



הגורם האנושי באיסוף נתונים – 096275

סמסטר חורף תשפ"ג 2022-2023

תרגיל 1

קראו את הכתבות הבאות:

<https://onlinedegrees.mtu.edu/news/ways-big-data-affects-your-personal-life>

<https://www.forbes.com/sites/forbestechcouncil/2018/05/16/big-data-is-changing-the-way-people-live-their-lives/?sh=54464c1c3ce6>

וצפו בסרטון הבא:

<https://www.youtube.com/watch?v=nogE5tOt3g8>

בחרו באחד מהתחומים הבאים ומצאו עבורו קובץ נתונים רלוונטי

- Entertainment: Music, Shows, Gaming and Movies
- Healthcare and Medical Services
- Shopping and Marketing
- Travel and Transportation
- Public Policy and Safety
- News and Information
- Education and Employment

תוכלו (אבל אין חובה) להיעזר במאגרי הנתונים המפורסמים ב-Kaggle.

ענו על השאלות הבאות לגבי התחום שבחרתם:

1. כיצד ניתן לעשות שימוש בנתוני עתק כדי לשפר את החוויה של אנשים בתחום זה? ספקו 2 דוגמאות.

2. תארו את קובץ הנתונים: התחום שבחרתם, מקור (האיסוף, לא האתר), מספר רשומות (שורות), אלו משתנים כלולים בקובץ (טורים שונים) ומהם הפיצ'רים (, ווג – קטגוריאלי, בינארי, float, int, וכו'), לינק לקובץ.

3. בחרו 3 משתנים ותארו אותם כמותית (ממוצע, סטיית תקן, ערכים שונים) אם ניתן, אם לא דווחו ערך שכיח (בכמה אחוזים מופיע), ערך נדיר, מספר ערכים שונים.
4. המשיכו באקספלורציה של קובץ הנתונים, הציגו 3 גרפים מעניינים לדעתכם המתארים תופעה מתוך קובץ הנתונים. חשוב לכלול כותרות לכל גרף ושמות לצירים.
5. נסחו שאלה שעליה תוכלו לענות בעזרת קובץ נתונים זה. קשרו בין השאלה לגרפים שהצגתם (או חלק מהם).
6. תנו דוגמא לשאלת המשך שלא ניתן לענות עליה בעזרת קובץ הנתונים הקיים.

הוראות הגשה

1. יש להגיש את העבודה בקובץ PDF
2. היקף העבודה לא יעלה של שני עמודים.
3. ההגשה בזוגות בלבד! (רק אחד מבני הזוג מגיש).
4. ההגשה במודל בלבד עד השעה 23:00 ב 30.10.22
5. שם קובץ ההגשה הינו HW1_ID1_ID2

בהצלחה!