

PlayStudios – קטלוג סרטונים

שאלה עסקית

כיצד נחלק בין הסרטונים כדי להגדיל את הסיכויים לכניסה ותשלום באתר, ובפרט: איך לשפר את בחירת המיקום להצגת הסרטונים (Home/Bottom Page)?

משימות

- קטלוג 100 הסרטונים הקיימים לקטגוריות הבאות:
 - "Hot" ○
 - "Stable and Popular" ○
 - "Everything else" ○
- דקירת הפיצ'רים המתאימים לכל קטגוריה
- שימוש באופן הקטלוג ליצירת שיטה חכמה ומסודרת מוכוונת פרודקט

הנחות עבודה

- צוות המוצר ירצה לדרג את הסרטונים העתידיים עפ"י נתוני מסגרת, ללא נתונים על ביצועי עבר (לדוג' שפה או איכות, אך לא כמות צפיות)
- אין יכולת להתאמה אישית של סרטונים עבור כל משתמש (לדוג' להציע סרטונים בשפה שמתאימה לו על בסיס שפות סרטונים אחרים בהם צפה)

מסקנות

מאפייני סרטון מוצלח לפי משקל (בסדר יורד):

- אורך קצר, כ-17 דקות
- איכות גבוהה
- שפה אנגלית\סינית
- עלה לאחרונה

תהליך עבודה

שלב א': הגדרת מדדי הצלחה ובנייתם

- הגדרת מדדים להצלחה – בהסתמך על נתוני הצפייה בלבד
 - כמות צפיות כוללת
 - ממוצע צפיות ליום
 - כמות ימים בהם הסרטון נצפה
 - ההפרש (באחוזים) בין הממוצע השבועי לכמות הצפיות השבועית (Avg Offset)
 - אחוז הימים בהם הסרטון עקף את הממוצע השבועי
 - כמה מהר הגיע הסרטון לכמות קבועה של צפיות
- לאחר ניתוחים של כלל השדות, הוחלט שהשדות שהכי משקפים את הצלחת הסרטון הם:

- כמות צפיות כוללת (כלל הסרטונים נמדדים על אותה כמות ימים, בדיוק 120 יום מאז העלאתם)
 - Avg Offset
 - אחוז הימים בהם הסרטון עקף את הממוצע השבועי
3. בניית שדות מחושבים לפי המדדים הנ"ל בתוך טבלת Video Features (דרך Python, מצורף כ-"data configuration")
4. Scaling ע"י דירוג פנימי עבור כל מדד במספר רץ (Rank, 1 - מקום ראשון, 2 - שני וכו')

שלב א': תוצאות

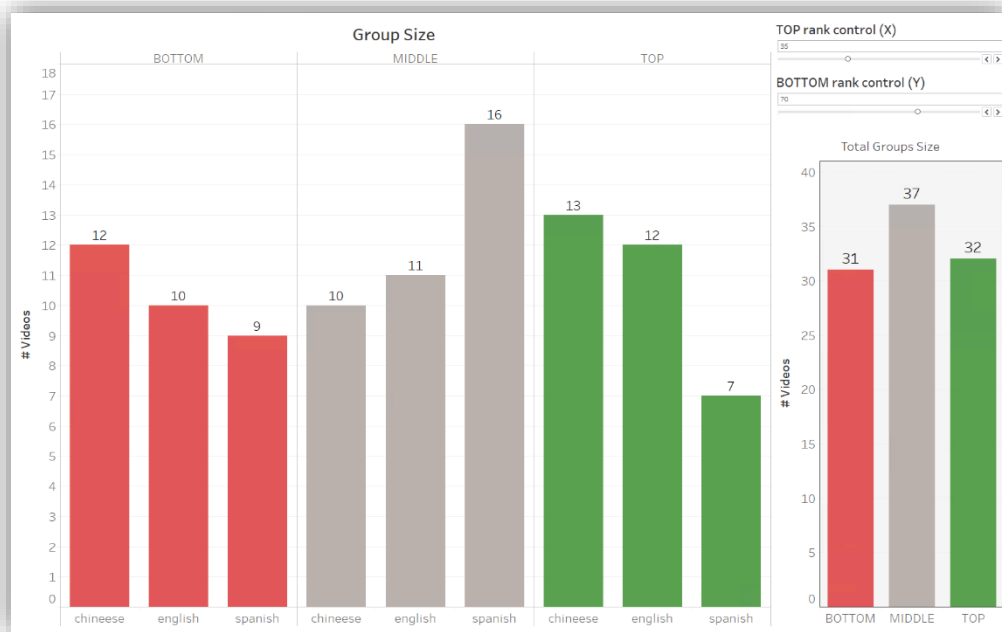
מצ"ב הנתונים המועשרים כ-"video_features_enriched.xlsx"

שלב ב': ניתוח נתוני הצפיות וחלוקה לקבוצות

1. יצירת כלי לפילוח לקבוצות לפי הדירוגים עפ"י ההגדרות הבאות:
 - כאשר כל המדדים גדולים מ-X, הסרטון יסווג **כמוצלח** (כדי למנוע מצבים חריגים בהם אחד המדדים גבוה והשני נמוך, מעיד על volatility) - **TOP**
 - אם לא, כאשר כל המדדים גדולים מ-Y, הסרטון ידורג **כיצב ופופולרי** (ע"מ לתת דגש על היציבות) - **MIDDLE**
 - אחרת, יסווג הסרטון כ"אחר" - **BOTTOM**
2. יצירת פילוחים שונים פילוח הנתונים לפי השדות השונים כדי למצוא דפוסים וקורלציות בין הפיצ'רים לנתוני הצפיות

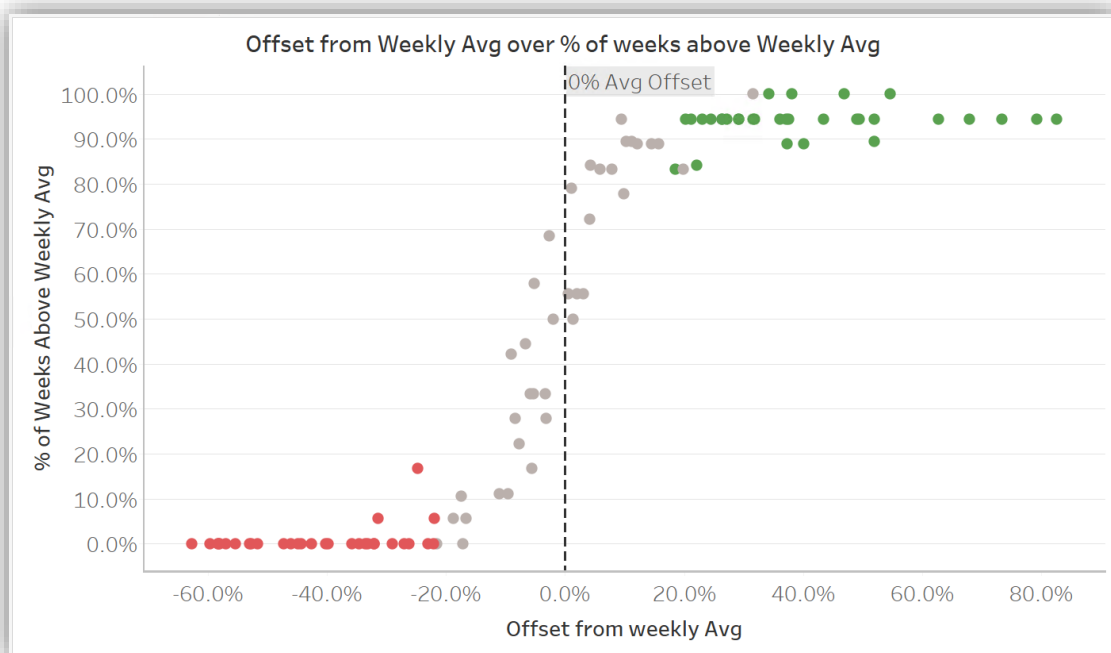
שלב ב': תוצאות

1. **בניית הקבוצות -** שאיפה לקבוצות בעלות גודל בערך זהה כדי לקבל מגוון פרמטרים
 - הדירוג החלש ביותר (בין 1 ל-100) עבור הקבוצה העליונה - 35 || **32% מהסרטונים**
 - הדירוג החלש ביותר (בין 1 ל-100) עבור הקבוצה האמצעית - 70 || **37% מהסרטונים**
 - כל השאר || **31% מהסרטונים**



מסקנה נוספת: סרטונים בספרדית פחות נפוצים בקבוצת הלהיטים

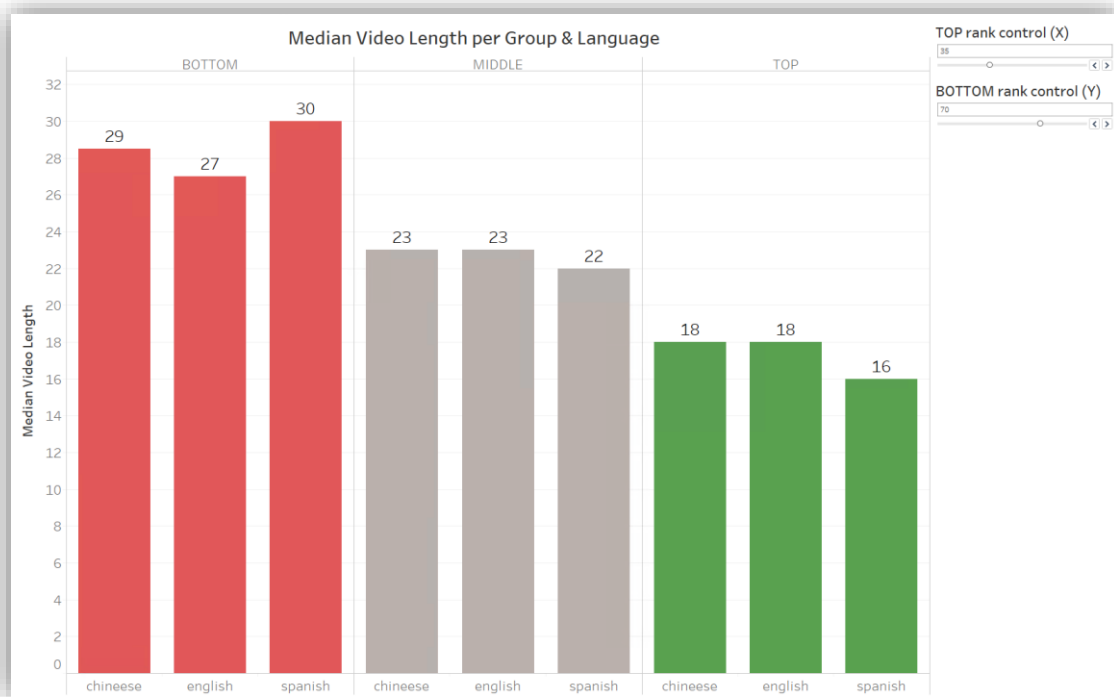
- בתרשים הבא ניתן לראות את פיזור על פני שני המדדים המחושבים: **הפרש מממוצע שבועי ואחוז שבועות** בהם הסרטון עקף את הממוצע השבועי



2. אורך הסרטון

- להלן חישוב אורך הסרטון החציוני עבור כל שפה וכל קבוצה

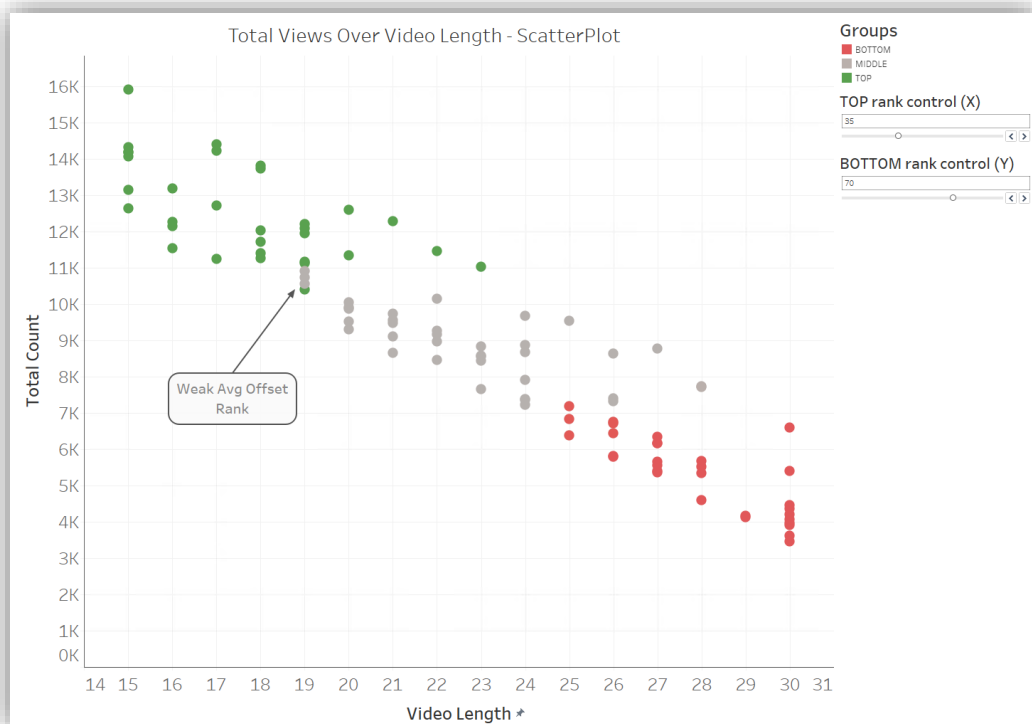
- מסקנה: הצלחת הסרטון - ביחס הפוך לאורכו!



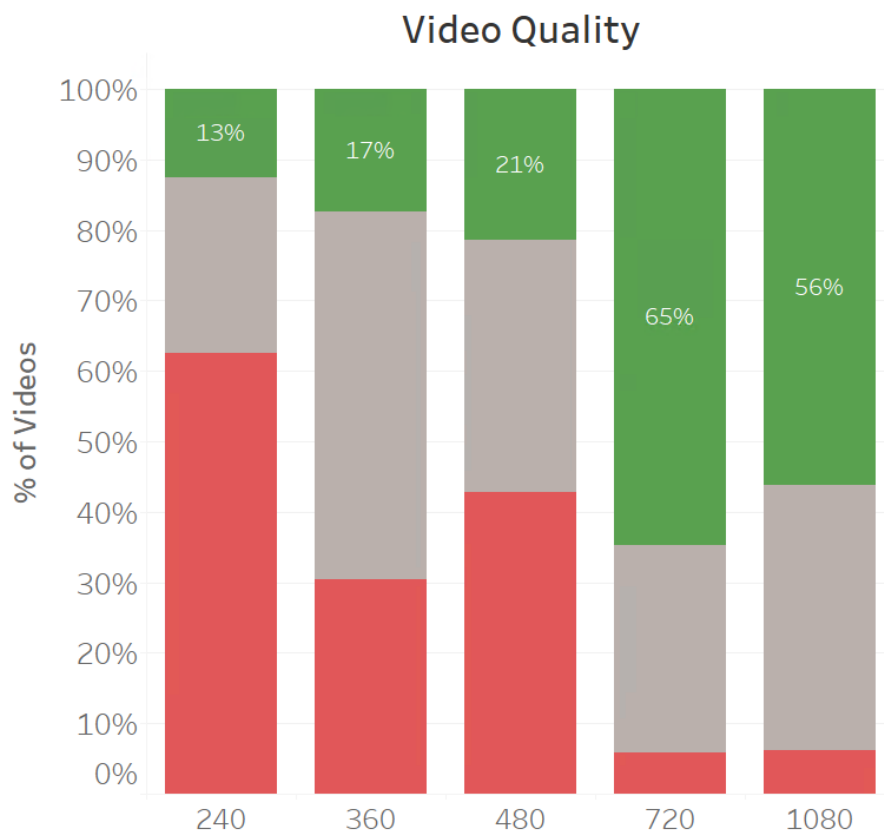
מסקנה נוספת: אין שוני מהותי בין השפות השונות בהקשרי האורך

- גרף פיזור - מספר צפיות על אורך הסרטון

- ניתן לראות את המגמה מהסעיף הקודם בבירור

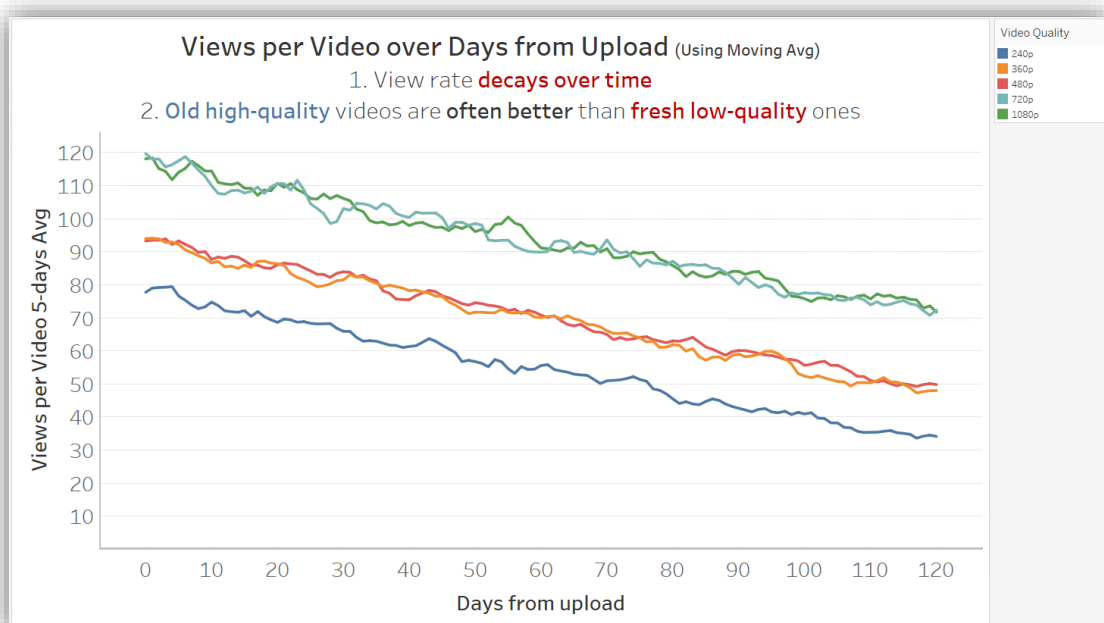


Distribution of Classes per Video-Quality

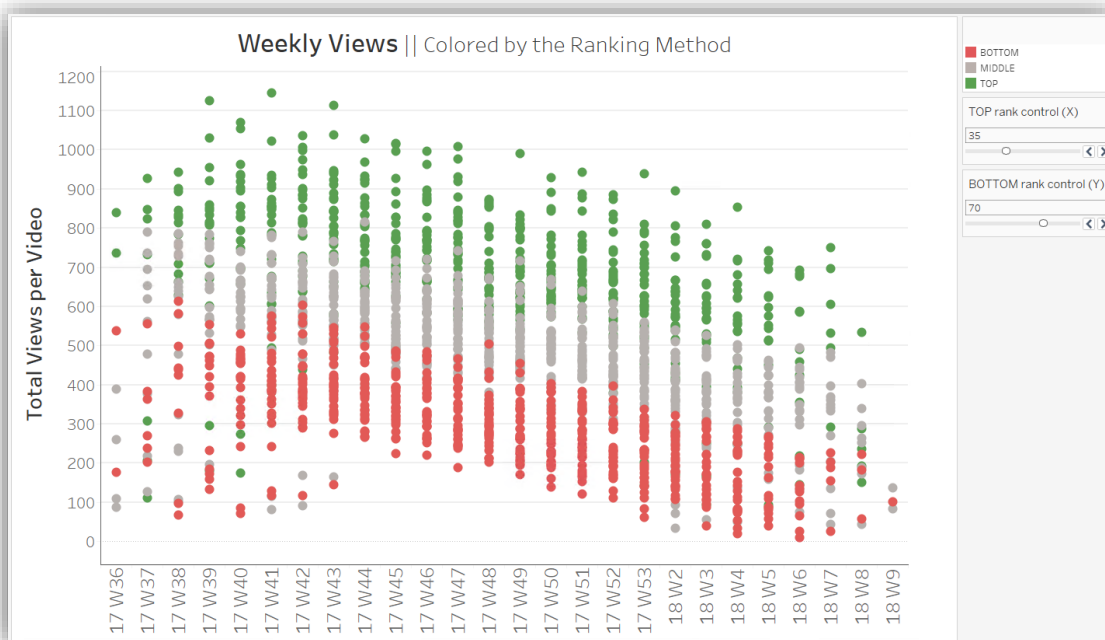


4. זמן העלאה:

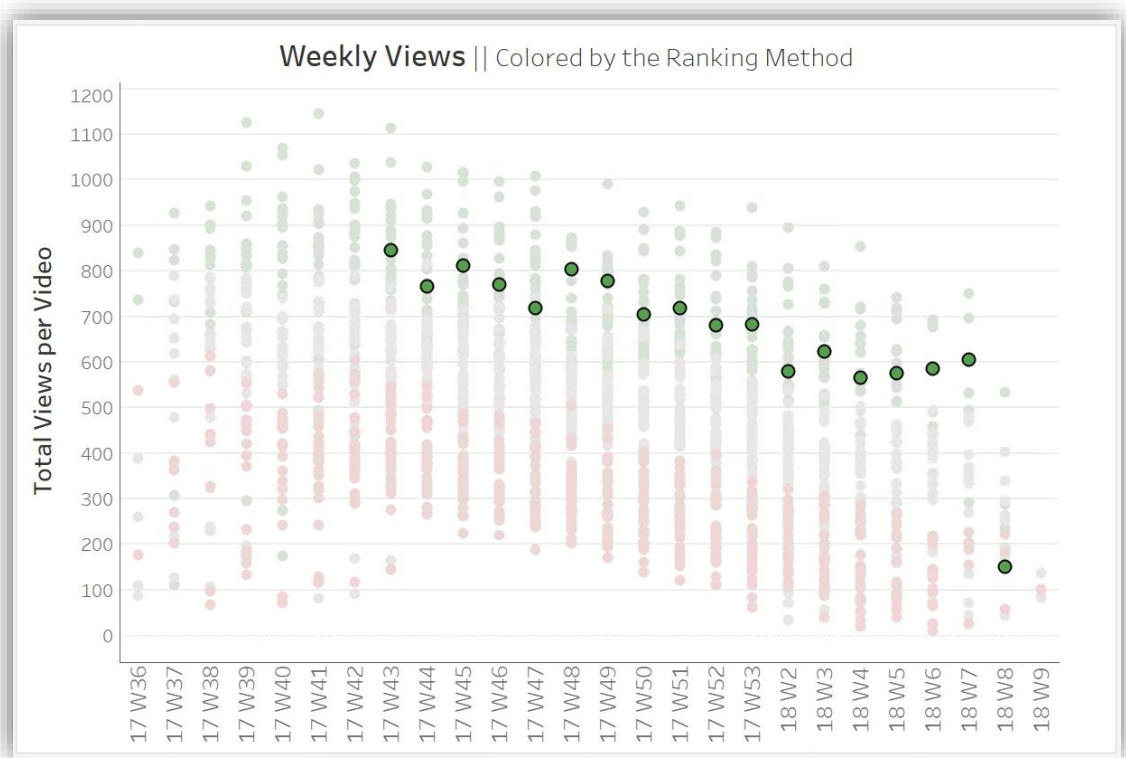
- מסקנה: ככל שעובר הזמן מהעלאה - יצפו פחות בסרטון



- **מסקנה נוספת: לפרמטר איכות הסרטון משקל רב יותר מאשר זמן היותו "באוויר"**
- ניתן לראות מסקנה זו גם בתרשים הבא:



מסקנה נוספת: גם כאן נראה הדיוק הרב בשיטת הסיווג. בתרשים נראית בבירור הסיווג של הסרטונים הנצפים ביותר לאורך זמן כקבוצה המובילה. וגם במקרים בודדים בהם סרטון המסווג גבוה (ירוק) נראה בתחתית הדירוג השבועי, הוא אינו משקף את מאפייניו השולטים בשאר הזמן של הסרטון, כפי שמודגם להלן:



בתרשים: סימון הופעותיו של סרטון במזהה 62. מהעשירונים העליונים של כמעט כל השבועות מלבד חריגה אחת בלתי משקפת