון |
| 30.2285 | סטיית תקן |
| 1941.75 | רבעון ראשון |
| 1986 | רבעון שלישי |
| -0.3428 | א-סימטריה |

**8X-שנת בנייה**

ניתן לראות כי ערך הממוצע וערך החציון יחסית קרובים זה לזה, אך הממוצע מעט קטן יותר. עובדה זאת יחד עם ערך א-סימטריה שלילית מביאים לכך שאנחנו נצפה לקבל התפלגות סימטרית שלילית עם זנב שמאלי. בשל שהפרש בין הרבעונים גדול יחסית מצביע על כך שהפיזור יהיה רחב יותר.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 0.0948 | ממוצע |
| 0 | חציון |
| 0.2942 | סטיית תקן |
| 0 | רבעון ראשון |
| 0 | רבעון שלישי |
| 2.7302 | א-סימטריה |

**9X – האם הדירה שופץ**

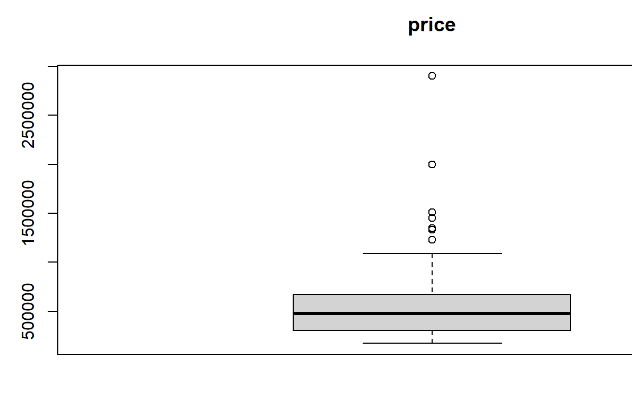
המשתנה הוא קטגוריאלי, ולכן אי אפשר להסיק על פי density כי ההתפלגות סימטרית או לא סימטרית.

|  |  |
| --- | --- |
| תוצאה | נתון סטטיסטי |
| 560967.7586 | ממוצע |
| 475000 | חציון |
| 382559.3258 | סטיית תקן |
| 307000 | רבעון ראשון |
| 670250 | רבעון שלישי |
| 2.7872 | א-סימטריה |
|  |  |

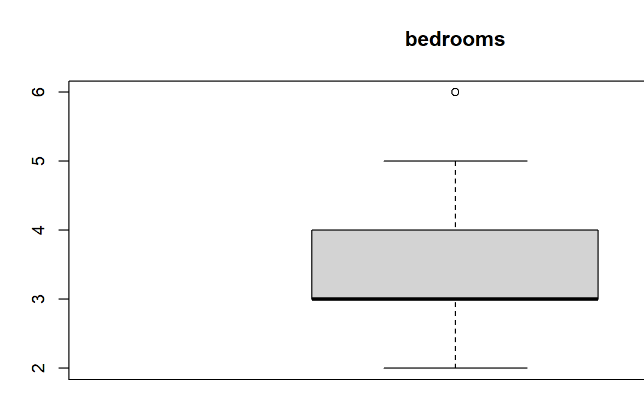
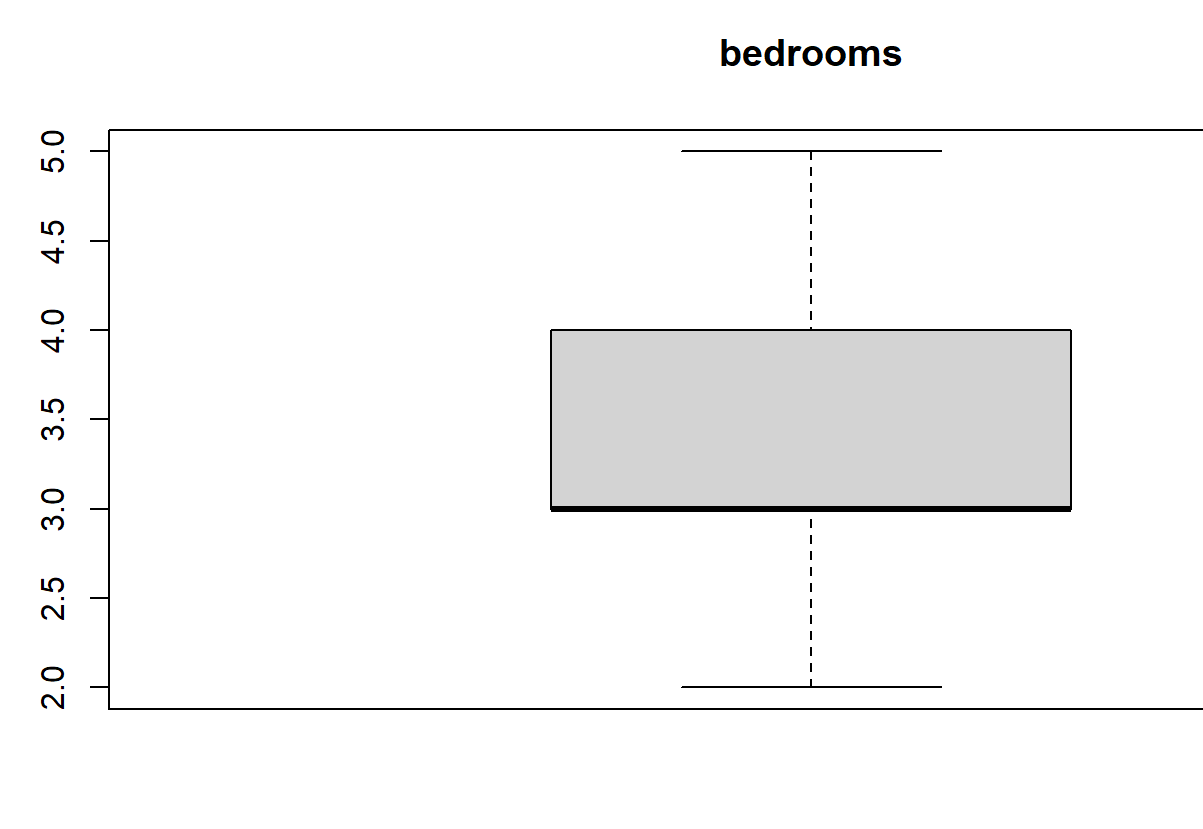
**Yמחיר המשתנה המוסבר –**

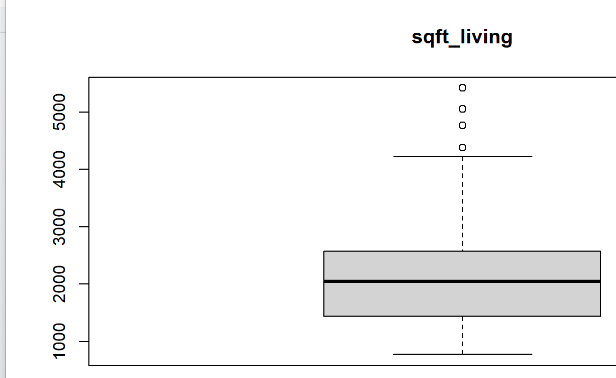
ניתן לראות כי ערך הממוצע גדול בהרבה מערך החציון. עובדה זאת יחד עם ערך א-סימטריה חיובי מביאים לכך שאנחנו נצפה לקבל התפלגות א-סימטרית חיובית עם זנב ימני ארוך. סטיית התקן גדולה יחסית ביחס לנתונים, מה שמצביע

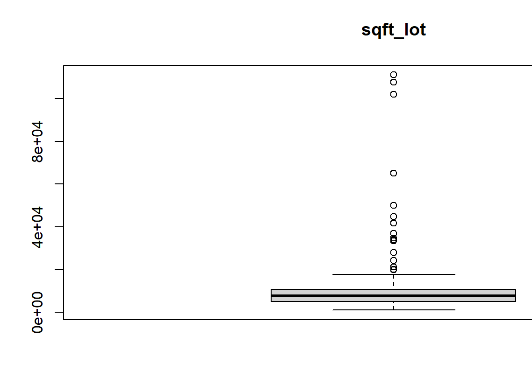
על כך שהפיזור רחב מאוד.

1. ****  
   **Y המשתנה המוסבר, המחיר של נכס-**  
   גבול תחתון –171800 גבול עליון – 1090000 חציון- 475000

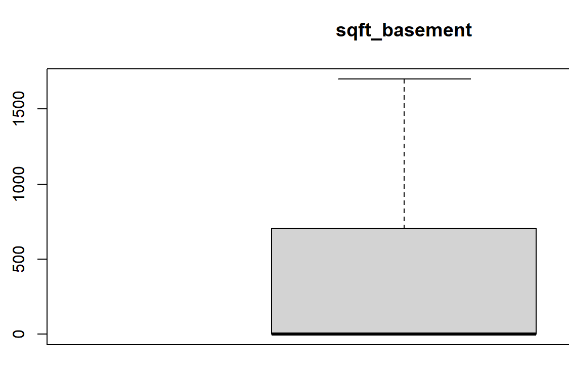
התקבלו 7 תוצאות כולן מעל האחזון ה75, לפי בדיקה של הנתונים שמנו לב כי מדובר על נכסים שגודל הבית שלהם יחסית גדול ורצינו להראות קשר בין מחיר הנכס לבין השטח, דבר שגם תאם את הנתונים ונרצה להראות את הקשר בין השניים לכן לא נרצה לנפות חריגים במקרה זה.

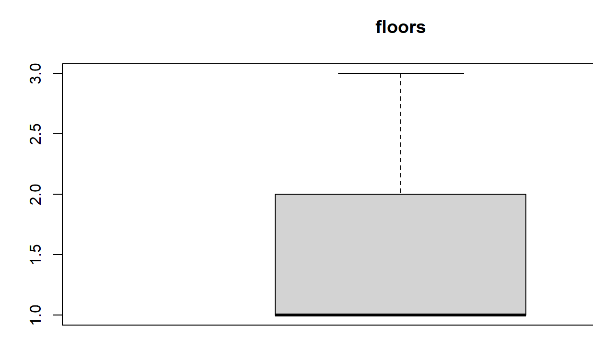
****  
**2X -מספר חדרים**  
גבול תחתון –2 גבול עליון –5 חציון-3  
התקבלה תוצאה חריגה אחת שהיא 6 חדרים, החלטנו להוציא את התוצאה מכיוון שמבחינת הנתונים היא עלולה לסמן על מגמה של 6 חדרים בבית אולם מדובר רק על תוצאה אחת שלפיה לא ניתן באמת לראות מגמות כלשהן ולא נרצה להטות את המודל.  
אחרי ההוצאה התרשים נראה כך:  


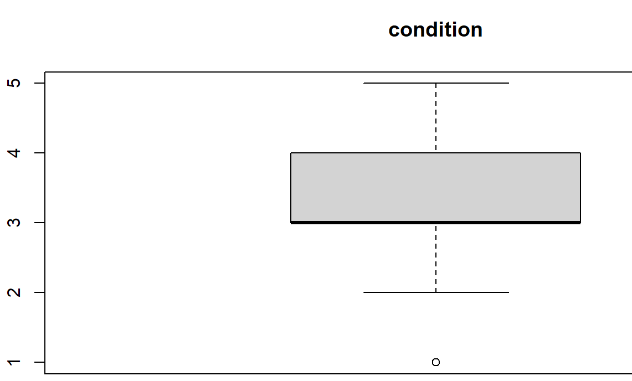
  
  
**3X – גודל הבית**  
גבול תחתון –770 גבול עליון –4220 חציון-2045  
התקבלו 4 חריגים כולם מעל האחוזון ה75, לפי שכבר הצגנו למעלה חריגים אלה מתייחסים גם לחריגים במחיר של בנכס ומראים את הקשר בין המחיר לבין הגודל של השטח הבנוי ולכן לא נרצה להוריד אותם.

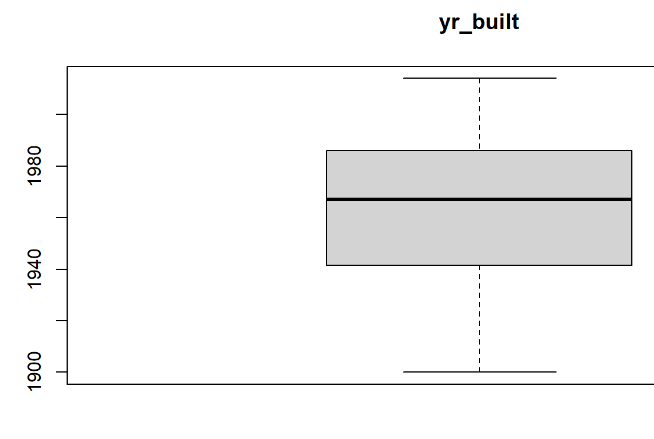
****  
**4X -גודל המגרש**  
גבול תחתון – 1044.0גבול עליון – 17789.0 חציון-7680.0  
התקבלו 17 חריגים כולם מעל האחוזון ה75, מדובר על כמות יחסית גדולה של חריגות יחסית וניתן לראות גם התפלגות כלשהי מבין החריגות, שיש כאלה שהן הרבה מעל כל שאר החריגות האחרות.  
בחרנו גם לנפות את 3 החריגות העליוניות מכיוון כמות הגיונית לנפות ו17 עלול להשפיע על המדגם שלנו, בנוסף לכך בחרנו לנפות דווקא אותן מכיוון שלא ראינו קשר משמעותי בניהם לבין הנתונים שלנו שהיה ניתן להסיק אליהן קשר הגיוני.  
לאחר הניפוי התרשים נראה כך:  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

****  
  
**5X -גודל המחסן**  
גבול תחתון – 0 גבול עליון –1700 חציון-0   
התקבלו 0 חריגות ולכן לא נשנה דבר

****  
**6X – מספר קומות**  
גבול תחתון –1 גבול עליון – 3 חציון-1  
אין חריגים לכן אין מה להוציא

****  
**7X – מצב הנכס**  
גבול תחתון – 2 גבול עליון – 5 חציון-3   
ישנה חריגה אחת על נכס שקיבל את הערך 1, אולם מבחינת הנתונים לא ראינו סיבה הגיונית אולם מדובר על חריגה אחת מתוך 116 ולכן החריגה היא מינורית ובחרנו שלא להוציא אותה

  
**8X -שנה בה הנכס נבנה**  
גבול תחתון –1900.0 גבול עליון – 2014.0 חציון-1967  
אין חריגות ולכן אין צורך להוציא אותן

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

**9X -האם הנכס שופץ**  
גבול תחתון – 0 גבול עליון –0 חציון-0  
התקבלו 11 חריגות שכולן מצביעות על העובדה כי הנכס שופץ זוהי חריגה, אולם אנו יודעים כי זאת אינה חריגה ומצב זה הגיוני ולכן התייחסנו כך לנתונים ובחרנו שלא להוציא את החריגות

1. **פונקציית צפיפות והתפלגות מצטברת:**

**3X – שטח אזור מחייה:**

**פונקציה צפיפות ההסתברות:**

Chart, histogram

Description automatically generated

ניתן לראות כי מדובר בפונקציית צפיפות א-סימטרית חיובית בעלת זנב ימני התואם את ערך הא-סימטריות החיובי 1.1184.זנב זה מעיד על מגוון של נכסים עם שטח אזור מחייה גדול מהרגיל . שיא פונקציית הצפיפות הוא בכ-2000 מ"ר מה שמצביע על כך שממוצע של שטח אזור מחייה היינו כ-2000 מ"ר.  
נתון זה מתיישב עם הממוצע שמצאנו שערכו הוא 2136.3103 מ"ר.  
בעזרת ההיסטוגרמה נוכל לראות כי רוב שטח המגרש נע בין כ-1000 מ"ר עד כ-3000 מ"ר.

**פונקציה התפלגות מצטברת:**

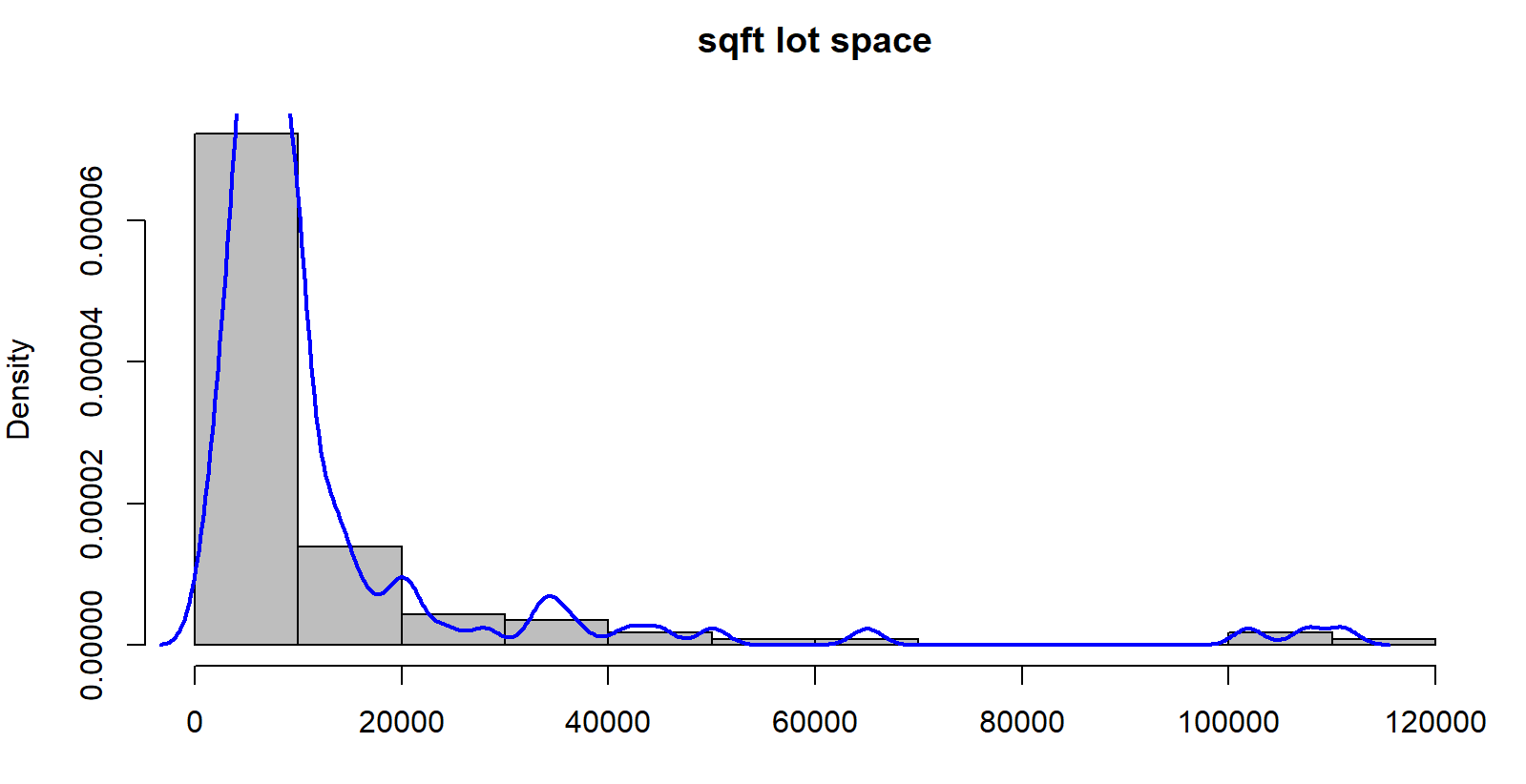
Chart, line chart

Description automatically generated

ניתן לראות שקיים שיפוע יחסית מתון של הגרף הנוצר בעקבות מגוון רחב של נכסים שהשטח הממוצע שלהם תואם את המסקנות מפונקציית הצפיפות. נסתכל על הגרף בשני תחומים, הראשון התחום בין כ-1000 מ"ר עד 3000 מ"ר בו ניתן לראות שכ-80% מאוכלוסיית המדגם נמצאים בתחום זה. התחום השני הוא מ3000 מ"ר עד 5000 מ"ר, תחום זה כולל את כל הנכסים ששטח הבית שלהם גדול יחסית מהרגיל. ניתן להסיק מהגרף כי כ-80% מהמבנים הינם נעים סביב גודל הממוצע של 2136.3103 מ"ר וה-20% הנותרים מפוזרים בתחום של 3000 מ"ר עד 5000 מ"ר.

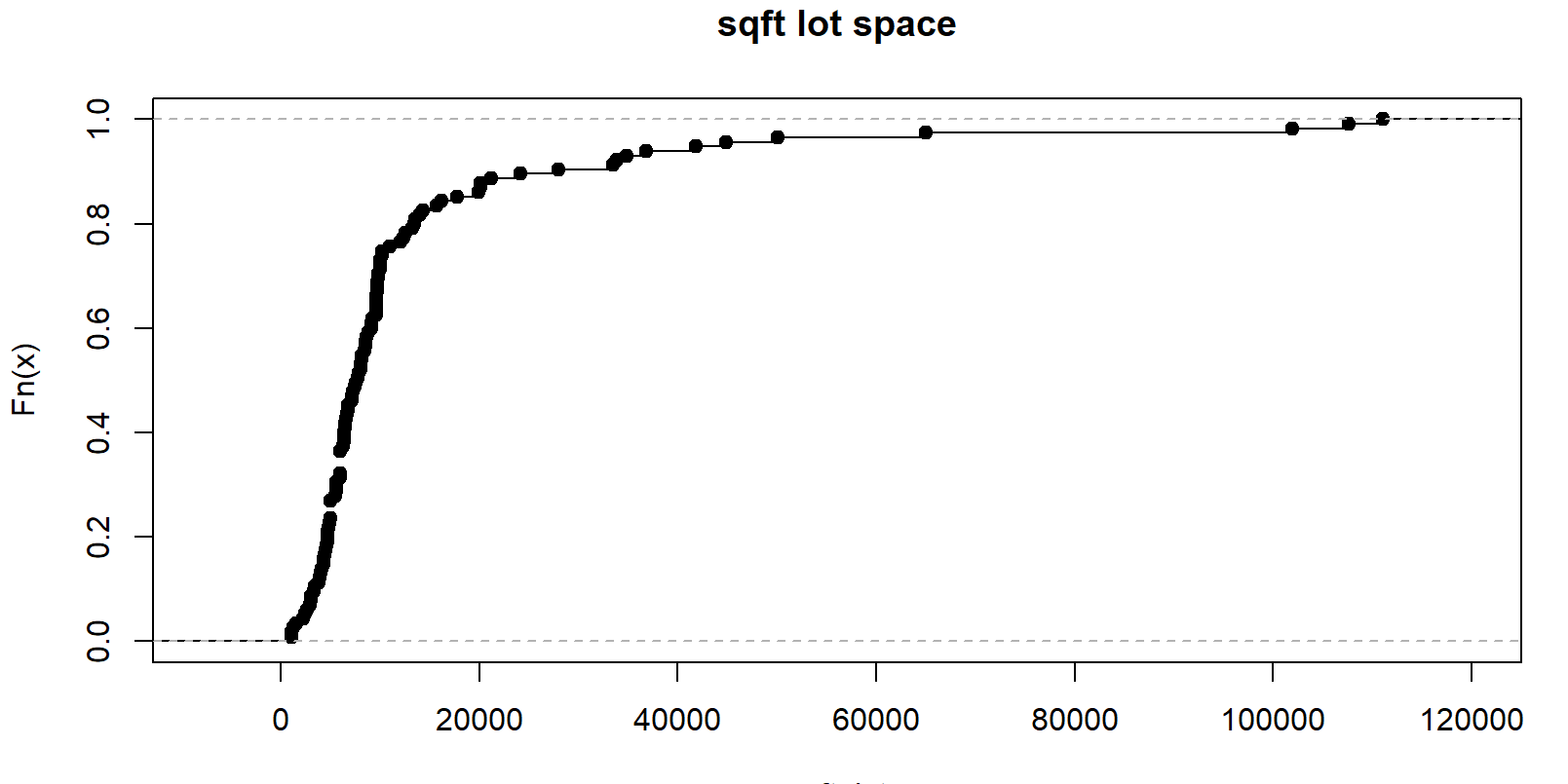
**4X – שטח המגרש:**

**פונקציה צפיפות ההסתברות:**



ניתן לראות כי מדובר בפונקציית צפיפות א-סימטרית חיובית בעלת זנב ימני גדול יחסית התואם את ערך הא-סימטריות החיובי 3.7481 . זנב זה מעיד על מגוון של נכסים בעלי שטח מגרש גדול מהרגיל.   
בנוסף לכך, קיים גם אחוז קטן של מגרשים אשר גדולים פי 9 מ"ר מהממוצע. נתון זה מצביע על כך שקיים פיזור גדול במדגם.  
שיא פונקציית הצפיפות הוא כ-11000 מ"ר מה שמצביע על כך שממוצע שטח אזור מחייה היינו כ-11000 מ"ר, נתון זה מתיישב עם הממוצע שמצאנו מתוך סעיף 5 13037.1206 מ"ר.  
בעזרת ההיסטוגרמה נוכל לראות כי במספר רב של נכסים עם שטח מגרש אשר שיטחם נע בין כ-0 מ"ר עד כ-20000 מ"ר.  
בנוסף, כאשר נתבונן בזנב השמאלי נראה שאין מגרשים ששיטחם בתחום בין כ-7000 מ"ר עד כ-10000 מ"ר.

**פונקציה התפלגות מצטברת:**



ניתן לראות שקיים שיפוע יחסית תלול של הגרף הנוצר בעקבות מגוון רחב של נכסים בעלי שטח מגרש ממוצע אשר תואם את המסקנות מפונקציית הצפיפות. התחום בין כ-0 מ"ר עד 15000 מ"ר בו ניתן לראות שכ-80% מאוכלוסיית המדגם נמצאים בתחום זה. כאשר נתבונן ב-20% הנותרים נראה שקיים מדרגות בפונקציית ההתפלגות המצטברת, הם מעידים על כך שאין מגרש במדגם שגודלו נמצא בתחום זה. המסקנות אלו תואמות את המסקנות מפונקציית הצפיפות. לכן נאמר, כ-80% ממגרשי המדגם נעים סביב שטח הממוצע של 13037.1206 מ"ר ו-20% הנותרים הם בין 20000 מ"ר עד 110000 מ"ר כאשר תחום זה מכיל תחומים ריקים, ז"א תתי תחומים שאין מגרש אשר מוכלים בהם .

**8X – שנת בנייה:**

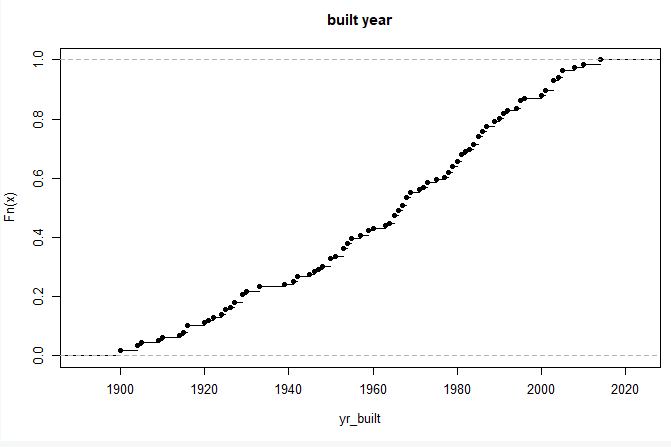
**פונקציה צפיפות ההסתברות:**

Chart, histogram

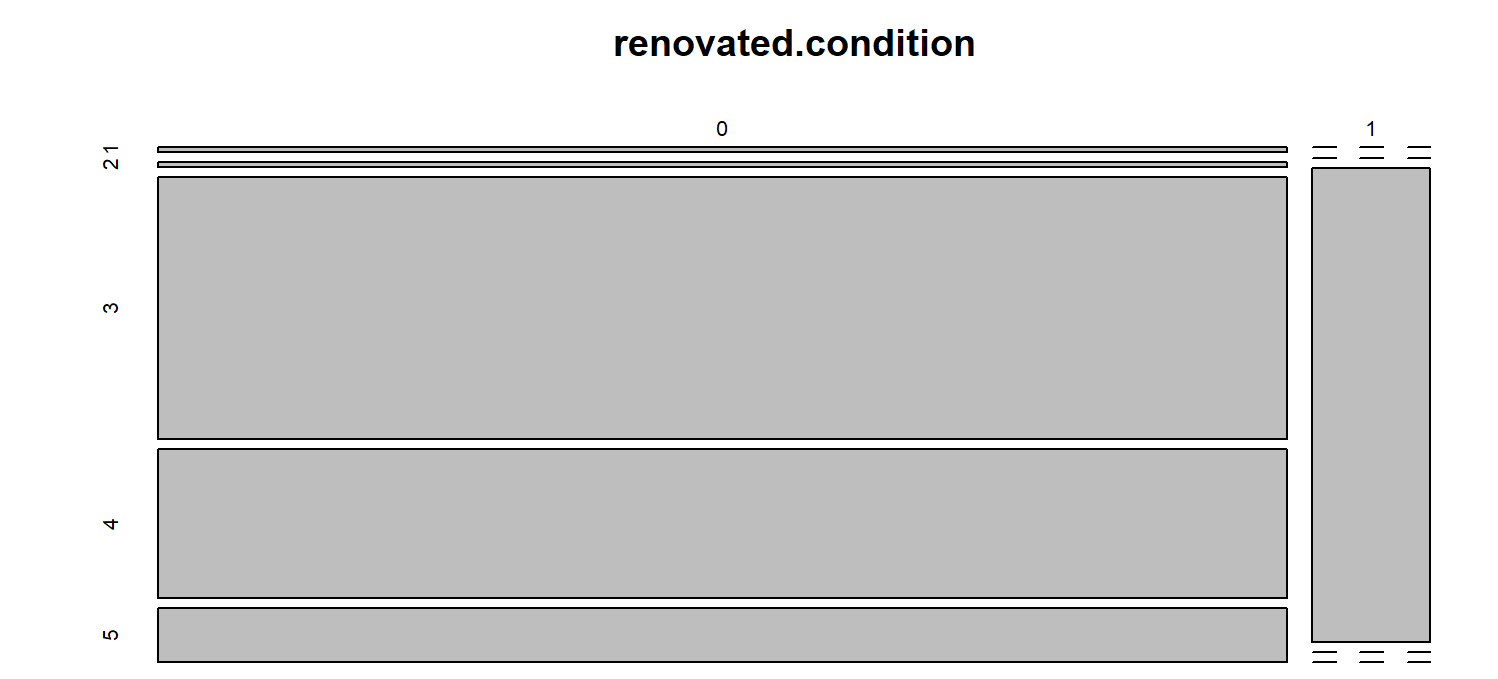
Description automatically generated

ניתן לראות כי מדובר בפונקציית צפיפות א-סימטרית שלילית בעלת זנב שמאלי יחסית קטן התואם את ערך הא-סימטריות השלילי-0.3428.  
זנב זה מעיד על מגוון של מבנים שנבנו בתחילת שנות ה-90 . פונקצית צפיפות ההסתברות מזכירה לנו התפלגות פעמונית בהיסטוגרמה. שיא פונקציית הצפיפות הוא בשנת 1980 מה שמצביע על כך שנת הבנייה הממוצעת הינה ב-1980, נתון זה מתיישב עם הממוצע שמצאנו 1962. בנוסף ניתן לראות כי כל המבנים נבנו בין השנים 1900 עד 2020, ממצא זה מתיישב לנו עם סטיית התקן של 30.2285 . בעזרת ההיסטוגרמה נוכל לראות כי ישנו מספר רב של מבנים אשר נבנו בין השנים 1940 ל-2000 . אם נתבונן בזנב השמאלי נראה שבין השנים 1900 ל-1940 בהם הבנייה התחילה ולכן הכמות קטנה יחסית ולאחר שנת 2000 קצב הבנייה ירד וזאת מפני שעברנו את השיא.

**פונקציה התפלגות מצטברת:**



ניתן לראות שקיים שיפוע יחסית מתון של הגרף הנוצר בעקבות מגוון רחב של מבנים שנבנו בשנת הממוצע אשר תואם את המסקנות מפונקציית הצפיפות. נתבונן בתחום בין שנת 1900 עד 2000 בו ניתן לראות שכ-80% מאוכלוסיית המדגם נמצאים בתחום זה. ב-20% הנותרים נראה את כל הבנייה משנת 2000 עד 2020. לכן נאמר, כ-80% ממבני המדגם נבנו סביב השנה הממוצעת של 1962ו-20% הנותרים נבנו בין 2000 עד 2020 .

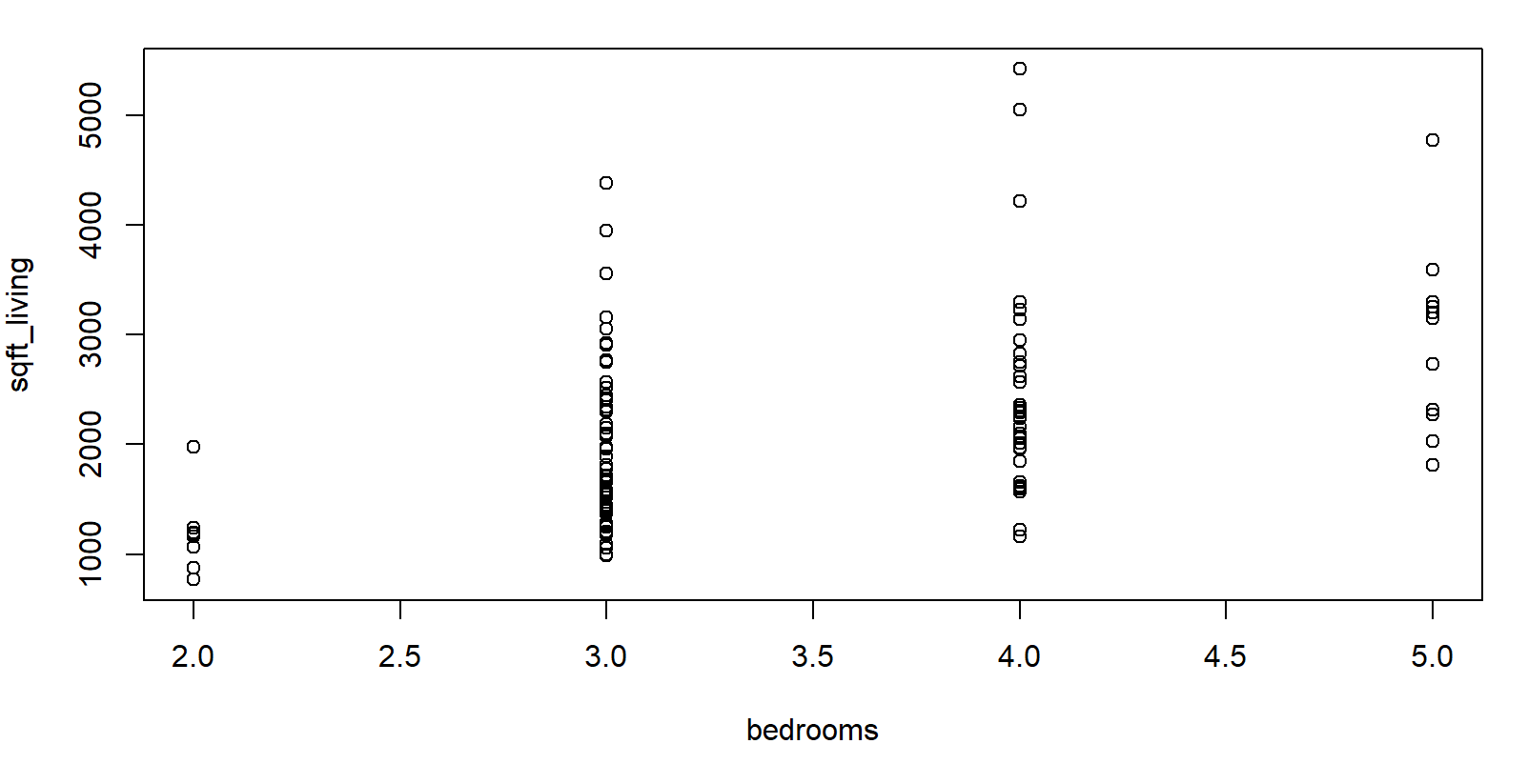
1. **ייצוג קשרים בעזרת תרשימים:  
   קשר בין מצב הנכס 8 Xלבין האם הוא שופץ 9X -**בחרנו לתאר קשר זה באמצעות תרשים פסיפס כיוון שמדובר במשתנים קטגוריאליים וניתן לראות בצורה טובה את הקשרים.  
   זהו קשר מעניין שכבר הזכרנו אותו לאורך העבודה, ניתן לראות נתון מפתיע כי כל הדירות שעברו שיפוץ קיבלו את הדירוג 3 וזה מפתיע כי לפי הערכה שלנו דירות שעוברות שיפוץ יהיו בדירוג גבוה יותר ומכן אנו יכולים להסיק שדווקא דירות ששופצו הן במצב יחסית פחות טוב בניגוד למה שחשבנו, לגבי דירות שבמצב 1 ו2 ניתן לראות שהם פחות משמעותיים ולכן לא נתייחס אליהם.  
   

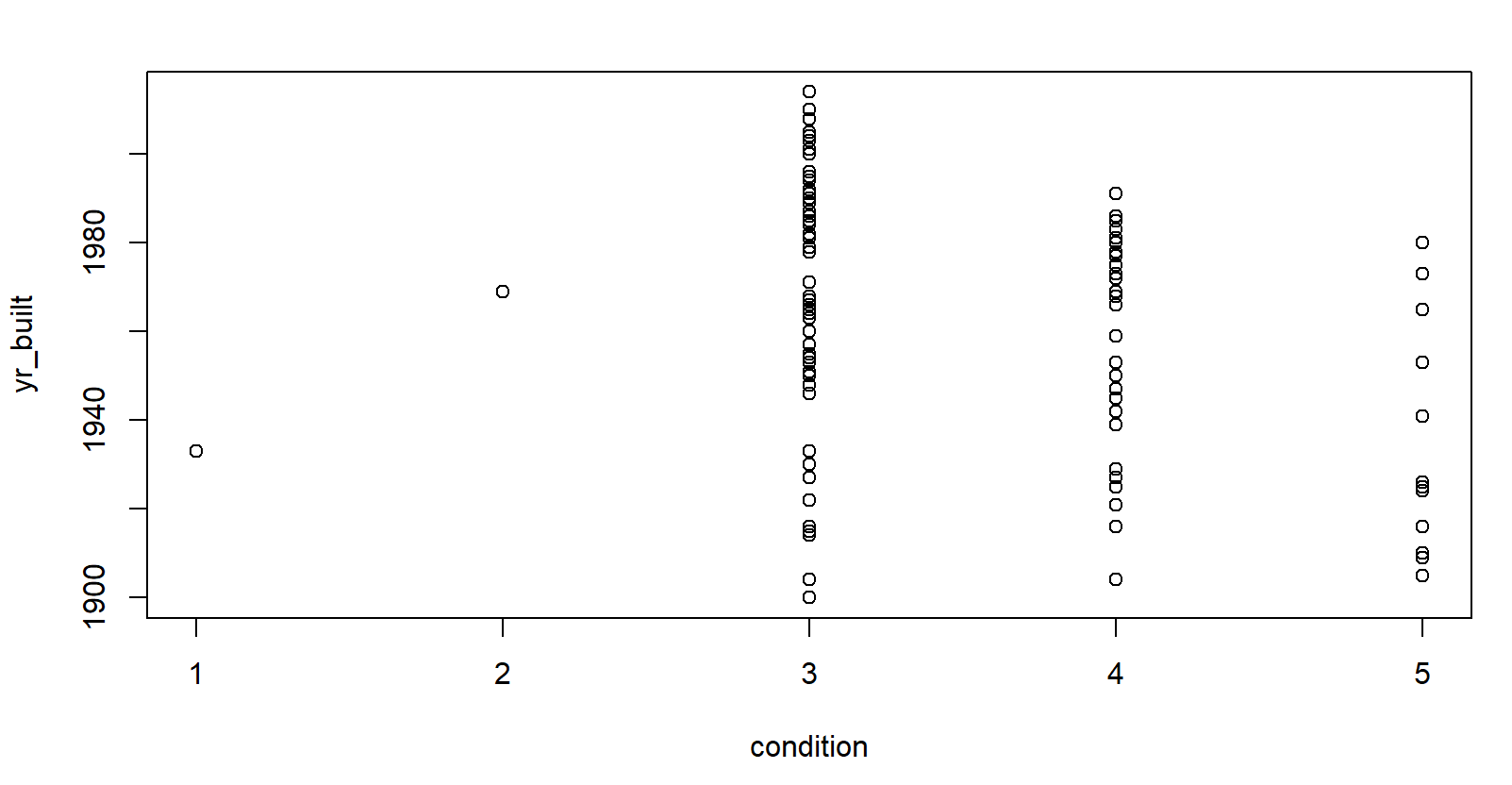
**קשר בין מספר חדרי השינה 2 Xלבין מספר הקומות 9X -**בחרנו לתאר קשר זה באמצעות תרשים פסיפס כיוון שמדובר במשתנים קטגוריאליים וניתן לראות בצורה טובה את הקשרים.  
ניתן לראות כי רוב הנתונים מתרכזים בכך שמספר חדרי השינה הוא בין 3 ל4 ושם השטח הגדול, בנוסף עוד אנו רואים כי עבור קומה 1 גם נמצא השטח הגדול של הנתונים מבחינת נתוני הקומות אולם ניתן גם לראות גידול במספר הקומות ביחס למספר החדרים.

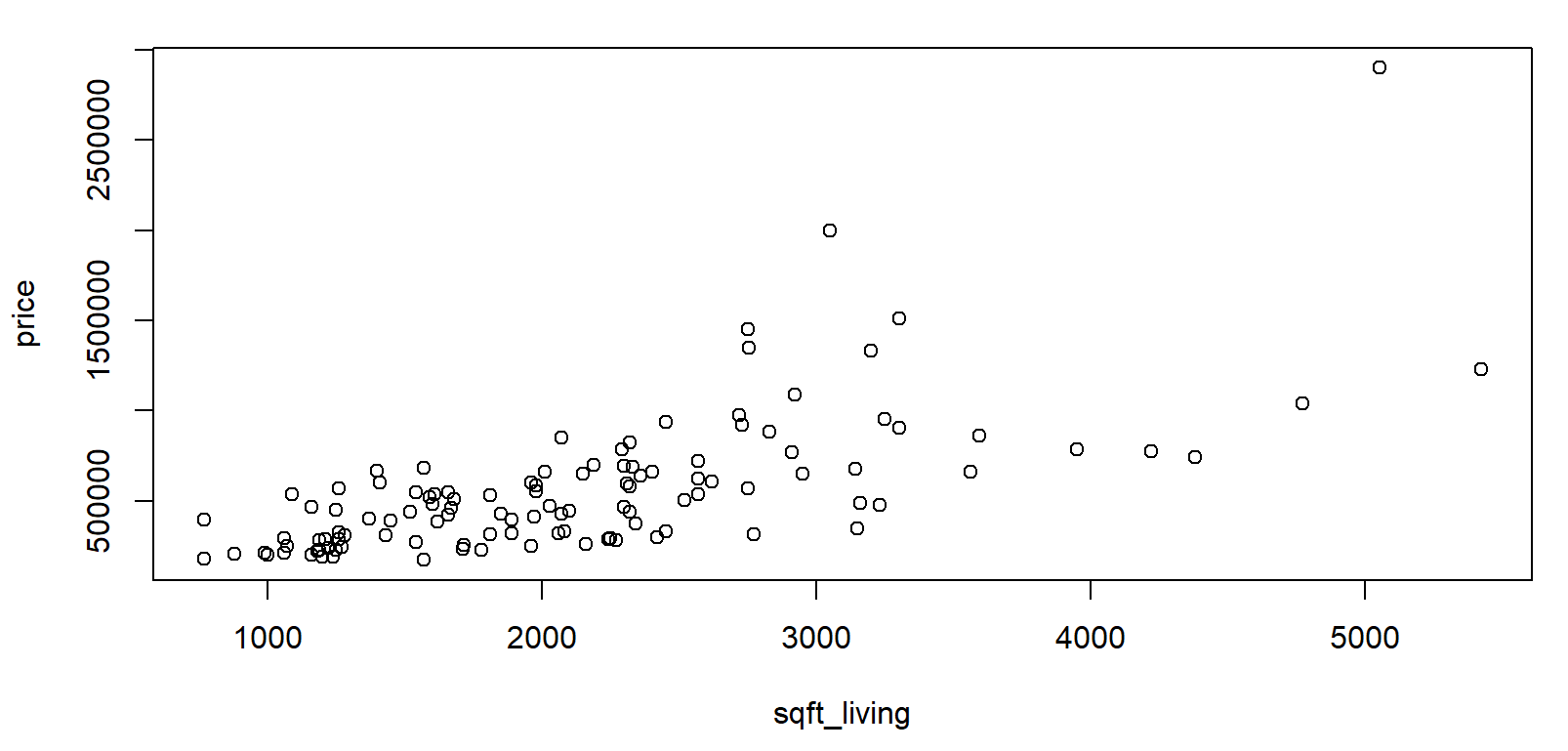


**קשר בין מספר חדרי השינה 2 Xלבין מצב הנכס 7X -**בחרנו לתאר קשר זה באמצעות תרשים פסיפס כיוון שמדובר במשתנים קטגוריאליים וניתן לראות בצורה טובה את הקשרים.  
מתרשים זה אנו למדים כי הגדלים העיקריים נמצאים במרכז ולא בשוליים, כלומר עבור מספר חדרי שינה של 3 מצב הנכס הוא 3 בהתאם וזה השטח הגדול וכך גם לגבי שאר הגדלים, כלומר השטח המשמעותי נמצא בממוצעים של מספר חדרי השינה ושל המצב.



**קשר בין מספר החדרים 2 Xלבין גדול הבית 3X -**בחרנו לתאר קשר זה באמצעות תרשים פיזור כיוון שמדובר במשתנה אחד רציף ואחד בדיד ולכן חיפשנו מגמות שינוי כלשהו על הגרף. בגרף המוצג ניתן לראות כי ישנה מגמת עלייה בגודל הבית ככל שמספר הקומות גדול יותר אבל לא בהרבה וגם עבור 3-5 חדרים ניתן לראות כי ישנו שטח חופף גדול יחסית עבור הנתונים הללו-כלומר ניתן למצוא גודל זהה של בית עבור 3-5 חדרים לעומת 2 חדרים ששם הגודל יורד משמעותית, בנוסף ניתן לראות שרוב הבתים הם עם 4 או 3 חדרים.  


**קשר בין מצב הנכס 7 Xלבין השנה שבה נבנה 8X -**בחרנו לתאר קשר זה באמצעות תרשים פיזור כיוון שמדובר במשתנה אחד רציף ואחד בדיד ולכן חיפשנו מגמות שינוי כלשהו על הגרף. אנו צפינו כי יהיה קשר ישיר בין המשתנים כיוון שעבור נכס במצב טוב אנו צופים שיהיה חדש יותר (שנה מאוחרת יותר) אך מתוך התרשים ניתן לראות שהדבר אינו כך, הזמן שבנו נבנה הנכס מתפרס כמעט באופן שווה בין המצבים של הנכס כאשר רוב הנתונים נמצאים במצבים 3-5 ובמצב 1-2 לא ניתן להסיק מגמה כלשהי מתוך 2 נתונים.  
כלומר, מתוך הגרף לא ניתן לקבוע משהו על השני של הנכס לפי המצב שלו.  


**קשר בין המחיר Y לבין גדול הבית 3X -**ניתן לראות לכי ישנו קשר ליניארי כלשהו בין המשתנים , אומנם השיפוע קטן יחסית אך הקשר קיים, בנוסף, ניתן לראות כי רוב הנתונים נמצאים כשגודל הבית נע סביב 1000-3000 מ"ר ושמחירו 500,000 ₪ .  
מתוך הגרף ניתן להסיק כי קיים קשר בין גודל הבית לבין מחירו וכלל שהמחיר גדול יותר כך גם ייתכן כי גודל הבית יהיה גדול יותר.  


1. **טבלאות שכיחויות:**  
   **טבלאות חד ממדיות:**  
   X2-מספר חדרי שינה:

|  |  |  |
| --- | --- | --- |
| אחוזים | תדירות | מספר חדרים |
| 0 | 0 | 1 |
| 0.07826087 | 9 | 2 |
| 0.52173913 | 60 | 3 |
| 0.30434783 | 35 | 4 |
| 0.09565217 | 11 | 5 |

טבלה זאת מציגה את ההסתברות להיות בטווח ערכים מסויים(מספר חדרים בבית) ואת מספר הפעמים התצפיות להיות בטווח הזה.  
ניתן לראות כי הטווח של החדרים הם בין 2-5 כי הכמות המשמעותית של מס החדרים זה 3 חדרים שאלה הדירות הכי שכיחות ולאחר מכן 4 חדרים אולם ישנם גם בתים עם 2 ו5 חדרים אולם השכיחות שלהם נמוכה בהרבה ביחס לשאר.  
בנוסף, ניתן לראות שרוב הערכים הם במרכז (4-3) ולא בקצוות של הערכים.

8X-שנת בנייה:

|  |  |  |
| --- | --- | --- |
| אחוזים | תדירות | שנת בנייה |
| 0.09649123 | 11 | (1900,1920] |
| 0.13157895 | 15 | (1920,1940] |
| 0.19298246 | 22 | (1940,1960] |
| 0.22807018 | 26 | (1960,1980] |
| 0.22807018 | 26 | (1980,2000] |
| 0.12280702 | 14 | (2000,2020] |

טבלה זאת מציגה את טווחי השנים האפשריים שבהם נבנו בתים.  
לפי הטבלה ניתן לראות כי האחוזים פרוסים על פני טווח רחב יחסית ואין קפיצות גדולות באחוזים או בדהירות של הופעת הנתונים למעט בקצוות ששם ישנם קצת פחות נכסים.  
המשמעות היא שנכסים מחולקים לפי טווח רחב של שנים וניתן לראות שכמות הדירות בכל 20 שנה שנבנו יחסית קבועה בשנים שבין 1940-2000, לפני כן ואחרי כן הכמות הזאת ירדה וניתן לשייך את הירידה הזאת אולי למחסור במקום לבנות דירות בשנים המאוחרות ובשנים המקדמות אולי לכך שעוד ההתיישבות בארץ לא הייתה משמעותית.

הצלבה בין האם הדירה שופצה 9X לבין מצב הדירה X8:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| מצב הדירה  / האם שופץ | 1 | 2 | 3 | 4 | 5 |
| 1 – עבר שיפוץ | 0 | 0 | (0.09483%)11 | 0 | 0 |
| 0 – לא שופץ | (0.00863%)1 | (0.00863%) 1 | (0.5%)58 | (0.28448%)33 | (0.103448%)12 |

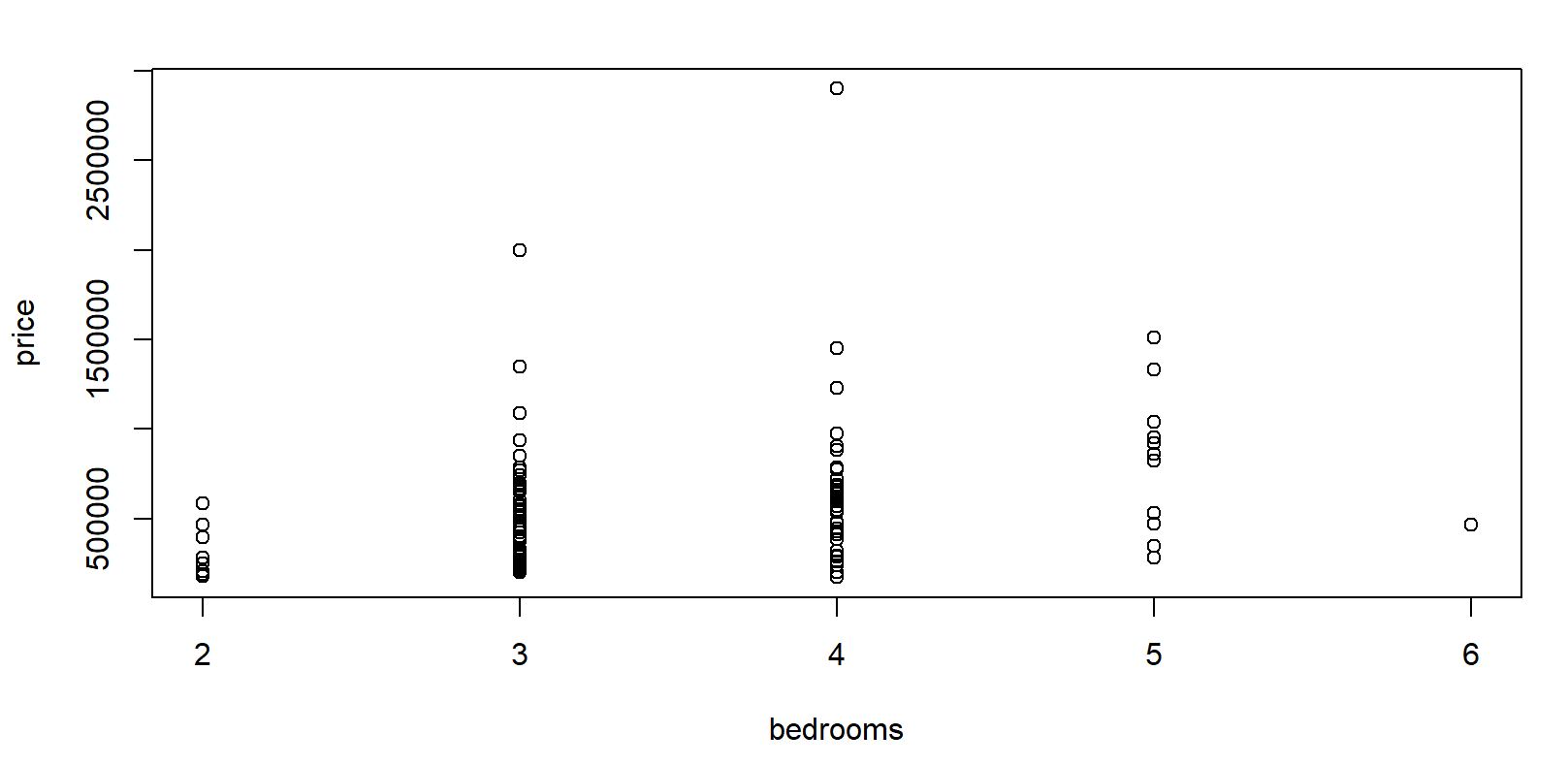
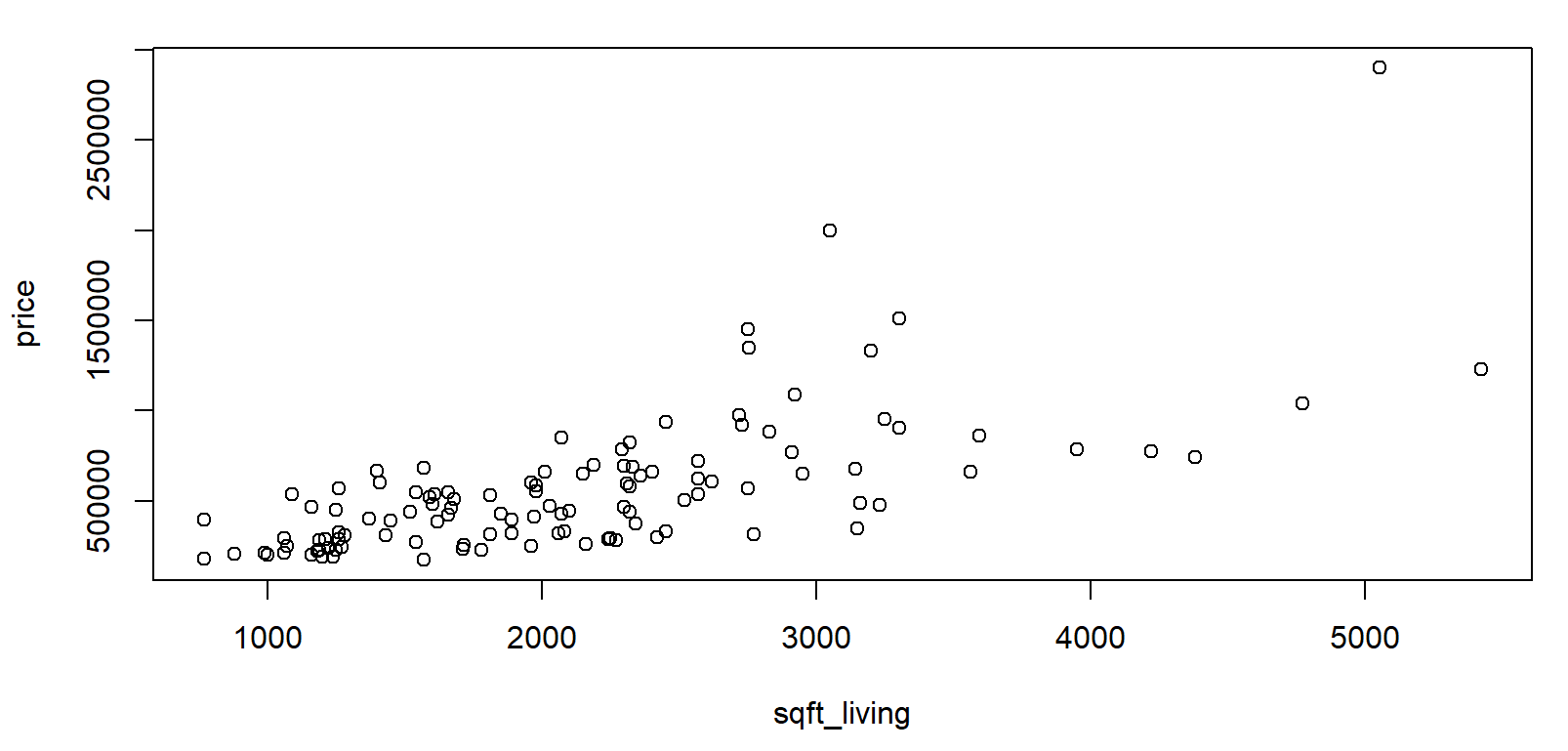
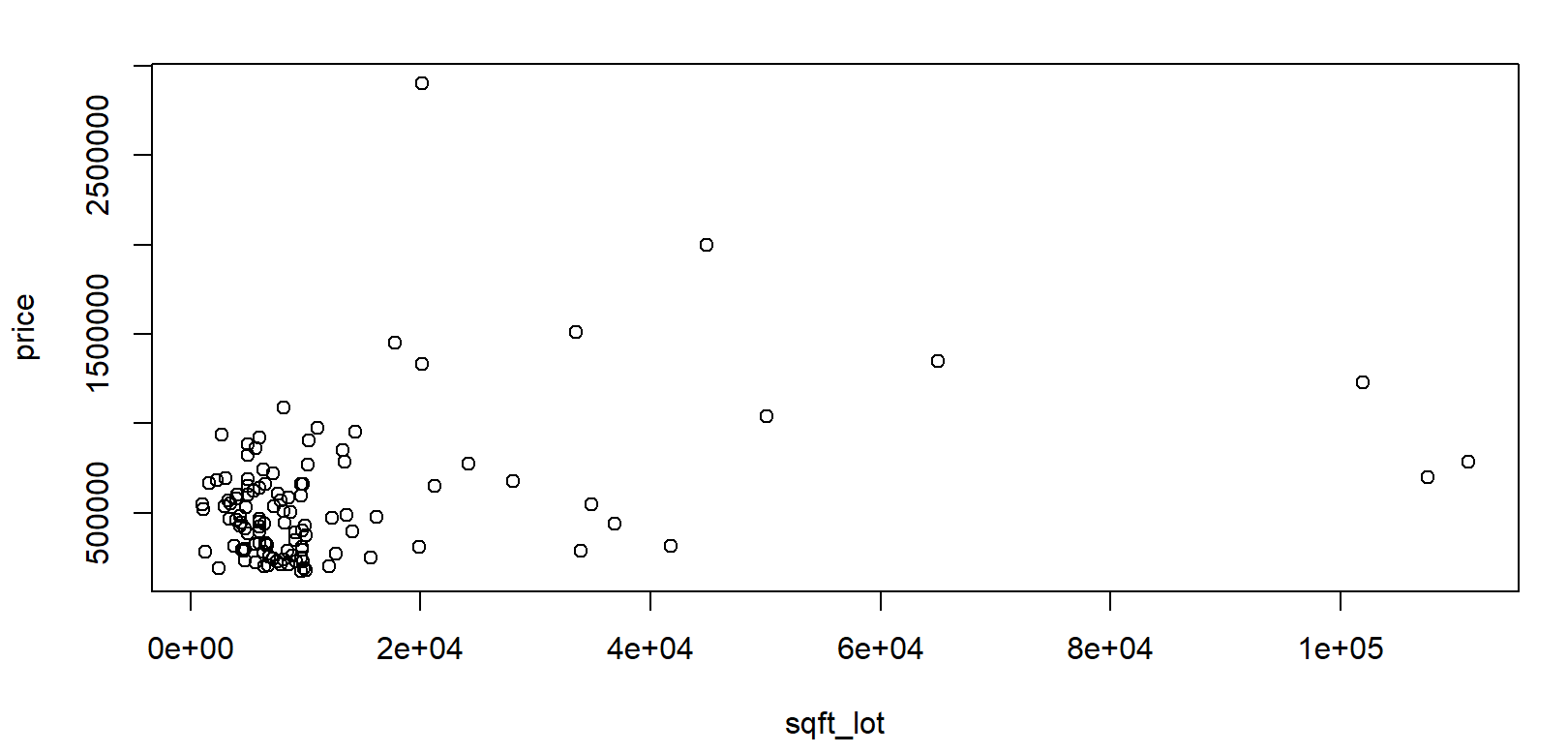
ניתן לראות כי מתוך הנתונים כמות הדירות שעברו שיפוץ היא קטנה ביחס לכמות הדירות שלא עברו שיפוץ, בנוסף ניתן לראות כי כל הדירות שעברו שיפוץ הן ברמת דירה 3 שהיא נמוכה יחסית ביחס לפיזור של נתוני מצב הנכסים. מתוך כל הנתונים הכמות הגדולה ביותר שמהווה חצי מכל הנתונים היא עבור תא ספציפי שהוא רמה 3 שלא עברו שיפוץ. ולאחר מכן רמה 4 שלא עבר שיפוץ, כלומר ניתן להסיק מכל שרוב הדירות הן במצב בינוני ולא עברו שיפוץ.

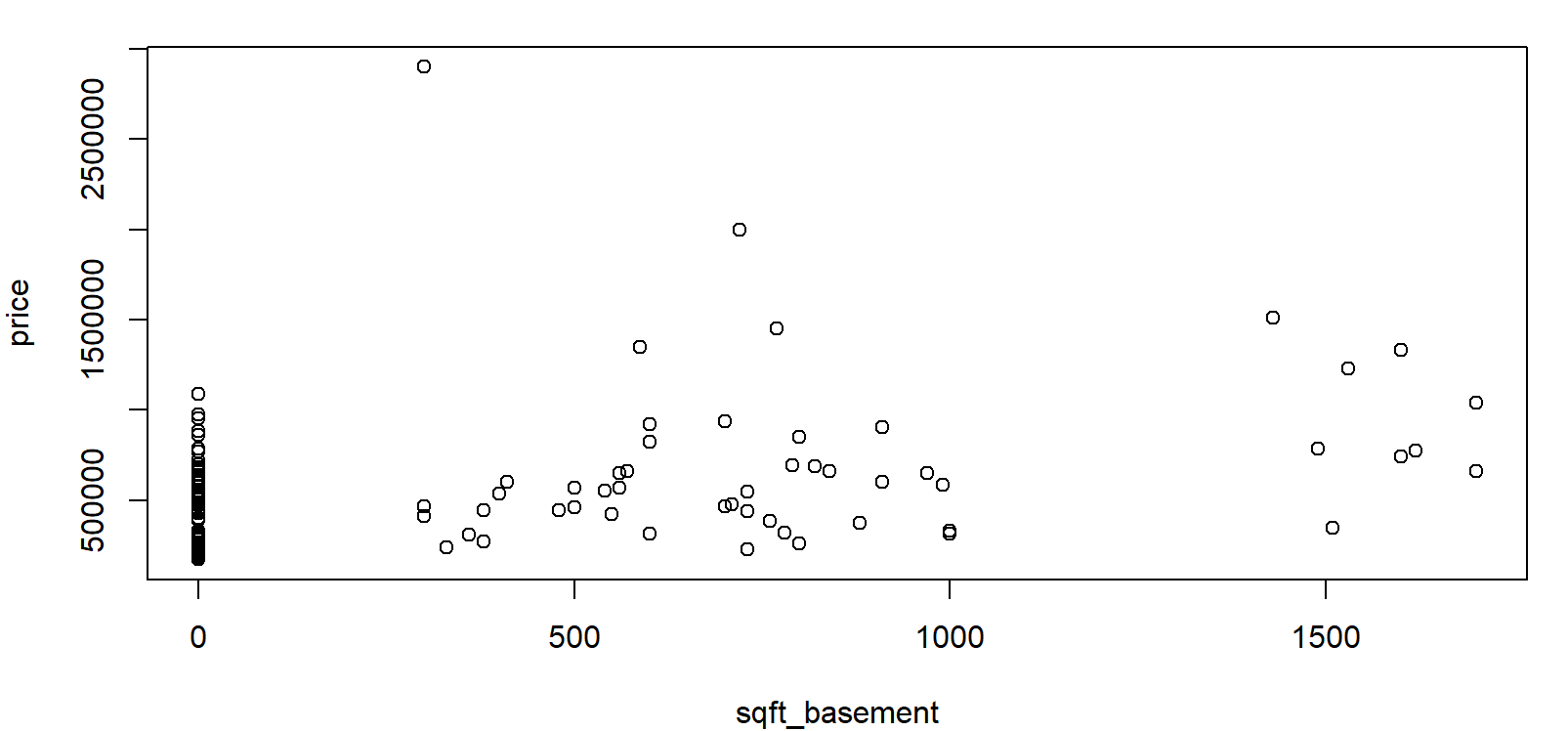
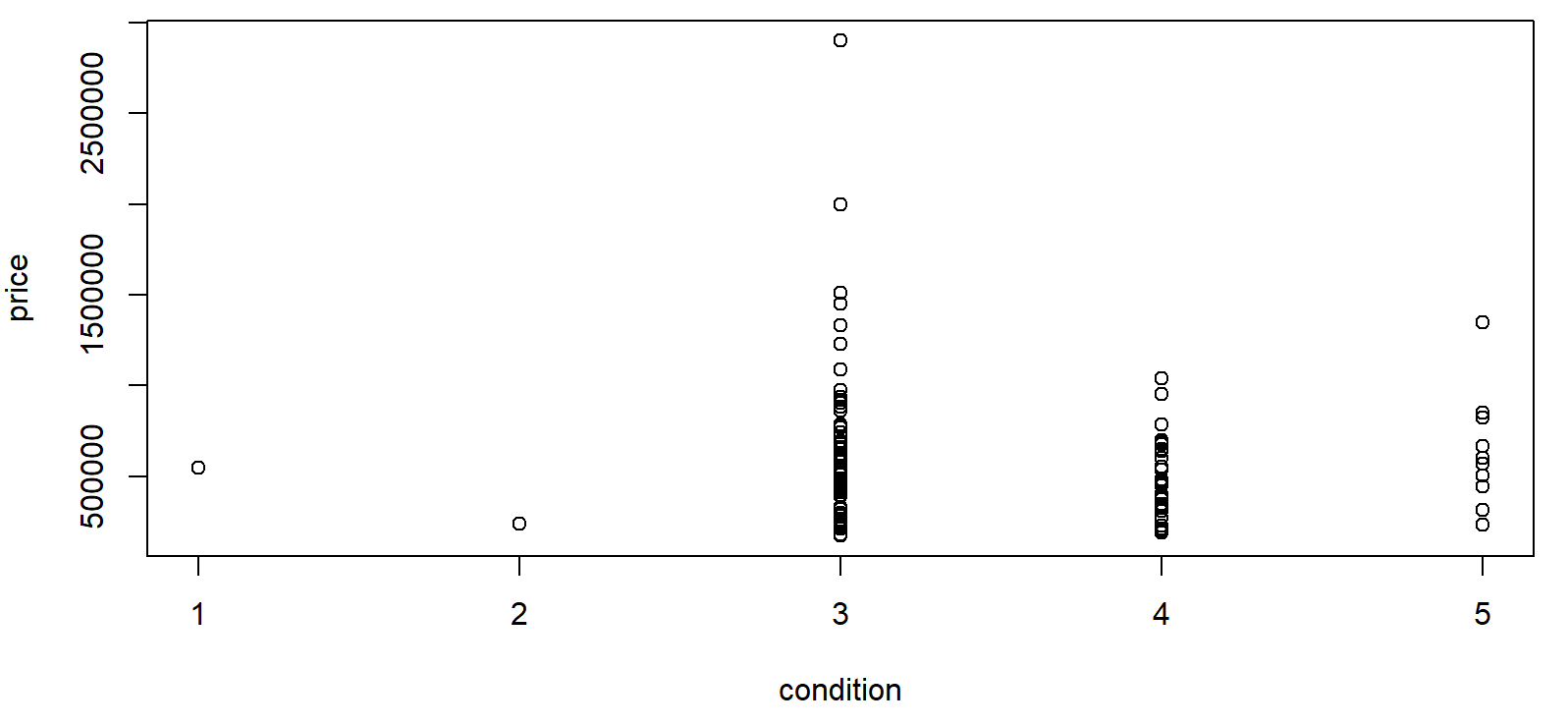
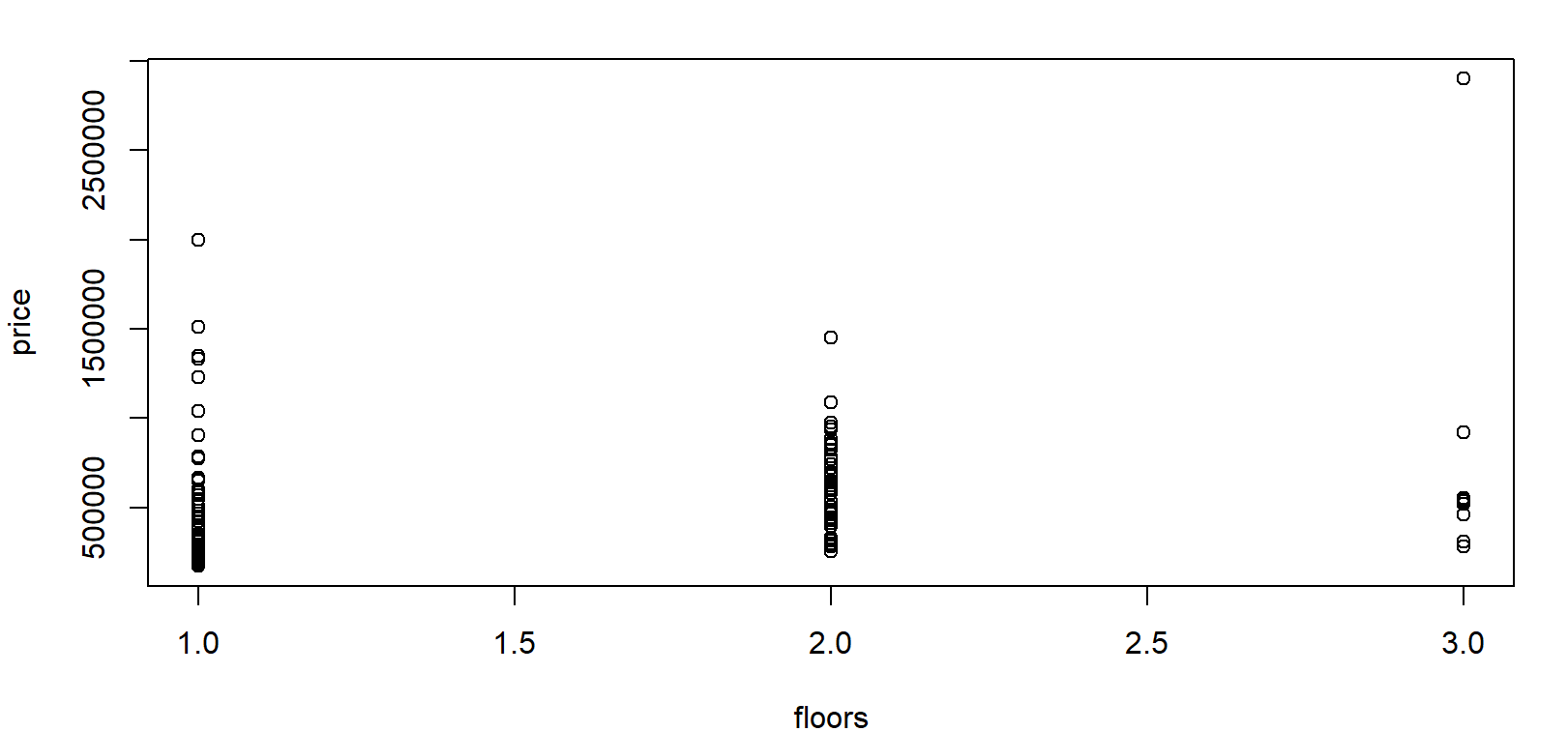
הצלבה בין מספר הקומות 6X לבין המחיר Y:

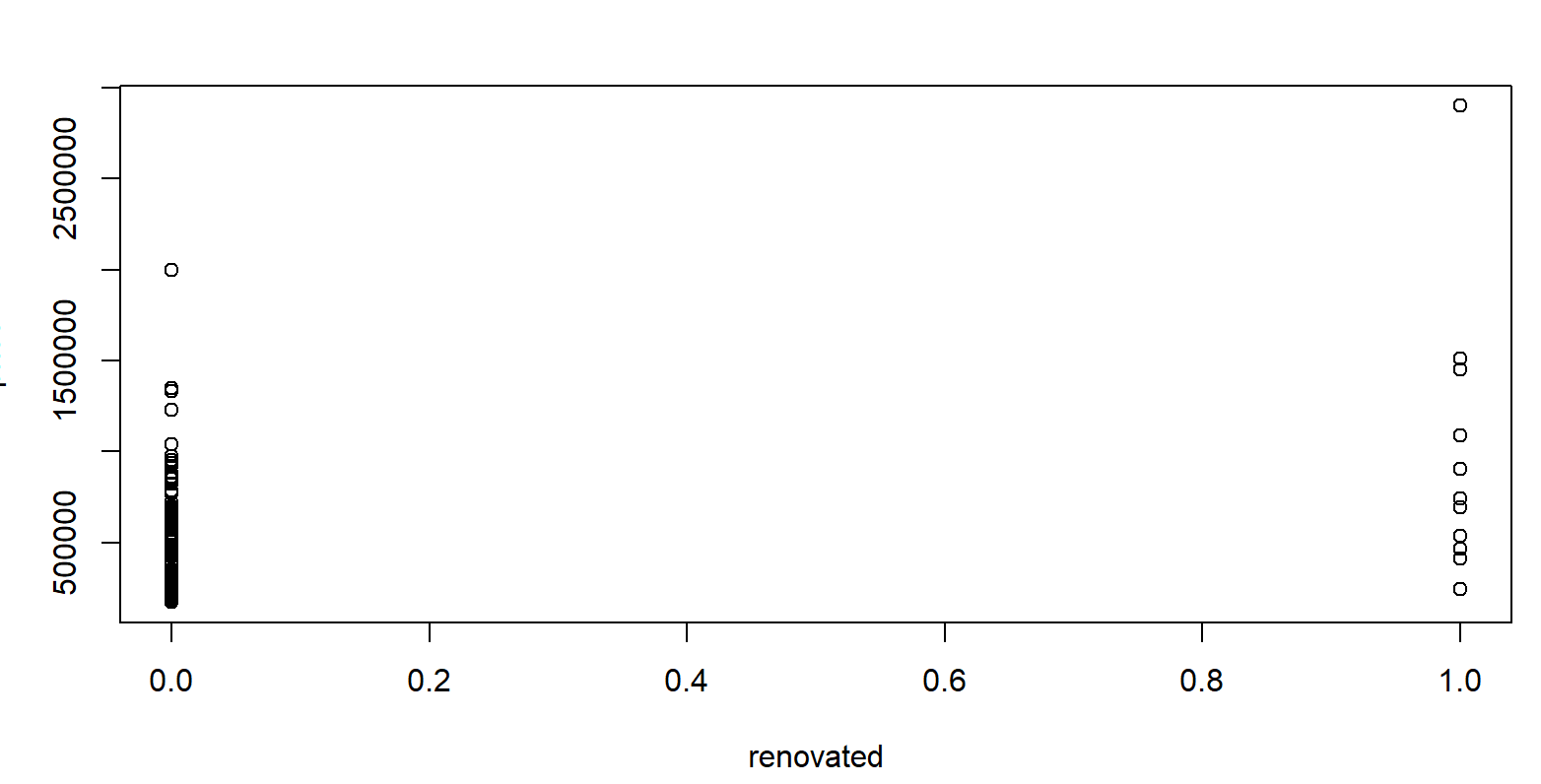
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| מחיר דירה/  מספר קומות | (170k-670k] | (670k-1170k] | (1170k-1670k] | (1670k-2170k] | (2170k-2670k] | (2670k-3170k] |
| 1 | (0.45217%)52 | (0.03478%)4 | (0.03478%)4 | (0.00869%)1 | 0 | 0 |
| 2 | (0.22608%)26 | (0.14782%)17 | (0.00869%)1 | 0 | 0 | 0 |
| 3 | (0.06956%)8 | (0.00869%)1 | 0 | 0 | 0 | (0.00869%)1 |

מתוך הנתונים ניתן לראות התא שבו יש הכי הרבה דירות (כמעט מחצית מהנתונים) הוא דירות עם מחיר נמוך יחסית וקומה אחת, בנוסף רוב הנתונים מתרכזים בכך שיש דירות עם קומה 1 או 2 ומחירן נמוך יחסית. תחילה חשבנו שיש קשר בין מספר הקומות לבין מחיר הדירה אולם פה ניתן לראות באופן מובהק כי אין קשר בין השניים. זאת לפי הנתונים 6% עם דירות של 3 קומות אבל הן זולות באופן יחסי ומתוך כל הדירות היותר יקרות 1170K ומעלה מתוך 5 דירות יש 0 עם 3 קומות ו1 עם 2 קומות שזה אחוז נמוך ביחס לדירות היקרות יותר

**נספחים:**

4.  
  
  


  
  
הקורלציה ביו המשתנים:  
  
 id bedrooms sqft\_living

id 1.00000000 0.16904954 0.1267828

bedrooms 0.16904954 1.00000000 0.4696185

sqft\_living 0.12678275 0.46961853 1.0000000

sqft\_lot 0.08072488 0.03197713 0.4276966

sqft\_basement 0.09070811 0.24284430 0.5857622

floors 0.07578736 0.18977359 0.1853100

condition 0.05153644 -0.07037642 -0.1137537

yr\_built -0.04182751 -0.02539303 0.1835277

renovated 0.27987700 0.15279987 0.2833402

price 0.14891526 0.31154500 0.6796191

sqft\_lot sqft\_basement floors

id 0.08072488 0.09070811 0.07578736

bedrooms 0.03197713 0.24284430 0.18977359

sqft\_living 0.42769664 0.58576224 0.18531004

sqft\_lot 1.00000000 0.32092352 -0.16899463

sqft\_basement 0.32092352 1.00000000 -0.25198811

floors -0.16899463 -0.25198811 1.00000000

condition -0.01161200 -0.01870895 -0.04153861

yr\_built 0.16957236 -0.15958983 0.03765950

renovated -0.03466388 0.19596663 0.12896087

price 0.34059721 0.34942302 0.21199796

condition yr\_built renovated

id 0.05153644 -0.04182751 0.27987700

bedrooms -0.07037642 -0.02539303 0.15279987

sqft\_living -0.11375370 0.18352770 0.28334020

sqft\_lot -0.01161200 0.16957236 -0.03466388

sqft\_basement -0.01870895 -0.15958983 0.19596663

floors -0.04153861 0.03765950 0.12896087

condition 1.00000000 -0.32488627 -0.20801426

yr\_built -0.32488627 1.00000000 -0.31410348

renovated -0.20801426 -0.31410348 1.00000000

price -0.06131727 -0.03384942 0.36923680

price

id 0.14891526

bedrooms 0.31154500

sqft\_living 0.67961911

sqft\_lot 0.34059721

sqft\_basement 0.34942302

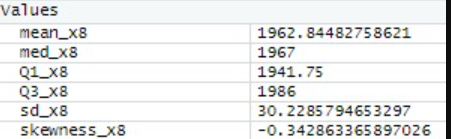
floors 0.21199796

condition -0.06131727

yr\_built -0.03384942

renovated 0.36923680

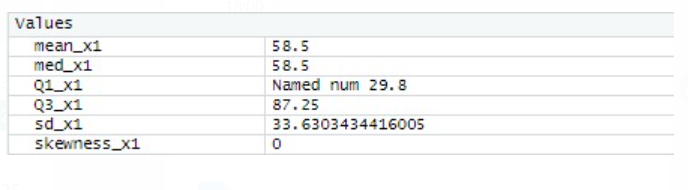
price 1.00000000

5.  
  
  
  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי  
תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי  
 תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

6.  
Y-   
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

X2-  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
X3-  
תמונה שמכילה טקסט, שולחן

התיאור נוצר באופן אוטומטי

X4-  
תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

X5-  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

X6-  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי  
X7-  
תמונה שמכילה שולחן

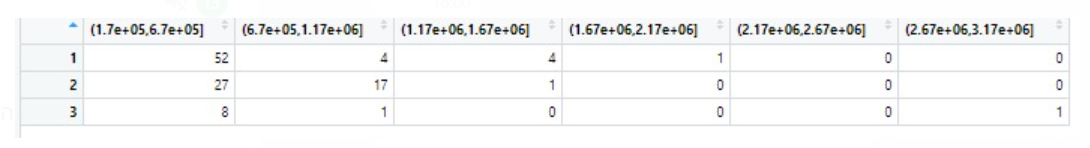
התיאור נוצר באופן אוטומטי

X8-  
תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

X9-  
תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

9.  
תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי