**חלק ג' – Design**

במסמך זה נציע תיכון למסד הנתונים בהתבסס על התובנות אליהן הגענו בחלק הקודם של הפרויקט. כפי שציינו בחלק הקודם, התובנות שלנו היו מבוססות ומונחות בעיקר על הערים, מתוך תקווה למצוא תובנות מעניינות שיהיו רלוונטיות לתכנון החלוקה האופקית והאנכית של הטבלאות במסד. בחלק הקודם מצאנו לא מעט תובנות מעניינות על המסד, שברובן נשתמש לתיכון.

בשורות הבאות נציג את התיכון למסד, טבלה-טבלה. מפאת קוצר היריעה, לא נוכל להרחיב על כל עמודה ושורה בכל טבלה, לכן נתייחס לחלוקה בקווים מעט יותר כלליים, בדגשים, וככל שלא נציין פרט מסוים, ניתן יהיה להניח שהוא אינו קריטי לתיכון בהתבסס על התובנות שמצאנו בחלק הקודם.

**הטבלה movies:**

חלוקה אופקית - דגשים:

* **ז'אנרים**: מהתובנות שמצאנו, עולה כי ז'אנר האקשן הוא ז'אנר פופולרי במיוחד בתל אביב (במקום הראשון ברשימת הז'אנרים הפופולריים בת"א, עם 21731 שאילתות כאשר הז'אנר הבא עם כ-4000 שאילתות בלבד). באופן דומה, הסקנו כי ז'אנר הדרמה פופולרי בחיפה. לפיכך, נרצה למקם את כל הרשומות של סרטי אקשן בת"א, ואת סרטי הדרמה בחיפה.
* **שחקנים**: נשים רשומות של סרטים עם ג'וני דפ, טום הנקס ובראד פיט בתל אביב וגם בירושלים – שחקנים אלו מאוד פופולריים באתרים אלה, ופופולריים בדיוק כמו שחקנים אחרים בשאר האתרים.
* **שפה**: סרטים בעברית פופולריים כמעט אך ורק בתל אביב ובירושלים, לכן נמקם את הרשומות של סרטים בעברית בתל אביב ובירושלים (שכפול). כמו כן, סרטים באנגלית הם פופולריים בכל האתרים ומהתבוננות קצרה בטבלה נראה כי כמעט כל הסרטים הם בין היתר באנגלית. לכן לא נתייחס לשפה האנגלית כפרדיקט מפריד בין האתרים. רשומות של סרטים שאינם בעברית או באנגלית, אשר נחשבים למאוד לא פופולריים ומספר השאילתות הנוגעות אליהן מועט מאוד, נמקם באתר אקראי אחד.
* **מדינה**: מכיוון שהסרטים הישראליים פופולריים מאוד, ואך ורק בירושלים ובתל אביב, נמקם את הרשומות בהן מופיעים סרטים ישראליים בצורה משוכפלת בתל אביב ובירושלים.
* **חברות הפקה**: נשים את כל הסרטים של חברות ההפקה: Pixar Animation Studios, Walt Disney Pictures, Warner Bros באילת, שכן חברות ההפקה מאוד פופולריות באילת ופופולריות כמו כל חברת הפקה אחרת בשאילתות הנוגעות לערים אחרות.
* **שנה**: מהתבוננות קצרה על הדאטה ניתן לראות כי ישנם סרטים שהופקו לפני שנת 1990, ואולם כפי שראינו בחלק הקודם, השאילתות מתייחסות רק לסרטים משנת 1990 ואילך. לכן, נמקם רשומות של סרטים שיצאו לפני 1990 באחד האתרים באקראי שכן אף שאילתה לא ניגשת אליהם.

חלוקה אנכית - דגשים:

* את העמודות: overview, revenue, tagline, title, נשים באחד האתרים באקראי שכן אף שאילתה לא ניגשת אליהן. את עמודת מפתח הטבלה movie\_id נשים בכל אחד מהאתרים.
* בהתאם לתובנות מהחלק הקודם (ראינו שהעמודה actors לא ריקה בחלק גדול מאוד מהשאילתות הנוגעות לירושלים ולתל אביב, העמודה director לא ריקה בחלק גדול מאוד מהשאילתות הנוגעות לחיפה, והעמודה country לא ריקה בחלק גדול מאוד מהשאילתות הנוגעות לאילת), נמקם את העמודה actors בירושלים ובתל אביב, את העמודה director בחיפה, ואת העמודה country באילת.

**הטבלה credits:**

חלוקה אופקית - דגשים:

* **שחקנים**: נשים רשומות של סרטים עם ג'וני דפ, טום הנקס ובראד פיט בתל אביב וגם בירושלים, מאותם שיקולים שצוינו לעיל.

חלוקה אנכית - דגשים:

* את עמודת מפתח הטבלה id נשים בירושלים, בתל אביב ובחיפה.
* את העמודה actors נשים בירושלים, את העמודה director נשים בחיפה, מאותם שיקולים דלעיל.
* את כל שאר האינפורמציה בטבלה מקורית אפשר לחלק באקראי בין האתרים – אף שאילתה לא ניגשת אליה.

**הטבלאות tickets ו-users**:

הטבלאות tickets ו-users שונות מהטבלאות movies ו-credits: הן קטנות יותר מבחינת כמות המידע (ניתן אף לחזות בכך אם נתבונן בגודל הקבצים של הטבלאות שסופקו) ולמעט הערים השונות שמופיעות בכל רשומה (ב-tickets תחת השדה city וב-users תחת השדה user\_location) לא ניתן למצוא פרדיקטים ששאילתות מפרידות רשומות בטבלאות הללו על פיהם. לכן נחלק את הטבלה חלוקה אופקית בלבד, ונשים בכל אתר את הרשומות שבהן מופיעה העיר המתאימה לאתר.