

# פרוייקט בקורס וויזואליזציה של מידע

תום תלמי, עידו דהן

## 1. מבוא:

נושא הוויזואליזציה: התפלגות העדפותיהם המוסיקליות של משתמשי אפליקציית "Spotify" ומאפייניהם. על אילו נתונים אתם מסתמכים, מה מקורם ומה היקפם: מערך הנתונים שבחרנו מציג נתונים אודות 2000 רצועות מוסיקליות מאפליקציית "Spotify". הנתונים מתייחסים למאפיינים המוסיקליים של הרצועה וגם ציון פופולריות לכל רצועה. רצועות אלו הם להיטים נבחרים שיצאו לאור בין השנים 1999-2019. לקחנו את מערך הנתונים מאתר "Kaggle". הנתונים נאספו במקור מספריה של "Spotify" לפייתון.

## 2. נתונים:

### א. תיאור כללי:

כל רשומה במערך הנתונים מייצגת רצועה אחת שנחשבת "להיט" מבחינת "Spotify". הנתונים הם מאפיינים מוסיקליים של הרצועה + ז'אנרים אליהם משתייכת הרצועה + ציון הפופולריות של הרצועה. ציונים אלו נאספו מאפליקציית "Spotify" שהוקמה בשנת 2008 וזמינה ברוב אירופה, אמריקה, אוסטרליה, ניו זילנד ובחלקים מאסיה.

### ב. מיפוי לפי הטופולוגיה האבסטרקטית של Munzner:

Dataset type: Table

Data type - Item: Musical track

2000 items

Attribute	Attribute type	טיפוס נתונים	ערכים בנתונים	תיאור
אומן	categorical	string	835 artists	שם האומן
שיר	Key - categorical	string	2000 songs	שם השיר
משך	quantitative	int64	113000-484146	משך השיר במילי-שניות
האם התוכן פוגעני	categorical	bool	True / False	האם מילות השיר או הוידאו קליפ מכילים אחד או יותר מהקריטריונים שיכולים להחשב פוגעניים או לא תואמים לילדים.
שנה	ordinal	int64	1998-2020	השנה בה יצא לאור השיר
פופולריות	quantitative	int64	0-89	ככל שגבוהה יותר כך השיר יותר פופולרי. מחושבת לפי אלגוריתם המתבסס בעיקר על מספר השמעות וכמה הן תכופות.
יכולת ריקוד	quantitative	float64	0.129-0.975	מתארת עד כמה רצועה מתאימה לריקוד. ערך של 0.0 הוא הכי פחות ניתן לריקוד ו-1.0 הוא הכי ניתן לריקוד.
אנרגיה	quantitative	float64	0.0549-0.999	מייצגת מדד תפיסתי של עוצמה ופעילות. מדד מ-0.0 ל-1.0.
מפתח	categorical	float64	0-11	המפתח המוסיקלי של השיר.

עוצמה	quantitative	float64	-60-0	העוצמה הכוללת של השיר בדציבלים (dB). ערכי עוצמת הקול מוערכים בממוצע על פני כל השיר.
מצב	categorical	int64	0/1	מציין את המודאליות (מז'ור או מינור) של השיר. מז'ור מיוצג על ידי 1 והמינורי הוא 0.
דיבור	quantitative	float64	0.0232-0.576	מזהה נוכחות של מילים מדוברות ברצועה. ככל שכולל יותר דיבור הערך יהיה קרוב ל-1.
אקוסטיות	quantitative	float64	0-0.976	מדד ביטחון בין 0.0 ל-1.0 האם המסלול הוא אקוסטי. 1.0 מייצג ביטחון גבוה שהמסלול הוא אקוסטי.
אינסטרומנטליות	quantitative	float64	0-0.985	נבא אם רצועה אינה מכילה שירה. ככל שערך האינסטרומנטליות קרוב יותר ל-1.0, כך גדל הסיכוי שהרצועה לא מכילה תוכן ווקאלי.
הופעה חיה	quantitative	float64	0.0215-0.853	מזהה נוכחות של קהל בהקלטה. ערכים גבוהים יותר מייצגים סבירות מוגברת שהמסלול בוצע בשידור חי.
ערכיות	quantitative	float64	0.0381-0.973	מדד בין 0.0 ל-1.0 המתאר את החיוביות המוזיקלית שמעבירה רצועה. רצועות עם ערכיות גבוהה נשמעות חיוביות יותר (למשל שמחות, עליזות, אופוריות), בעוד רצועות עם ערכיות נמוכה נשמעות שליליות יותר (למשל עצובות, מדוכאות, כועסות).
טמפו	ordinal	float64	0-220	הקצב הכולל המשוער של רצועה בפעימות לדקה (BPM). בטרמינולוגיה מוזיקלית, קצב הוא המהירות או הקצב של יצירה נתונה, והוא נובע ישירות ממשך הפעימה הממוצע.
ז'אנר	categorical	string	14 genres	הז'אנר של הרצועה. יתכנו רצועות המשתייכות ליותר מז'אנר אחד..

• אין ערכי null במערך הנתונים.

### 3. מטלות:

א. מיפוי לפי המונחים האבסטרקטיים בהתאם לטופולוגיה של Munzner:

תיאור	action	target
מה מאפיין את המוסיקה שאנשים אוהבים לשמוע מבחינת מאפיינים מוסיקליים, תוכן ופונקציונליות?	Present, discover	Trends
מה מאפיין כל אחד מהז'אנרים? כלומר, למה באמת מתכוונים כשאומרים "פופ" או "רוק"?	Present, discover	Trends, dependencies
איך נראית ההתפלגות של הטעם המוסיקלי? האם יש רוב ברור לז'אנר או למאפיין ספציפי? או שבכלל ההתפלגות יותר קרובה לאחידה?	Present, discover	distribution
אילו מאפיינים מוסיקליים מופיעים בדרכ יחד? כלומר, האם קיימת קורלציה בין המאפיינים המוסיקליים?	derive	correlation
באילו מגמות ניתן להבחין על ציר הזמן בהקשר העדפות מוסיקליות של משתמשי האפליקציה? אולי ניתן, דרך מוסיקה, להגיד משהו על שינויים הקורים בחברה?	discover	Trends
מי חתום על היצירות המוסיקליות הפופולריות ביותר, ומה מאפיין את אלו?	discover	features

#### 4. רעיונות ועיצובים אלטרנטיביים:

- לשים תמונה של האומן או פרטים כגון מין, גיל, מדינת מוצא.
- לאיזה ז'אנר משוייכים זמרים? גרף שמראה עבור כל אמן pie chart של מספר שירים שהוא שר מכל ז'אנר.
- קורלציה בין כל שני מאפיינים – מפת חום שבקליק ימני מציגה גרף של התכונות בשני צירים ואולי גם עם קו רגרסיה.
- להשוות בין הז'אנרים בכל המאפיינים, כדי לקבל אפיון עבור ז'אנר בודד.

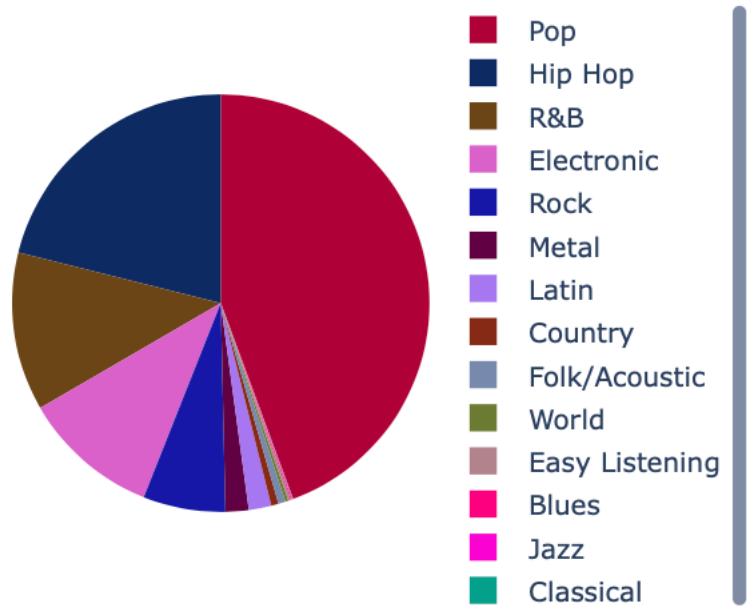
#### 5. הסבר על היישום:

א. עיבוד מקדים:

- את עמודת "האם התוכן פוגעני" המרנו מערים בוליאניים True/False לערכים נומריים 0/1.
- את הערכים הנומריים שאינם ביחידות מידה מוכרות המרנו לטווח בין 0-1.
- נרמלנו את התכונות הנומריות עם minmaxscaler
- עבור שאילתות מסוימות להן הזדקקנו בוויזואליזציה, יצרנו מסט הנתונים סט חדש בו עמודת הז'אנר מפוצלת. בהתחלה בעמודה זו היו ערכים מרובים, יצרנו רשומה עבור כל ערך של ז'אנר.  
למשל:  
 $(\text{Its My Life} \rightarrow \text{rock, metal}) \Rightarrow (\text{Its My Life} \rightarrow \text{rock}), (\text{Its My Life} \rightarrow \text{metal})$



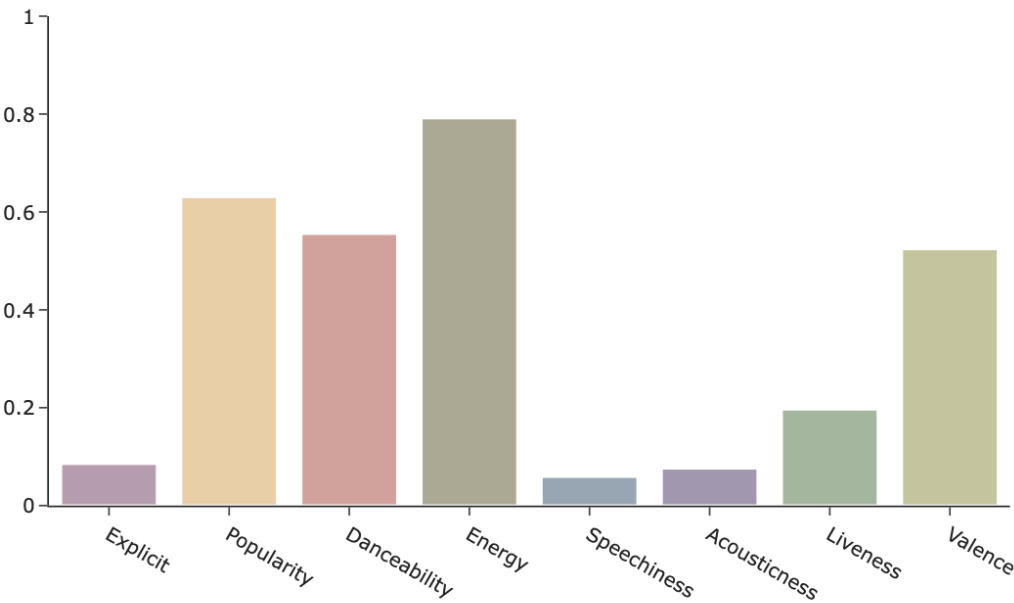
## Genre Distribution:



Rock

Rock Statistics:





Rock

Pop

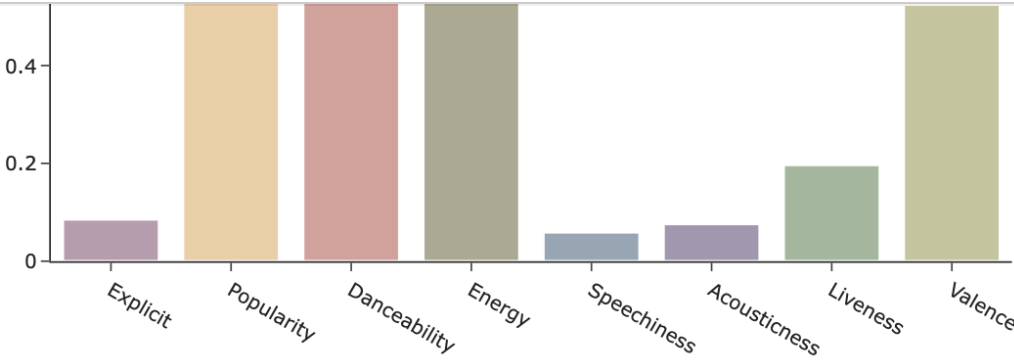
**Rock**

Country

Metal

Hip Hop

R&B

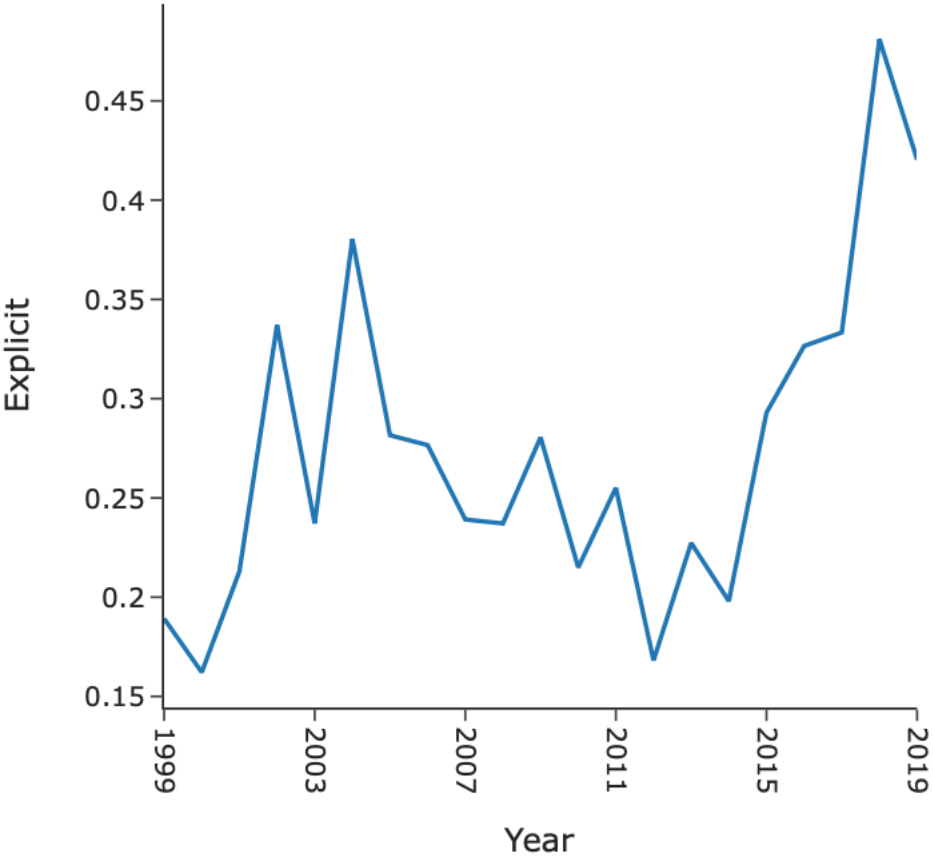


Explicit

×

▼

Avrage Hits Explicit Over The Years:



Explicit Presenae In Genre:

