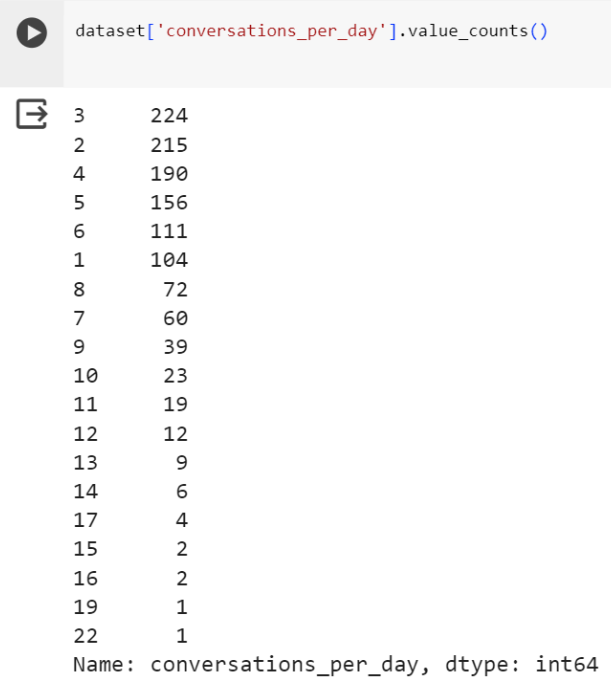
HW1

**שאלה 1:**

טבלת הנתונים מכילה 1,250 שורות ו-26 עמודות.

**שאלה 2:**

****

כיוון שה-data שלנו מכיל מידע על אנשים שחלו, התכונה conversation\_per\_dayעשויה לייצג את כמות השיחות הממוצעת ביום שמנהל אדם שחלה באופן פרונטלי.

התכונה הזו היא תכונה אורדינלית כיוון ש:

1. היא ניתנת לסידור- כלומר ניתן למיין את התכונה לפי סדר כלשהו – למשל לפי סדר עולה.
2. היא בת מניה כיוון שהערכים האפשריים שתכונה זו יכולה לקבל הם בהכרח מספרים טבעים.

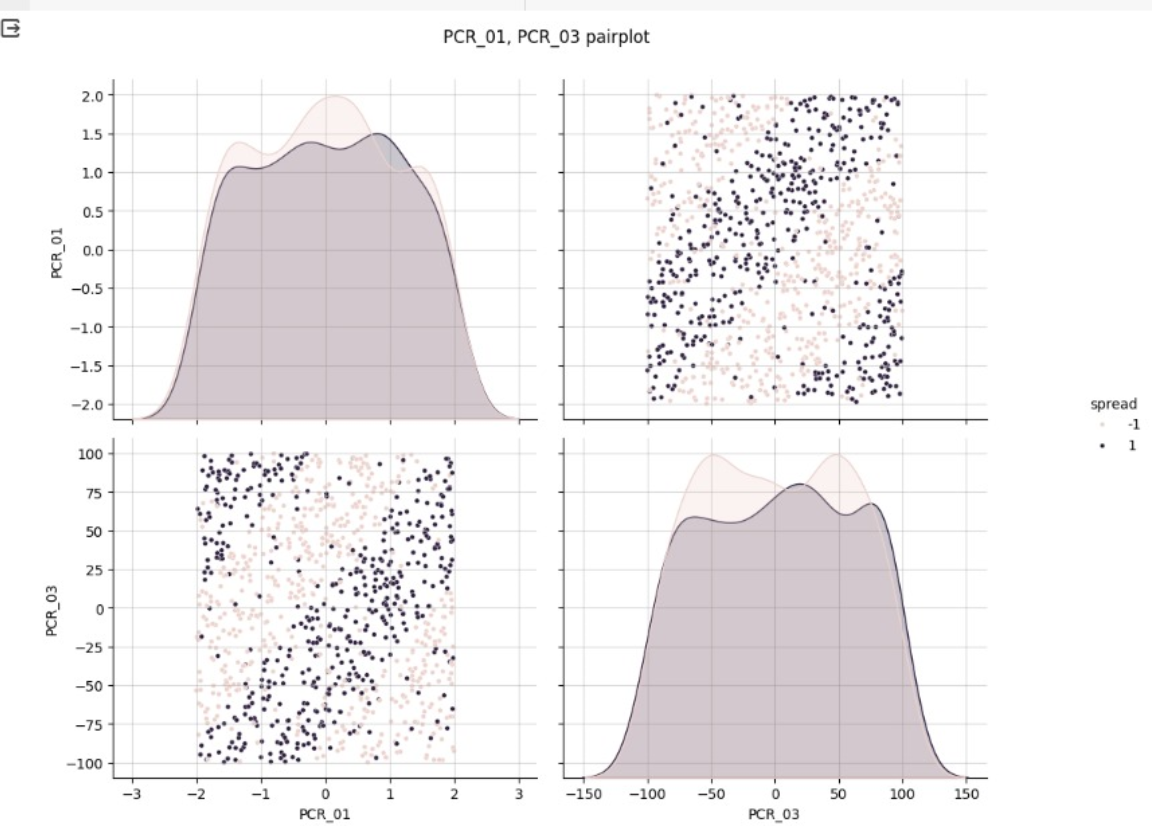
**שאלה 3:**

|  |  |  |
| --- | --- | --- |
| **תכונה** | **תיאור** | **סוג** |
| patient\_id | תעודת הזהות של החולה | other |
| age | גיל החולה | ordinal |
| sex | מין החולה | Categorical |
| weight | משקל החולה | Continuous |
| blood\_type | סוג הדם של החולה | Categorical |
| current\_location | קואורדינטות שמייצגות את מיקומו הנוכחי של החולה | Other |
| num\_of\_siblings | מספר האחים שיש לחולה | ordinal |
| happiness\_score | מספר בין 0 ל-8 המתאר את רמת האושר שהחולה חש | ordinal |
| household\_income | ההכנסות הכוללת של משפחת החולה (ביחידות של 1,000 ₪) | Ordinal |
| conversations\_per\_day | כמות השיחות הפרונטליות אותן מנהל החולה בממוצע ביום | Ordinal |
| sugar\_levels | כמות הסוכר שיש לחולה בדם | Ordinal |
| sport\_activity | כמות הפעמים בהן החולה נוהג מתאמן בשבוע | Ordinal |
| symptoms | מהם הסימפטומים שהחולה חווה | Categorical |
| pcr\_date | התאריך בו ביצע החולה את בדיקת ה-PCR | Ordinal |
| PCR\_01- | תוצאת דגימת ה-PCR הראשונה | Continuous |
| PCR\_02 | תוצאת דגימת ה-PCR השנייה | Continuous |
| PCR\_03 | תוצאת דגימת ה-PCR השלישית | Continuous |
| PCR\_04 | תוצאת דגימת ה-PCR הרביעית | Continuous |
| PCR\_05 | תוצאת דגימת ה-PCR החמישית | Continuous |
| PCR\_06 | תוצאת דגימת ה-PCR השישית | Continuous |
| PCR\_07 | תוצאת דגימת ה-PCR השביעית | Continuous |
| PCR\_08 | תוצאת דגימת ה-PCR השמינית | Continuous |
| PCR\_09 | תוצאת דגימת ה-PCR התשיעית | Continuous |
| PCR\_10 | תוצאת דגימת ה-PCR העשירית | Continuous |

שאלה 4:

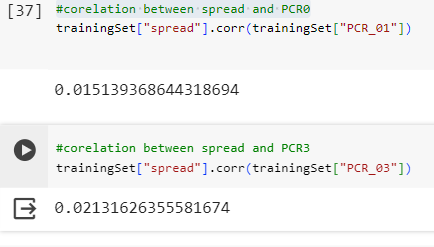
**בסוף**

**שאלה 5:**



התכונות הללו שימושיות בחיזוי ההתפשטות כיוון שהן מאפשרות חלוקה של דיאגרמת הפיזור לתחומים מופרדים כאשר כול תחום מכיל רוב מובהק של דוגמאות מסוג מסוים.

**שאלה 6:**



לפי התוצאות שקיבלנו ניתן לראות שלא קיים קשר לינארי בין הפרמטרים הנבחרים.

התוצאה הזאת לא סותרת את התוצאות בשאלה 5, כיוון שקורלציה מודדת הקשר הלינארי בין פרמטרים- מה שלא מייצג את התלות של משתנה המטרה בשני המשתנים.

**סופי -שלך**