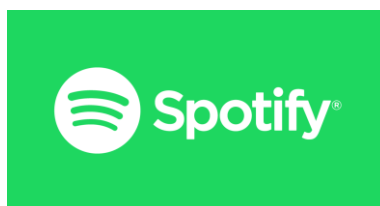


Sentiment-Based Spotify Playlists



Eran Kan, Amotz Levy

תוכן עניינים

2	• הקדמה
2	• מסדני הנתונים
2	• איך מחליטים מה הוא שיר שמח ?
3-6	• ניתוח מאפייני השירים
6	• מציאת דמיון ובניית הפלייליסט
6	• תיאור הניסוי
7	• תוצאות הניסוי
8	• תובנות מתוצאות הניסוי
8	• עבודה עתידית
8	• סיכום
9	• הפלייליסטים שייצרנו בספוטיפיי
9	• קישורים

הפופולריות של שירותי צריכת מוזיקה ישירה (Streaming) כמו Spotify ו-Apple Music נמצאות בעליה בשנים האחרונות. שירותים אלו מאפשרים למשתמשים לבנות פלייליסטים לפי קריטריונים שונים שלאחר מכן ניתן לשתף. כמו כן, בספוטיפיי, לכל שיר יש מספר מאפיינים (features) המתארים את תכונות השיר [קישור לתיאור המאפיינים בסוף], כגון טמפו, אנרגיה, עוצמה וכדומה. מטרת הפרויקט שלנו היא לבצע ניתוח של השירים בספוטיפיי לפי המאפיינים השונים כדי לאפיין שירים על פי רגשות שונים כמו שמחה, עצב וכעס, ועל סמך הניתוח הזה לבנות פלייליסטים חדשים המבוססים על רגש מסוים.

מסדי הנתונים

בשביל הפרויקט השתמשנו בשני מסדי נתונים שונים של ספוטיפיי:

- ספוטיפיי מאפשר חיפוש ומעבר על פלייליסטים לפי קטגוריות שונות. על מנת להבין אלו פיצ'רים מאפיינים שיר ברגש מסוים ביצענו crawling לפלייליסטים של ספוטיפיי בארה"ב הנמצאים תחת הקטגוריה של Mood וסיננו אותם לפי מילות מפתח (הסבר מפורט בהמשך). באופן זה יצרנו 3 קבצים המכילים שירים ברגש מסוים: הקובץ עם השירים השמחים מכיל 1059 שירים, עם השירים העצובים מכיל 474 שירים וזה עם השירים הכועסים 210 שירים.
- על מנת לבנות פלייליסטים חדשים התואמים רגש מסוים לקחנו מסד נתונים מ Kaggle המכיל כ- 116,000 שירים שונים ועודכן לאחרונה בדצמבר 2018. כל שיר במסד מכיל את העמודות הבאות: Artist_name, track_id, track_name, acousticness, danceability, duration_ms, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time_signature, valence, popularity.

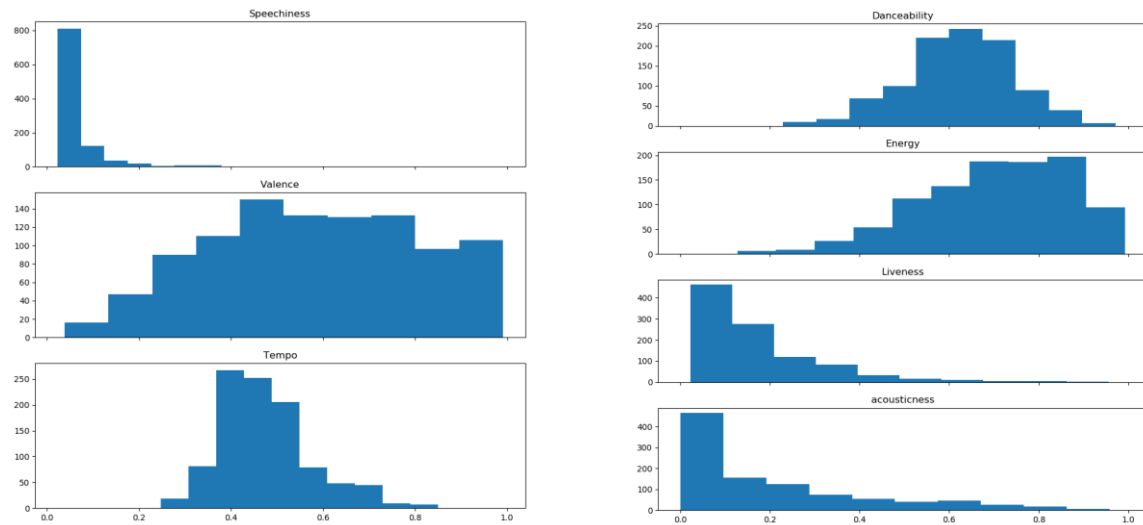
איך מחליטים מה הוא שיר שמח ?

על מנת שנוכל לבצע ניתוח של שירים ולהבין מה מאפיין שיר ברגש מסוים, לקחנו פלייליסטים הנמצאים תחת הקטגוריה של "מצב רוח" (Mode) בארה"ב וסיננו אותם לפי מילות מפתח. כך עבור אפיון שיר שמח, חיפשנו פלייליסטים המכילים את המילים: {happy, happier, happiest, good, great, joy}. עבור פלייליסט אפיון שיר עצוב חיפשנו פלייליסטים המכילים את המילים {'Sucks', 'Down', 'Dark', 'Sad', 'Loss', 'low'} ועבור אפיון שיר כועס חיפשנו פלייליסטים המכילים את המילים {rage, rocking, bits, storm, angry, anger}.

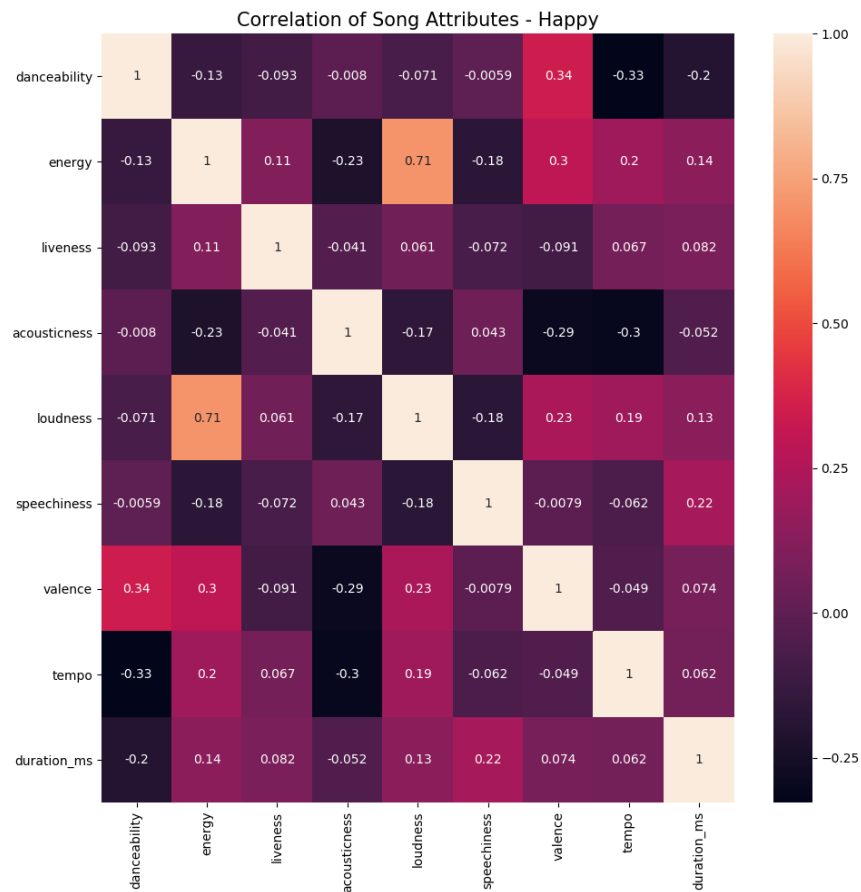
לאחר מכן, לקחנו את כל השירים עבור כל רגש ובדקנו האם יש קורלציה מסוימת בין המאפיינים השונים של השירים עבור כל אחד מהרגשות, ובנוסף בנינו היסטוגרמה של כמות השירים בכל פיצ'ר ובכל ערך.

ניתוח מאפייני השירים

עבור השירים השמחים



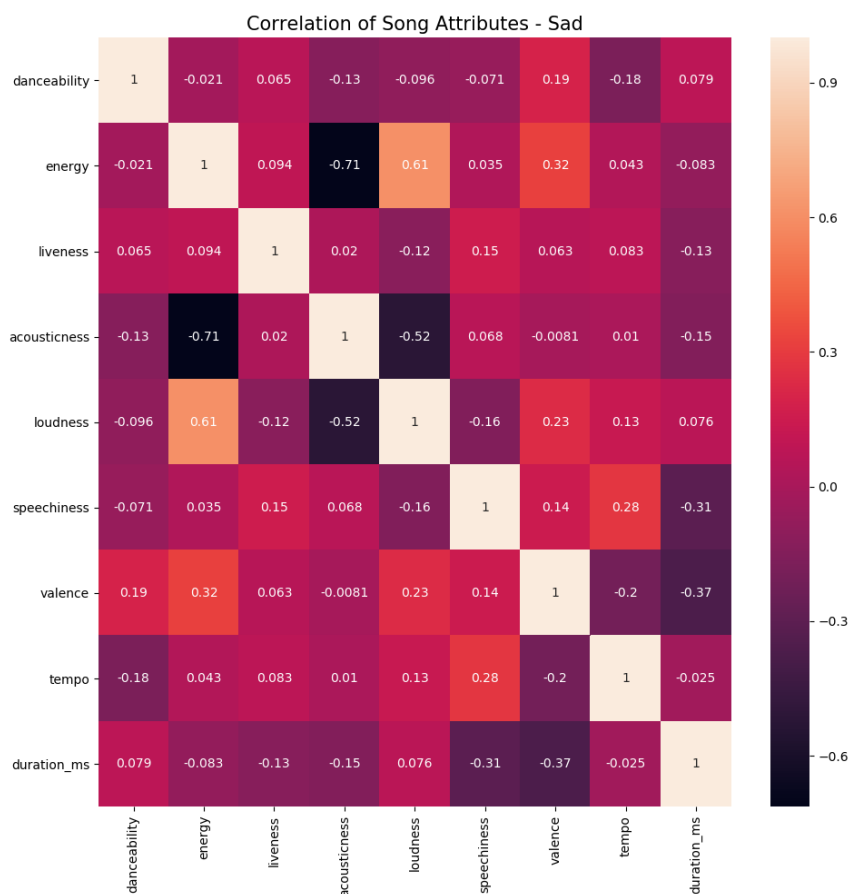
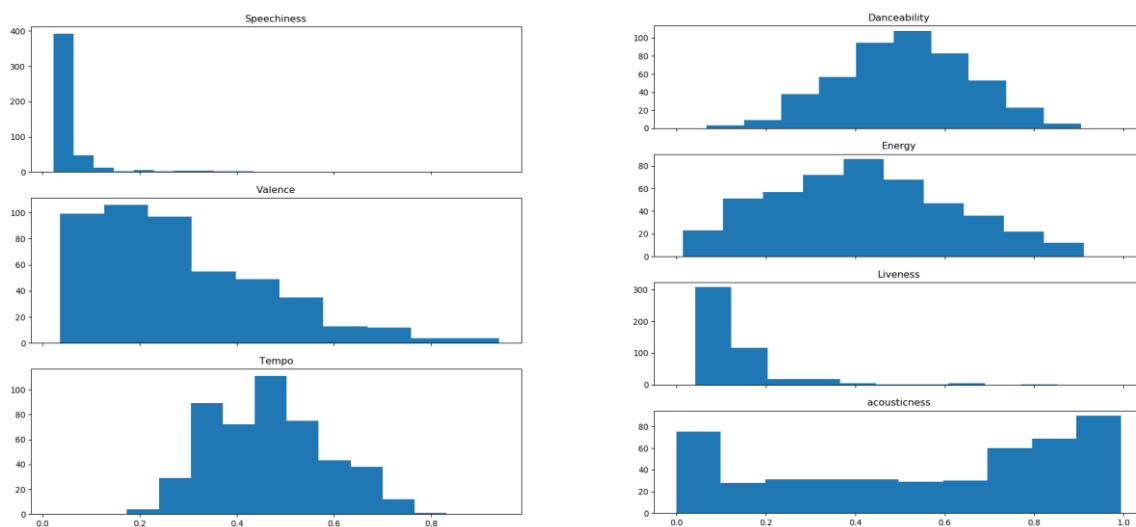
הסבר לגרף: ערכי כל הפיצ'רים נעים בין 0 ו-1 וגובה העמודה מתאר את מספר הישירים עבור הערך של הפיצ'ר, השם כל פיצ'ר מתואר מעל ההיסטוגרמה שלו.



הסבר לגרף: הגרף מתאר את מידת הקורלציה בין פיצ'רים של 50 שירים "שמחים" רנדומליים לבין עצמם. כך שכל שצבע המשבצת לבן יותר כך יש התאמה גבוהה יותר עבור שני מאפיינים.

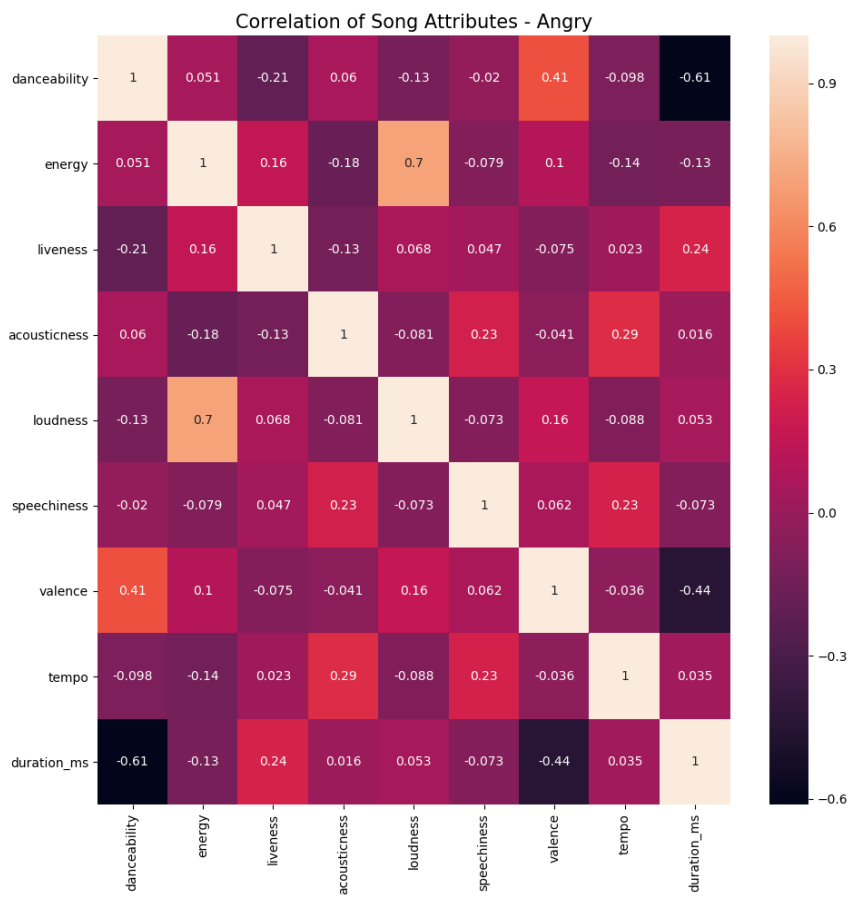
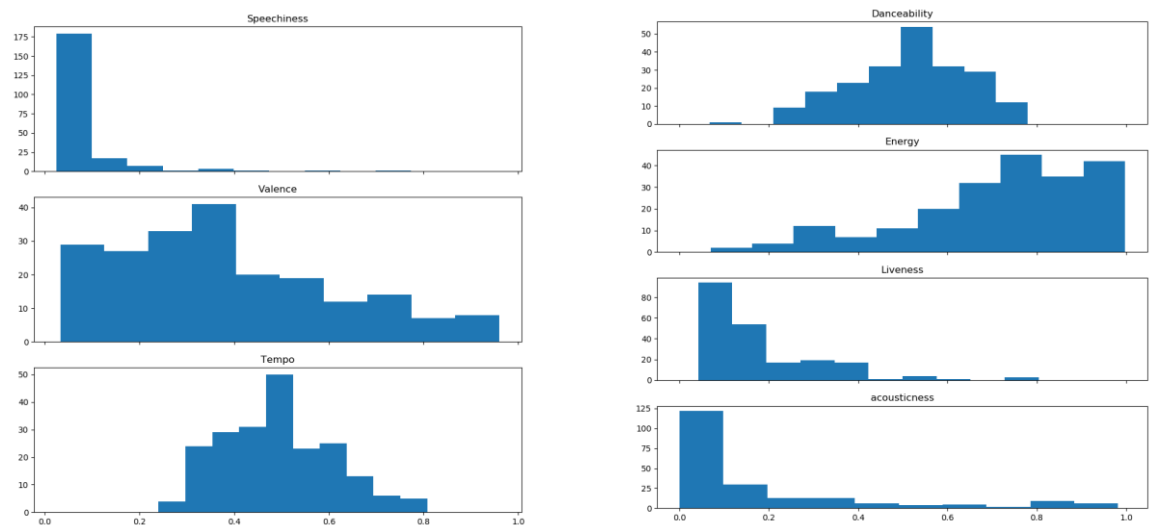
כך למשל, ניתן לשים לב כי ישנה התאמה יחסית גבוהה בין energy ו- valence, danceability ו- energy וכן רוב השירים השמחים מתאפיינים ברמות גבוהות יחסית של ערכים אלו לכן הווקטור המתאר המצומצם של שירים שמחים יכלול את הפיצ'רים האלו. (ניתן לראות התאמה גבוהה בין energy ו- loudness אך מכיוון ש- loudness נמדד ב-db ולא נמצא בין 0 ו- 1 החלטנו שלא להשתמש ב- loudness בווקטור המתאר המצומצם).

עבור השירים העצובים



גם עבור השירים העצובים ניתן לראות שרמת ההתאמה בין valence, energy ו-danceability נשמרת, אך כמו שניתן לראות בהיסטוגרמות של פיצ'רים אלו, הערכים שלהם נמוכים בהרבה בהשוואה לשירים השמחים. לכן גם במקרה זה הווקטור המצומצם יכול ללכוד את הפיצ'רים האלו.

עבור השירים הכועסים



עבור השירים הכועסים ניתן לראות כי ה-energy גבוה מאוד (אפילו יותר מהשמח) אבל ה-valence דומה יותר לזה של השירים העצובים. מכיוון שהערך של ה-energy כל כך גבוה החלטנו להשתמש בו בוקטור המצומצם, אפילו שההתאמה שלו עם ה-valence לא אופטימלית.

כתוצאה מכך, לקחנו את כל השירים מפלייליסטים בעלי רגש מסוים ובנינו שני ווקטורי פיצ'רים המתארים את הרגש: אחד הכולל את הערך הממוצע של כל השירים עבור כל פיצ'ר ואחד את הערך ממוצע רק עבור פיצ'רים נבחרים שמצאנו קשר דומיננטי בינם לבין עצמם עבור הרגש הנתון [valence, energy, danceability]. נציין שבדקנו גם ווקטורים מתארים שבמקום לקחת את הערך הממוצע של כל פיצ'ר לקחו את החציון של כל ערכי הפיצ'ר של כל השירים, אך רמת ההתאמה הייתה נמוכה יותר כאשר חיפשנו שירים שתואמים למתאר ולכן בחרנו בוקטור הממוצעים.

מציאת דמיון ובניית הפלייליסט

לאחר שבנינו שני ווקטורים מתארים לכל רגש, ביצענו מעבר על מסד הנתונים הגדול: בשלב הראשון סיננו שירים על ידי פופולריות וזאת מכיוון שרצינו ליצור פלייליסטים איכותיים יחסית, באופן דומה יכולנו דווקא לבחור שירים עם פופולריות נמוכה כדי להציג למשתמש שירים שבסבירות גבוהה יותר הוא לא מכיר. לאחר מכן ביצענו בדיקת דמיון בין הווקטורים המתארים של הרגש ובין וקטור הפיצ'רים של כל שיר. בדקנו שתי שיטות של דמיון: Cosine similarity ו-Manhattan similarity (cosine נתן לנו רמת התאמה טובה יותר ולכן בחרנו בו) ובחרנו את 10 השירים שהיו בעלי ההתאמה הגבוהה ביותר.

תיאור הניסוי

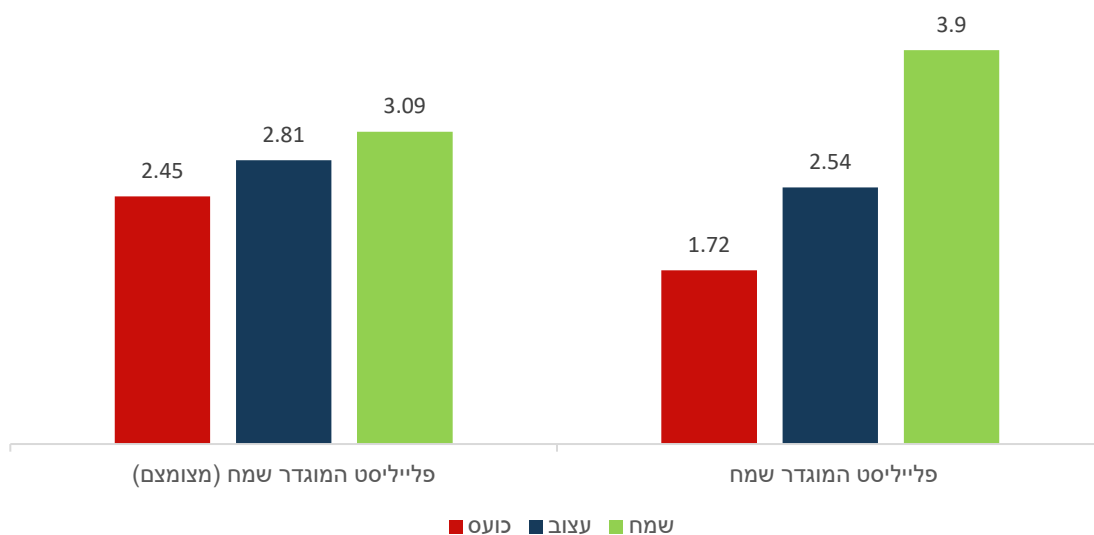
מכיוון שאנחנו לא מכירים אף מדד איכותי לכמה שיר הוא "שמח" או "עצוב" החלטנו לערוך סקר ולבדוק. בנינו 6 פלייליסטים של 10 שירים כל אחד: לכל רגש בנינו שניים, אחד של ההתאמה עבור וקטור מתאר הכולל את כל הפיצ'רים ואחד הכולל רק את הפיצ'רים הדומיננטיים עבור כל רגש וביקשנו שאנשים ידרגו בין 1 ל-5 עד כמה הפלייליסט מתאר עבורם שמחה, עצב וכעס.

כדי למדוד הצלחה, בדקנו עבור כל פלייליסט מה הערך הממוצע שקיבל כל רגש והאם הערך האמיתי אכן יותר גבוה מדירוג הרגשות האחרים (לדוגמה, האם הדירוג הממוצע של "עצב" גבוה יותר מזה של "שמחה" וכעס" עבור פלייליסט שלפי השיטה שלנו מוגדר כעצוב).

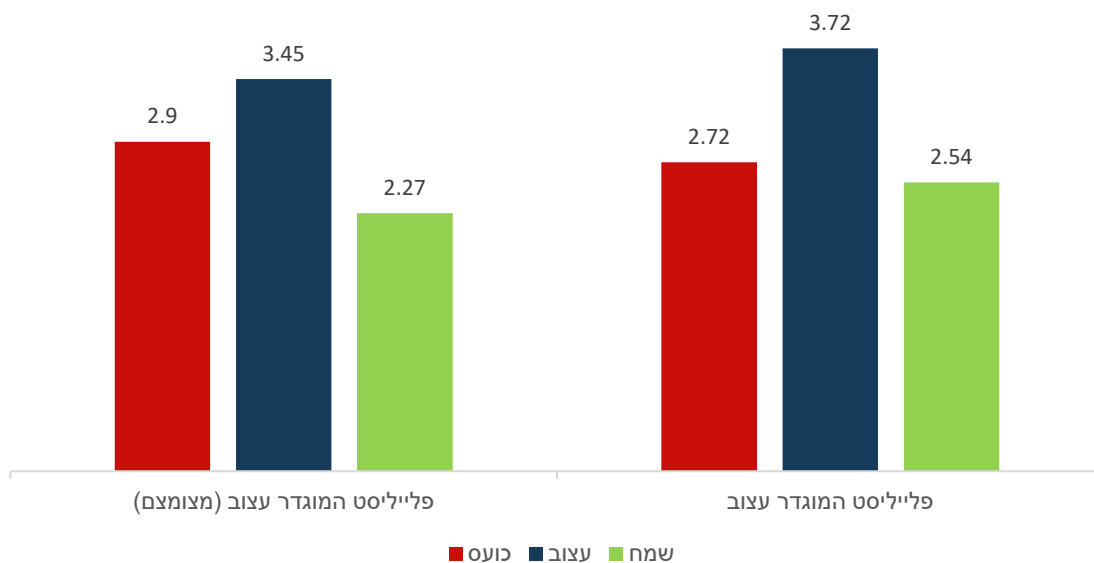
על מנת לוודא שהתוצאות שלנו לא יהיו מקריות, הפלייליסטים עורבבו ולא הוצגו לפי סדר מסוים ובנוסף ניתנו להם שמות כלליים של `playlist_x` כאשר x הוא מספר.

בחרנו להציג את הערך הממוצע של כל רגש בצבע אחר: ירוק עבור הערך הממוצע של רמת ה"שמחה", כחול עבור "עצב" ואדום עבור "כעס". כמו כן בחרנו בתרשים של עמודות המאפשר להבחין בהבדלים של הערכים השונים עבור כל פלייליסט, כך שגובה העמודה מייצג את הערך הממוצע שהרגש קיבל בפלייליסט. בנוסף, שמנו כל שני פלייליסטים המתארים רגש מסוים אחד ליד השני וזאת על מנת שיהיה קל יותר לראות את ההבדל בין שתי השיטות שתיארנו קודם.

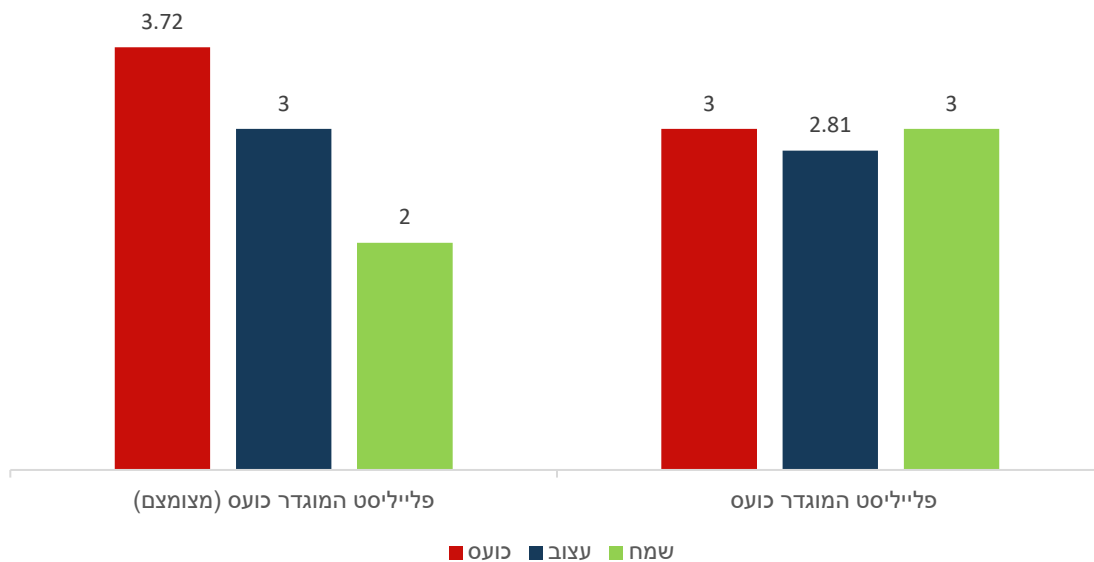
פלייליסטים שהשיטה שלנו הגדירה כ"שמחים"



פלייליסטים שהשיטה שלנו הגדירה כ"עצובים"



פלייליסטים שהשיטה שלנו הגדירה כ"כועסים"



תובנות מתוצאות הניסוי

ניתן לראות כי עבור הפלייליסטים השמחים והעצובים, ממצעי הסקר תואמים את הרגש שהשיטה שלנו נתנה לכל אחד מהם ואכן הערך הממוצע של "שמחה" היה הכי גבוה בפלייליסט השמח והערך הממוצע של "עצב" היה הכי גבוה בפלייליסט העצוב. למרות זאת ניתן לראות כי בדיקת ההתאמה עם הוקטור המצומצם של הפיצ'רים (לעומת זו הכוללת את כל הפיצ'רים) דווקא הביאה לירידה בערך הממוצע של הרגש, שכן ערך ה"שמחה" ירד מ-3.9 ל-3.09 בין הפלייליסטים השמחים ואילו ערך ה"עצב" ירד מ-3.72 ל-3.45 בין הפלייליסטים העצובים. לעומת זאת, ניתן לראות כי השיטה המורחבת בפלייליסט הכועס לא הצליחה להפריד בין שמחה וכעס, אך למרות זאת ערך הכעס של פלייליסט זה (3) עדיין גבוה יותר מערכי הכעס בפלייליסטים השמח (1.72) והעצוב (2.72), כלומר הוא דורג ככועס יותר אפילו שהשיטה לא הצליחה להביא למצב חד משמעי כמו במקרים הקודמים. לעומת זאת, ניתן לראות כי בניגוד לשמחה ועצב, השיטה של הוקטור המצומצם דווקא הביאה לבניית פלייליסט שהאנשים שדירגו את הסקר מצאו כ"כועס" יותר מאשר זה של השיטה שהתחשבה בכל הפיצ'רים.

עבודה עתידית

בפריקט זה ניסינו לעשות ניתוח ולסווג שירים המעבירים רגשות כמו שמחה, עצב וכעס ובעתיד ניתן לבצע ניתוח של רגשות רבים נוספים. כמו כן, על ידי ניתוח הרגש ניתן לבנות פלייליסט הלוקח אותו בחשבון, כך למשל ניתן לבנות פלייליסט שמתחיל עצוב ולאט לאט עובר לשירים שיש להם התאמה גדולה יותר לשמחה וזאת על מנת לשמח את המאזין. כיוון אפשרי נוסף הוא לבצע ניתוח רגש לשירים שיש למשתמש בפרופיל שלו ולהמליץ לו על שירים נוספים, התואמים את הרגשות של שירים אלו, אך שאינו מכיר.

סיכום

בפריקט זה ניסינו להגדיר שיר על פי רגש מסוים, תוך כדי התחשבות במאפייני השירים של פלייליסטים קיימים בספוטיפיי התואמים למצב רוח מסוים. לאחר שבנינו פרופיל לכל רגש, ניסינו למצוא על ידי דמיון את השירים שהכי טועמים את הרגש ולהרכיב פלייליסט שלם ברוח אותו הרגש. בנוסף, ביצענו סקר שבו אנשים דירגו את רמת הרגש של כל פלייליסט, וזאת על מנת ליצור מדד איכות לשאלה: "כמה הפלייליסט משקף את הרגש?".

הפלייליסטים שייצרנו בספוטיפיי (האזנה מהנה D:)

- שמח (כל הפיצ'רים):
<https://open.spotify.com/playlist/2HyP0eQIzE5VCMsfXdjrd3>
- שמח (וקטור פיצ'רים מצומצם):
<https://open.spotify.com/playlist/68PsQVivYl8ZYbp3HoERfj>
- עצוב:
<https://open.spotify.com/playlist/6rCVTIaDybhqXcwihMxyzi>
- עצוב (מצומצם):
<https://open.spotify.com/playlist/6tmzdhcBROnwSPLInNN9Vq>
- כועס:
<https://open.spotify.com/playlist/5H8EiFyDN9U6l1DWpvaZM4>
- כועס (מצומצם):
<https://open.spotify.com/playlist/7MPxl7pi6C6vnGG5SGjESA>

קישורים

- קישור למסד הנתונים של כלל השירים שלקחנו מ Kaggle -
<https://www.kaggle.com/tomigelo/spotify-audio-features>
- הסבר על הפיצ'רים השונים של שיר בספוטיפיי -
<https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features>