

תרגיל pandas

אתם מקבלים קובץ נתונים על לקוחות שביצעו רכישות שונות בחנות ובאונליין

1. קראו את הקובץ והפכו אותו לטבלה.
2. השתמשו בפונקציה `unique` על העמודה בטבלה בשם: `marital_status` שמתארת את מצבו האישי של האדם (האם חוק או נשוי וכו') והדפיסו את הערכים הקיימים בעמודה הזו
3. על פי התוצאות מהשאלה הקודמת, אם יש שני ערכים שמשמעותם זהה, אחדו אותם לאותו ערך בעזרת הפונקציה `replace`
4. השתמשו בפונקציה `value_counts` עם האופציה של `normalize=True` (שהופכת את כמות הפעמים שכל ערך מופיע לכמות באחוזים) על העמודה `educational_level` כדי לגלות לכמה אחוזים מהאנשים יש תואר שני לפחות.
5. השתמשו ב `loc` וב `isin` כדי לקחת רק את הרווקים מתוך הטבלה. לאחר מכן מיינו את ההכנסה השנתית של הרווקים בסדר יורד בעזרת `sort_values` והדפיסו את זה.
6. הוסיפו עמודה חדשה של סך הרכישות באופן כללי (זאת אומרת הסכום של הרכישות אונליין ובחנות).
7. הוסיפו עמודה של נמוך, בינוני וגבוה לפי מספר הרכישות הכולל שהאדם רכש. אם רכש מתחת 10 - יחשב נמוך, אם בין 10 ל 20 - יחשב בינוני, ואם קנה מעל 20 - יחשב גבוה.
8. הציגו גרף בצורת פאי לפי העמודה של נמוך, בינוני וגבוה. הוסיפו תוויות לכל חלק בפאי.
9. השתמשו ב `groupby` כדי להציג את הטבלה על פי עמודת הנמוך-בינוני-גבוה, והציגו את הממוצע של מספר הרכישות באונליין, בחנות וסך ההכנסה השנתי.
10. הציגו גרף המורכב משני גרפים בעזרת `subplot`: הראשון - ציר ה X יהיה ההשכלה של האדם (הפכו את הערכים למספרים כאשר השכלה בסיסית יהיה 0 ודוקטור יהיה 4) וציר ה Y תהיה ההכנסה השנתית שלו. השני - ציר ה X יהיה ההשכלה של האדם (הפכו את הערכים למספרים כאשר השכלה בסיסית יהיה 0 ודוקטור יהיה 3) וציר ה Y יהיה מס' הרכישות שלו. הוסיפו כותרות.

ניתן להיעזר בתרגיל בקישור הבא: [https://towardsdatascience.com/pandas-in-practice-](https://towardsdatascience.com/pandas-in-practice-64be9228062a)

[64be9228062a](https://towardsdatascience.com/pandas-in-practice-64be9228062a)