**מטלה מסכמת-בינה עסקית**

קבוצה F

מגישים: שגיב טורה ועמית פרידמן

1. **שאלה עסקית+:KPIs**

מטרת הפרויקט שלנו היא להבין מה משפיע על הצלחת סרטים בכפוף לנתונים שמופיעים במסד הנתונים. בעזרת ניתוח של הנתונים: ציוני הסרטים, תקציבים, רווח, חברת הפקות והמדינה בה צולם הסרט ובעזרת שאילתות SQL ננסה לקבל תוצאות לגבי המדינה שבה הסרטים הכי מצליחים או אילו חברות ההפקה המוצלחות ביותר.

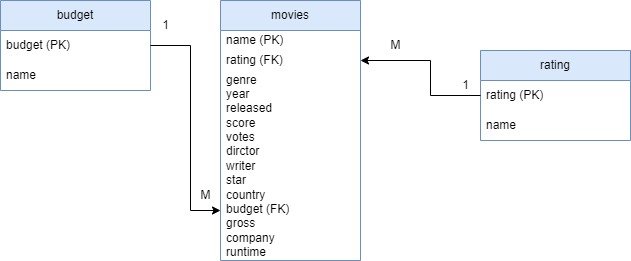
KPIs:

* כמה הכנסות ברוטו הסרט מרוויח
* מה הדירוג של הסרט באתר .IMDB

1. **הגדרת Data Warehouse:**
2. עבור הדאטה שבחרנו, החלטנו להשתמש בסכמת כוכב.

בחרנו בסכמה זו כי הדאטה שלנו פשוטה וקלה להבנה ולנו יש צורך לחקור בפרויקט את הממדים השונים ולבצע שאילתות אנליטיות.

1. תיאור ה-Data Warehouse לאחר שילוב הסכימה (באמצעות ERD):



1. אנחנו רוצים לדעת איך הדירוג והתקציב של סרט משפיע על ההכנסה שלו. סכמת הכוכב יכולה לעזור לחברת ההפקות להבין איך לתקצב את ההפקה ע"י שליפת נתונים מהירה ושימוש קטן בJoin.
2. **תהליך ETL:**

* שלב ה-Extraction:  
  ממסד הנתונים הוצאנו את הטבלאות הבאות:

1. שמות הסרטים , חברות ההפקה שצילמו את הסרט, המדינות בהם צולמו הסרטים, תקציב, ורווח.

* שלב ה-Transform:  
  השתמשנו בשיטות ההמרה הסטנדרטיות שלמדנו ב- STTM על מנת, לנקות את הנתונים ולהבטיח את התיקוף שלהם בכדי שנוכל לבצע את הניתוח כראוי.

תהליכים עיקריים בעיבוד הנתונים הידני:

1. מחיקת שורות בעלות ערכים ריקים באחד מהקריטריונים שהוגדרו תחילה להצלחתו של סרט.
2. העברת פורמט שמות הסרטים וחברות ההפקה כstr
3. מחיקת כפילויות - לדוגמה: חברות הפקה שעובדות כחברות בת וסרטים שהועלו פעמיים עם גרסת במאי.

לאחר הניקוי הידני ביצענו ניקיון סופי על ידי שיטת ההמרה הסטנדרטית ב-STTM. תחילה הכנסנו אל מחלקת ההמרה פונקציה לניקיון השמות של הסרטים על מנת שיהיו קריאים וברורים. הפונקציה מורידה רווחים מיותרים, משנה את האות הראשונה למילה לאות גדולה ואת שאר האותיות לקטנות.  
בנוסף הציונים של הסרטים מגיעים אל המערכת כ-str, בעזרת המיפוי אנו שינינו אותם ל-int.

1. **ניתוח Data Warehouse:**

* **נוציא את השם את שם החברה ואת הציון שלהם ממוין לפי שם של חברת הפקה.**

. SELECT name, company, rating

FROM movies

ORDER BY company ;

* **לכל חברת הפקה נחשב את ממוצע הציונים של הסרטים שלהם**

SELECT company, rating,

AVG(score) OVER (PARTITION BY company) AS average\_score

FROM movies

ORDER BY average\_score DESC ;

* **ממוצע תקציב של חברות ההפקה**

SELECT company, rating,

AVG(budget) OVER (PARTITION BY company) AS average\_budget

FROM movies

ORDER BY average\_budget DESC ;

* **שם הסרט המדינה שבה הוא צולם והציון שלו ממוין לפי מדינות**

SELECT name, country, score

FROM movies

ORDER BY country;

* **מדינות והממוצע ציוני סרטים בהם**

SELECT name, country, score,

AVG(score) OVER (PARTITION BY country) AS average\_score

FROM movies

ORDER BY average\_score DESC;

* **מדינה וממוצע תקציב לסרטים שצולמו בה**

SELECT name, country, score,

AVG(budget) OVER (PARTITION BY country) AS average\_budget

FROM movies

ORDER BY average\_budget DESC ;

1. To get the name of the movie, the country where it was filmed, and its gross sorted by the name of the country:

SELECT name, country, gross FROM movies ORDER BY country;

1. To get the name of the country with the average gross sorted from the largest to the smallest:

SELECT country, AVG(gross) AS average\_gross FROM movies GROUP BY country ORDER BY average\_gross DESC;

1. To get the name of the company and their gross sorted by the name of the production company:

SELECT name, company, gross FROM movies ORDER BY company;

1. For each production company, the average gross of their films is calculated, sorted from the largest to the smallest:

SELECT name, company, gross, AVG(gross) OVER (PARTITION BY company) AS average\_gross FROM movies ORDER BY average\_gross DESC;

1. **מסקנות:**
   1. תמונה שמכילה טקסט, קבלה, גופן, לבן

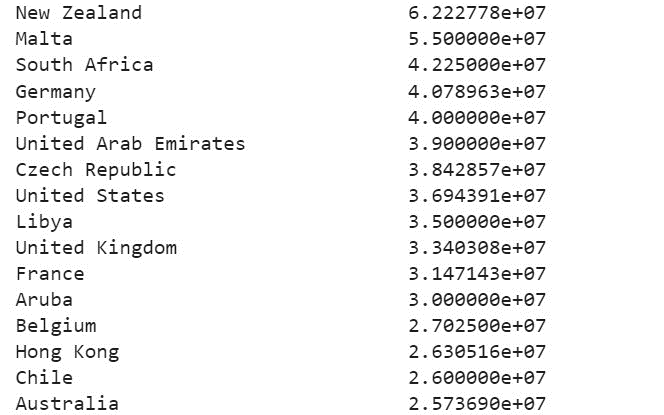
      התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, קבלה, גופן, צילום מסך

      התיאור נוצר באופן אוטומטיניתן לראות בטבלה את חברות ההפקה בעלות התקציבים הגבוהים ביותר. בטבלה שמתחתיה מצוינים הציוניים הממוצעים של חברות ההפקה. לא ניתן לחזות את הציונים לפי התקציב של חברת ההפקה בעקבות חברות הפקה שהפיקו סרטים בודדים שקיבלו ציון גבוה באופן יחסי ולכן הממוצע שלהן יהיה מאוד גבוה ביחס לחברות גדולות שמפיקות הרבה סרטים.
   2. בדומה למסקנה הראשונה, גם ציון ממוצע של סרט לפי מדינות יכול להיות מוטה בעקבות מיעוט בנתונים של מדינות מסוימות.

תמונה שמכילה טקסט, צילום מסך, תפריט, גופן

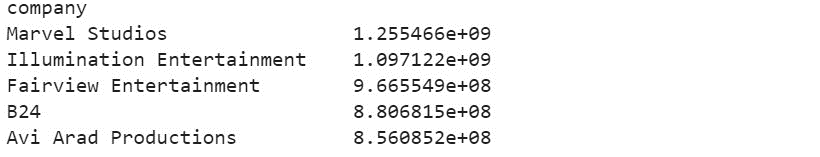
התיאור נוצר באופן אוטומטי

* 1. בטבלה שמוצגת מתחת ניתן לראות את התקציב הממוצע לפי מדינה. ובטבלה שמתחתיה אפשר לראות את הרווח הממוצע לפי מדינה וניתן להבחין שמדינה כמו ארה"ב מייצרת הכנסה גדולה בהרבה ביחס להשקעה שלה ולכן ניתן להמליץ להפקות של סרטים להפיק את הסרטים שלהם במדינה כמו ארה"ב.



תמונה שמכילה טקסט, צילום מסך, גופן

התיאור נוצר באופן אוטומטי

* 1. ניתן לראות בטבלה את הרווח הממוצע ממוין לפי חברות ההפקה. ניתן מן הרשימה שמארוול זאת חברת ההפקה הרווחית ביותר. בנוסף ניתן לראות בטבלה הראשונה (שמוצגת מתחת למסקנה הראשונה) שזאת חברה שנמצאת במקום שני ברשימת התקציבים הגבוהים שמניבות רווח גבוה.

**לסיכום:**

ניתן לראות באופן מובהק שמבחינת רווחיות הפקות שיצולמו בארה"ב תחת מארוול יהיה להם את הסיכוי הגבוה ביותר להצלחה מבחינת רווחיות כלכלית.

מבחינת ציונים להפתעתנו, השוק הלבנוני מקבל ציונים גבוהים וניתן להגיד בערבון מוגבל שסרט שיצולם בהפקה לבנונית יקבל ביקורות טובות.