ניתוח מאגרי מידע - מטלת סיכום

דנה אדלר - [dana.adler1@mail.huji.ac.il](mailto:dana.adler1@mail.huji.ac.il), dana.adler

ליה שטראוס - [lia.strauss@mail.huji.ac.il](mailto:lia.strauss@mail.huji.ac.il), liastrauss

מאיה שטראוס - [maya.strauss@mail.huji.ac.il](mailto:maya.strauss@mail.huji.ac.il), maya.strauss

תיאור המחקר

מטרת המחקר שלנו **מתמקדת באפיון הצלחת חברת וולט דיסני.** חברת וולט דיסני היא חברת בידור שנוסדה בשנת 1923 על ידי האחים וולט ורוי דיסני. מאז, החברה הפכה למעצמה בתעשיית הבידור: היא הפיקה סרטי אנימציה ולייב אקשן אייקוניים, יצרה דמויות אהובות ומוכרות, והקימה פארקים מצליחים. בעבודה זו, שאפנו לנתח ולקבל תובנות ביחס להצלחת החברה, בהתבסס על מאגרי נתונים המתייחסים לסרטים שהופקו על ידיה ולביקורות וחוות דעת על הפארקים שלה. זאת, על מנת לשפוך אור על גורמי ההשפעה של הצלחת החברה שהפכה לחלק בלתי נפרד מהילדות של כולנו, ואולי אף להצליח לגזור תובנות חשובות עבור חברות דומות.

מדידת ההצלחה נעשתה במספר מישורים:

1. **רווחים:** מהם הסרטים הרווחיים ביותר של החברה? מהם הפרמטרים המשפיעים על הרווחים?
2. **יוצרים:** כיצד המפיקים נמדדים ביחס לאחרים? האם הרכב מסוים הפיק סרטים מוצלחים/ אהובים יותר?
3. **היסטוריה:** האם ניתן לזהות השפעה של אירועים היסטוריים בעולם ובתוך החברה על פעילותה של דיסני?
4. **פארקים:** מהם הפרמטרים החשובים למבקרים? מהי החלוקה הרגשית של הביקורות השונות והאם ניתן לזהות הבדלים לפי המוצא של המבקר?

מקורות המידע

בעבודתנו הסתמכנו על שלושה מאגרי נתונים שמקורם באתר Kaggle, המציע מערכי נתונים בנושאים מגוונים.

1. **Disney Movie Dataset** - מאגר מידע הכולל נתונים אודות 431 סרטים שהפיקה חברת וולט דיסני בין השנים 1937-2021. המאגר כולל שדות רבים וביניהם שם הסרט, חברת ההפקה, תאריך יציאה, אורך, תקציב, רווח הקופה, דירוג IMDb, דירוג Metascore (כיום Metacritic), דירוג Rotten Tomatoes, במאים, מפיקים ומלחינים.
2. **Disneyland Reviews** - מאגר מידע הכולל 42,656 ביקורות של מבקרי דיסנילנד בפארקים בקליפורניה, פריז והונג קונג, בין השנים 2011-2019. המאגר כולל את השדות הבאים: דירוג הפארק, חודש ושנת הביקור, מדינת מגורים של המבקר, טקסט הביקורת, הפארק אותו ביקרו. הביקורות נאספו מאתר TripAdvisor.
3. **US CPI** - מאגר המתאר את מדד המחירים לצרכן בארה"ב לכל יום בין התאריכים 01/01/1913-07/01/2021. המאגר Disney Movies Dataset כולל נתונים כספיים, המושפעים מאינפלציה. לפיכך, ביצענו נרמול של השדות המוניטריים, בכדי שנוכל לבצע השוואה ביניהם.

תיאור הפתרון

על מנת לקרוא, לעבד, לנקות ולבצע ויזואליזציה של מאגרי המידע השתמשנו בכלים הבאים:

**ספריות** matplotlib, textblob, numpy sklearn, nltk; **חבילות** re wordcloud, pandas, networkx;

**מודול** collections.

הויזואליזציות בהן השתמשנו הן סקאטר פלוט, גרף עמודות ורשתות חברתיות וענני מילים.

הצלחה יכולה להימדד לפי פרמטרים שונים, ולכן במהלך עבודתנו עם מאגרי מידע מגוונים הגדרנו הצלחה בדרכים שונות. ברוב המקרים יצאנו מנקודת הנחה שהצלחה יכולה להימדד באופן מספרי.

1. **רווחים:** השתמשנו במדד המחירים לצרכן בארה"ב (מיקום חברת וולט דיסני והמטבע בו מוגדרים הערכים), והנוסחה בה השתמשנו היא , כאשר i היא שנת הסרט ו-r הוא שיעור האינפלציה בין שנה i לשנת 2021, השנה העדכנית ביותר במאגר המידע.

הגדרנו ממוצע CPI שנתי בכדי לחשב את שיעור האינפלציה בכל שנה. הגדרה זו הכרחית לנוכח עדכניותו החלקית של המאגר בשנת 2021. נוסחת הנרמול:

1. **יוצרים:** מדדי ההצלחה שנמצאים ברשותנו הם נתוני רווח קופה ודירוג הסרטים באתרים IMDb, Metascore ו-Rotten Tomatoes. במאגר המידע, חסרים רווחי הקופה של 18% מהסרטים ולכן נאלצנו להוריד סרטים אלה במענה על שאלות הכוללות נתוני רווח קופה. עבור סרטים שאחד או יותר מהדירוגים חסר להם, חישבנו את הממוצע של הדירוגים הנותרים והכנסנו אותם בתור הדירוג החסר, מתוך הנחה שהדירוג של כל סרט נמצא בטווח דומה של מספרים ושואף לממוצע. בנוסף, ביצענו התאמה בין סולמי הדירוג השונים.
2. **היסטוריה:** הצלחה נמדדת לפי מספר הסרטים שיצאו בכל שנה. השתמשנו באלגוריתם שסוכם עבור כל שנה את מספר הסרטים שיצאו בה, תוך הוספת שנים שאינן מופיעות במאגר ונמצאות בטווח השנים 1937-2021, עם הערך 0. לאחר יצירת הגרף, ביצענו השוואה מול מאורעות היסטוריים עולמיים ומול נקודות מפתח בחברה עצמה, בניסיון להסביר מגמות חריגות ולאפיין השפעת אירועים חיצוניים על הצלחתה.
3. **פארקים:** מדד ההצלחה במאגר מידע זה מתבטא בדירוג מספרי וביקורות מילוליות, עליהן ביצענו ניתוח סנטימנט והצגנו אותן בענני מילים. לשם כך, ניקינו את הביקורות ממילות קישור ועיצור, ובדקנו את המילים כשכולן באותיות קטנות. לאחר הרצה ראשונית שמנו לב שהמילים one ו- even קיבלו הרבה מקום על אף חוסר חשיבותן העקרונית. בנוסף, הגרף העלה שימוש רב באותיות הבודדות S ו-U וn't שכן להבנתנו U היה קיצור ל-you ו- S n't נוצרו כחלק מהסינון. לכן החלטנו למחוק את אלו גם כן. בנוסף, לא ביצענו stemming על מנת להימנע מעיוותים ואיבוד המשמעות של המילה.

בכדי להבין טוב יותר את המשמעות של אופי הטקסט החלטנו לבחון את המתאם של ניתוח זה לרייטינג שניתן לפארק. במטרה לאפשר השוואה נוחה בין השניים, בחרנו להתאים את תוצאות ניתוח הסנטימנט לנתונים של הרייטינג, כלומר לבצע חישוב שיתאים את תוצאות הניתוח לטווח של הרייטינג (0-5).

הערכה ומסקנות

**רווח:** רצינו למדוד את הצלחתם של הסרטים באמצעות נתוני **רווח קופה**, שיכולה לתת אינדיקציה לגבי ההצלחה של הסרט בעת יציאתו לאקרנים. כאמור, ביצענו נרמול לרווח לשנת 2021 לפי שער האינפלציה. כפי שציינו לעיל, במאגר המידע קיימים נתוני רווח הקופה עבור 82% מהסרטים, ולכן רק הם נכללו בניתוחינו. כך מצאנו את עשרת הסרטים בעלי רווח הקופה הגבוה ביותר. להוסיף הסתייגות לנוכח התעלמות מ18%.

על מנת לוודא את נכונות הגרף, הצלבנו את התוצאות עם ערך הויקיפדיה שמתאר את הסרטים עם רווחי הקופה הגבוהים ביותר. שמנו לב כי אלאדין (2019) אמור להימצא במקום התשיעי ברשימה זו, אך כיוון שרווח הקופה שלו היה חסר במאגר המידע, החסרנו אותו.

שיטה נוספת למדידת ההצלחה של סרטים היא לפי הרווח השולי שלהם, הלוקח בחשבון את התקציב. כלומר, אם לסרט יש רווח קופה גבוה אך גם תקציב גבוה, יכול להיות שהוא לא היה רווחי ועל כן נחשב לפחות מוצלח מסרט עם רווח קופה נמוך יותר. התוצאות מוצגות בגרף הבא (במאגר המידע חסרים נתוני התקציב של 37% מהסרטים, אותם נאלצנו להוריד לשם הניתוח): **גרף**

ממצא מעניין שעולה מהגרף הוא ששמונה מהסרטים הרווחיים ביותר הם גם הסרטים עם הרווח השולי הגבוה ביותר. סרטים אלו שומרים גם על מיקומם היחסי מבחינת רווחיות בין שני הגרפים. כמו כן, הרוב המוחלט של הסרטים המופיעים בגרפים הם סרטי המשך או עיבוד מחדש. שמנו לב כי למעט מלך האריות שיצא ב-1994, כל הסרטים יצאו ב-15 שנים האחרונות (כאשר צעצוע של סיפור 3 הוא המוקדם מביניהם ויצא ב-2010). הסבר אפשרי לכך הוא שככל שהחברה השקיעה בסרטים שהקהל הכיר ואהב, כך הרווחים היו גבוהים יותר. עם זאת, נציין כי הרגלי הצריכה השתנו מאוד ב-20 השנים האחרונות, דבר שעשוי להשפיע על הממצאים.

רצינו לבדוק האם מאגר המידע יכול להצביע על גורמים אפשריים שהביאו להצלחתם הכבירה של הסרטים שהצגנו. הגורם הראשון שרצינו לחקור לפיו הוא התקציב - האם כמות הכסף שתוקצב להפקת הסרט ופרסומו מביא לרווחי קופה גבוהים יותר? לשם כך, בדקנו מהם עשרת הסרטים עם התקציב הגבוה ביותר: **גרף**

נשים לב, ישנם רק שלושה סרטים שגם התקציב וגם רווח הקופה שלהם נמצאים בעשרת הגבוהים, מה שמרמז על קשר חלש בין תקציב להצלחת הסרט, שמתבטאת ברווח הקופה. את נכונות המידע על התקציב וידאנו באמצעות דגימה של סרטים מהמאגר והשוואתם למידע הקיים באינטרנט.

יצרנו גרף סקאטר-פלוט, שמציג את התקציב ורווח הקופה של הסרטים במאגר עבורם נתונים אלו קיימים. בחרנו בויזואליזציה זו כדי למפות את כל הסרטים ולבדוק התקיימות קשר בין תקציב לרווח הקופה.

בגרף זה ניתן לראות את המיפוי של כל הסרטים ביחס לתקציב (ציר X) ולרווח הקופה שלהם (ציר Y), במיליוני דולרים. שרטטנו ישר המכונה "Profitability line" - נקודות שנמצאות על ישר זה מייצגות סרטים שתקציבם שווה ערך לרווח הקופה. בהסתכלות ראשונית נשים לב כי הרוב המוחלט של הנקודות נמצאות מעל לישר, כלומר **ש**חברת וולט דיסני הפיקו בעיקר סרטים רווחיים. כמו כן, שרטטנו את ישר הרגרסיה על גרף זה. ניתן לראות כי הקשר הלינארי בין תקציב לרווח קופה הוא חיובי, וככל שמשתנים אלה גדלים, הקשר הלינארי נחלש. זאת לפי הצפיפות ההולכת וקטנה של הנקודות מסביב לישר הרגרסיה.

כדי לאמוד את טיב הקשר בין התקציב לרווח קופה, הוצאנו את נתון ה-R^2 העולה מנתונים אלו, וקיבלנו כי R^2=0.54. כלומר, 54% מהשונות במשתנה רווח קופה ניתנת להסבר על ידי המשתנה התקציב עבור סרטי דיסני. ערך זה מרמז שבממוצע, סרטים עם תקציב גבוה יותר נוטים להיות בעלי רווחי קופה גבוהים יותר. הנכונות של דיסני להשקיע בהפקה, שיווק והפצה תורמת ככל הנראה להצלחה הכוללת של הסרטים שלה. בעוד שהתקציב הוא גורם חשוב, הערך 0.54 מצביע גם על כך שחלק ניכר מהשונות בקופות לא מוסבר על ידי תקציב בלבד. גורמים אחרים כמו הרכב הבמאים והמלחינים או תאריך היציאה יכולים להשפיע גם הם על הצלחת סרטי דיסני.

נבחן את השפעת גורם חודש יציאת הסרט על הצלחתו בקופות. בשביל ניתוח זה, יצרנו סקאטר פלוט הממפה את כל הסרטים לפי החודש בו יצאו ורווח הקופה שלהם. כיוון שרווח הקופה נוצר בסמוך לחודש יציאת הסרט כאשר הוא מוצג בקולנוע, ניתן להסיק על קשר בין השניים. כמו כן, שרטטנו קו המתאר את הרווח הממוצע לכל חודש, במטרה לראות את המגמה הכללית של הרווחים. יש לציין כי בניתוח זה לא קיים הקשר לשנה בה יצאו הסרטים ולכן לא מגלם בתוכו הבדלים שנובעים מכך.

כפי שניתן לראות בגרף, החודשים עם ממוצעי הרווח הגבוהים ביותר הם מאי ויוני. הסבר אפשרי לכך הם הסמיכות של חודשים אלה לחופש הגדול, בו משפחות נמצאות בחופש ועל כן מבלות יותר בקולנוע, מה שתורם ישירות לרווחי הקופה. כמו כן, אלה החודשים בהם עונת התיירות נמצאת בשיאה. החודשים עם **ממוצעי הרווח הנמוכים ביותר אוגוסט, ספטמבר וינואר.** חודשים אלה מביאים איתם חזרה לשגרה לאחר חופשת המולד והחופש הגדול, והם מהווים הסבר אפשרי לממוצע הרווח הנמוך. ממצא נוסף שניתן לראות, הוא **שהסרטים עם רווחי הקופה הגבוהים ביותר מתפרסים על פני כמה חודשים: יולי, נובמבר, פברואר ויוני.** נשים לב כי מלבד חודש יוני, אלו לא חודשים עם ממוצע רווח גבוה. מכאן עולה כי הקשר בין חודש יציאה לבין רווח הקופה הוא חד כיווני. כלומר, סרטים שיצאו בחודשים מסוימים נוטים להיות רווחיים יותר, אך עם זאת ישנם סרטים רווחיים שיצאו בחודשים אחרים (לדוגמה, מלך האריות (2019) שיצא בחודש יולי הוא בעל רווח הקופה הגבוה ביותר).

מפיקים: עד כה, המדד להצלחה שבדקנו הינו רווח הקופה. זהו מדד זמני, התחום בזמן בו הסרט מוצג בקולנוע, ולא מציג את חווית הקהל לאורך זמן. לכן בחרנו גם למדוד הצלחת סרטים לפי **הדירוג** שלהם, נתון עדכני יותר שמשקף את הערכת הסרטים זמן רב לאחר שיצאו. לרשותנו נתוני דירוג משלושה אתרים שונים: IMDb ו-Rotten Tomatoes, האוספים דירוגים של משתמשים באתר, כאשר הדירוג ב-IMDb הוא מספר עשרוני בין 1-10 והדירוג ב-Rotten Tomatoes הוא באחוזים; ו-Metascore, אתר המבצע אגרגציה לדירוגים של מבקרים ומומחים בתחום הקולנוע ונותן דירוג מספרי בין 1-100. כיוון שסולמות המדידה שונים בין שלושת האתרים, ביצענו נרמול של כל הדירוגים למספרים עשרוניים בין 1-10 על מנת שיהיו ברי השוואה ולשם ויזואליזציה אינטואיטיבית ואחידה.

אחד מהגורמים שמביאים להצלחה של סרט הם המפיקים. בתור אלו שמקבלים החלטות מרכזיות ומנהלים את תהליך יצירת הסרט, ניתן לייחס הצלחה אליהם במידה רבה. זאת, גם במטרה למדוד את השפעתו של וולט דיסני עצמו על הצלחת החברה, בתור העומד בראשה. לשם כך, מיפינו את המפיקים של הסרטים וניסינו למדוד האם ישנם הרכבים של מפיקים אשר מביאים לסרטים מוצלחים במיוחד. בבואנו למפות את הרכב המפיקים, סיננו את אלו שהפיקו מספר זניח של סרטים ויטו את הדירוג לטובתם. לכן החלטנו להתמקד במפיקים שלקחו חלק ב-6 סרטים ומעלה, והגענו ל-19 המפיקים המשמעותיים בחברת דיסני. מתוכם, עשרת המפיקים עם הדירוג המשולב הממוצע הכי גבוה הינם (בסוגריים נמצאים מספר הסרטים של אותו מפיק):

ראוי לציין את המפיקים וולט דיסני ורון מילר, שבנוסף להיותם בעשרת המפיקים הכי מצליחים של החברה, גם הפיקו מספר עצום של סרטים ותרמו למורשת של החברה.

לאחר מכן, מיפינו את רשת שיתופי הפעולה של 19 המפיקים המשמעותיים על מנת לזהות הרכבים. כלומר, הגרף יוצר חיבור בין מפיקים שעבדו יחד על אותם סרטים. עשרת המצליחים שהוצגו בגרף לעיל סומנו באדום.

ניתן לזהות שישה מקבצים, כאשר הגדול והמקושר ביותר מורכב ממפיקים שעבדו בדיסני ב-40 שנותיה הראשונות (1937-1985) ויצרו את הסרטים הקלאסיים שכולנו מכירים, כגון "ספר הג'ונגל", "פיטר פן", ו"במבי". כמו כן, מקבץ זה מורכב ברובו המוחלט ממפיקים הנמצאים בעשרת המצליחים, כאשר במרכזם נמצאים וולט דיסני ורון מילר עם 6 ו-5 קשרים בהתאמה. ממצא זה תואם את הגרף הראשון, שמראה כי מפיקים אלו הפיקו מספר סרטים גבוה בהרבה ממפיקים אחרים. הקשר החזק ביניהם נטוע גם בקרבה משפחתית - מילר הוא חותנו של דיסני, מה שמסביר את היותו במרכז פעילות החברה ואת תרומתו הגדולה לסרטים החביבים על הקהל עד היום.

המקבץ השני בגודלו מורכב ממפיקים בתקופה המאוחרת יותר של דיסני (1992-2017), כאשר שליש מתוכם הם מעשרת המצליחים. מבין הסרטים שהמפיקים במקבץ זה השתתפו בהפקתם נמצאים "היפה והחיה" (2017), "אליס בארץ הפלאות" (2010), ו"מליפיסנט" (2014), כאשר המשותף להם הוא היותם סרטי המשך או עיבוד מחדש לסרטים אהובים ומוכרים. נראה שסרטים מסוג זה זוכים להצלחה הן מבחינת רווחים והן מבחינת דירוג.

שאר המקבצים הינם קטנים יותר ומורכבים ממפיקים שגם הם מהתקופה המאוחרת של דיסני. ניתן לראות שככל שהחברה צמחה, כך רשת שיתופי הפעולה של המפיקים המשמעותיים הפכה לדלילה יותר.

בשונה מהסרטים בעלי רווחי הקופה הגבוהים, נראה שהסרטים בעלי הדירוגים הגבוהים הם דווקא הסרטים המוקדמים יותר של החברה (ניתן לראות זאת לפי המפיקים עם הדירוג המתאים). ניתן להבין דיסוננס זה בהקשר של שינוי התנהגות צרכנית וגישות שיווק אסטרטגיות. ניתן לייחס את הדירוג הגבוה של הסרטים הקלאסיים של דיסני למשיכה הנוסטלגית המתמשכת שלהם וליכולתם להדהד את רגשות הקהל לאורך זמן. נזכור כי הרבה מן הסרטים עם רווחי הקופה הגבוהים הינם סרטי המשך ועיבוד מחדש, מה שמראה את היכולת של דיסני לשמור על רלוונטיות תוך הישענות על המודל המוכר והאהוב.

היסטוריה: רצינו לבחון התכנות של השפעת אירועים היסטוריים, בדגש על אירועים שקרו בתוך החברה, על מספר הסרטים שיצאו באותה שנה. בחרנו בתצוגת גרף עמודות, כדי לאתר מגמה כללית. כפי שציפינו, ככל שהחברה גדלה והתפתחה, במקביל לשינויים הטכנולוגיים בתחום הקולנוע, החברה הוציאה יותר סרטים בכל שנה. עם זאת, ישנן מספר קפיצות בגרף, שאינן מתיישבות עם מגמה זו. את הסטיות האלו בחנו את מול אירועים היסטוריים ואירועים שקרו בחברה באותן השנים, בתקווה להצליח לאתר השפעות אפשריות.

**1953** - גידול משמעותי במספר הסרטים שיצאו ביחס לשנים קודמות (פי שניים ואף יותר). בתקופה זו ארה"ב מתאוששת מזוועות מלחמת העולם השנייה, שבמהלכה פלטפורמת הטלוויזיה נוצלה לצרכי תעמולה. כך למשל הפקת הסרט המצליח פיטר פן אמנם החלה עוד לפני פרוץ המלחמה, אך הסרט יצא רק בשנה זו.

**1964-1965** - מגמת ירידה ללא אירועים היסטוריים חריגים. ייתכן שניתן להסביר ירידה זו בעקבות ניתוב משאבי החברה באותה התקופה לצורך הקמת פארק Disney World בשנת 1965.

**1978-1987** - ירידה מגמתית כללית. באותה התקופה הנהלת החברה השתנתה ולמרות הצלחתו של ערוץ דיסני החברה הייתה במצב כלכלי פגיע.

**1995-2001** - מגמת ירידה חריגה ומתמשכת. העלייה שלאחריה נבעה מפועלו של מייקל אייזנר כמנכ"ל החברה.

**2012-2013**- ירידה קלה. לא מצאתי סיבה

**2021** - ירידה שנובעת מכך שהמאגר עדכני לינואר 2021. באופן מפתיע, נראה שמשבר הקורונה לא השפיע משמעותית על פעילות החברה.

**פארקים:** בכדי להמשיך ולאפיין את הצלחתה של דיסני ראינו לנכון לבחון מישור נוסף בו דיסני משקיעה משאבים רבים- דיסנילנד. לצורך כך, השתמשנו במאגר המכיל ביקורות על שלושה פארקים מרכזיים- קליפורניה, פריז והונג-קונג. על מאגר זה ביצענו מספר ניתוחים כאשר המוקד העיקרי היה סביב ניתוח הביקורות עצמן ואיפיונן.

ענני מילים

ראשית, רצינו לאפיין בצורה כללית את הביקורות של המבקרים בפארקים. החלטנו ליצור ענן מילים שכן זו ויזואליזציה המציגה בצורה אינטואיטיבית את מידת השימוש בכל מילה. על מנת ליצור ענן זה, ניקינו את הביקורות ממילות קישור ועיצור, ובדקנו את המילים כשכולן באותיות קטנות. לאחר הרצה ראשונית שמנו לב שהמילים one ו- even קיבלו הרבה מקום על אף חוסר חשיבותן העקרונית. בנוסף הגרף העלה שימוש רב באותיות הבודדות S ו-U וn't שכן להבנתנו U היה קיצור ל-you ו- S n't נוצרו כחלק מהסינון. לכן החלטנו למחוק את אלו גם כן. בכדי לוודא את מהימנות הגרף, ביצענו חיפוש של המילים בטקסט וראינו שהכמות תואמת את הגודל (queue-14155, kid=16914, people- 11262).

בהתבוננות על ענן המילים של כלל הביקורות יחד, מספר מילים מושכות את תשומת הלב- fast-pass, queue- מציגות את החשיבות שהמבקרים נותנים לשהות בתור. atrraction, space mountain מביאות לידי ביטוי את העיסוק בדברים האטרקטיבים של הפארק בכלל ובאופן ספציפי במתקן ה- space mountain. מתקן זה הינו רכבת הרים יחודית הנמצאת בחלל סגור, המלווה באורות והדמיות מרשימות של החלל החיצון. לפי התגובות עושה רושם שהמתקן מרכזי בביקוריהם של האורחים בפארק.

את ענן המילים הרצנו גם על כל פארק בנפרד על מנת להבחין בהבדלים ביניהם, כאשר אלו המילים הבולטות בכל אחד מהפארקים:

הונג-קונג**:** ride, park, hong-kong, kid, day and time

קליפורניה:ride, park, disneyland, fast-pass, line, day and kid

פריז:ride, park, queue and day

מתוך ניתוח זה ניתן לראות כי אנשים מרבים להתייחס **למתקנים** (ride) **ולתור** (line, queue). באופן מעניין נראה כי לפארק בהונג-קונג יש את ההתייחסות הנמוכה ביותר לתור עצמו (יתכן והמילה זמן מתייחסת לזה בחלק מהפעמים אך סביר שפחות מהמילה queue למשל). בנוסף, בפארק שבקליפורניה ישנה התייחסות מרכזית לfast pass והתייחסות נמוכה למילה תור, לעומת הפארק שבפריז. הדבר יכול לתת בסיס לבדיקת ההשפעה של fast-pass על התייחסותם של המבקרים לתור.

נסתייג ונאמר שניתוח מסוג זה לוקה בחסר שכן חסר ההקשר של המילים הנאמרות.

מתוך כך החלטנו לבצע ניתוח סנטימנט על הביקורות.

ניתוח סנטימנט

ניתוח סנטימנט מאפשר לנו להבין טוב יותר את הטקסט שלפנינו כמו למשל עד כמה ביקורת הייתה שלילית או חיובית. במקרה שלנו הביקורות דורגו באמצעות פונקציה של ניתוח סנטימנט בטווח [1, 1-], כאשר 1- זאת ביקורת שלילית ביותר ו 1+ זו ביקורת חיובית ביותר. על נתונים אלו ביצענו ניתוח סטטיסטי. מתוך ניתוח זה ניתן לראות כי הממוצע גבוה מהמצופה (במצב של חלוקה סימטרית הממוצע =0) כלומר ניתן להבין כי התגובות חיוביות יחסית.

בחרנו בסקאטר פלוט בכדי להראות בצורה טובה את התפלגות הדירוגים תוך מתן דגש על הממוצע (באדום).

אם נסתכל על הנקודות האדומות המיצגות את ממוצעי הנתונים ניתן לפצל את ניתוח הגרף לשלוש הנקודות העליונות (חלק 1) ושתי הנקודות התחתונות (חלק 2). בחלק 1, ניתן להסיק כי ממוצע הסנטימנט נמוך באופן יחסי לממוצע הדירוג. למשל, ממוצע הדירוגים שקיבלו בניתוח סנטימנט 3 הינו 4. בחלק 2 ההפך הוא הנכון, שכן ממוצע הדירוגים שקיבלו ניתוח סנטימנט של 2.5 הינו 1. מכאן נראה כי הדירוג עצמו הוא יותר חד משמעי (סולם סדר), אבל הביקורת המילולית מתונה יותר ושואפת לממוצע. חשוב לציין שמגבלה עיקרית של גרף זה הינה שהגדרנו בעצמנו את החישוב, דבר שיכול ליצור הטייה.

**עבודה עתידית**

כיוון שדיסני היא חברה המורכבת מרבדים רבים אנו סבורות שיש בסיס לבחון את הצלחתה גם בתחומים נוספים.

1. דמויות:לדעתנו הצלחתו של סרט יכולה להיות קשורה במאפיינים הנוגעים לדמות הראשית בו. הנחתנו היא כי תחושות של קרבה והזדהות יכולות לעורר אהדה כלפי הדמות, הנאה מהסרט ומכאן לייעל את הצלחתו. מחקר מעניין על דמויות הסרטים יכול לבחון מאפייני דמות כמו מאפייני אישיות, מצב סוציו-אקונומי, מגדר, מין (חיה/ אדם) וכד'.
2. שירים: מוטיב מרכזי בעולמה של דיסני הוא המוסיקה. לדידנו, שיר מוצלח הזוכה לאהדת הקהל ולהשמעה חוזרת בערוצי רדיו, יכול להעיד על הצלחה של סרט מסויים ואף לעורר בצופים רצון לראות אותו פעם נוספת. בניתוח מסוג זה ניתן למשל לאפיין את הרכבי המילים של השירים המוצלחים והסנטימנט המועדף. בנוסף, ניתן לנתח את המנגינות שהצליחו ולאפיין רצפי תווים נעימים ו"מדבקים".
3. השוואה למתחרים: דיסני אינה החברה היחידה הפועלת בתחומי הקולנוע ופארקי השעשועים. מעניין לבצע השוואה בין הצלחת דיסני להצלחת חברות אחרות, בדגש על חברות הפקה המתחזקות פארקים (לדוגמה יוניברסל).
4. ניתוח כולל של החברה:בעבודתנו ניתחנו את הצלחתה של דיסני בתחומים שונים בנפרד. מעניין לבחון את הצלחתה במקביל במישורים השונים. ייתכן שבניתוח זה ניתן לראות תנועות מעניינות בין העולמות (למשל, האם במקביל להקמה של פארק חדש בפריז ניתן לראות שינוי בהכנסותיה של דיסני מסרטים בפריז)

סיכום

בשל נוכחותה המרכזית של דיסני בתעשיית הבידור ובלבבותיהם של הצופים, ראינו חשיבות בחקירת רבדי הצלחתה השונים. במהלך המחקר מצאנו מתאמים מעניינים הנוגעים בין היתר להשפעתם של תקציב ותזמון על הכנסות הסרט, ולקשר בין הרכב המפיקים לדירוגו של הסרט. בנוסף, ניסינו להסביר שינויים בכמות הסרטים שיצאו בשנה מסוימת ביחס לאירועים היסטוריים ופנימיים, ואפיינו את הצלחת הפארקים באמצעות ניתוח טקסטואלי. מתוך אלו אנו מסיקות כי אצל חברת דיסני הקסם המתמשך של העבר מתלכד עם האסטרטגיה החדשנית של ההווה, כדי לעצב את המסלול הכפול של דיסני לשבחי ביקורת יחד עם הצלחה מסחרית.