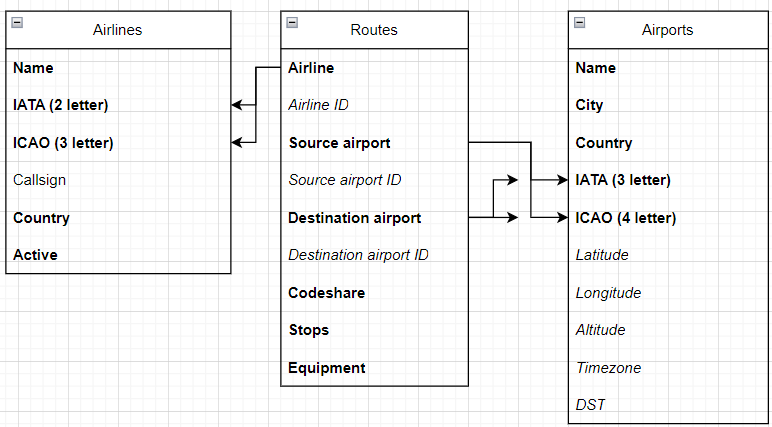
**פרויקט בינה עסקית**

מגישות:

ליאורה סיאק – 205444813 ,הדר מוזס - 313528614

קבוצה T

1. **שאלה עסקית + KPI**
   1. קישור לדאטה: <https://www.kaggle.com/datasets/elmoallistair/airlines-airport-and-routes>
   2. מטרת הפרויקט: ניתוח דפוסי סגירה של קווי תעופה המשוייכות לחברות תעופה, תוך התייחסות למדינה אליה משוייך קו/חברת התעופה ומספר העצירות בקו כדי לנסות ולמצאות מכנה משותף ולהבין את הסיבות לסגירת הקווים.
   3. שאלת מחקר: מהו הגורם המשותף לקווי התעופה שנסגרו? בחינת הסיבות הפונציאליות לסגירתם.
   4. KPIs:
      1. בדיקת מספר העצירות שהיו בקווי התעופה שנסגרו. היעד רלוונטי לשאלת המחקר, ספציפי, מדיד, ישים לפי הנתונים שברשותנו.
      2. מדידת מספר/אחוז קוי התעופה שנסגרו לכל חברת תעופה. זה יעד רלוונטי כדי לפרוט את שאלת המחקר, ספציפי מדיד ובר השגה.
      3. בחינת היעדים בעולם אשר בהם נסגרו יותר קווי תעופה וחיפוש דפוסים. עוזר לשאוב עוד מידע בנוגע לשאלת המחקר, רלוונטי לנושא, מדיד ובר השגה.
2. **הגדרת Data warehouse**
   1. הסכמה הנבחרת Star Schema:
      1. בחרנו בסכמה זו מכיוון שזיהינו שאחת הטבלאות שלנו מכילה שדות המהווים מפתחות זרים בטבלאות האחרון. בתרשים טבלת הrouts היא הטבלה המרכזית ושתי הטבלאות האחרות ((airports,airlines מקושרות אליה ומכילות רשומות נוספות. הטבלות הנוספות מחזיקות מידע מדויק ומלא על הנתונים הרלוונטיים לכל תת-נושא. הסכימה הזו יעילה במקרים שבהם יש קשרים ישירים ומובנים בין השדות, ונדרשת יכולת דיוק גבוהה על הנתונים.
   2. Datawherhouse לאחר שילוב הסכמה:
      1. 
   3. Use Case:
      1. תרחיש: ניתוח תנועת התעופה של חברות התעופה
      2. מנהל מחלקת תפעול בחברת התעופה מעוניין לנתח את תנועת התעופה של החברה בכדי לקבל מידע חשוב לקבלת החלטות תפעוליות ואסטרטגיות. הוא משתמש במערכת Data Warehouse ובכלי אנליטיקה לצורך איתור מדדים מרחביים וזמן על תנועת התעופה.   
         סכימת הכוכב (Star Schema) מאפשרת למנהל התפעול לבצע את התהליך הבא:
      3. איתור טבלת שדות התעופה: בעזרת טבלת השדות התעופה המכילה את נתוני תחנות התעופה, המיקום, הקודים והמדדים הגיאוגרפיים, המנהל יכול לזהות נתונים רלוונטיים על תחנות התעופה השונות כגון המיקום, הגובה, הקודים IATA ו-ICAO.
      4. איתור טבלת חברות התעופה: בעזרת טבלת חברות התעופה, המכילה מידע על שמות החברות, קודי IATA ו-ICAO ופרטים נוספים, המנהל יכול לבצע חיפושים ואיתור חברות התעופה הרלוונטיות לניתוח התנועה.
      5. איתור טבלת נתיבי התעופה: בעזרת טבלת נתיבי התעופה, המכילה מידע על נתיבי הטיסה, מזהים של חברות התעופה, קודי המקור והיעד של הנתיבים ופרטים נוספים, המנהל יכול לזהות תופעות ומגמות בתנועת התעופה בין חברות התעופה ותחנות התעופה.
      6. בסכמת הכוכב, המרכזיות של הנתונים והיכולת לבצע שאילתות מצומצמות מביאים לשיפור בביצועים ותפוקה, ומאפשרים למשתמש לקבל תובנות מהירות ומדויקות שיכולות לשפר את קבלת ההחלטות במערכת התעופה
3. **תהליך ה-ETL**
   1. Load (טעינה): ייבאנו את הנתונים מטבלת מרכזת של טבלאות שדות תעופה, חברות תעופה ונתיבי תעופה למסד הנתונים המרכזי (Data Warehouse).
   2. Transform (עיבוד): בשלב זה, ביצענו השמטת רווחים מיותרים, העברנו את הטקסט לתצורה אחידה בעזרתם פעולות lower ו-Title, הסרת ערכים חסרים. הגדרנו מקור ויעד לעמודות השונות.
4. **ניתוח הData warehouse**
   * 1. סכימת קווי התעופה הלא פעילים ולאיזה מדינות הן שייכות

SELECT DISTINCT

AirlineCountry,

COUNT(\*) OVER (PARTITION BY Country\_x) AS NotActiveAirlinesCount

FROM flightsJoin

WHERE Active = N;

* + 1. מתוך קווי התעופה הלא פעילים, מהי המדינת יעד שבה נסגרו הכי הרבה קווים -

SELECT DISTINCT Destination Airport, COUNT(\*) AS InactiveAirlinesCount

FROM flightsJoin

WHERE Active = N

GROUP BY Destination Airport

ORDER BY InactiveAirlinesCount DESC

LIMIT 1;

* + 1. עבוד מדינת united states כמה קווי תעופה נסגרו וכמה עדיין פעילים בסכימה

SELECT

Country\_x,

SUM(CASE WHEN Active = N THEN 1 ELSE 0 END) AS InactiveAirlinesCount,

SUM(CASE WHEN Active = Y THEN 1 ELSE 0 END) AS ActiveAirlinesCount

FROM

flightsJoin

WHERE

Country = ‘United States’

GROUP BY

Country\_x;

* + 1. שאילתה שמוצאת את שדות תעופה(מקור) שבהם הכי הרבה קווי תעופה נסגרו -

SELECT

"Source Airport",

NotActiveAirlinesCount

FROM (

SELECT

"Source Airport",

COUNT(\*) AS NotActiveAirlinesCount,

RANK() OVER (ORDER BY COUNT(\*) DESC) AS rank

FROM

flightsJoin

WHERE

Active = N

GROUP BY

"Source Airport"

) subquery

WHERE

rank = 1;

* + 1. עבוד שדה תעופה ROA כמה קווים סגורים יש עבור ROAבתור שדה תעופה היעד וכמה בתור שדה תעופה המקור -

SELECT

MAX(CASE WHEN "Source Airport" = 'RAO' THEN SourceAirportCount END) AS SourceAirportCount,

MAX(CASE WHEN "Destination Airport" = 'RAO' THEN DestinationAirportCount END) AS DestinationAirportCount

FROM (

SELECT

"Source Airport",

"Destination Airport",

COUNT(\*) OVER (PARTITION BY "Source Airport") AS SourceAirportCount,

COUNT(\*) OVER (PARTITION BY "Destination Airport") AS DestinationAirportCount

FROM

flightsJoin

) subquery;

­

* + 1. קווי התעופה הלא פעילים ומדינותיהן -

SELECT Airline, Country\_x

FROM flightsJoin

WHERE Active = N;

1. **מסקנות**
   1. ניתן לראות כי מרבית קווי התעופה השתייכו למדינת המקור ארצות הברית. באופן יחסי לארצות הברית יש הרבה יותר קווי תעופה אותם היא מפעילה ביחס למדינות אחרות, כמובן בגלל הגודל שלה, ולכן זה תואם את מה שציפינו לראות.
   2. לא נמצא קשר מובהק בין כמות העצירות של קו התעופה לסגירה שלו.
   3. בדאטה סט שנבחר אין עמודות עם מידע מספרי, דבר שמקשה על ניתוח מעמיק וזיהוי דפוסים.
   4. כמו שציפינו, ישנו פער גדול בכמות קווי התעופה שנסגרו עבור מדינה מסויימת (ארצות הברית) לבין קווי התעופה הפעילים (כ- 7 אחוז מכלל קווי התעופה שתועדו הם קווים שנסגרו, השאר מוגדרים פעילים).