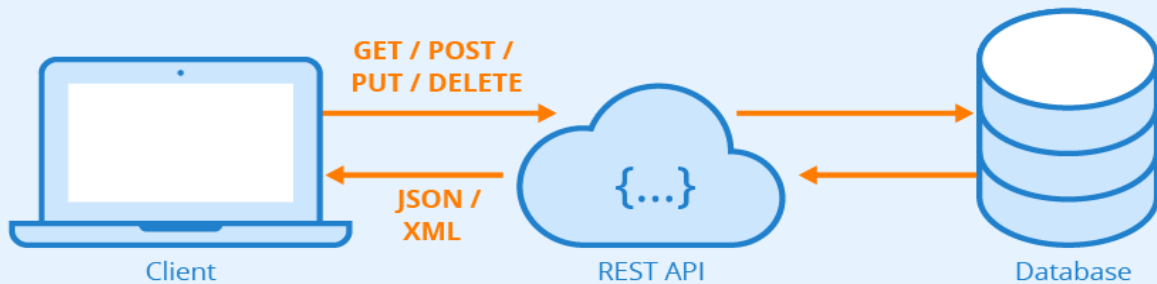


# הרכשת נתונים - API



שלבי העבודה של מדען הנתונים:

- 1) תכנון - לאחר שכבר העלנו את שאלת המחקר, עלינו להגדיר אילו נתונים נזדקק להם. כלומר, אילו נתונים אנחנו רוצים שיהיו בטבלה שלנו כדי שישמשו אותנו לאנליזה, הניתוח.
- 2) הרכשת הנתונים – בשלב הזה אנחנו צריכים להשיג את הנתונים, לשם כך אנחנו נלמד 2 שיטות.

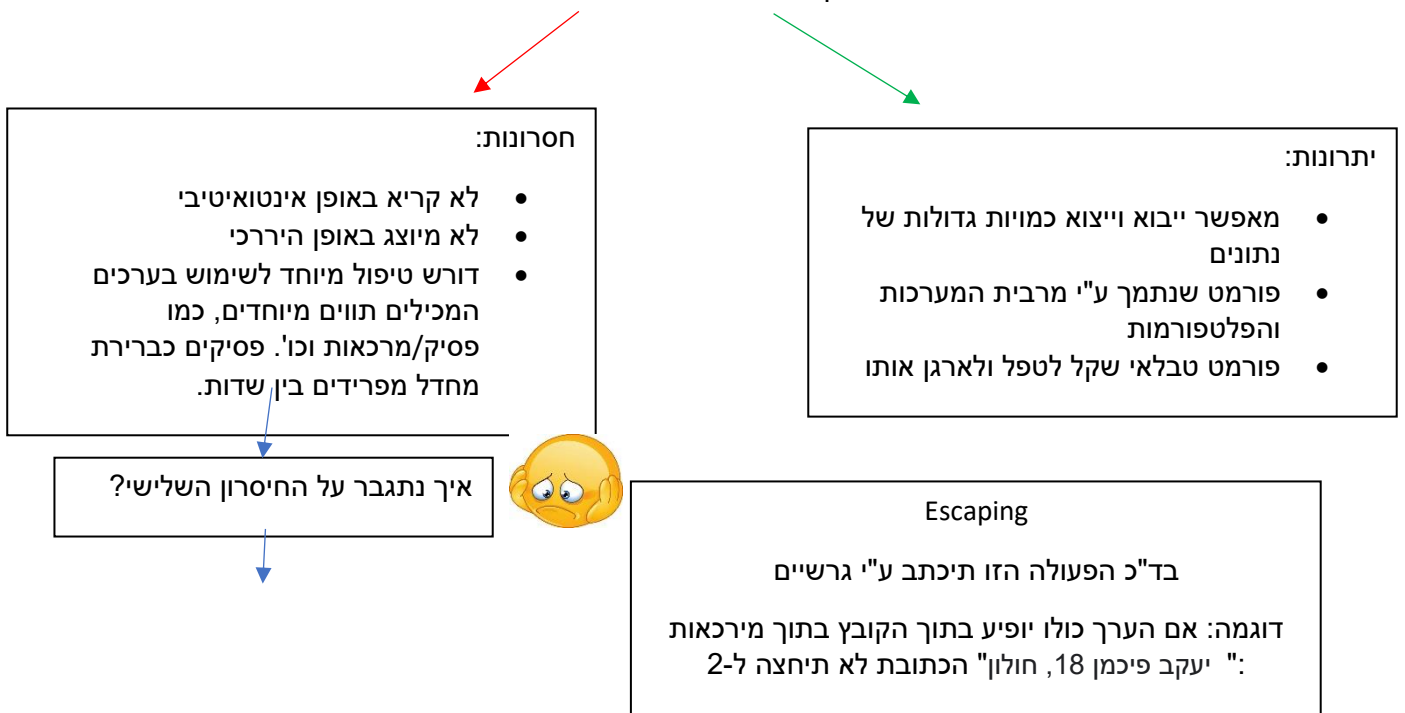
\* API - Application programming Interface

ממשק שתוכנן מראש לצורך הרכשת הנתונים, הוא מאפשר גישה פשוטה לנתונים של גופים שמעוניינים בחשיפה כמו למשל: פייסבוק, IMDB, TWITTER

Crawling - "זחלן רשת". דרך מתוחכמת אך מורכבת יותר להשגת הנתונים

- 3) שמירת הנתונים- קבצי CSV כמו Excel שכבר ראינו, או Jason , קובץ היררכי.

לקבצי CSV יש יתרונות וחסרונות



פורמט Jason :

- השימוש העיקרי בו הוא העברת אינפורמציה בין מערכות.
- מאפשר להעביר מבנים מורכבים
- קריא

בנוי מזוגות סדורים של key ו- value .

Key : תמיד מוצג בתור מחרוזת בתוך גרשיים  
Value : יכול להיות מחרוזת/ מ'ס רשימה אובייקט