README

מבוא:

בתרגיל זה אנו מנתחים דאטה סט של תיאורי משרות, אשר מספק אוסף מקיף של פרסומי דרושים, הניתוח מתבצע באמצעות שאילתות SQL על הדאטה סט כאשר כל שאילתה מחזירה כתוצאה דאטה סט "קטן" המספק תובנות על מצב שוק העבודה בהתאם לשאלה שנשאלה.

בתרגיל זה אנו עובדים באמצעות פייתון עם DuckDB אשר בתוכו אנו מייצרים טבלאות.

הטבלאות שיצרנו בתוך DuckDB:

1. טבלה עבור הדאטה סט המקורי
2. טבלה המכילה 500 שורות מדגמיות מהדאטה סט המקורי
3. טבלה עבור כל דאטה סט קטן המתקבל מכל תוצאה של שאילתה

לאחר מכן באמצעות קוד פייתון אנו מעתיקים רק את הטבלאות הקטנות (סעיפים 2 ו-3) אל תוך קובץSQLITE המוכר.

לאחר מכן באמצעות אנו טוענים את הטבלאות ששמרנו בקובץ SQLITE ובאמצעות הכלי streamlit אנו מייצרים דאשבורד אשר יהווה ממשק ויזואלי ואינטראקטיבי עבור המשתמש, בדאשבורד יוצגו השאילתות והתובנות שלמדנו מהשאילתות וכן גרפים הממחישים את הנתונים המתקבלים מהשאילתות כמו כן מוצג עבור המשתמש גם הטבלה המכילה את 500 השורות המדגמיות מתוך הדאטה סט המקורי

תיאור הקבצים המצורפים:

קובץ word "תרגיל גדול מסכם" :

הקובץ עונה על כל סעיפי התרגיל ומכיל בתוכו לינק לדאטה סט שאיתו אנו עובדים, תיאור הדאטה סט , הסברים על השאילתות ששאלנו, תובנות , צילומי מסך של הדאשבורד ותיאורים סכמתיים (ארכיטקטורת התרגיל, מידול הנתונים כתצוגת GRAPH DB)

קובץ csv “500\_rows\_job\_description.csv”:

קובץ זה מכיל 500 שורות מהדאטה סט המקורי (ללא קשר לדאטה סט עם 500 השורות המדגמיות שלקחנו מהדאטה סט ויצרנו עבורו טבלה)

קבצי Python:

קובץ פייתון duckdb\_sqlite.py: מריץ את החלק הראשון של התרגיל מבחית הרצת השאילתות על הדאטה סט וטעינת טבלאות לתוך DuckDB , העתקת הטבלאות הקטנות לתוך קובץ SQLITE .

קובץ פייתון dashboard\_plots.py: קובץ פייתון אשר מקבץ את כל הפלוטים והויזואליזציות שבחרנו לכל הנתונים שהתקבלו מהשאילתות

קובץ פייתון dashboard\_jobs\_data: קובץ פייתון אשר מציג את חלונות הדאשבורד עבור המשתמש וכן הוא גם משתמש בקובץ dashboard\_plots.py על מנת לקרוא לויזואליזציות המתאימות לכל חלון

קובץ SQLITE:

הקובץ בו שמרנו את הטבלאות הקטנות ( 500 השורות המדגמיות ותוצאות השאילתות )

קובץ requirments.txt:

קובץ המתאר אילו חבילות בפייתון צריך להוריד על מנת שהקבצי פייתן שהגשנו ירוצו

קובץ README:

קובץ המספק הסבר כללי על התרגיל וקבצים שהגשנו בתוך הקובץ ZIP