**פרויקט סיום בקורס 'כריית נתונים'**

*ניתוח שיח ברשתות חברתיות של אנשים חירשים וכבדי שמיעה בעת מלחמה""*

מרצה: ד"ר אבי יוסיפוף

מגישים:

205660525 שמעון טורצק

318717535 אורטל סלמן

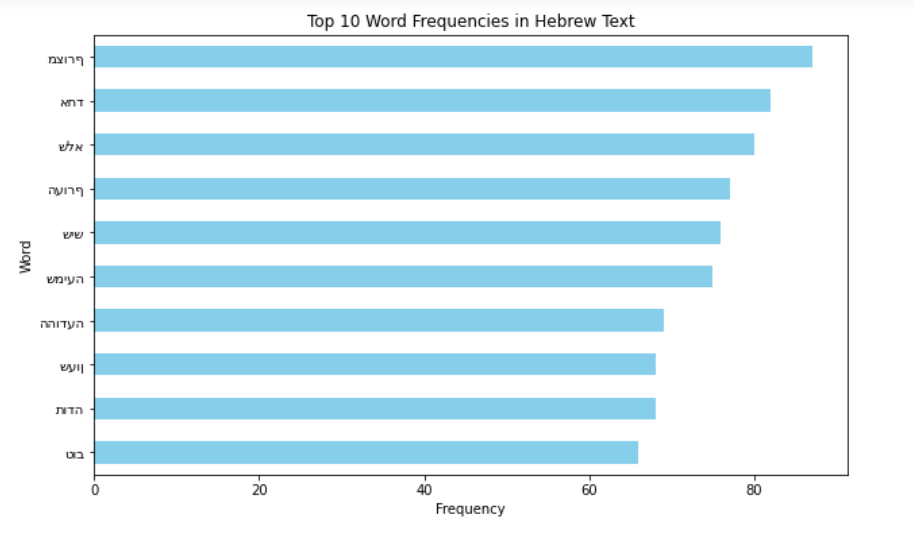
313197014 נתנאל חג'בי

**תובנות:**

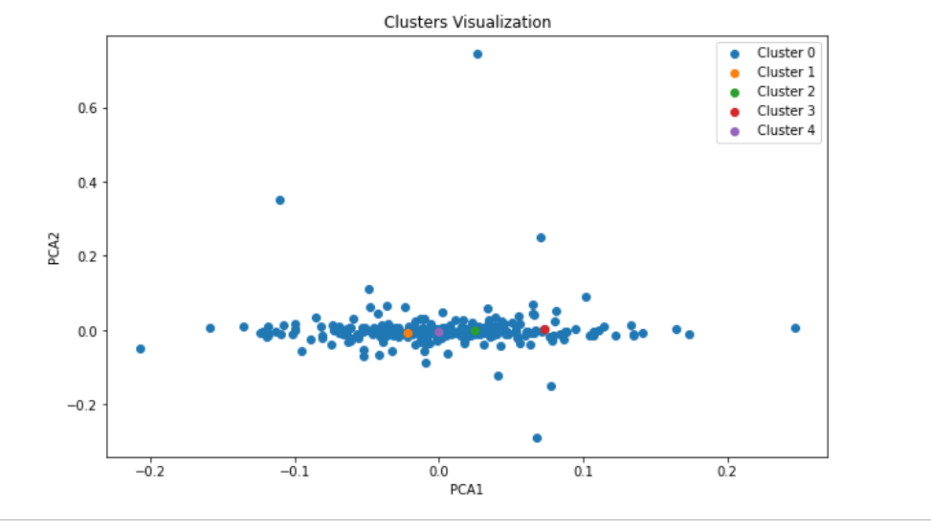
1. מילים שכיחות בטקסט
2. נושאי שיחה עיקריים.
3. רשימת האנשים הכי נוכחים בקבוצה.
4. שעות פעילות הקבוצה.
5. תעבורת הודעות גבוהה על פי תאריכים שונים.
6. אבחון משפטים על פי קריטריון רגשי.

**הרחבה:**

1. **מילים שכיחות בטקסט:**מצאנו את 20 המילים הכי שכיחות בטקסט לאחר הורדת מילות חסרות משמעות (כמו מילות קישור): טוב, שעון, תודה, ההודעה, שמיעה, שיש, העורף, שלא, אחד, מצורף.

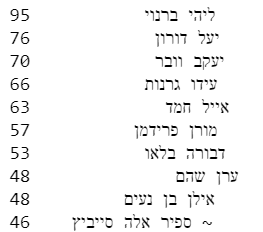


1. **נושאי שיחה עיקריים:**   
   רצינו לדעת מה היו התכנים הכי נפוצים בקבוצת הוואטסאפ ולכן השתמשנו בשיטות כמו k-means, PCA ו-clustering כדי לנסות לחלק את המידע שקיבלנו לקבוצות מרכזיות על פי התוכן. העברנו את הטקסט שקיבלנו ל'קורפוס' וביצענו עליו וקטוריזציה למילים ולמשפטים. בתמונה למטה ניתן לראות כי האלגוריתם חילק את הטקסט לכמה קבוצות מרכזיות ועל פי wordCloud ייצגנו את הקבוצות והמילים שבתוכן:





בתמונה הנ"ל ייצגנו cluster מסוים וניתן לראות כי המילים המרכזיות (מיוצגות בגדול) מעידות על כך שהתקשורת היא סביב ביטחון ושמיעה של האזעקה וההתרעות באפליקציות מסוימות.

1. **תמונה שמכילה טקסט, גופן, קו, צילום מסך

   התיאור נוצר באופן אוטומטירשימת האנשים הכי נוכחים בקבוצה:**   
   לאחר הבנה בסיסית של בסיס הנתונים והכנתו, מצאנו מי הם האנשים הכי 'מקובלים' בקבוצה, כלומר מי שולח הכי הרבה הודעות ועל מה. אלו התוצאות שקיבלנו:

כפי שניתן לראות, אישה בשם 'ליהי ברנוי' נמצאה כמי ששולחת הכי הרבה הודעות בצ'אט. המשכנו להעמיק ולבדוק מהו התוכן עליו היא כותבת ואלו מסקנות ניתן ללמוד מתוכן זה:

ליהי ברנוי מדברת על ביטחון, שעון, אחריות, סייבר. מרבה להשתמש בביטויים רגשיים כמו: ביטויים "הרע במיעוטו" או "בגלל האויב לא פשוט". היא מדברת על כך שההתרעות, כנראה בפלאפון, עובדות טוב אבל בשעונים פחות (בהמשך ניתן לראות כי קיימת שיחה ארוכה מאוד על איזה שעון כדאי לקנות ומה המחיר שלו).

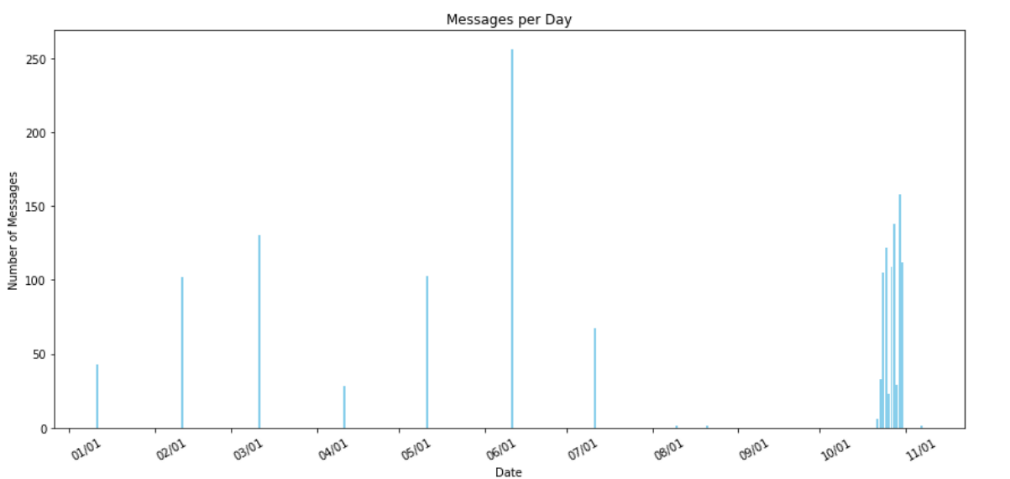
ליהי מצטיירת כאישה מבוגרת יותר מהשאר עקב ביטויים שונים ושפה גבוהה-"הרי". הסקנו כי היא אישה דעתנית, חזקה, טעונה ומובילה חברתית, חשוב לה שהקבוצה תהיה פעילה ושכל אחד ידע ויממש את זכויותיו. היא משמשת כעוגן ממשי לשאר חברי הקבוצה וניתן לראות זאת על פי משפטים כמו: "גם אני אוהבת את עידו ומצדיעה לעידו על פועלו למען הקהילה הוא האיש שאפשר באמת לסמוך הוא האיש הכי נגיש זכינו בך", או: "גאה ביעקוב אולאיל על היוזמה... וזה מבורך כן ירבו כמוהו....הפריע לי שברווחה לא מבינים שמגיע לכל חירש גם אם יש אדם שומע בסביבה....עובדה התלוננו על זה. וזה לא באשמתך".

1. **תמונה שמכילה טקסט, קו, עלילה, צילום מסך

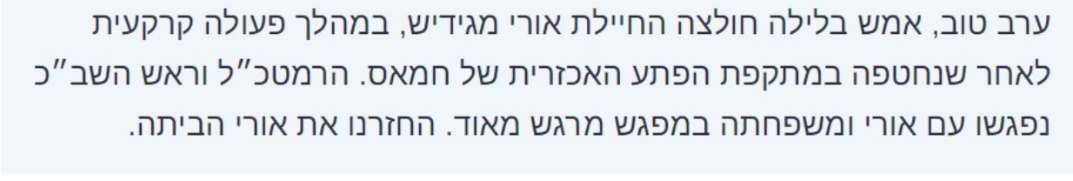
   התיאור נוצר באופן אוטומטישעות** **פעילות הקבוצה:**

בדקנו את פיזור ההודעות לפי שעות היממה והגענו למסקנה כי הקבוצה פעילה יותר בשעות הערב, ניתן לשער כי אנשי הקבוצה עסוקים במהלך שעות הבוקר והצהרים וככל הנראה נמצאים במעגל העבודה, למרות היותם כבדי שמיעה או חירשים.

1. **תעבורת הודעות גבוהה על פי תאריכים שונים:**

לאחר התייעצות עם אבי, יצרנו time-stamp שבו בדקנו את מספר ההודעות ביחס לכל יום. בגרף הבא נוכל לראות כי הימים: 30.10 ו-6.11 היו ימים עם התעבורה הגבהה ביותר:

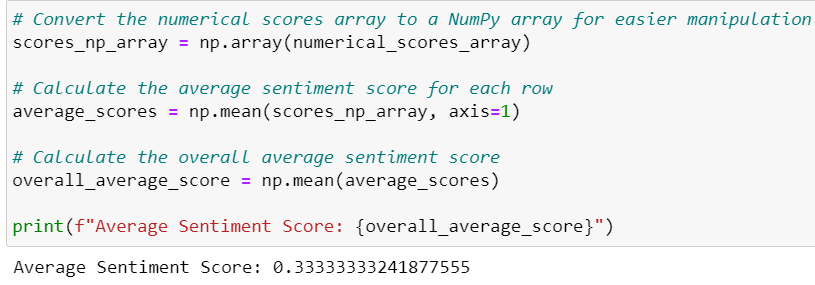
בעקבות כך בדקנו מה התרחש ב-30.10 החיילת אורי מגידיש שבה הביתה ב"ה:



בנוסף לכך ב-6.11 התרחשו שיחות רבות על השעון המדובר.

1. **אבחון משפטים על פי קריטריון רגשי:**

השתמשנו ב- heBERT\_sentiment\_analysis כאלגוריתם שמצבע 'חתימה רגשית' לכל משפט בקורפוס (כמובן לאחר tokenization) והסקנו כי רוב התוכן שעלה מן הצ'אט סומן כתוכן שלילי. מסקנה זו הייתה הגיונית עקב המלחמה.



לסיכום, ניתן להסיק כי האנשים מהמאגר עסוקים במה שרוב הישראלים בימנו עסוקים – המלחמה והחיים לצידה. הם דואגים אחד לשני, תומכים ומשתמשים בקבוצה כמקור למידע על זכויות, התרעות וחברה.

פרויקט זה היווה צוהר לעולמם של החירשים וכבדי השמיעה בארצנו בזמן מלחמת 'חרבות ברזל'. העבודה עם מאגר הנתונים ועיבודו תרמה לנו בהבנת אלגוריתמים, הקשורים ללמידת מכונה, באופן מעמיק יותר. נהנו מאוד לעבוד יחד.

[**קישור לקוד**](https://github.com/OrtalGitHub/DataMiningWarTime.git)