

## Homework 2 – Business Intelligence

### חלק 1 – הגדרת Data Warehouse

1. מתוך הסכמות: Star, Snowflake, Galaxy, בחרו סכמת נתונים עבור הדאטה שבתחילת בית הקודם. הסבירו במשפט אחד מדוע בחרתם בסכמה זו עבור אוסף הנתונים שלכם.
2. תארו את ה Data Warehouse לאחר שילוב הסכימה. (ניתן לבצע סעיף זה באמצעות ERD).
3. ממשו את ה Data Warehouse באמצעות אחד מהכלים הבאים: Excel, SQL. הוסיפו להגשה את הטבלאות החדשות (data-sources, fact table/s)
4. הציגו use-case אחד המתאר מדוע הבחירה שלכם היא יעילה ועוזרת בקבלת החלטות למשתמש. למשל: הפחתת joins, שליפה מהירה יותר, וכו'.
5. הוסיפו להגשה את ה DW שיצרתם (טבלאות sql / אקסל)

### חלק 2 – הגדרה ומימוש ETL

1. הגדירו את תהליך ה ETL עבור אוסף הנתונים שבחרתם. עבור כל שלב, E-T-L, הסבירו במשפט אחד מה ברצונכם לבצע.
2. הגדירו את תהליך ה ETL Pipeline (6 שלבים) והסבירו מה נדרש לבצע בכל אחד מהשלבים עבור אוסף הנתונים שבחרתם.
3. ממשו את תהליך ה ETL Pipeline באמצעות אחד מהכלים הבאים: Excel, Python, SQL.

### חלק 3 – הבנת הנתונים

- השתמשו בכלי לניתוח סטטיסטי פשוט (למשל אקסל, R, python) והציגו תחקור נתונים עבור ה Data Warehouse שיצרתם (לא ה data-sources, אלא ה fact-tables שיצרתם).
1. מדדי פיזור של התכונות (ממוצע, סטיית תקן, רבעונים)
  2. תלויות וקשרים (קורלציה)
  3. כל מדד סטטיסטי נוסף אשר יכול לעזור לכם לקבל הבנה טובה יותר על הנתונים.