**משימת הקלאסטרינג**

מה שהתבקשנו: "בסעיף זה עליכם לקלסטר את החברות המפעילות את הטיסות על פי הדמיון ביניהן במדיניות התמחור. בסעיף זה תרצו לסנן החוצה מהדאטה את הטיסות שאינן מופעלות על ידי אותה חברה   
בהלוך ובחזור. הציעו דרכים שונות להציג את הקלאסטרים של החברות."

הסבר: תחילה נסנן רק את הטיסות שהחברת הלוך והחברת חזור בה הן אותו חברה. לאחר מכן, נבין אילו מדיוניות תמחור שונות קיימות עבור כל סוגי החברות (זיהוי **הגורמים המאפיינים** את מדיניות התמחור של כל חברה). לאחר מכן נבצע אלגוריתם K-means כאשר כל נקודה תייצג חברה וכל קלאסטר יהווה מדיניות תמחור.

איך נעשה את זה בפועל: נסנן את המאגר שלנו לפי הדרישה. כעט, נבדוק את נתחי החברות שקיבלנו ונוריד את אלו שמהווים פחות מאחוז 1 מהטיסות הרלוונטיות (לא שוות התעסקות). עבור כל חברה בנפרד, נמצא סטטיסטיקות ונתונים שונים כדי להבין את מדיניות התמחור (לדוגמא ממוצע עלות טיסה ללא כבודה כנגד עם כבודה, מחיר כנגד משך זמן טיסת ההלוך/חזור, מחיר כנגד שעת המראה (בוקר/צהריים/ערב) וכו...). בעזרת הנתונים והחישובים ננסה להבין את מדיניות התמחור של אותה חברה (יתכן שיש כמה). לאחר מכן, נרכז את כל הגורמים ל - df כאשר כל שורה מייצגת חברה וכל עמודה מאפיין תמחור. נשתמש באלגוריתם כדי לחלק את החברות לקלאסטרים. בעזרת הקטנת מיימדים (PCA או TSNE) נייצג את הקלאסטרים כך שכל נקודה תהוו חברה והצבע יציין לאיזה קלאסטר היא שייכת.