חלק 3 – הגדרה ומימוש ETL –

**Extract**: שלב חילוץ הנתונים כולל שליפת מערך הכושר ממקורותיו. מקורות אלו יכולים לכלול מסדי נתונים, גיליונות אלקטרוניים, סקרים או כל מאגר נתונים אחר שבו מאוחסנים נתוני הכושר. תהליך המיצוי אוסף את הנתונים הגולמיים הקשורים לתכונות הכושר, כגון שם, מין, גיל, העדפות פעילות גופנית, רמת כושר, הרגלי אכילה וגורמי מוטיבציה.  
במקרה שלנו, הדאטה סט נאסף מאתר "קאגל", בצורת גיליון אקסל אלקטרוני שמכיל נתונים גולמיים הקשורים לעולם הכושר.

**Transform**: לאחר חילוץ הנתונים, הם עוברים שינויים שונים כדי להבטיח את האיכות, העקביות והתאימות שלהם. השינויים הספציפיים שלנו למערך הכושר כוללים:

ניקוי ואימות: הסרת רשומות לא רלוונטיות או חלקיות, תיקון שגיאות או חוסר עקביות בנתונים ואימות שלמות הנתונים.

סטנדרטיזציה: הבטחת פורמטים ויחידות עקביים עבור שדות כגון גיל או זמן.

גזירת משתנים חדשים: חישוב מדדים נוספים על סמך נתונים קיימים.

טרנספורמציות אלו עוזרות להכין את הנתונים להמשך ניתוח ודיווח.

**Load**: הנתונים שעברו טרנספורמציה נטענים לאחר מכן אל היעד, שיכול להיות מחסן נתונים, מסד נתונים או כלי אנליטי. שלב הטעינה כולל:

הגדרת פורמט היעד: יצירה או עדכון של המבנה של מסד הנתונים של היעד או מחסן הנתונים כדי להתאים לנתוני הכושר שעברו שינוי. זה כרוך בהגדרת הטבלאות, העמודות וסוגי הנתונים המתאימים.

מיפוי: מיפוי שדות נתוני הכושר שהשתנו לתכונות המתאימות בפורמט היעד. לדוגמה, מיפוי השדה "שם" לעמודה "שם" בטבלת היעד.

טעינת הנתונים: ביצוע פעולת הטעינה בפועל, כגון הכנסת הנתונים שעברו טרנספורמציה למסד הנתונים היעד או למחסן הנתונים.

על ידי ביצוע תהליך ETL זה המותאם למערך הכושר המסופק, ניתן לחלץ את הנתונים, לשנות אותם על ידי יישום ניקוי, סטנדרטיזציה ולאחר מכן לטעון ליעד יעד להמשך ניתוח, דיווח או קבלת החלטות הקשורות לכושר. ותובנות הקשורות לתרגיל.