Business Intelligence

**Homework 1 – Group AB**

מגישים:

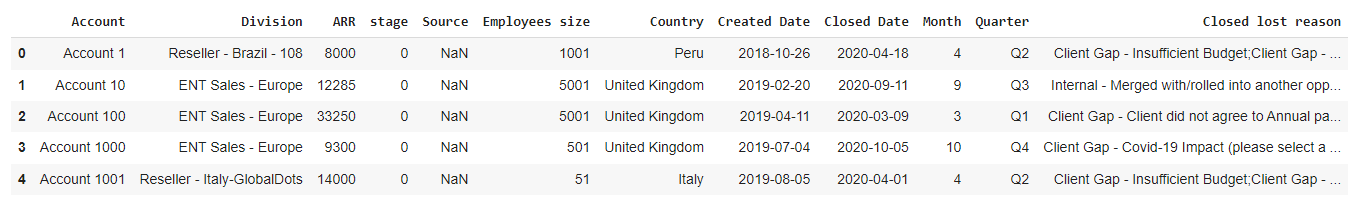
יואב ברק – 315311993

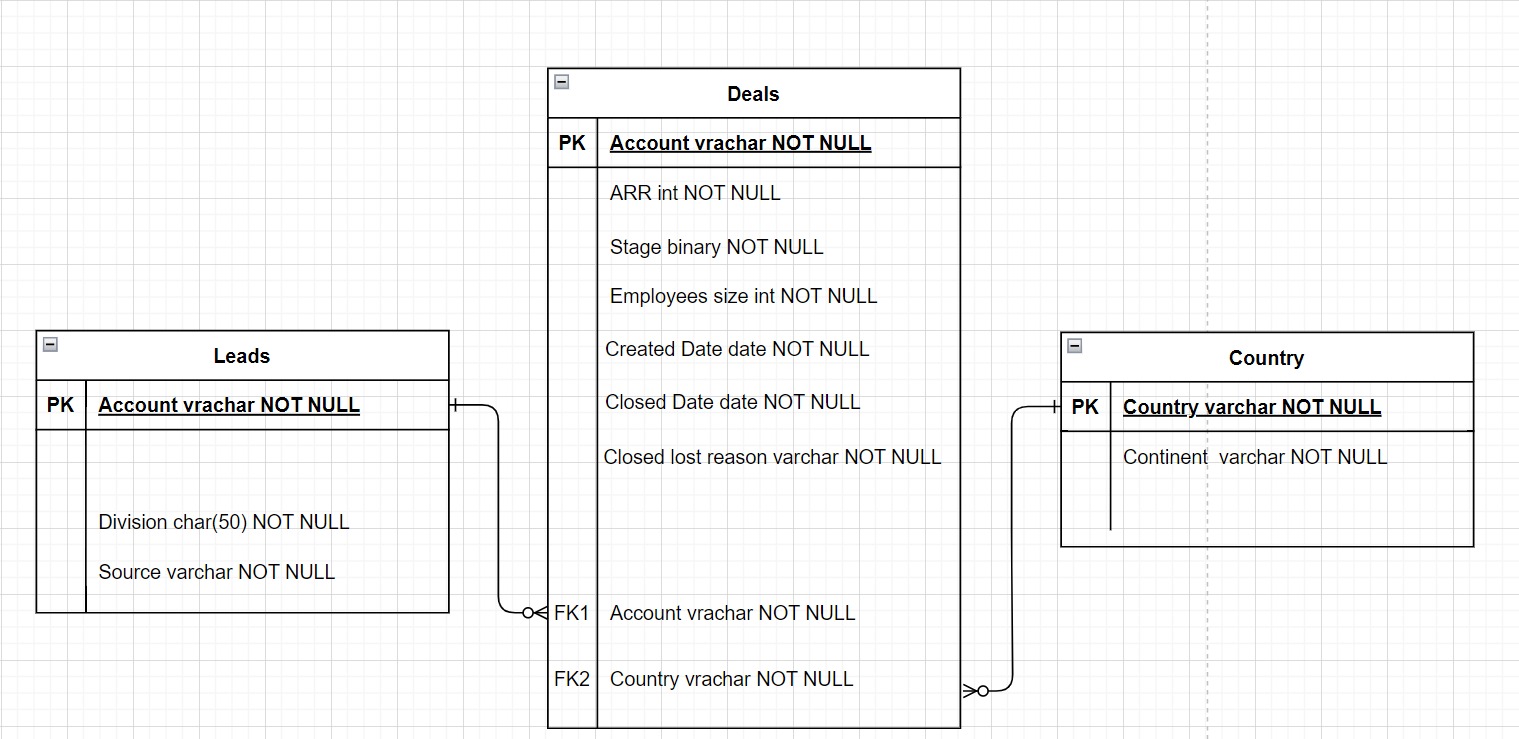
ולדיסלב פשניצ'ני – 317152569

חלק א'- הגדרת Data Warehouse

הדאטה שבחרנו במטלה 1 מציג עסקאות שבוצעו בחברה מסוימת.

1. על מנת להגדיר את ה Data Warehouse נבחר בסכמת כוכב (Star Schema) מכיוון שנרצה לחלק את הנתונים כך שטבלת העסקאות תהיה במרכז הכוכב. כך נוכל לשלוף מידע עסקי בצורה מהירה יותר וללא צורך בחיבורים מורכבים.

תזכורת- דגימת 5 רשומות מתוך מאגר הנתונים:

1. תיאור Data Warehouse באמצעות ERD:
2. -USE CASE נרצה להציג למנהל הכספים את התפלגות הרווחים לפי יבשות.

SELECT distinct continent ,SUM(arr)

OVER (partition by continent) as ContinentRev

FROM SMWBDeals SD

JOIN SMWBCountry SC ON SD.country = SC.Country

order by continentrev DESC

בגלל החלוקה של הנתונים נוכל לשלוף את הנתון בצורה מהירה ויעילה יותר מאשר

הייתה לנו רק טבלה אחת מכיוון שכמות המידע קטנה ובעקבות כך זמן הריצה יקטן .

מצורף קובץ SQL לדוגמא של השליפה מה DW.

1. **הגדרה ומימוש ETL**

* Extract- הנתונים בפרויקט ניתנו בפורמט אחיד של 3 טבלאות אקסל.

היבטים חשובים –הבנת מודל הנתונים של הקבצים הרלוונטיים והבנת תלות בין מקורות שונים העשויים להשפיע על פלט תהליך הוצאת הנתונים.

* Transform- איחוד הנתונים אשר ניתנו בפורמט של 3 טבלאות אקסל לDB אחד.

בנוסף המרנו את עמודת stage מטקסט לבינארי (0= closed lost |closed won=1 ). מחקנו 2 עמודות חישוביות-רבעון וחודש כי אפשרויות לחישוב באופן פשוט.

* Load- את הדאטה המעודכן העלנו ל GitHub.