**תרגיל – feature engineering ולמידת מכונה**

תרגיל זה הוא המשך לתרגיל הקודם ונוגע לאותו נושא ולאותם מקורות מידע.

לאחר שביצעתם את תהליך ניקוי המידע נאמנה והעברתם בעיות של חוסר במידע ללקוח, יש כעת ברשותכם מסד נתונים נקי ואיכותי.

כעת עליכם לבצע את המשימה אליה נקראתם: דירוג החשודים לפי סיכויים להיות טרוריסטים (בראש הרשימה יופיעו חשודים בעלי סיכוי גבוה יותר להיות טרוריסטים).

**המשימה**

השתמשו ב-train\_data המתוייג ובכל מקורות המידע האחרים (db1 עד db5) בכדי ללמוד מודל פרדיקציה עבור כל חשוד האם הוא טרוריסט או לא (מודל שנותן הסתברות ולא רק תיוג).

השתמשו במודל הנ"ל כדי לדרג את החשודים.

השתמשו ב-labels הנתונים לכם על ה-test data בכדי להעריך את איכות הדירוג שלכם (על ידי הסתכלות על ה-roc ו-auc).

נסו לייצר פיצ'רים המבטאים את הידע העסקי שנמסר לכם על הבעיה, ובנוסף יצרו פיצ'רים יצירתיים נוספים שמבטאים את ההיגיון שלכם לגבי הבעיה.