**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**

1. קטלוג 100 הסרטונים הקיימים לקטגוריות הבאות:
   * "Hot"
   * "Stable and Popular"
   * "Everything else"
2. דקירת הפיצ'רים המתאימים לכל קטגוריה
3. 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
* איכות גבוהה

**תהליך עבודה**

* + - 1. 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
      2. בניית שדות מחושבים לפי המדדים הנ"ל בתוך טבלת Video Features
      3. אקספלורציה – פילוח הנתונים לפי השדות השונים כדי למצוא דפוסים וקורלציות בין מימד לנתוני הצפיות
         * יחס בין אורך לצפיות
         * יחס בין איכות לצפיות
         * ממוצע צפיות לסרטון לאורך זמן מעלייתו
      4. מציאת תכונות משותפות בתוך קבוצת הסרטונים הבולטים בנתוני הצפיה, וכן בתוך בעלי הנתונים החלשים
         * קיבוץ שבועי, דירוג הסרטונים לכל שבוע
         * פיזור הסרטונים על ממוצע סטייה שבועית (כמות צפיות שבועית לעומת הממוצע השבועי לסרטון) על כמות צפיות כוללת
      5. מתוך התוצאות, נדרג את הסרטונים שלנו ונאמן מודל סיווג שיקבל נתוני מסגרת ויתן פלט של סיווג לפי הקטגוריות הרצויות
         * Random Forest
         * לאמן עם קלט הכולל נתוני צפיה
         * לאמן עם קלט ללא נתוני צפיה