

פרויקט - בינה עסקית

I. שאלה עסקית + KPIs

בשלב ראשון, נרצה להבין מה היא מטרת הפרויקט ואיזה תובנה עסקית נרצה לקבל ממנה. למשל, חברה העוסקת בפרסומות במדיה השונה, אשר אוספת נתונים על הפרסומות ורוצה לדעת "מה היא הפרסומת הרווחית ביותר?".

- המטרה העיקרית בשלב זה הוא לסמן את הנקודות החשובות, כך שבמהלך התהליך נוכל להתכנס ולענות על נקודות אלו. כתבו **פסקה אחת** המתארת את מטרת הפרויקט.
- הגדירו לפחות שני KPIs עבור השאלה העסקית.

II. הגדרת Data Warehouse

1. מתוך הסכמות: Star, Snowflake, Galaxy, בחרו סכמת נתונים עבור הדאטה שבחרתם. 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. הציגו שני use-case המתארים מדוע הבחירה שלכם היא יעילה ועוזרת בקבלת החלטות למשתמש. למשל: הפחתת joins, שליפה מהירה יותר, וכו'.

III. תהליך הETL

- הסבירו בקצרה כל אחד משלבי הETL בפרויקט שלכם.
- הציגו את מיפוי העמודות משכבת הODS אל DataWarehouse. תשובתכם צריכה לכלול גם את פעולות הTransform עבור כל עמודה.
- ממשו תהליך זה ע"י STTM במתודולוגיית תכנות מונחה עצמים (כמו במטלה 2).
- הערה: יש לממש את כלל הטבלאות - טבלת Fact וכל המימדים.
- שמרו את הData Warehouse שהתקבל בקבצי excel/csv.

IV. ניתוח הData Warehouse

כתבו (לפחות) 10 שאילתות SQL אשר מטרתם לחלץ מידע מה Data Warehouse שבניתם ועונות על השאלה העסקית אשר הוגדרה בחלק 1. השאילתות נדרשות להיות יעילות הן מבחינת זמן ריצה והן מבחינת ניהול זיכרון.

V. מסקנות

רשמו 3-5 מסקנות עבור חלק 4 (תוצאות השאילתות).

VI. ניהול גרסאות

חבר צוות אחד יבצע את המשימות הבאות:

1. צרו repo חדש בשם "BI Project".
2. העתיקו את repo לסביבה מקומית (git clone).
3. תחת הקובץ README.md, הוסיפו את שמות חברי הקבוצה.
4. הוסיפו אל הפרויקט את קבצי ההגשה של מטלה 1 ואוסף הנתונים שלכם.
5. שמרו את השינויים והוסיפו אותם אל staging area בעזרת git add.
6. בצעו commit של השינויים הנמצאים בstaging area.
7. בצעו push ובדקו שאכן ה remote branch ב GitHub התעדכן.
8. לחצו על לשונית "settings" ולאחר מכן על "collaborators" בתפריט מצד שמאל.
9. הוסיפו את חברי הקבוצה כשותפים לפרויקט זה.

שאר חברי הקבוצה:

1. בדקו שקיבלתם הזמנה (דרך המשתמש GitHub שלכם) לפרויקט.
2. צרו branch חדש עם השם שלכם.
3. פתחו את סביבת העבודה והריצו את הפקודה git clone עם הקישור לפרויקט.
4. הריצו את פקודת checkout והחליפו את סביבת העבודה לbranch שיצרתם בסעיף 2.
5. בדקו שכל הקבצים אכן נמצאים בפרויקט.
6. בצעו שינוי כלשהו בקובץ README.md.

7. שמרו את השינויים והוסיפו אותם אל ה staging area בעזרת git add.

8. בצעו commit של השינויים הנמצאים ב staging area.

9. בצעו push ובדקו שאכן ה remote branch ב GitHub התעדכן

VII. Final Report

יש להגיש את הפרויקט בליווי דו"ח מסכם. ניתן לעשות זאת במקביל ב jupyter notebook, ולהשתמש בתאי markdown כדי לכתוב טקסט פשוט. הדו"ח צריך להכיל 5 חלקים, עפ"י שלבי פרויקט זה.

בהצלחה !