

Homework 2 – Business Intelligence

חלק 1 – פתיחת פרויקט משותף ב-GitHub

חבר צוות אחד יבצע את המשימות הבאות:

1. צרו repo חדש בשם "BI Project".
2. העתיקו את הrepo 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 ב-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 staging area בעזרת git add.
8. בצעו commit של השינויים הנמצאים בstaging area.
9. בצעו push ובדקו שאכן ה remote branch ב-GitHub התעדכן.

חלק 2 – הגדרת Data Warehouse

1. מתוך הסכמות: Star, Snowflake, Galaxy, בחרו סכמת נתונים עבור הדאטה שבחרתם בתרגיל 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. הציגו use-case אחד המתאר מדוע הבחירה שלכם היא יעילה ועוזרת בקבלת החלטות למשתמש. למשל: הפחתת joins, שליפה מהירה יותר, וכו'.
4. הוסיפו את חלק זה ל Git של הפרויקט.

חלק 3 – הגדרה ומימוש ETL

1. הגדירו את תהליך ה ETL עבור אוסף הנתונים שבחרתם. בתשובתכם, יש להסביר את הפעולות המתרחשות בשלב ה Extract, Transform, Load.
2. ממשו את תהליך ה ETL שלכם ע"י STTM במתודולוגית תכנות מונחה עצמים. כפי שהוסבר בשיעור, הוסיפו למחלקה Database את האפשרויות הבאות:
 - a. משתנה מחלקה מסוג מילון המאחסן את הערך המוחזר מ `get_data_source_target_mapping`.
 - b. פונקציות אשר מחליפות את המימוש ה"דני" של STTM לפונקציות חיצוניות:
 - i. `add_source` – מקבלת את הפרמטרים המתאימים ומוסיפה את ה source אל המשתנה הגלובלי.
 - ii. `add_destination` – כנ"ל.
 - iii. `add_transform` – כנ"ל.
 - iv. `add_mapping` – כנ"ל.
 - c. ניתן להוסיף מחלקות / פונקציות כרצונכם.
3. הריצו ובדקו את תהליך ה ETL.
4. הערה: אין צורך בשלב זה לאכלס את כל ה Data Warehouse אלא את טבלת Fact בלבד.
4. הוסיפו את חלק זה ל Git של הפרויקט.

הגשה:

יש להגיש קובץ txt בלבד המכיל קישור ל GIT של הפרויקט.