**חלק 3 – הגדרה ומימוש ETL**

**1**.  
  
**Extract:**

חילוץ הנתונים הגולמיים אשר מופקים מקובץ CSV, המכיל את פרטי עסקאות המכירה.

זיהוי מקור: הגיליון הספציפי בתוך קובץ CSV המכיל את הנתונים הרלוונטיים מזוהה, והנתונים נשלפים מאותו גיליון.

**Transform:**

ניקוי נתונים: שלב זה כולל הסרת כפילויות, טיפול בערכים חסרים והבטחת שלמות הנתונים.

עיצוב נתונים: וידוא כי השדה "InvoiceDate" הוא בפורמט אחיד DD/MM/YYYY   
(פורמט נפוץ לצורך עקביות וקלות ניתוח).

המרת נתונים: המרת "Quantity" מסוג מחרוזת לסוג int.

מיפוי נתונים: מיפוי עמודות או שדות בנתונים שחולצו לשדות המתאימים בסכמת הנתונים. זה מבטיח שהנתונים שעברו טרנספורמציה יתיישרו עם הסכימה שהוגדרה.

**Load:**

יצירת Data Warehouse יעד: נוצר כדי לאחסן את הנתונים שעברו שינוי.   
מבנה הסכימה, כגון סכימת ה-Star במקרה שלנו, מיושם ב- Data Warehouse בכדי לייעל את השאילתות והניתוח.

טעינת נתונים: הנתונים שעברו את הטרנספורמציה והניקוי נטענים ל-Data Warehouse. זה כולל מיפוי של שדות הנתונים שעברו טרנספורמציה לעמודות המתאימות בסכימת היעד.

על ידי ביצוע תהליך ETL זה, ניתן לחלץ בהצלחה את נתוני עסקאות המכירה מקובץ ה-CSV, להמיר אותם ולטעון אותם לתוך ה- Data Warehouse. זה מאפשר ניתוח יעיל, דיווח וקבלת החלטות על בסיס נתונים נקיים ומובנים.