**בינה עסקית – מטלה 2**

מגישות :

דין חחיאשווילי 311558183

אלמוג דרור 209139682

גל שטיינברג 312407604



חלק 1 – פתיחת פרויקט משותף GitHut

להוסיף קישור

חלק 2 – הגדרת Data Warehouse

1. בחרנו בסכמת Snowflake עבור אוסף הנתונים שלנו משום שהיא מספקת מבנה פשוט ומאורגן לנתונים, ידועה באופטימיזציה של ביצועי שאילתות וניתנת להרחבה ולהתאמה לצרכים עסקיים משתנים.
2. תיאור ה Data Warehouse-

**האם סרטים מקוריים של נטפליקס שיצאו עד שנת 2021 מדורגים גבוה יותר ע"י משתמשים מאשר סרטים שנרכשו ע"י נטפליקס עד שנת 2021.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **dimMovies** | | | | | |
| **MovieID** | **Netflix Link** | **IMDb Link** | **Image** | **Poster** | **Trailer** |
| Title | Netflix Link | IMDb Link | Image | Poster | IMDb Trailer |

שורה ראשונה – שם העמודה

שונה שניה – שם העמודה בדאטה המקורי

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Fact** | | | | | | |
| **isOriginal** | **Time Period** | **Awards** | **IMDB Votes** | **IMDb Score** | **Movie Title** | **MovieID** |
| Production House | Release Date | Awards Received | IMDb Votes | IMDb Score | Title | Title |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **dimInformation** | | | | | |
| **MovieID** | **Genre** | **Language** | **MovieLength** | **RottenScore** | **MetaScore** |
| Title | Genre | Languages | **Runtime** | Rotten Tomatoes | Metacritic Score |

Series or movies – ניקח רק את הסרטים ונזרוק את העמודה

**MovieLength –**

**הערך 0 לסרטים בין 30-60 דקות, הערך 1 לסרטים בין שעה לשעתיים, ערך 2 לסרטים מעל שעתיים**

Metacritic Score, Rotten Tomatoes – לחלק את העמודות ב10

לוודא שורות ריקות

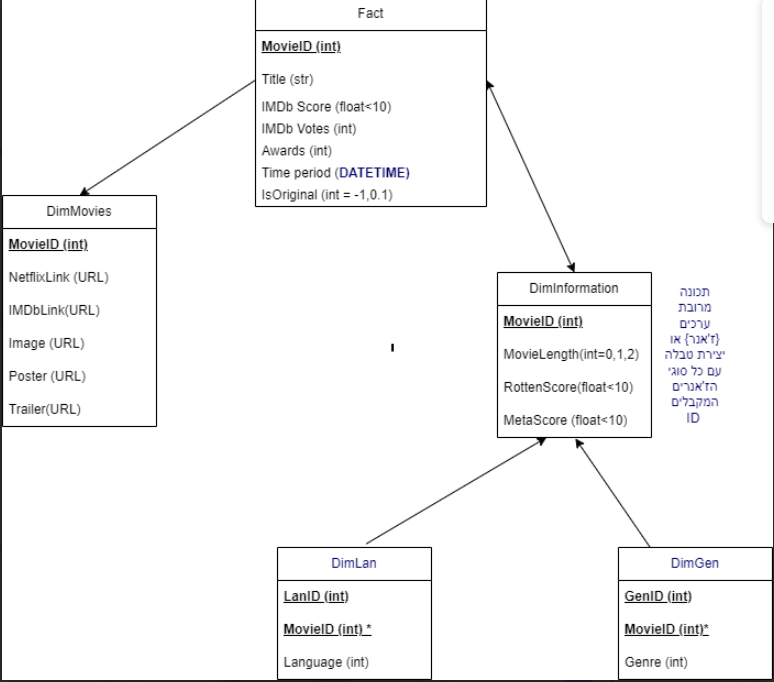
|  |  |  |
| --- | --- | --- |
| **Source** | **Target** | **Mapping** |
| Title | Movie Title | * **הסרת רווחים מיותרים strip** * **המרה לאותיות קטנות lower** * **הפעלת פונקציית title** |
| IMDb Score | IMDb Score | * להעביר כמו שהוא * לשים -1 במקום. ואז לשלוף ערכים חיוביים. |
| IMDB Votes | IMDB Votes | * להעביר כמו שהוא |
| Awards Received | Awards | * להעביר כמו שהוא * האם לכתוב אין מידע? - לשים -1 במקום. ואז לשלוף ערכים חיוביים. |
| Production House | isOriginal | **ערך 1-** לתאים ריקים (לא ידוע)  **ערך 0** לתאים בהם נטפליקס לא שותפה   * **ערך 1** לתאים בהם נטפליקס שותפה |
| Release Date | TimePeriod | * נמיר את התאריך לפורמט:   **DD/MM/YYYY**   * נפצל לפי ״/״ וניקח את הערך האחרון (YYYY) * המרה לbins |

טבלת DimMovies

|  |  |  |
| --- | --- | --- |
| **Source** | **Target** | **Mapping** |
| Netflix Link | Netflix Link | * להשאיר בינתיים ולטפל בשלב הדשבורד. |
| IMDb Link | IMDb Link | * **להשאיר כמו שהוא** |
| Image | Image | * **להשאיר כמו שהוא** |
| Poster | Poster | * **להשאיר כמו שהוא** |
| IMDb Trailer | Trailer | * **להשאיר כמו שהוא** |

DimInformation טבלת

|  |  |  |
| --- | --- | --- |
| **Source** | **Target** | **Mapping** |
| Genre | Genre | * שליפת כל סוגי הז'אנרים הקיימים בדאטה פעם אחת * הסרת רווחים מיותרים strip * המרה לאותיות קטנות lower * הפעלת פונקציית title * יצירת טבלה ונתינת אינדקס לכל ז'אנר |
| Language | Language | * שליפת כל סוגי השפות הקיימים בדאטה פעם אחת * הסרת רווחים מיותרים strip * המרה לאותיות קטנות lower * הפעלת פונקציית title * יצירת טבלה ונתינת אינדקס לכל שפה |
| RunTime | MovieLength | * הערך 0 לסרטים בין 30-60 דקות, הערך 1 לסרטים בין שעה לשעתיים, ערך 2 לסרטים מעל שעתיים |
| RottenScore | Rotten Tomatoes | * לחלק את העמודות ב10 |
| MetaScore | Metacritic Score | * לחלק את העמודות ב10 |



1. הבחירה שלנו במודל נתונים snowflake היא יעילה ועוזרת בקבלת החלטות למשתמש- כאשר יצרנו חלוקה לתתי טבלאות- בצורה היררכית, טבלת ז'אנרים, וטבלת שפות, שהיא תחת טבלת על מנת להקל על משתמש לבצע שליפה מהירה יותר של הסרטים באמצעות סינון ושליפה מהירה לפי פרמטרים אלו.