# פרויקט - בינה עסקית קבוצה AB

מגישים:

יואב ברק – 315311993

ולדיסלב פשניצ'ני – 317152569

קישור לGitHub:

https://github.com/yoavbrk1/BIProject.git

## מטרת הפרויקט + KPIs

**מטרת הפרויקט -**

GlobalTech Solutions היא חברת טכנולוגיה בינלאומית מובילה ברחבי העולם.

החברה ידועה במוצרים החדשניים ובשירות הלקוחות יוצא הדופן שלה, החברה ביססה את עצמה כשחקנית מפתח בתעשיית הטכנולוגיה.

עם המטה הראשי בעמק הסיליקון, קליפורניה, GlobalTech Solutions פועלת ביבשות מרובות, ומספקת מגוון רחב של לקוחות.

בפרויקט זה, נרצה לבחון איזו יבשת הינה הרווחית ביותר, כל זאת על מנת להסיק מסקנות שיועילו וישפרו את שאר היבשות על מנת להביא למיקסום הפוטנציאל שלהן.

**KPIs -**

האם ניתן לחזות ששוני במדינה/גודל הארגון/רבעון תביא ל ARR יותר גבוה? במילים אחרות האם עסקה תניב מעל $100,000 ARR באמצעות מדינת המקור וגודל הארגון והרבעון?

מדד ה- KPI הראשון שלנו הוא ARR - זה הוא אומדן של ההכנסה שתהיה לחברה השנה בהתבסס על המצב הנוכחי שלה. לכן, זה לוקח בחשבון את המנוי השנתי שיש לה עד עכשיו ואת אסטרטגיית התמחור שלה, בין היתר. מדד זה יכול לתת כמה לומדות מעניינות על הצמיחה ולאן החברה צועדת עם האסטרטגיה הנוכחית שלה.

מדד ה- KPI השני שלנו הוא Monthly Sales Growth - שימוש בצמיחה חודשית במכירות כ-KPI למכירות נותן לך תובנות מעשיות בהן תוכל להשתמש כדי לייעל את תהליכי המכירה, האסטרטגיות ו סדרי העדיפויות של המוצר שלך.

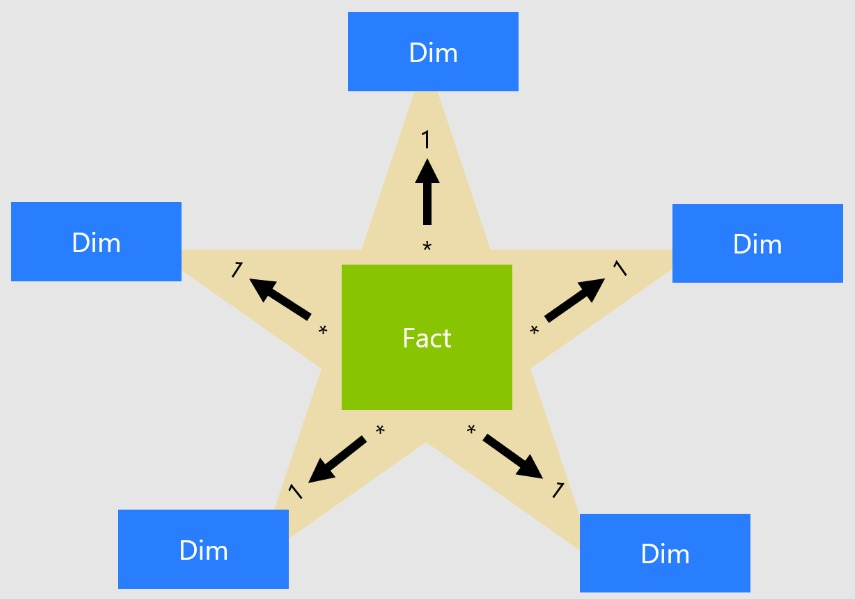
## הגדרת Data Warehouse

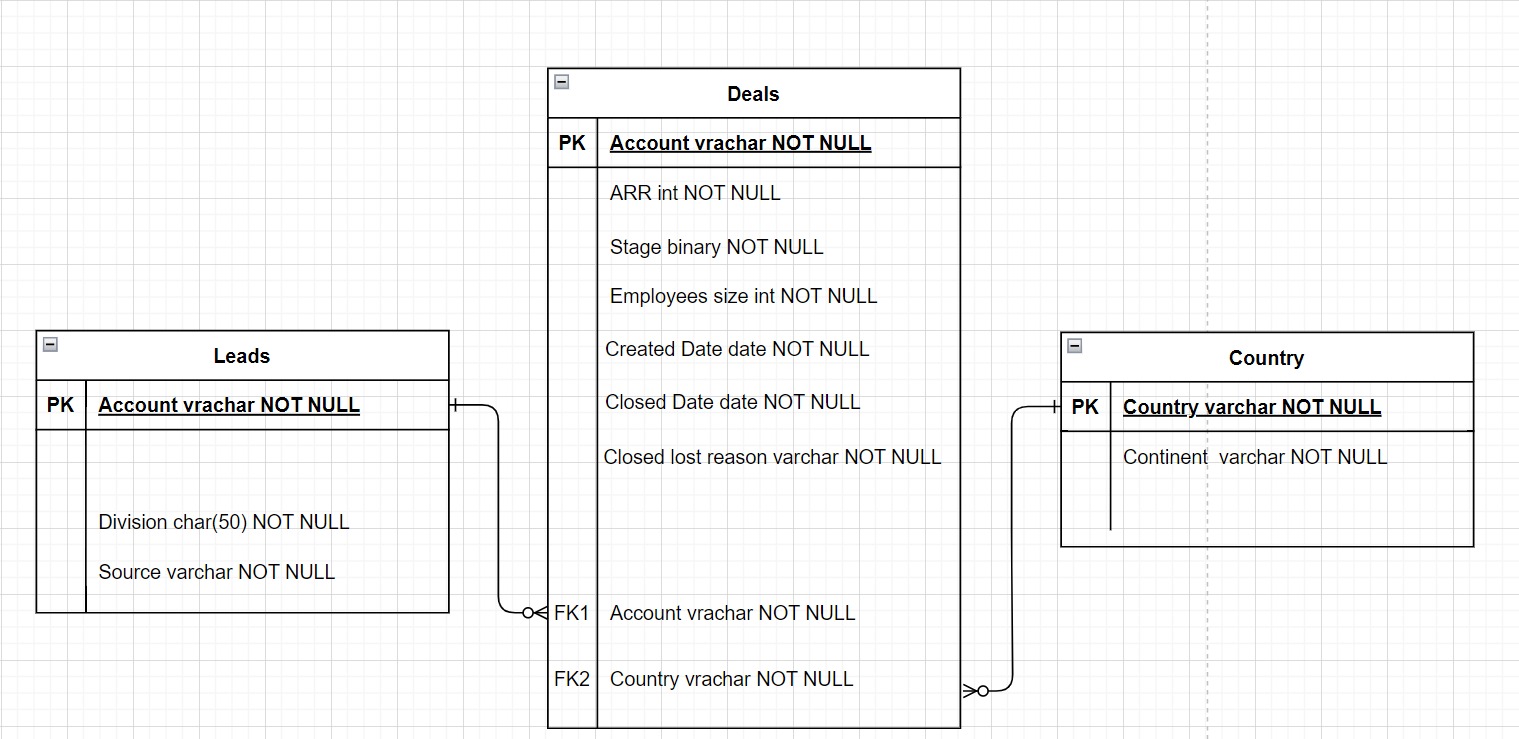
נבחר בסכמת כוכב (Star Schema) מכיוון שנרצה לחלק את הנתונים כך שטבלת העסקאות תהיה במרכז הכוכב. כך נוכל לשלוף מידע עסקי בצורה מהירה יותר וללא צורך בחיבורים מורכבים.

סכימת "כוכב" של מחסן נתונים היא טכניקת מודל נתונים פופולרית המשמשת במחסני נתונים כדי לארגן ולייצג נתונים לצורך שאילתות וניתוח יעילים.

בסכימה זו, טבלת העובדות מכילה את הנתונים המספריים או הניתנים למדידה המייצגים את המדדים העסקיים או מדדי הביצועים העיקריים (KPIs) של ארגון. טבלת העובדות מורכבת בדרך כלל ממפתחות זרים המקשרים למפתחות הראשיים של טבלאות הממדים הקשורות.

עיצוב הסכימה מפשט אחזור וניתוח נתונים על ידי דה-נורמליזציה של הנתונים. דה-נורמליזציה כוללת ביטול קשרים מורכבים בין טבלאות על ידי שכפול נתונים ומיזוג תכונות קשורות לטבלה אחת. מבנה דה-נורמל זה משפר את ביצועי השאילתות מכיוון שהוא מקטין את מספר ההצטרפות הנדרשים לגישה לנתונים.



**תיאור את ה Data Warehouse באמצעות ERD**

**-USE CASE**

נרצה להציג למנהל הכספים את התפלגות הרווחים לפי יבשות.

SELECT distinct continent ,SUM(arr)

OVER (partition by continent) as ContinentRev

FROM Deals SD

JOIN Country 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 .

## תהליך הETL

**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.

## ניתוח הData Warehouse

* + - חישוב אחוז גידול ההכנסה לפני רבעי שנה:

SELECT Division,

SUM(ARR) over (partition by Division,Quarter order by Quarter DESC ) as "Revenue in Current period",

Quarter,

LAG(SUM(ARR)) over (partition by Division,Quarter order by Quarter DESC) as "Revenue in Previous period",

SUM(ARR) over (partition by Division,Quarter) - LAG(SUM(ARR)) over (partition by Division,Quarter)/LAG(SUM(ARR)) over (partition by Division,Quarter)\*100

as "KPI: % Groth of Revenue"

FROM Deals 2020 inner join Leads

ON Deals 2020.Account = Leads.Account

* + - חישוב ה- CVR (Conversion Rate) לפי "לידים":

SELECT Division,

sum(case Stage when 'Closed Won' then 1 else 0 end) AS "Total # of won leads",

COUNT(Stage) OVER (PARTITION BY Division) AS "Total # of leads",

sum(case Stage when 'Closed Won' then 1 else 0 end)/COUNT(Stage) OVER (PARTITION BY Division)\*100

AS "KPI: Convertion Rate"

FROM Deals 2020 inner join Leads

ON Deals 2020.Account = Leads.Account

* + - אחוז ה-Referrals מכל ה"לידים" לפי division:

SELECT Division, Source,

COUNT(Source) OVER (PARTITION BY Division) as "# of total Leads",

COUNT(CASE WHEN Source = 'Referral Conversation' THEN 1 END) OVER (ORDER BY Division) AS "# of Refferals per Division"

COUNT(Source) OVER (PARTITION BY Division)/COUNT(CASE WHEN Source = 'Referral Conversation' THEN 1 END) OVER (ORDER BY Division) \*100

AS "% of refferals from total leads"

FROM Leads

* + - מציאת היבשת שמביאה את הרווח המקסימלי לפי ריבעון:

SELECT Continent,

Quarter,

SUM(ARR) OVER (PARTITION BY Continent, Quarter ORDER BY Continent)

FROM Deals INNER JOIN Country

on Deals.country = Country.Country

* + - מציאת היבשת עם הפוטנציאל הרווחי ביותר:

SELECT Continent,

Quarter,

SUM(ARR) OVER (PARTITION BY Division),

SUM(case Quarter when 'Q1' then 1 else 0 end) AS "Revenue in Quarter 1",

LAG(SUM(case Quarter when 'Q1' then 1 else 0 end) AS "Revenue in Previous Quarter",

SUM(case Quarter when 'Q1' then 1 else 0 end) - LAG(SUM(case Quarter when 'Q1' then 1 else 0 end)\*100

AS "KPI: % of growth by Quarters ",

FROM Deals INNER JOIN Country

on Deals.Country = Country.Country

* + - חישוב אחוז הרווח משיטות גיוס הלקוחות:

SELECT Leads.Source,

SUM(ARR) OVER (PARTITION BY Account ORDER BY Account) AS "Revenue per method",

SUM(ARR) OVER (PARTITION BY Account ORDER BY Account)/SUM(ARR)\*100

AS "KPI: % Method Revenue"

FROM Deals INNER JOIN Leads

on Deals.Account = Country.Account

* + - ממוצע ההכנסות בשנה – ARR :

SELECT country, quarter, AVG(ARR)

FROM Deals

GROUP BY country, quarter

* + - מציאת הריבעון שמביא את כמות ההכנסות המקסימלי:

SELECT extract(quarter from quarter) as quarter , avg(quarter) over (partition by extract(quarter from quarter))

from

(SELECT DISTINCT date ((date), quarter) as quarter , sum(ARR) over (partition by (date) ) as profit\_ quarter from `Deals`)

## מסקנות

* היבשת שמביאה את הרווח המקסימלי-

|  |  |
| --- | --- |
| **Continent** | **ARR ($)** |
| Europe | 7757432 |
| North America | 5515454 |
| Asia & Pacific | 4733771 |
| South/Latin America | 962033 |
| Middle east | 355079 |

* היבשת עם פוטנציאל הרווח הגדול ביותר –

|  |  |
| --- | --- |
| **Continent** | **Sales Growth rate** |
| South/Latin America | 184% |
| #N/A | 131% |
| Middle east | 58% |
| Asia & Pacific | 52% |
| North America | 48% |
| Europe | 31% |
| Africa | 8% |
| Arab States | 0% |

* הרבעון שמביא את כמות ההכנסות הגדול ביותר–

|  |  |
| --- | --- |
| **Quarter** | **ARR ($)** |
| Q3 | 5722920 |
| Q4 | 5534622 |
| Q2 | 4480259 |
| Q1 | 3851721 |

* ממוצע ההכנסות בשנה הוא –

|  |  |
| --- | --- |
| **Continent** | **Average of ARR (By Deal)** |
| Arab States | 20000.0 |
| Middle east | 19726.6 |
| North America | 18953.5 |
| Europe | 14472.8 |
| Asia & Pacific | 13602.8 |
| Africa | 11024.3 |
| South/Latin America | 9525.1 |
| #N/A | 7031.8 |