**מטלה 2 - בינה עסקית**

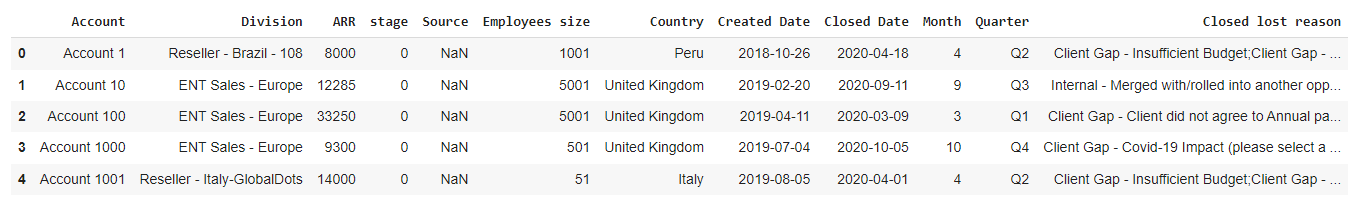
**קבוצה K**

מגישים:

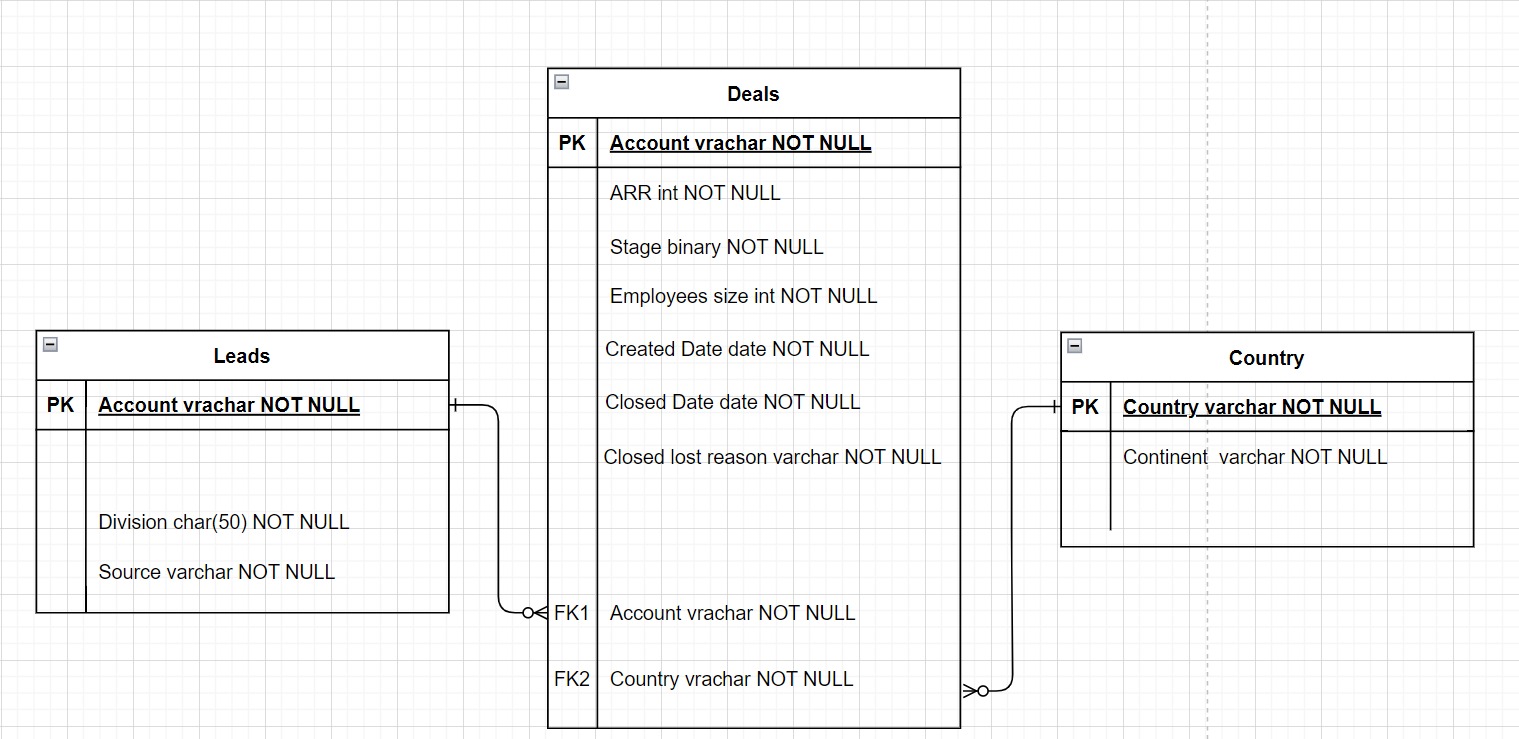
|  |  |
| --- | --- |
| שם | ת.ז |
| אמיתי אוירבך | 307885509 |
| רועי נדם | 311147763 |
| בן גורניצקי | 313580144 |
| אסף איסן | 313289506 |
| בר יעקבי | 203261144 |

**חלק א'- הגדרת Data Warehouse**

הדאטה שבחרנו במטלה 1 מציג עסקאות שבוצעו בחברה מסוימת.  
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:



1. מצורף לעבודה כקובץ SQL.( SQLite-Data Warehouse.sql)
2. -USE CASEנרצה 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 .

1. מצורף קובץ SQL

**חלק ב'-הגדרה ומימוש ETL**

1.

* **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 בפורמט SQL.

בנוסף המרנו את עמודת stage מטקסט לבינארי (0= closed lost |closed won=1 ). מחקנו 2 עמודות חישוביות-רבעון וחודש (ניתן לחשבן בצורה פשוטה על ידי פונקציית ""BEETWEN.

* **Load**- את הדאטה המעודכן העלנו לSQLite כקובץ מסוג DB.

2.

|  |  |
| --- | --- |
| **מה נדרש לבצע בכל שלב?** | **Step** |
| אין צורך בשלב זה (זהו שלב אופציונלי וניתן לשלול אותו אם אין צורך) | Step 1: Reference Data |
| כלל הקבצים בפורמט אחיד מסוג אקסל. | Step 2: Extract from Data Reference |
| שדה נתונים צריך להכיל תאריכים Created Date ,Closed Date תקפים ב-12 החודשים האחרונים במסד נתונים של עסקאות פיננסיות מהשנים 2019 ו 2020. כמו כן לכל מדינה צריכה להיות היבשת המתאימה לה בטבלת Country. | Step 3: Data Validation |
| הסרת נתונים זרים או שגויים: מחקנו 2 עמודות חישוביות-רבעון וחודש.  המרנו את עמודת stage מטקסט לבינארי.  בדיקת שלמות הנתונים. סכימה של הכנסות לפי רבעונים. | Step 4: Transform Data |
| הזנת נתונים במסד נתונים. | Step 5: Stage |
| טעינת הנתונים לטבלאות. | Step 6: Publish to Data Warehouse |

3. מצורף קישור לתהליך ETL שביצענו בGoogle Colab

<https://colab.research.google.com/drive/1uOJ00jKN39-TmREqOb41xSKUz4FJo8fC?usp=sharing>