



ניהול קמפיינים חכם: אופטימיזציה באמצעות למידת מכונה ומודלים בייסיאנים

שמות המנחים:

מנחה אקדמאי: ד"ר יונת בראון

מנחה מקצועי: מעיין בן לולו

מספר קבוצה:

39

שם הסטודנט/ית	תעודת זהות
גל בן לולו	314668393
איתי ראובן חבשה	207379447
רותם בראל	318223872



תוכן עניינים:

1.	תקציר מנהלים.....	3
2.	תיאור הארגון.....	3
2.1	רקע לפרויקט.....	3
2.2	מבנה ארגוני ותהליכי עבודה מרכזיים.....	4
2.3	המודל העסקי.....	4
2.4	מטרות הארגון.....	5
3.	הגדרת הבעיה.....	5
4.	מטרות הפרויקט.....	7
5.	תרשים גאנט.....	8
6.	סקירת ספרות.....	8
7.	תיאור הנדסי של המצב הקיים.....	13
7.1	תיאור פורמלי של תהליכי העבודה בארגון.....	13
7.2	מדדים כמותיים של המצב הקיים.....	15
7.3	תחזיות למצב הרצוי.....	16
8.	מתודולוגיה.....	17
9.	בחינת חלופות.....	25
9.1	כללי.....	25
9.2	תהליך הערכת חלופות.....	25
9.3	הצגת החלופות.....	25
9.4	השוואה בין חלופות.....	27
9.5	קביעת החלופה האופטימלית.....	28
10.	מימוש הפתרון.....	28
11.	הערכת הפתרון.....	40
12.	דיון ומסקנות.....	41
12.1	דיון.....	41
12.2	מסקנות.....	42
12.3	המלצות.....	42
12.4	תרומת חברי הצוות.....	43
13.	ביבליוגרפיה.....	44
14.	נספחים.....	45
15.	מאגר GitHub, קובץ דאטה ודשבורד אינטראקטיבי.....	50

1. תקציר מנהלים

פרויקט זה עוסק בשיפור אפקטיביות הקמפיינים הדיגיטליים של חברת ReasonLabs, חברת סייבר ישראלית הפונה לקהל B2C, באמצעות שילוב בין למידת מכונה למודלים בייסיאנים. מטרת הפרויקט היא להגדיל את שיעור ההמרה ולהפחית את עלות רכישת הלקוח (CAC), תוך התמודדות עם שני אתגרים מרכזיים: **דלילות נתונים**, ו**ניהול לא מיטבי של תקציבי הפרסום**. הבעיה הראשונה - דלילות הנתונים, נובעת מכך שאחוז ההמרה בפועל נמוך במיוחד (כ-2.2%), כלומר מספר הרכישות קטן ביחס לכמות החשיפה. כתוצאה מכך, אין מספיק תצפיות לצורך ביצוע A/B Testing מסורתי והחברה נאלצת לקבל החלטות שיווקיות על בסיס נתונים גולמיים ולא מובהקים. במקביל, החברה מטרגטת את הקמפיינים שלה לפי שלב הרכישה בלבד שזהו השלב היקר ביותר בפאנל השיווקי תוך התעלמות משלבים מוקדמים יותר, אשר עשויים להיות בעלי קשר חזק לרכישה. גישה זו מובילה לניצול תקציב לא יעיל ולתשלום גבוה על קליקים (CPC) שאינם מתורגמים לרכישה. לשם כך פותחה מתודולוגיה המשלבת שתי גישות משלימות:

- **למידת מכונה** - לזיהוי משתנים קריטיים בשלבים מוקדמים בפאנל על מנת לאפשר טרגוט חכם יותר של קמפיינים, שיפור אחוזי ההמרה והפחתת עלויות פרסום.
- **מודלים בייסיאנים** - מאפשר קבלת החלטות מושכלת גם כאשר כמות הנתונים דלה, באמצעות עדכון הסתברויות רציף בזמן אמת.

הפתרון ייושם באמצעות פיתוח דאשבורד אינטראקטיבי ב-Streamlit. הכלי יאפשר לצוות השיווק להעלות נתונים, לנתח ביצועי קמפיינים, לזהות משתנים משפיעים ולבצע השוואות לצורך אופטימיזציה מדויקת של הקמפיינים וניהול חכם של התקציב.

2. תיאור הארגון

2.1 רקע

ReasonLabs היא חברת סייבר ישראלית שנוסדה בשנת 2016 ומתמחה בפיתוח פתרונות אבטחת מידע והגנה דיגיטלית עבור משתמשים פרטיים (B2C). מטה החברה ממוקם בתל אביב, ולצידו סניפים נוספים בדובאי. החברה הוקמה מתוך מטרה להתמודד עם איומי הסייבר המתרבים בסביבה הדיגיטלית של המאה ה-21 ולהנגיש אבטחה מתקדמת לכל משתמש, גם ללא רקע טכנולוגי. החברה מפתחת מגוון פתרונות מבוססי בינה מלאכותית ולמידת מכונה, המאפשרים זיהוי ונטרול איומים בזמן אמת. מוצרי החברה מופצים למשתמשים בארה"ב תוך שימוש במודל עסקי מסוג Freemium ומשווקים באמצעות קמפיינים דיגיטליים ממוקדים בפלטפורמות כמו: Google, Facebook ו-YouTube.

2.2 מבנה ארגוני ותהליכי עבודה מרכזיים

החברה בנויה ממספר מחלקות הפועלות זו לצד זו, כאשר כל מחלקה אחראית על תחום התמחות שונה:

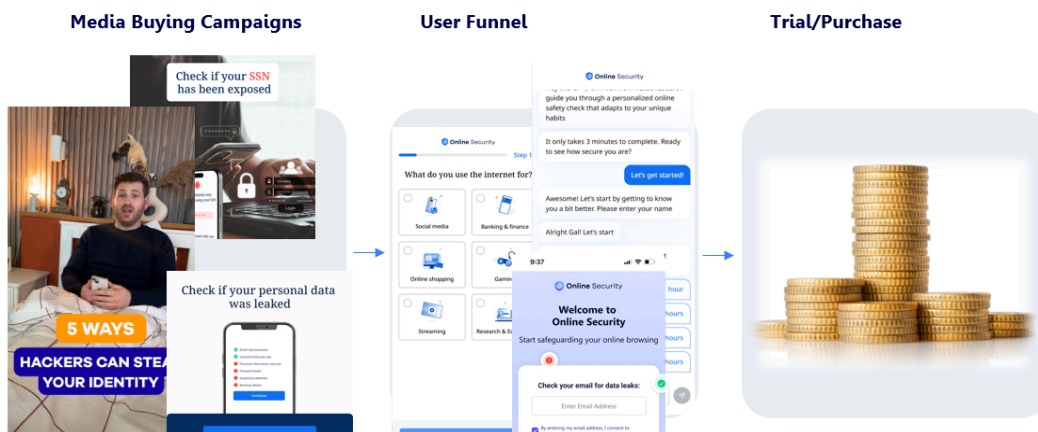
- **מחלקת הפיתוח** - אחראית על תכנון, פיתוח ותחזוקה של מוצרי האבטחה שהחברה מציעה.
- **מחלקת השיווק** - אחראית על ניהול הקמפיינים הדיגיטליים, כולל תכנון אסטרטגיות פרסום, טרגוט קהלים, ניתוח ביצועים, וניהול תקציבי הפרסום.
- **מחלקת דאטה** - מתמקדת באיסוף וניתוח נתונים הקשורים להתנהגות המשתמשים, מתוך מטרה להפיק תובנות שיווקיות ומוצריות שישפרו את ביצועי החברה.

ReasonLabs פועלת בסביבה עסקית תחרותית שמתאפיינת בקצב גבוה של שינויים טכנולוגיים, עלייה מתמדת בכמות איומי הסייבר, וצורך בהיענות לרגולציות שונות בתחום הגנת הפרטיות והמידע (למשל GDPR באירופה ו-CCPA בארה"ב). תהליכי העבודה המרכזיים של החברה כוללים פיתוח מתמשך של מוצרים, מעקב שוטף אחר מגמות שוק ואיומים טכנולוגיים, וכן הפעלת קמפיינים פרסומיים ממומנים בפלטפורמות דיגיטליות. החברה עושה שימוש בנתונים שנאספו מהקמפיינים השיווקיים, כדי לנתח את דפוסי ההתנהגות של המשתמשים ולבצע אופטימיזציה מתמדת לביצועים.

2.3 המודל העסקי:

ReasonLabs פועלת במודל Freemium, במסגרתו המשתמשים מקבלים גישה חינמית לפיצ'רים בסיסיים, בעוד שהגרסה המתקדמת זמינה בתשלום.

מודל עסקי



איור 1- מודל עסקי

תהליך ההמרה של המשתמשים מתבצע בשלושה שלבים:

1. Media Buying Campaigns - השקעה בפרסום בפלטפורמות דיגיטליות (Google Facebook YouTube, Instagram) להגברת החשיפה ולמשיכת משתמשים חדשים.
 2. User Funnel - תהליך מותאם אישית הכולל שאלות אבחון שמסייעות ביצירת חוויית שימוש ממוקדת.
 3. Trial/Purchase - אפשרות לניסיון חינמי או לרכישת מינוי הכולל כלים מתקדמים כגון VPN, אנטי-וירוס, ניטור דליפות מידע והגנה על ילדים ברשת.
- מודל זה מאפשר להמיר משתמשים חנימיים ללקוחות משלמים ולשפר את החזר ההשקעה בקמפיינים באופן מתמיד.

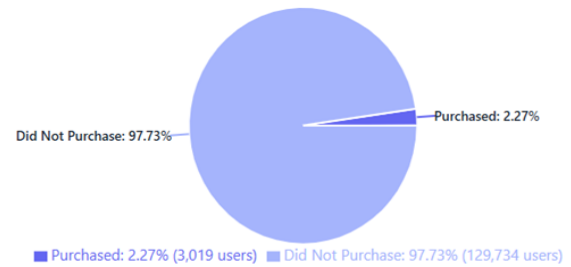
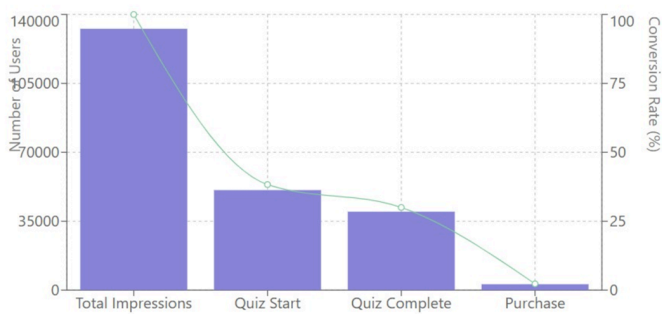
2.4 מטרות הארגון

המטרה המרכזית היא להגן על פרטיותם של משתמשים פרטיים ולאפשר להם גלישה בטוחה תוך שמירה על חוויית משתמש פשוטה ונגישה. החברה שואפת לפתח פתרונות חכמים, מהירים ואפקטיביים אשר מסוגלים להתמודד עם איומי סייבר מתקדמים ולהנגיש את טכנולוגיות ההגנה גם לאוכלוסיות לא טכנולוגיות. מבחינת השוק, ReasonLabs שואפת להגדיל את נתח המשתמשים המשלמים דרך יעול מערך השיווק הדיגיטלי, הגדלת יחס ההמרה והפחתת עלויות רכישה.

3. הגדרת הבעיה

ReasonLabs משקיעה תקציבים משמעותיים בקמפיינים שיווקיים ממומנים, במטרה להמיר משתמשים חנימיים ללקוחות משלמים. עם זאת, מחלקת השיווק מתמודדת עם שני אתגרים מרכזיים המשפיעים באופן ישיר על האפקטיביות של הקמפיינים ועל יכולת קבלת ההחלטות:

1. דלילות נתונים - אחת הבעיות המרכזיות בניהול הקמפיינים של החברה היא מגבלת הדאטה, הנובעת משיעור המרה נמוך במיוחד שגורם לכך שברוב המקרים אין לחברה מספיק תצפיות, כדי לבצע ניתוחים סטטיסטיים מובהקים, ובפרט - A/B Testing מסורתי.
- A/B Testing הוא כלי קריטי בקבלת החלטות שיווקיות, שכן הוא מאפשר השוואה בין שתי גרסאות של קמפיין, תוך ניתוח השפעת שינוי מסוים על התוצאה הרצויה. עם זאת, על מנת שההבדלים בין הקבוצות יהיו בעלי משמעות סטטיסטית, נדרש מדגם בגודל מספק - כלומר מספר גבוה של המרות. הגרפים שלהלן ממחישים את הירידה המשמעותית לאורך הפאנל השיווקי: מתוך כ-140,000 חשיפות, רק כ-3,000 משתמשים ביצעו רכישה בפועל - כלומר, שיעור המרה של 2.27% בלבד. נתון זה מדגיש את הפער בין נפח החשיפה לקמפיינים לבין כמות הרכישות בפועל וממחיש את קושי החברה להסיק מסקנות מהימנות בהתבסס על כמות כה קטנה של המרות.



איור 2- שיעור ההמרה ופיזור המשתמשים לאורך הפאנל - מהחשיפה ועד לרכישה

תחת מגבלה זו, כל שינוי קטן במספר הרוכשים עשוי להוביל לתנודות חדות במדדי הביצוע- תופעה המוכרת בשם שונות גבוהה. שונות זו מקשה על צוות השיווק להסיק מסקנות ברורות ולעיתים אף מובילה לקבלת החלטות שגויות:

- False Positives- Type I Error: כאשר נראה בטעות שקמפיין אחד עדיף על אחר, למרות שההבדל מקרי בלבד.
- False Negatives- Type II Error: כאשר קמפיין מוצלח נפסל מוקדם מדי, רק משום שטרם הספיק לצבור מספיק תצפיות כדי להוכיח את יתרונו.

יתרה מכך, קצב ההשקה הגבוה של קמפיינים חדשים מחייב תגובה מהירה, אך A/B Testing דורש פרקי זמן ארוכים להצטברות נתונים. עיכוב זה פוגע ביכולת של החברה להתאים את עצמה לשינויים בזמן אמת ולבצע אופטימיזציה שוטפת לקמפיינים. במציאות שבה התקציב מנוהל על בסיס נתונים חלקיים, ההחלטות מתקבלות לרוב תוך הסתמכות על מדדים שטחיים כמו: CTR (Click-Through Rate) או CPC (Cost Per Click), במקום תובנות עמוקות המבוססות על ביצועים בפועל.

2. אסטרטגיית טרגוט ממוקדת בשלב הרכישה בלבד - מחלקת השיווק מטרגטת את הקמפיינים הדיגיטליים בשלב הרכישה, השלב הסופי בפאנל השיווקי. זהו שלב בעל עלות גבוהה במיוחד, שבו התחרות על שטחי הפרסום מוגברת והעלות לקליק גבוהה. אסטרטגיה זו אמנם שואפת להגיע למשתמשים פוטנציאליים בסמוך לרגע ההמרה, אך בפועל היא יוצרת מצב שבו התקציב השיווקי מנוצל בצורה פחות יעילה, עלות רכישת הלקוח עולה והחזר ההשקעה נפגע. התמקדות בשלב הסופי בלבד מתעלמת מהעובדה ששילבים מוקדמים בפאנל, כגון: התחלת שאלון, או מענה על שאלות ספציפיות עשויים לשקף רמות שונות של מעורבות, עניין או נכונות לרכישה ולעיתים אף לקיים קשר ישיר או עקיף להמרה בפועל. כאשר אין התחשבות שלבים אלו, החברה מחמיצה תובנות קריטיות ומוותרת בפועל על האפשרות לזהות ולטרגט משתמשים בעלי פוטנציאל גבוה כבר בשלבים הראשונים, בהם גם העלות הפרסומית נמוכה יותר.

4. מטרת הפרויקט

הפרויקט נועד לספק פתרון אנליטי יישומי שישפר את קבלת ההחלטות במחלקת השיווק, תוך התמודדות עם אתגרים של דלילות דאטה, שיעורי המרה נמוכים וניצול לא מיטבי של התקציב. המטרות המרכזיות:

- שיפור אפקטיביות הקמפיינים הדיגיטליים באמצעות הפחתת עלות רכישת הלקוח והגדלת שיעור ההמרה.
- חיזוק תהליך קבלת ההחלטות השיווקיות גם בתנאים של דלילות נתונים, באמצעות מעבר מהחלטות אינטואיטיביות להחלטות מבוססות-דאטה.
- בניית דשבורד אינטראקטיבי תומך החלטה, שישמש את צוות השיווק ככלי יומיומי לניתוח קמפיינים.

מדדי הצלחה:

הצלחת הפרויקט תיבחן על סמך מדדים כמותיים ברורים, אשר נגזרים ישירות מהאתגרים שהוגדרו. מדדים אלו נבחרו משום שהם משקפים שיפור בפעילות השיווקית הדיגיטלית של ReasonLabs - הן ברמת האפקטיביות של הקמפיינים והן ביכולת קבלת ההחלטות של הארגון.

- שיפור יחס ההמרה מ-2.2% ל-3% לפחות
- הפחתת עלות רכישת הלקוח מ-\$500 ל-\$350
- לפחות 70% מההחלטות השיווקיות יתקבלו על בסיס ניתוחים והמלצות שהופקו מהדשבורד

אילוצים:

במהלך הפרויקט זוהו מספר אילוצים המשפיעים על אפשרויות הפיתוח, היישום וההטמעה של הפתרון בארגון. אילוצים אלה נובעים מהמציאות הארגונית, מהסביבה העסקית בה פועלת ReasonLabs. האילוצים המרכזיים שנלקחו בחשבון הם:

- מגבלות תקציב שיווקי- הקצאה מוגבלת של משאבים לפרסום דיגיטלי, המגבילה את היכולת לבצע ניסויים בקנה מידה גדול או להריץ קמפיינים במקביל.
- דלילות נתונים - שיעור ההמרה הנמוך יוצר מאגר נתונים מצומצם, בעיקר בשלב הרכישה, מה שמקשה על בניית מודלים מדויקים ומפחית את היכולת לבצע A/B Testing מסורתית.
- פערי ידע בתחום הדאטה והניתוחים ההסתברותיים בקרב המשתמשים ידרשו בנייה של דשבורד נגיש ואינטואיטיבי, אשר ישים דגש על הצגה חזותית של תחזיות והמלצות, כך שגם משתמשים ללא היכרות מוקדמת עם מודלים בייסיאניים או התפלגויות סטטיסטיות יוכלו להפיק ממנו תובנות ישימות לצורך קבלת החלטות שיווקיות.

5. תרשים גאנט:

תהליך העבודה	2024 Q4			2025 Q1			2025 Q2			2025 Q3		
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
זיהוי והגדרת הבעיה												
איסוף ועיבוד נתונים												
מחקר ובדיקת גישות												
פיתוח המערכת												
הטמעת המערכת												
סיכום												

איור 3- תרשים גאנט

6. סקירת ספרות

בעולם השיווק הדיגיטלי המתאפיין בקצב מהיר, נפחי דאטה משתנים ותחלופה גבוהה של קמפיינים, צוותי השיווק נדרשים להפיק תובנות מהירות אודות הקמפיינים, גם כאשר המידע חלקי או מרובה רעש. הקושי מתעצם כאשר מדובר בתרחישים של דלילות נתונים- תופעה שכיחה בקמפיינים בעלי שיעור המרה נמוך, אשר אינם מאפשרים ניתוח סטטיסטי קלאסי המבוסס על נפח תצפיות גבוה. מגבלה זו פוגעת ביכולת לבצע A/B Testing מסורתי, שמהווה שיטת ניתוח נפוצה אך אינה מותאמת למצב בו הנתונים מתכנסים באיטיות. לנוכח מגבלות אלו, הספרות המחקרית מציעה גישות חדשניות המתמודדות עם תנאי אי-ודאות ודאטה מוגבל, תוך שילוב בין הסקה הסתברותית ובין אלגוריתמים של למידת מכונה. גישות אלו מאפשרות לזהות משתנים מנבאים מוקדמים להמרה, להעריך הסתברויות גם כאשר גודל המדגם מצומצם ולבצע אופטימיזציה של התקציב השיווקי באופן מושכל.

סקירה זו מציגה גישות מתקדמות שנחקרו בשנים האחרונות ומתמקדת בחמישה נושאים שעליהם נשען הפרויקט הנוכחי:

1. מגבלות ה-A/B Testing הקלאסי בעולמות של דלילות דאטה
2. יתרונות השימוש בגישה בייסיאנית להסקה במצבים של אי-ודאות
3. מיפוי שלבי הפאנל השיווקי ונקודות מגע במסע הלקוח כבסיס לניתוח התנהגות צרכנית

4. יישום למידת מכונה לזיהוי שלבים קריטיים במסלול ההמרה
5. פיתוח ממשקים חכמים לתמיכה שוטפת בקבלת החלטות מבוססות-נתונים

באמצעות שילוב בין כלים הסתברותיים, ניתוח תבניות התנהגותיות ויישום של פתרונות ויזואליים מותאמים לארגון, הפרויקט מציע גישה אינטגרטיבית להתמודדות עם האתגרים המרכזיים של ReasonLabs, מתוך מטרה להוביל לשיפור אפקטיביות הקמפיינים, הפחתת עלויות הרכישה, ודיוק גבוה יותר בטרגוט קהלי יעד.

בעיית דלילות הנתונים בקמפיינים שיווקיים

גישת ה-A/B Testing נתפסת כסטנדרט המקובל לבחינת יעילותם של קמפיינים שיווקיים דיגיטליים. השיטה מבוססת על הקצאה אקראית של משתמשים לשתי קבוצות, קבוצת ניסוי וקבוצת ביקורת, והשוואת הבדלים מובהקים סטטיסטית במדדים כמו שיעור המרה או ביצוע רכישה. אך בשוק הדיגיטלי הדינמי עולות מגבלות משמעותיות ביישום השיטה, במיוחד כאשר הדאטה מצומצם ורועש.

מאמר [1] מצביע על כך שכאשר שיעור ההמרות נמוך או גודל המדגם מוגבל, כפי שקורה לעיתים קרובות בקמפיינים דיגיטליים ממוקדים, שיטת A/B Testing מתקשה להגיע למובהקות סטטיסטית בטווח זמן סביר. ההמתנה להצטברות של מספר מספיק של רכישות, לשם הסקת מסקנות אמינות, עלולה לארוך שבועות ואף חודשים דבר שפוגע ברלוונטיות הניתוח. בנוסף, המחקר מדגיש את הבעייתיות של השיטה בהקשרים בהם ההשפעות אינן מיידיות, אלא נמדדות לאורך זמן, או כאשר קיים שוני מהותי בין קבוצות המשתמשים, המערער את תנאי ההקצאה האקראית.

מעבר לכך, המחקר מציג את האתגר של דלילות אפקט, מצב שבו גם אם המדגם גדול ההשפעה עצמה קטנה יחסית, כך שנדרשת עוצמה סטטיסטית גבוהה במיוחד כדי לאתר אותה. במצבים כאלה, נדרשת, על פי [1], גישה שמסוגלת לשלב מידע מצטבר ממקורות שונים, להתעדכן עם הזמן ולהתמודד עם תנאי אי-ודאות, תכונות שלא קיימות בשיטת A/B Testing הקלאסית.

יתרונות השימוש בגישה בייסיאנית להסקה במצבים של אי-ודאות

בהמשך לבעיית דלילות הנתונים, ובפרט מגבלותיה של גישת ה-A/B Testing שהוזכרו במאמר [1], במאמר [2] מוצעת מסגרת מתודולוגית בשם Bayesian Additional Evidence, שפותחה כדי להתמודד עם הקשיים שבניתוח תוצאות ממדגמים קטנים. מטרת המסגרת היא לאפשר קבלת החלטות מושכלת גם כאשר אין די נתונים להסקה מובהקת בשיטות קלאסיות. במקום להישען על ערכי p ואינטרוולים סימטריים שרגישים מאוד לגודל המדגם, הגישה הבייסיאנית מעריכה את האפקט דרך הסתברות פוסטרירורית, ומספקת תחום אמינות שמתאר את רמת אי-הוודאות האפיסטמית הנובעת מהמדגם הקיים. אחת התרומות המרכזיות של המודל היא רעיון "נקודת ההיפוך הבייסיאנית"-כלומר, החישוב של כמה ראיות נוספות נדרשות כדי לשנות את מידת האמון בהשערה. הגישה מיושמת על מקרה של השוואת טיפולים רפואיים במדגם קטן. החוקרים מראים

שכאשר התוצאה הראשונית מצביעה על יחס סיכון של 0.31 עם רווח סמך קלאסי 95% (0.09, 1.1) שהיא תוצאה לא מובהקת, חישוב בייסיאני מצביע על כך שרק תוצאה נוספת עם יחס סיכון של 0.54 או נמוך יותר תספיק כדי להפוך את המסקנה לבעלת תוקף פוסטריורי. משמעות הדבר היא שגם כאשר אין מובהקות, ייתכן שההשפעה כן קיימת, אך פשוט לא נמדדה בדיוק מספק.

גישה זו מדגישה כי תוצאה שאינה מובהקת אינה שקולה להיעדר אפקט, אלא עשויה לשקף מגבלות סטטיסטיות בלבד. השיטה מאפשרת קבלת החלטות אינפורמטיביות בזמן אמת, ומקטינה את הסיכון לפרשנות שגויה הנובעת ממדגם קטן או רעש גבוה בדאטה.

בהקשר לפרויקט, מאמר [2] סיפק את הבסיס התיאורטי לאימוץ הגישה הבייסיאנית כחלופה מתודולוגית ל-A/B Testing, אשר אינו ישים בתנאים של דלילות נתונים ואי־ודאות גבוהה. תרומתו המרכזית של המאמר הייתה בהצגת מסגרת הסתברותית שמאפשרת להסיק מסקנות גם במדגמים קטנים, תוך חישוב הסתברויות ותחומי אמינות במקום הסתמכות על מובהקות קלאסית. גישה זו הושמה בפרויקט באמצעות שימוש בהתפלגות Beta לצורך הערכת הסתברות ההמרה, והיותה מרכיב מרכזי בהתמודדות עם האתגר של ניתוח והשוואת ביצועי קמפיינים בסביבה עתירת אי־ודאות.

מיפוי מסע הלקוח והשלבים הקריטיים בפאנל השיווקי

האתגר בזיהוי שלבים משמעותיים במסלול ההמרה השיווקי מקבל התייחסות נרחבת בספרות, בין היתר במאמר [3], המציג מסגרת מתודולוגית חדשנית למיפוי וניתוח "נקודות מגע" במסע הלקוח הדיגיטלי. המאמר מתמקד בהבנה של תרומת שלבים שונים בתהליך ההתקשרות עם המשתמש, לרבות אינטראקציות מוקדמות לזיהוי כוונות רכישה והשפעה על המרה בפועל.

החוקרים מציעים שילוב בין נתונים איכותניים וכמותיים כאסטרטגיה מחקרית: מצד אחד, ראיונות עומק עם מנהלים ולקוחות לזיהוי תפיסות סובייקטיביות של ערך הנקודות השונות במסע, ומצד שני, ניתוח התנהגותי המבוסס על מידע בפועל ממערכות CRM. ממצאי המחקר מעלים כי קיימים פערים בין תפיסת המנהלים לגבי השלבים הקריטיים במסלול ההמרה לבין הנתונים האובייקטיביים, אשר לעיתים מצביעים דווקא על אינטראקציות מוקדמות, כגון חשיפה ראשונית, מילוי שאלון או ביקור חוזר באתר כגורמים המשפיעים ביותר על סבירות הרכישה.

המחקר מדגיש את הצורך בבחינה מקיפה של מסע הלקוח, תוך התמקדות בקשרים בין שלבים מוקדמים להמרות עתידיות. במקום להתמקד רק בתוצאה הסופית, הוא מציע לאמץ תפיסה הרואה במסע השיווקי רצף מצטבר של אינטראקציות, כאשר כל שלב בו עשוי לשמש אות משמעותי לכוונת רכישה ולכן ראוי למדידה והערכה.

בהקשר לפרויקט, הממצאים במאמר [3] תומכים בגישה לפיה אין להתמקד אך ורק בשלב הרכישה הסופי כמדד יחיד להצלחת הקמפיין. בהתאם לכך, הפרויקט מאמץ פרספקטיבה רחבה יותר, הבוחנת את ההשפעה המצטברת של שלבי הפאנל השיווקי על סבירות ההמרה. הניתוח מתמקד בתשובות המשתמשים לשלבים מוקדמים במסלול, כמו הרגלי שימוש באינטרנט, תפיסת סיכונים דיגיטליים, רמת מודעות והיסטוריית חשיפה כמנבאים אפשריים לרכישה בהמשך. תובנות אלו מאפשרות גיבוש אסטרטגיות טרגוט חכמות יותר, המבוססות על זיהוי אותות מוקדמים להתעניינות, עוד בטרם התקבלה החלטת רכישה מפורשת.

יישום למידת מכונה לזיהוי שלבים קריטיים במסלול ההמרה

בעוד שמאמר [3] מדגיש את החשיבות האסטרטגית של מיפוי שלבי הפאנל השיווקי והבנת תרומתם להמרה, מאמר [4] מרחיב את הממד האנליטי של גישה זו באמצעות יישום שיטתי של אלגוריתמים של למידת מכונה לצורך זיהוי לידיים בעלי פוטנציאל המרה גבוה. המאמר מציע מסגרת מקיפה הכוללת עיבוד נתונים, פיתוח מודלים והשגת תובנות אופרטיביות, כל זאת לשם תמיכה בקבלת החלטות שיווקיות מבוססות-נתונים.

החוקרים התבססו על מערך נתונים הכולל למעלה מ-9,000 רשומות לידיים, שביניהן מאפיינים דמוגרפיים, התנהגותיים ומידע על אינטראקציות באתר. תהליך העבודה כלל שלבים מוקפדים של עיבוד מקדים, ניקוי שדות חסרים, הסרת ערכים קיצוניים, קידוד משתנים קטגוריאליים, נרמול משתנים רציפים ולאחר מכן חלוקה לסטים של אימון, ולידיה ובדיקה. לבחינת ביצועי המודלים, נבחנו חמש טכניקות נפוצות של למידת מכונה: Random Forest, XGBoost, CatBoost, Logistic Regression ו-Desicion Tree, כאשר כל אחת מהן עברה כוונן פרמטרים באמצעות Grid Search.

הממצאים הצביעו על כך שמודלים מתקדמים כגון CatBoost ו-Random Forest סיפקו את התוצאות המדויקות ביותר, תוך איזון בין ביצועים לבין יכולת פרשנות של חשיבות משתנים. הדגש במאמר לא היה רק על דיוק הסיווג, אלא גם על הפקת תובנות שניתן ליישמן הלכה למעשה. תובנות אלה הוצגו באמצעות דשבורדים אינטראקטיביים, אשר אפשרו זיהוי פרופילי משתמשים, הבנת משתנים מנבאים והשוואה בין קבוצות שהומרו לאלו שלא.

בהקשר לפרויקט, ממצאי מחקר [4] מדגישים את הפוטנציאל של למידת מכונה ככלי לזיהוי משתנים בעלי השפעה גבוהה על סבירות ההמרה, כבר בשלבים מוקדמים במסלול ההמרה. בהתאם לכך, יושמו בפרויקט מודלים מונחים לצורך ניתוח דפוסי התנהגות של משתמשים, במטרה לשפר את הדיוק בטרגוט ולחזות הסתברות רכישה עתידית. התובנות שהופקו מהמודלים ישולבו בפלטפורמה אינטראקטיבית, שתנגיש מידע אנליטי לאנשי השיווק ותשמש כבסיס לקבלת החלטות מבוססות-נתונים בזמן אמת.

פיתוח דשבורדים שיווקיים ככלי אסטרטגי לקבלת החלטות

הדשבורד השיווקי הפך לכלי קריטי המגשר בין עולם הדאטה לבין תהליכי קבלת ההחלטות בארגונים. מאמר [5] מציע מסגרת תיאורטית ומעשית מקיפה לבניית דשבורד המייצר ערך אסטרטגי, תוך שימת דגש על תחזיות מדויקות, אינטגרציה בין-מחלקתית, ומיקוד בהחלטות עסקיות הנובעות מתוך ניתוח מושכל של נתונים. המאמר מדגיש את המעבר מדשבורדים סטטיים, המתארים את העבר בלבד, למערכות דינמיות ופרואקטיביות המשלבות יכולות תחזית, הערכת תרחישים עתידיים והצעת אסטרטגיות פעולה בזמן אמת. דשבורדים אלה מתבססים על בחירה קפדנית של מדדים בעלי ערך, המייצגים קשרים סיבתיים בין משתנים שיווקיים לתוצאות עסקיות. הם מבוססים על מודלים אמפיריים המשלבים סדרות זמן, למידת מכונה וניתוחים של עקומות תגובה לא ליניאריות.

המחברים מדגישים את חשיבות ההתאמה המתמשכת של הדשבורד לשינויים בשוק ולתגובות המתחרות, לצד הצורך בהנגשתו לכלל בעלי העניין בארגון לא רק אנליסטים או אנשי דאטה באמצעות ממשק אינטואיטיבי, מפורש ומבוסס נתונים איכותיים. בתוך כך, הם מתריעים מפני הטיות ניהוליות העלולות להתבטא בבחירת מדדים מגמתית, ומציעים עקרונות לבניית דשבורד אובייקטיבי, מהימן ומבוסס מדידה אחידה לכלל המחלקות.

בהקשר לפרויקט, מאמר [5] סיפק תשתית לבניית דשבורד שמטרתו אינה רק להציג נתונים, אלא להנגיש תובנות פעולה בזמן אמת. הרעיונות המרכזיים במאמר, ובעיקר הדגש על תחזיות, סימולציות והבנת קשרים סיבתיים בין משתנים, הובילו לאימוץ גישה שממקדת את הדשבורד ככלי אסטרטגי לקבלת החלטות, כזה שמזהה אותות מוקדמים להתנהגות משתמשים, מעריך הסתברויות המרה, ומאפשר בחינה של הצעות שיווקיות גם תחת מגבלות דאטה. בכך, הדשבורד שפותח בפרויקט מהווה יישום מעשי של עקרונות הליבה שהוצגו במאמר, ותורם ליכולת לנהל קמפיינים ביעילות גבוהה יותר תוך הסתגלות לתנאים משתנים.

סיכום תובנות מסקר הספרות

חמשת המאמרים שנבחנו במסגרת סקירת הספרות סיפקו תשתית תיאורטית ומתודולוגית מוצקה, אשר סייעה לעצב את מבנה הפרויקט ב-ReasonLabs, הן בהגדרת האתגרים המרכזיים והן בבחירת הכלים האנליטיים שנועדו לתת להם מענה. בבסיס ההחלטה לבחירה במודל הבייסיאני עומדת התרומה של מאמר [1], שהציג את המגבלות של A/B Testing מסורתי במצבי דלילות נתונים ובקמפיינים שהשפעתם מתפרסת על פני זמן. מתוך כך, אומצה בפרויקט מסגרת בייסיאנית חלופית, שזכתה לחיזוק נוסף במאמר [2], אשר פירט את היתרונות של שימוש במודל בייסיאני כתחליף לשיטות המסורתיות.

מאמר [3] הציע תפיסה רחבה להבנת מסע הלקוח דרך מיפוי נקודות מגע ושלבים מוקדמים, תובנה שהובילה למיקוד אנליטי בשלבים השונים של הפאנל השיווקי ולא רק ברכישה עצמה. מאמר [4] העמיק את ההיבט היישומי והדגים כיצד ניתן להשתמש בלמידת מכונה לזיהוי משתנים קריטיים לאורך הפאנל ולתרגם תובנות אלה לדשבורד אינטראקטיבי תפעולי, עקרונות שיוטמעו בפרויקט בניתוח החשיבות של שלבי הפאנל ובפיתוח תובנות פרסונליות.

לבסוף, תרומתו של מאמר [5] ניכרת בגיבוש התפיסה האסטרטגית של הדשבורד בפרויקט. המאמר מדגיש את החשיבות בהבחנה בין מדדים תפעוליים לבין מדדים עסקיים ואת הצורך בזיהוי קשרים סיבתיים בין פעולות שיווקיות לתוצאות כגון ROI ו-CAC. כמו כן, החוקרים מצביעים על הערך בהנגשת תובנות אנליטיות גם למקבלי החלטות שאינם בעלי רקע טכני. עקרונות אלה עיצבו את אופן תכנון הדשבורד בפרויקט ככלי ניהולי תומך החלטה, שמאפשר שימוש פשוט ומושכל במידע, גם בתנאים של אי-ודאות ודלילות דאטה. סך התובנות העולות מן הסקירה מהוות בסיס מוצק לפיתוח פתרון המשלב בין הסקה הסתברותית, אלגוריתמים מונחים וממשק משתמש אפקטיבי, מתוך שאיפה לייצר תשתית אמינה, מדידה וברת-שימוש לניהול מושכל של קמפיינים שיווקיים תחת מגבלות דאטה מובהקות.

7. תיאור הנדסי של המצב הקיים

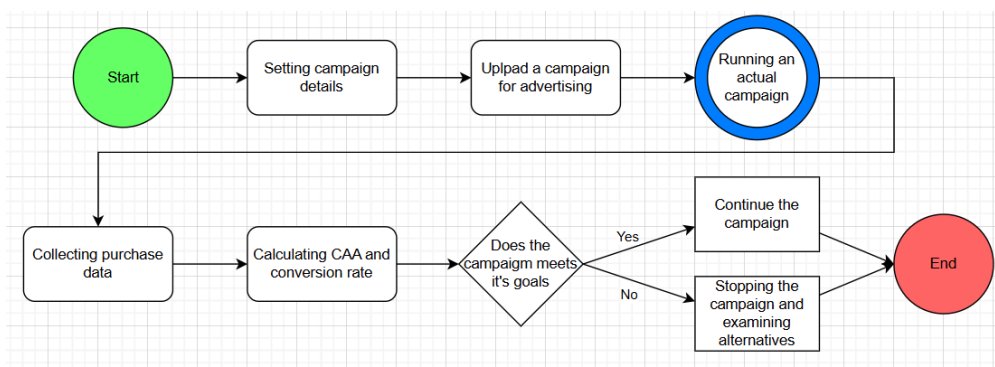
7.1 תיאור פורמלי של תהליכי העבודה בארגון

תהליך ניהול הקמפיינים הדיגיטליים ב-ReasonLabs מתבצע בגישה אינטואיטיבית הנשענת בעיקר על ניסיון מצטבר ונתוני רכישה בסיסיים. לאחר הגדרת הקמפיין, קביעת תקציב והעלאתו לפלטפורמות הפרסום השונות, נאספים נתונים ישירות מתוך הממשקים הדיגיטליים של אותן פלטפורמות. נתונים אלה כוללים בעיקר את כמות הרכישות שבוצעו כתוצאה מהקמפיין ואינם כוללים ניתוח מעמיק של שלבים מוקדמים יותר במסלול הלקוח, כגון חשיפות, הקלקות, מעורבות עם תוכן או דפוסי גלישה באתר. הערכת ביצועי הקמפיין נעשית לרוב לאחר תקופה מוגדרת מראש או לאחר הצטברות כמות תצפיות מספקת, באמצעות חישוב מדדים כמותיים כמו שיעור ההמרה ועלות רכישת לקוח. מדדים אלו מחושבים על ידי צוות ה-Performance באמצעות גיליונות Excel. בהתאם לתוצאות, מתקבלת החלטה אם להמשיך את הקמפיין, לעצור אותו או לבצע בו התאמות.

תיאור פורמלי של תהליך העבודה בחברה:

- Start - פתיחת תהליך והגדרת קמפיין חדש.
- Setting campaign details - קביעת פרטי הקמפיין: קהל יעד, קריאייטיב, תקציב ויעדים.
- Upload a campaign for advertising - העלאת הקמפיין לפלטפורמות הפרסום (Google Facebook ועוד).

- Running an actual campaign - הפעלת הקמפיין בזמן אמת.
- Collecting purchase data - איסוף נתוני רכישה בלבד.
- Calculating CAC and conversion rate - חישוב מדדים כגון עלות רכישה ממוצעת (CAC) ושיעור המרה (Conversion Rate).
- Evaluation - השוואת ביצועים מול היעדים.
- Decision - החלטה אם להמשיך בקמפיין או לעצור אותו ולבחון חלופו.

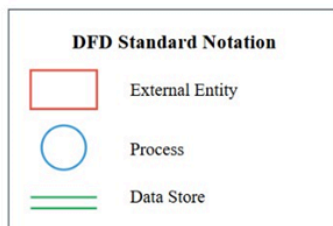
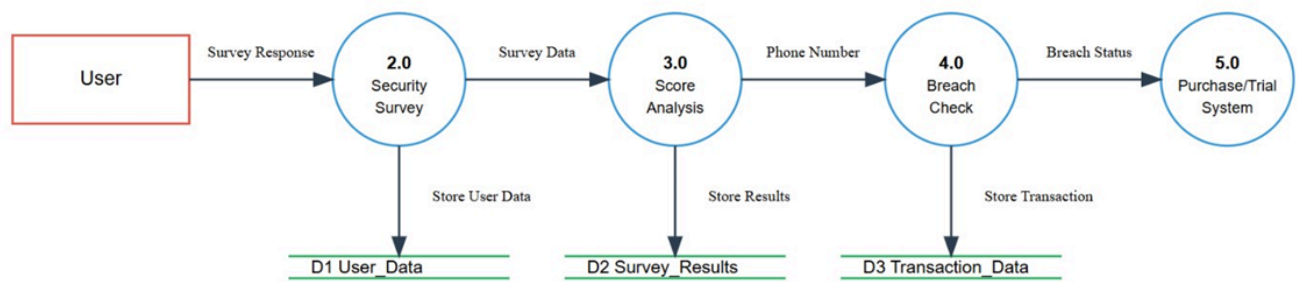


איור 4- תרשים זרימה של תהליך ניהול קמפיינים דיגיטליים במצב הקיים (AS IS)

תיאור פורמלי של זרימת הנתונים במערכת (DFD):

התרשים מציג את מסלול הפעולה של המשתמש בקמפיין, משלב האינטראקציה הראשונית ועד להצעת הרכישה. הוא מדגיש את זרימת הנתונים בין רכיבי המערכת לצורך קבלת החלטות שיווקיות.

רכיב	תיאור הפעולה
User	המשתמש החיצוני מתחיל את התהליך על ידי מילוי שאלון אבטחה.
Security Survey	המשתמש משיב על שאלון אבטחה, והתשובות נשלחות למערכת לעיבוד ראשוני.
User_Data	נתוני המשתמש נשמרים במאגר נתונים ייעודי.
Score Analysis	המערכת מחשבת ציון אבטחה אישי (Security Score) על בסיס הנתונים שנאספו.
Survey Results	תוצאות הניתוח נשמרות במאגר תוצאות נפרד.
Breach Check	נבדקת התאמה בין פרטי המשתמש (כגון מספר טלפון) לבין מאגרי פריצות, לצורך איתור דליפות מידע.
Transaction Data	תוצאות בדיקת הפריצה נשמרות במאגר עסקאות.
Purchase\ Trail System	בהתאם לממצאים, המשתמש מועבר להצעה לרכישה או להתחלת גרסת ניסיון (Freemium/Premium).

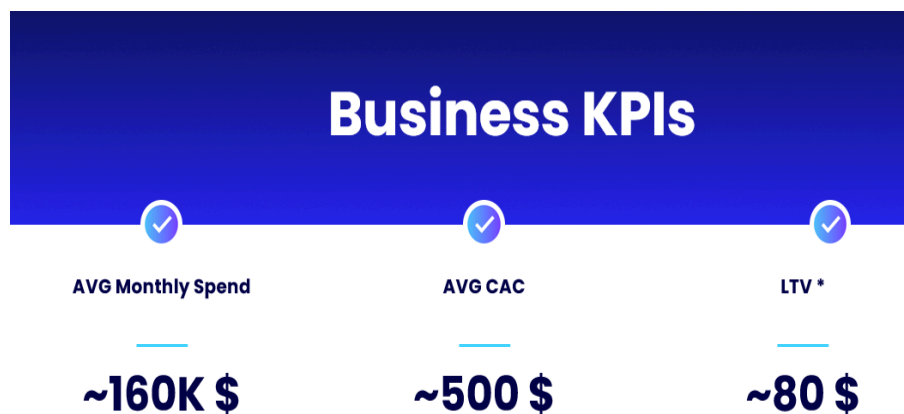


תרשים 5- תרשים DFD המתאר את שלבי עיבוד המידע

7.2 מדדים כמותיים של המצב הקיים

הנתונים מציגים שלושה מדדים מרכזיים:

- הוצאה חודשית ממוצעת על פרסום (AVG Monthly Spend)
- עלות רכישת לקוח (AVG CAC)
- שווי חיי לקוח (LTV).



איור 6- ממדים כמותיים במצב הקיים

מהנתונים הקיימים ניתן לראות פער משמעותי בין עלות רכישת לקוח לבין שווי חיי הלקוח, דבר המעיד על מודל שיווקי שאינו רווחי:

- עלות רכישת לקוח (CAC) עומדת על כ-500 דולר. מדובר בסכום גבוה, המשקף את ההשקעה הנדרשת כדי להמיר משתמש חנימני ללקוח משלם.
- שווי חיי לקוח (LTV) הוא כ-80 דולר בלבד כלומר, ההכנסה הכוללת הצפויה מכל לקוח לאורך כל תקופת השימוש שלו נמוכה משמעותית מעלות הרכישה.
- היחס בין CAC ל-LTV עומד על פי 6.25, כלומר החברה מוציאה פי שישה ויותר ממה שהיא מרוויחה על כל לקוח חדש.
- הוצאה חודשית כוללת על פרסום עומדת על כ-160,000 דולר, אך השקעה זו אינה מניבה החזר השקעה חיובי (ROI), לאור העלות הגבוהה של רכישת לקוחות שאינם מצדיקים כלכלית את ההוצאה עליהם.

המשמעות היא שהחברה מפסידה כסף על כל לקוח חדש, מה שמוביל למסקנה ברורה: כל עוד ה-CAC גבוה משמעותית מה-LTV, המודל העסקי דורש שינוי מהותי בגישה השיווקית ובקבלת ההחלטות התקציביות.

נוסחת חישוב עלות רכישת הלקוח (CAC):

$$CAC(\text{עלות רכישת לקוח}) = \frac{\text{Cost (עלות קמפיינים)}}{\text{Subscriptions Amount (כמות לקוחות משלמים)}}$$

7.3 תחזיות למצב הרצוי

הפתרון שיפותח במסגרת הפרויקט צפוי לשנות באופן מהותי את האופן שבו מתקבלות החלטות שיווקיות בחברה הן ברמת ניתוח הנתונים והן מבחינת היעילות הכלכלית של הקמפיינים. שילוב של מודל הסתברותי, אלגוריתמים של למידת מכונה ודשבורד אינטראקטיבי יאפשר לארגון לפעול באופן מבוסס-נתונים, לזהות מוקדם מגמות בפאנל ההמרה, ולמקד את התקציב בקהלים בעלי פוטנציאל גבוה.

- שיפור שיעור המרה: מ-2.2% כיום לכ-3% לפחות, באמצעות טרגוט מדויק וניתוח דפוסים שימושי.
- הפחתת עלות רכישת לקוח (CAC): מכ-500 דולר ל-350 דולר בממוצע, תוך ניצול תקציבי יעיל יותר.

- שיפור יחס CAC/LTV: מ-6.25 כיום ל-3 לכל היותר, כך ששווי חיי הלקוח יעלה משמעותית ביחס לעלות הרכישה.

המערכת תאפשר קבלת החלטות מבוססות על מכלול הנתונים לאורך כל הפאנל השיווקי, במקום להסתמך על תצפיות גולמיות בשלב הרכישה בלבד. תחזית זו כוללת גם הפחתה בתלות ב-A/B Testing והחלפתו בגישות הסתברותיות גמישות יותר, מה שיאפשר ניהול קמפיינים מותאם-דינמיקה והגברת ה-ROI.

8. מתודולוגיה:

בתהליך הפרויקט נעשה שימוש בשיטות הסתברותיות, אלגוריתמים של למידת מכונה וכלים ויזואליים, במטרה להתמודד עם שני אתגרים מרכזיים בשיווק הדיגיטלי: ניתוח אפקטיביות קמפיינים בתנאים של דלילות נתונים, וזיהוי הגורמים המשפיעים על רכישה לצורך ניהול מדויק ויעיל של תקציב הפרסום.

המתודולוגיה נשענת על שני כיווני פעולה משלימים:

1. שימוש בלמידת מכונה לזיהוי משתנים קריטיים בשלבים מוקדמים בפאנל השיווקי, באופן שמאפשר טרגוט ממוקד יותר, שיפור שיעורי ההמרה והפחתת עלויות.
2. יישום מודלים בייסיאנים להסקת מסקנות בתנאי אי-ודאות, אשר מאפשרים קבלת החלטות מושכלת גם כאשר יש דלילות בנתונים.

תהליך העבודה כלל אפיון נתונים, עיבוד מקדים, בניית מודלים מתקדמים והקמת דאשבורד אינטראקטיבי באמצעות Streamlit.

הדאשבורד יספק תמיכה אפקטיבית לצוות השיווק של החברה- בהעלאת נתונים, ניתוח ביצועי קמפיינים, זיהוי משתנים משמעותיים והשוואת חלופות לצורך אופטימיזציה של התקציב.

להלן פירוט שלבי המתודולוגיה, תוך הצגת השיטות, הכלים ותרומתם לפרויקט.

שלב 1: קבלת הנתונים ואפיון ראשוני של התהליך

בשלב הראשון, טענו את הנתונים לסביבת העבודה בפורמט DataFrame, וביצענו אפיון ראשוני שנועד להבין את מבנה הדאטה, את סוגי המשתנים (כמותיים, קטגוריאליים ובינאריים), את שיעורי ההמרה ואת רמת הדלילות במידע. אפיון זה שימש בסיס להבניית שלבי הניתוח והמודלים בהמשך. לאחר מכן, עברנו לעיבוד מקדים של הנתונים, שכלל מספר שלבים: השלמת ערכים חסרים: זיהינו שדות חסרים, בעיקר בשלבים השונים בפאנל השיווקי. מתוך הבנה שערכים חסרים מייצגים נטישת תהליך, המרה לערך 0 שימשה לייצוג חוסר השתתפות או הפסקת מעורבות מצד המשתמש במענה על השאלות בפאנל.

פיצול משתנים מרובי ערכים: חלק מהמשתנים הכילו תגובות מרובות (multi-select) בפורמט טקסטואלי. פיצלנו אותם לעמודות בינאריות, כדי לאפשר ניתוח תקני בלמידת מכונה.

בניית משתני רוחב: יצרנו משתנים חדשים המסמנים האם המשתמש השיב לפחות על פריט אחד בכל קטגוריה (למשל: שימוש באינטרנט, פעילות ברשתות החברתיות), לצורך ניתוח עומק. המרת טיפוסים משתנים: טיפלנו בשדות לוגיים וקטגוריאלים תוך המרתם לערכים בינאריים/מספריים, כדי להבטיח עיבוד עקבי ונקי.

חלוקה לסטים: חילקנו את הדאטה לשני סטים - סט אימון (Train) לצורך בניית המודלים וכיול ההיפר-פרמטרים, וסט בדיקה (Test) אשר לא נחשף לתהליך הלמידה ושימש להערכת יכולת ההכללה של המודלים על נתונים חדשים.

שלב 2: ניתוחים ויזואליים ואפיון דפוסי התנהגות

לאחר עיבוד הנתונים, עברנו לביצוע ניתוחים ויזואליים שנועדו לזהות מגמות ודפוסים המשפיעים על שיעורי ההמרה בכל קמפיין ולהבין את הקשר בין מאפיינים כמו ציון בטיחות דיגיטלית ואיתור פרצות אבטחה לבין הנטייה להתחיל תהליך רכישה.

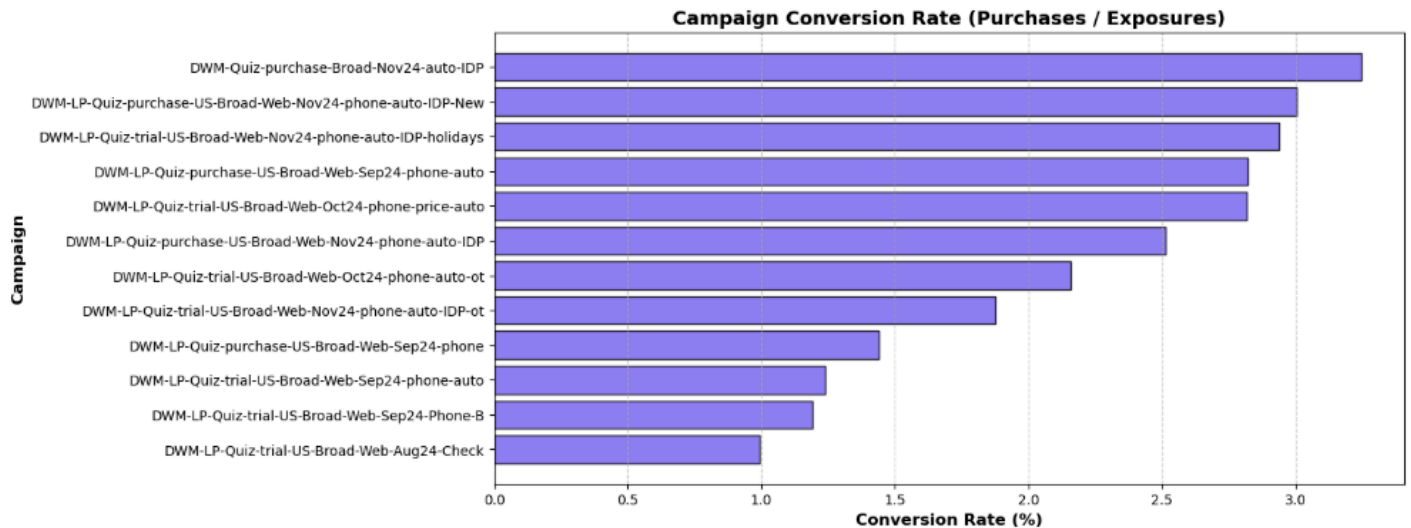
מטרת שלב זה הייתה לגבש תובנות ראשוניות מהנתונים לקראת תהליך בניית המודלים ולאפשר זיהוי מוקדם של בעיות פוטנציאליות או הזדמנויות חבויות. הניתוחים הוויזואליים סיפקו לנו פרספקטיבה איכותנית וכמותית כאחד: הם סייעו להבין אילו קמפיינים אינם אפקטיביים, ואילו פרופילים של משתמשים מציגים נטייה גבוהה יותר להתחיל תהליך רכישה.

לצורך כך נעשה שימוש בספריות matplotlib, seaborn ו-ipywidgets, אשר אפשרו יצירת גרפים אינטראקטיביים ודינמיים המתאימים לסביבות עבודה אנליטיות.

ניתוח שיעור ההמרה לפי קמפיין

באמצעות חישוב היחס בין רכישות לחשיפות לכל קמפיין, זיהינו אילו קמפיינים מציגים ביצועים טובים יותר, באופן יחסי לנפח החשיפה. מיזגנו את מספר החשיפות והרכישות לכל קמפיין וחישבנו את שיעור ההמרה. הנתונים הוצגו בגרף אופקי ממין לפי אפקטיביות.

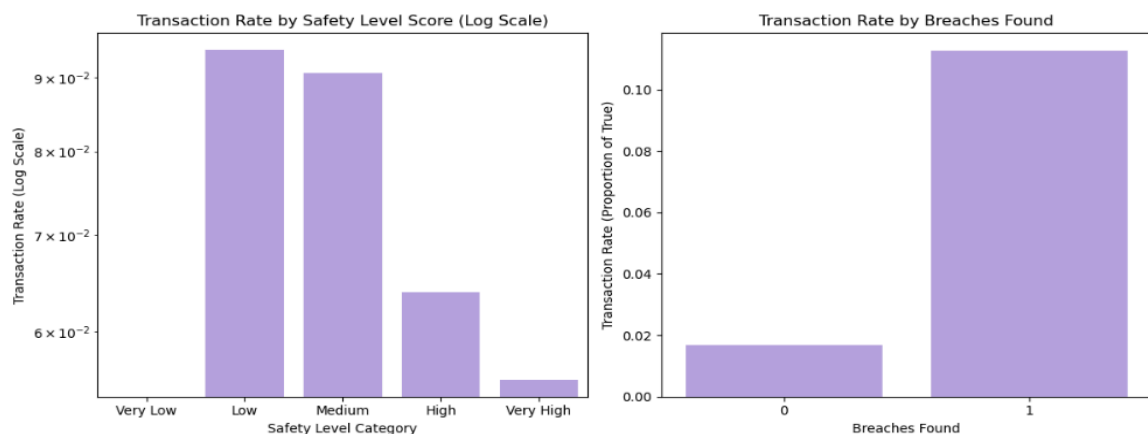
ניתן לראות שהפערים בין הקמפיינים משמעותיים, כאשר שיעור ההמרה בקמפיין המוביל גבוה פי כמה לעומת זה שבתחתית הרשימה. תובנה זו מאפשרת למקד משאבים בקמפיינים האפקטיביים ביותר ולזהות קמפיינים שיש מקום לבצע בהם שיפור או הסרה. גרף זה משמש בסיס להחלטות טקטיות בנוגע לחלוקת תקציב עתידית ומציב מדד כמותי ברור להצלחת קמפיין.



איור 7- אחוז ההמרה של הקמפיינים

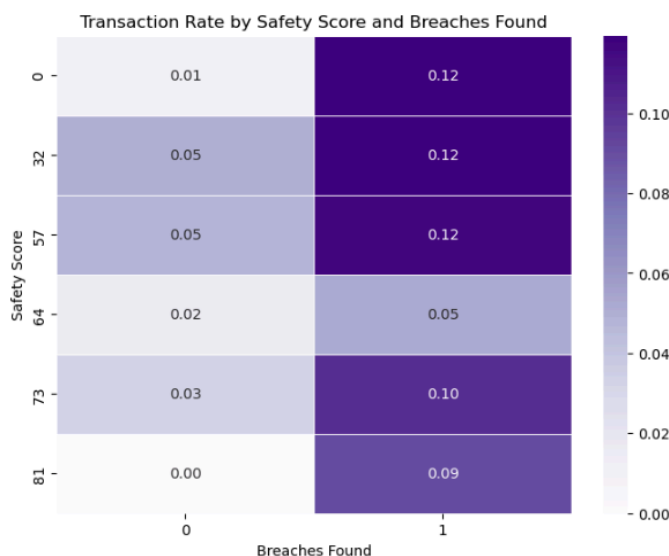
קשר בין רמת בטיחות דיגיטלית לבין התחלת תהליך רכישה

לשם בחינת ההשפעה של משתני בטיחות על התנהגות הרכישה, ביצענו חלוקה של ציון הבטיחות הדיגיטלית (safety score) לקטגוריות מדרגות (Very Low עד Very High), וניתחנו את שיעור התחלת תהליך הרכישה (transaction_start) בכל קטגוריה. תוצאות הניתוח הוצגו בגרף בעל סקלת לוגריתם, שאיפשרה זיהוי מגמות גם בקבוצות עם שיעורים נמוכים מאוד. מהממצאים עלה קשר חיובי ברור- ככל שרמת הבטיחות הדיגיטלית גבוהה יותר, כך עולה ההסתברות שהמשתמש יתחיל תהליך רכישה. בנוסף, ניתחנו את הקשר בין משתנה הפרצות (breach_found) לבין התחלת הרכישה, ומצאנו הבדלים מובהקים בין משתמשים שחוו פרצת אבטחה לבין אלו שלא. השוואה ויזואלית הצביעה על כך שפרצות אבטחה עשויות לתרום לתחושת דחיפות או למודעות מוגברת, אשר מובילות לתחילת תהליך רכישה בפועל.



איור 8 - השפעת ציון הבטיחות וזיהוי פריצה על הרכישה.

לניתוח המשולב בין ציון בטיחות לבין איתור פרצות יצרנו טבלת ציר תלת-ממדית על בסיס ממוצע התחלת רכישה, והוצגה באמצעות Heatmap, שמספקת תובנות אינטואיטיביות על קבוצות סיכון והזדמנות.



ה-Heatmap מציג שתי תובנות עיקריות:

1. משתמשים שחוו פרצת אבטחה נוטים להתחיל תהליך רכישה בשיעור גבוה יותר, ככל הנראה מתוך תחושת דחיפות או מודעות לסיכון.
2. גם משתמשים עם ציון בטיחות גבוה נוטים לרכוש, ייתכן מתוך רצון לשמר את רמת ההגנה שלהם.

איור 9 - שיעור ההמרה לפי רמות שונות של Safety Score

וקיום פריצה Breaches Found

שלב 3: ניתוח משתנים משפיעים באמצעות למידת מכונה

לצורך ניתוח ההשפעה של משתני הפאנל השיווקי על הסבירות לרכישה, נעשה שימוש באלגוריתם Random Forest Classifier. המודל אפשר להעריך את התרומה היחסית של כל שלב בתהליך ושל תשובות המשתמשים לשאלון, במטרה לזהות אילו פרמטרים משמשים כמנבאים החזקים ביותר להמרה. ממצאים אלה מהווים בסיס לשיפור אסטרטגיות הטרגוט והתאמת הקמפיינים לקהלי יעד בצורה מדויקת ואפקטיבית יותר. מודל זה נבחר בזכות עמידותו לדאטה רועש, יכולתו להתמודד עם משתנים מקטגוריות שונות, וכן בשל הביצועים הגבוהים שלו בהקשרים של ניבוי בינארי, במקרה של הפרויקט שלנו האם בוצעה רכישה או לא. המודל מאמן מספר רב של עצי החלטה על דגימות אקראיות שונות מתוך הנתונים ולאחר מכן מבצע חישוב משוקלל כדי לקבוע את התוצאה הסופית. תהליך זה מצמצם סיכון להתאמת יתר ומאפשר ניתוח מדויק של המשתנים משפיעים על הרכישה. המודל מפיק דירוג של המשתנים המשפיעים ביותר על רכישה בפועל, מהם ניתן היה להסיק מהם השלבים בפאנל שבהם המשתמשים נוטים לנטוש, ואילו מאפיינים מעלים את סיכויי ההמרה.

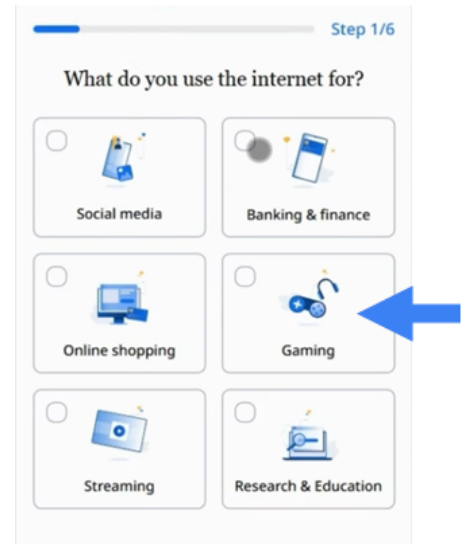
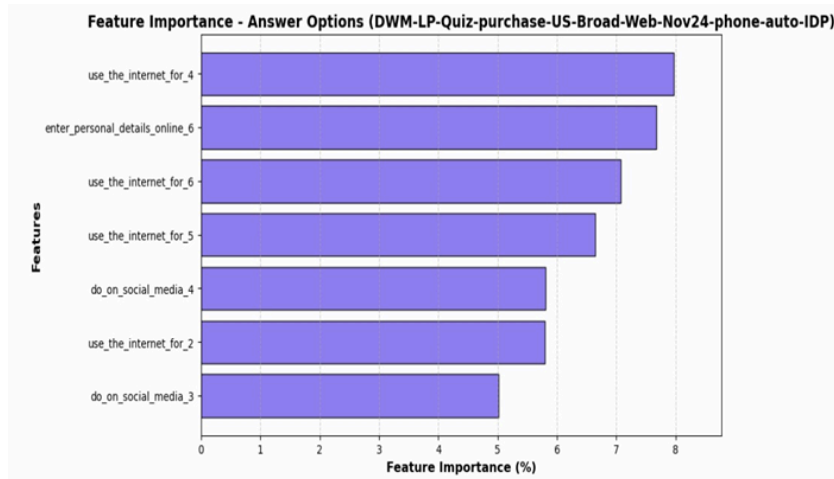
לצורך כך בוצעו שני סוגי ניתוח:

1. ניתוח עומק - בוחן אילו תשובות ספציפיות מתוך שאלות מרובות-בחירה (כגון העדפת שימוש

באינטרנט לגיימינג או לבנקאות) משפיעות ביותר על הרכישה.

תועלתו של ניתוח זה היא בכך שהוא מאפשר להבין אילו תחומי עניין הם גורמים לרכישה

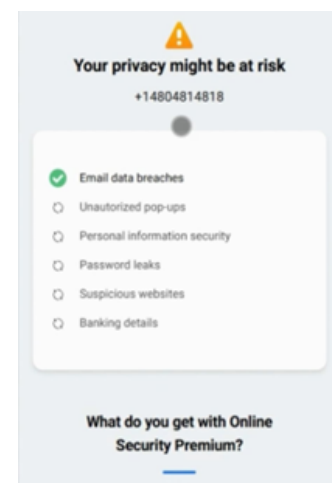
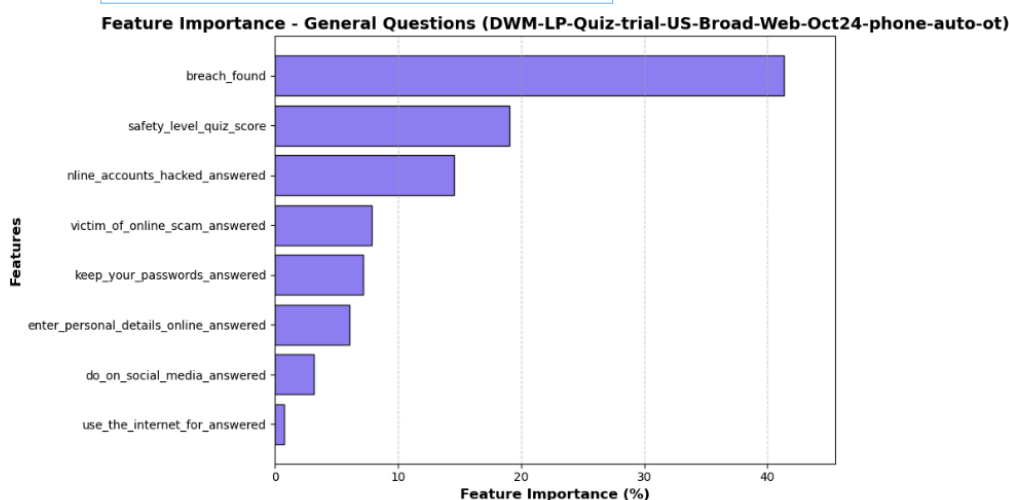
בפועל. תובנה זו יכולה לשמש לשיפור הקריאייטיב של הקמפיינים, להכנסת מסרים רלוונטיים (כגון אזכור של עולם הגיימינג לקהלים המתאימים), ולטרגוט מדויק יותר של תתי-קהלים לפי תחומי עניין מובהקים.



איור 10: דירוג חשיבות תשובות המשתמשים, כולל שאלות רבות-ברירה, כמשתנים מנבאים לרכישה

2. ניתוח רוחב - מתמקד בבחינת שלב מסוים בתוך הפאנל השיווקי, ללא תלות בתשובות הספציפיות שנבחרו בו. מטרת ניתוח זה היא להבין את רמת המעורבות של המשתמשים בשלב הנתון, ולאתר שלבים בתהליך שקיים בהם קשר חיובי או שלילי לאחוזי ההמרה. תובנות אלו מהוות בסיס לשיפור מבנה הפאנל כולו: חיזוק שלבים בעלי השפעה גבוהה על ההמרה, ושיפור חוויית המשתמש באותם שלבים שנמצאו כפחות אפקטיביים.

Select Campaign: DWM-LP-Quiz-trial-US-Broad-Web-Oct24-phone-auto-ot



איור 11 - דירוג חשיבות תשובות המשתמשים כמשתנים מנבאים לרכישה

שלב 4: בניית מודל הסתברותי בייסיאני

בשלב זה נעשה שימוש במודל הסתברותי בייסיאני לצורך חיזוי והערכת יחס ההמרה של קמפיינים שיווקיים. הבחירה בגישה בייסיאנית נובעת מהתאמתה למצבי אי-ודאות ודלילות נתונים, תופעה נפוצה בקמפיינים דיגיטליים שבהם נפח התצפיות מוגבל ולעיתים נצברות רכישות בודדות בלבד.

הגישה הבייסיאנית שונה מהגישה הקלאסית בכך שהיא אינה מנסה לאמוד את יחס ההמרה כערך נקודתי בלבד, אלא מייצגת אותו כהתפלגות הסתברותית הנבנית על סמך שילוב של ידע מוקדם (Prior) עם נתוני תצפית עדכניים (Likelihood).

תהליך זה מניב את ההסתברות המותנית (Posterior), שמעניקה הערכה מעודכנת ואמינה יותר ביחס להסתברות ההמרה.

לצורך היישום נעשה שימוש בהתפלגות מסוג Beta, אשר מתאימה באופן טבעי להערכת הסתברויות עבור משתנים בינאריים - כמו הצלחה או כישלון (רכישה לעומת אי-רכישה). יתרונה של התפלגות זו טמון בכך שהיא מוגדרת בתחום $[0, 1]$ טווח טבעי להסתברויות וכן בכך שהיא מאפשרת עדכון פשוט של הפרמטרים עם כל תצפית חדשה.

בהגדרת המודל:

- הפרמטר α מייצג את מספר הרכישות, בתוספת 1.
- הפרמטר β מייצג את מספר חשיפות שלא הובילו לרכישה, בתוספת 1.

באופן זה, לכל קמפיין נוצרת התפלגות מסוג $Beta(\alpha, \beta)$ שמתארת את ההסתברות להמרה.

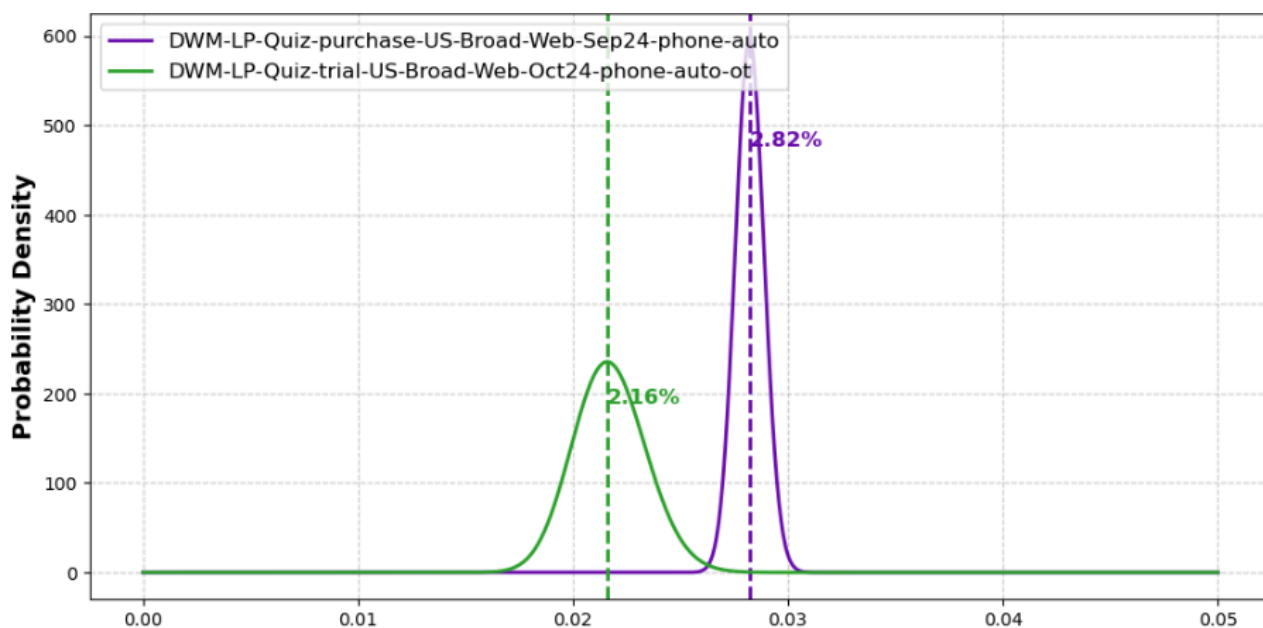
ממנה ניתן לחשב:

- ממוצע הסתברות ההמרה - שהוא תחזית המרכז של המודל.
- תחום אמינות (Credible Interval) - תחום סטטיסטי המציג את מידת אי-הוודאות של התחזית, בהתאם לגודל המדגם.

תחום זה חשוב במיוחד בקמפיינים בעלי מספר תצפיות קטן, מאחר שהוא מאפשר לזהות מתי אי-הוודאות כה גבוהה עד שאין מקום להסיק מסקנה מהנתונים. בנוסף, ניתן לבצע השוואה בין קמפיינים גם כשמספר התצפיות ביניהם אינו אחיד, דבר שמסייע בהקצאת תקציב מושכלת גם בשלב מוקדם של קמפיין.

האיור שלהלן מציג את ההתפלגות ההסתברותית של יחס ההמרה עבור שני קמפיינים שונים, כפי שחושבה באמצעות המודל הבייסיאני המבוסס על התפלגות Beta. כל עקומה מתארת את ההסתברות של יחס ההמרה האפשרי עבור קמפיין מסוים, תוך שקלול מספר החשיפות והרכישות שנרשמו בו בפועל. הקווים האנכיים מסמנים את ממוצע ההמרה עבור כל קמפיין - כלומר, התחזית

המרכזית שהמודל חזה. רוחב ההתפלגות מבטא את רמת אי-הוודאות: קמפיין עם מדגם גדול יציג התפלגות צרה וממוקדת יותר ואילו קמפיין עם פחות תצפיות יתאפיין בהתפלגות רחבה יותר, המעידה על חוסר ודאות גבוה. תצוגה זו מאפשרת לבצע השוואה בין קמפיינים גם כאשר מספר התצפיות שונה ביניהם, ולהעריך את ההסתברות שקמפיין אחד עדיף על אחר. בדרך זו ניתן לייצר בסיס החלטה מבוסס הסתברות, המאפשר להקצות תקציבים בצורה מושכלת עוד בשלב מוקדם של הקמפיין, גם כאשר כמות הנתונים אינה מספקת לניתוח קלאסי.



איור 12- השוואת תחזיות ההמרה בין שני קמפיינים באמצעות מודל בייסיאני

שלב 5: כיול מודל בייסיאני על בסיס נתוני אמת

מטרת הכיול: תהליך הכיול נועד לבדוק עד כמה ההתפלגות שחושבה על ידי המודל מתאימה לנתוני ההמרה שנמדדו בפועל.

אופן הבדיקה: שלב זה מבוסס על דגימות אקראיות מתוך ההתפלגות הבייסיאנית (לרוב התפלגות Beta). עבור כל קמפיין, נמשכים ערכים רבים מתוך ההתפלגות הפוסטרירורית של המודל, ובבדק מהו האחוז מתוך הדגימות שנמצאות מתחת לשיעור ההמרה האמיתי שנמדד בפועל.

קריטריונים להערכת איכות הכיול: אם כ-50% מהדגימות נמוכות מהשיעור האמיתי המודל נחשב מכויל היטב, שכן התחזיות מתפלגות באופן סימטרי סביב הערך האמיתי.

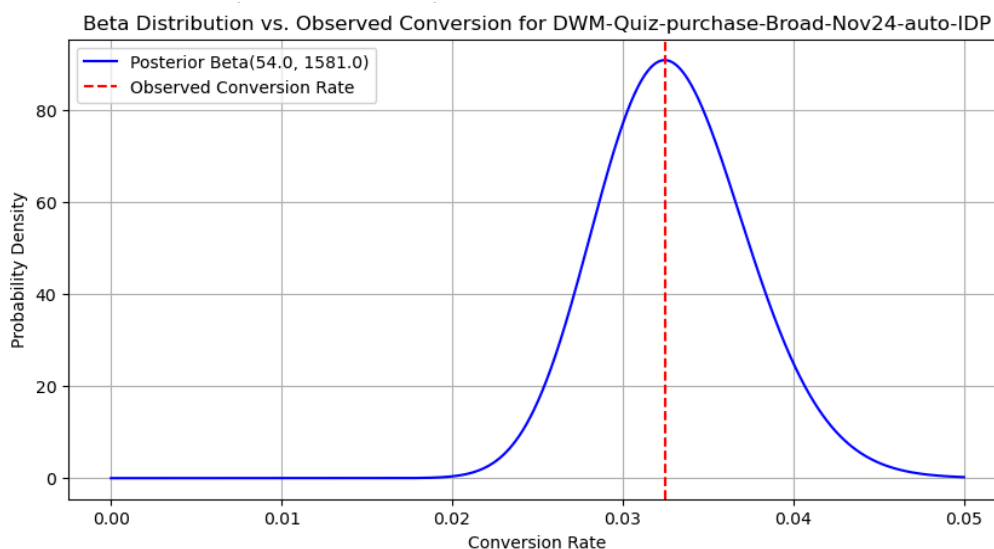
אם קיימת סטייה משמעותית, למשל פחות מ-30% או יותר מ-70% מהדגימות נמצאות מתחת לערך האמיתי הדבר מעיד על כך שהמודל אינו מייצג היטב את המציאות.

איזון התחזית בעזרת Prior היסטורי: במקרים כאלה, מבוצע תיקון באמצעות Prior היסטורי: מידע מצטבר מקמפיינים קודמים (למשל שיעור המרה ממוצע) מתורגם לערכי α_{prior} ו- β_{prior} , ונוסף לפרמטרים המקוריים של הקמפיין. בצורה זו מתקבלת התפלגות מאוזנת יותר, שמשקפת טוב יותר גם את הידע הקיים וגם את חוסר הוודאות.

כדי להמחיש את רמת הדיוק של המודל באופן אינטואיטיבי, נעשה שימוש בגרף השוואתי. גרף זה מאפשר לראות בצורה ברורה האם ההתפלגות הבייסיאנית מתאימה לשיעור ההמרה שנמדד בפועל, או שקיימת סטייה.

בגרף מוצגים המרכיבים הבאים:

- הקו הכחול: ההתפלגות הפוסטרירית (Posterior) של שיעור ההמרה שהתקבלה על סמך הנתונים שנאספו.
- הקו האדום המקווקו: שיעור ההמרה האמיתי שנמדד בפועל בקמפיין.
- ציר ה-X: טווח הערכים האפשריים של שיעור ההמרה.
- ציר ה-Y: צפיפות ההסתברות - עד כמה כל ערך נחשב סביר לפי המודל.



Model is already well-calibrated! No adjustments needed.

איור 13- כיול מודל בייסיאני, התפלגות פוסטרירית לעומת שיעור ההמרה בפועל

שלב 6: פיתוח דשבורד תומך החלטה

בשלב הסופי של הפרויקט יפותח דשבורד אינטראקטיבי באמצעות ספריית Streamlit, אשר ישמש את מחלקת השיווק ככלי תפעולי ותומך החלטה. הדשבורד יאפשר להזין נתוני קמפיינים חדשים ולקבל תובנות מיידיות בנוגע להצלחתם, תוך שימוש במודל הבייסיאני.

המערכת תשלב בין ניתוח הסתברותי לבין הצגת מידע ויזואלי ברור, ותכלול את הרכיבים הבאים:

- הסתברות המרה לכל קמפיין, כולל תחומי אמינות
- השוואת ביצועים בין קמפיינים גם בתנאים של דלילות נתונים
- הצגת מדדים שיווקיים כגון Conversion Rate, CPC ו-CAC
- ניתוח חשיבות משתנים המבוסס על Random Forest
- תצוגה גרפית של נטישת משתמשים לאורך הפאנל השיווקי

הדשבורד יעוצב כך שיוכל לשמש גם משתמשים לא טכניים, ויכולל יכולות סינון, אינטראקטיביות והצגת גרפים בזמן אמת. כלי זה צפוי להחליף תהליכי ניתוח ידניים ולהפוך את תהליך קבלת ההחלטות השיווקיות למהיר, מושכל ומבוסס נתונים. תוך התמודדות עם אתגרי אי-ודאות ודאטה דל.

9. בחינת חלופות**9.1 כללי**

בפרויקט שלנו נבחנו מספר חלופות הנדסיות לשיפור תהליך קבלת ההחלטות בקמפיינים שיווקיים, במטרה להתמודד עם מגבלות של דלילות נתונים, עלויות פרסום גבוהות, והצורך בטרנספרט יעיל יותר. כל חלופה מציעה יתרונות ואתגרים שונים, ולכן הערכתן נעשתה לפי קריטריונים מרכזיים: דיוק חיזוי, מורכבות, עלות יישום, יכולת פרשנות ועמידות לדאטה דליל.

9.2 תהליך הערכת החלופות

כל חלופה הוערכה על פי חמישה קריטריונים, כאשר לכל אחד מהם ניתן ציון בין 1 ל-5 (כאשר 5 הוא הציון הגבוה ביותר). משקלו של כל קריטריון, המבטא את חשיבותו היחסית, נקבע מראש על ידי צוות ההנהלה של ReasonLabs. הציון הסופי של כל חלופה חושב באמצעות ממוצע משוקלל: כל ציון הוכפל במשקלו הרלוונטי, והתוצאה סוכמה לקבלת ציון כולל.

9.3 הצגת החלופות**9.3.1 המצב הקיים**

כיום תהליך קבלת ההחלטות בקמפיינים השיווקיים מתבסס בעיקר על ניתוח ישיר של מדדי ביצוע בסיסיים, כגון מספר חשיפות, הקלקות ורכישות. ההחלטות בנוגע להקצאת תקציב וטרנספרט קהלים מתקבלות תוך הסתמכות על טבלאות ודו"חות סטנדרטיים, ללא תמיכה

במודלים הסתברותיים או בכלים חיזויים מתקדמים. שיטה זו מאפשרת תגובה מהירה לנתונים שמופיעים בזמן אמת, אך מוגבלת ביכולתה להפיק תובנות כאשר המידע חלקי או לא מובהק סטטיסטית.

9.3.2 חלופה א - רגרסיה לוגיסטית

מודל הרגרסיה הלוגיסטית מציע פתרון סטטיסטי פשוט אך אמין לבעיות סיווג בינאריות, באמצעות הערכת הסתברות התוצאה על בסיס משתנים מסבירים. המודל מתבסס על קשר ליניארי בין הפיצ'רים לבין הלוג-אודס של התוצאה, ומתאים במיוחד לבעיות שבהן נדרש מענה מהיר, פרשנות בהירה והתמודדות עם דאטה מוגבל. יתרונותיו המרכזיים כוללים יעילות חישובית גבוהה, קלות בהטמעה והבנה, ויכולת להסביר במדויק את תרומת כל משתנה לתוצאה - תכונה חשובה במיוחד במערכות קבלת החלטות שמערבות שיקול דעת אנושי. עם זאת, הרגרסיה הלוגיסטית מוגבלת ביכולתה להתמודד עם מבנים מורכבים או אינטראקציות לא ליניאריות בין פיצ'רים. לכן, במקרים שבהם קיימת תלות מורכבת בין משתנים או דרישה לביצועים תחזיתיים גבוהים, עשוי המודל שלא להספיק ולדרוש פתרונות מתקדמים יותר.

9.3.3 חלופה ב - מודלים מתקדמים בלמידת מכונה

חלופה זו מתבססת על מודלים מתקדמים מעולם למידת המכונה (Machine Learning), כגון Random Forest, LightGBM ו-XGBoost. מודלים אלו ידועים ביכולתם להתמודד עם מערכי נתונים מורכבים, בעלי ממדיות גבוהה ודפוסים לא ליניאריים, ובכך להשיג רמות דיוק תחזיתי גבוהות.

יתרונם המרכזי הוא ביכולת לזהות קשרים מורכבים בין משתנים ולהסתגל לשינויים בדאטה באופן שמודלים מסורתיים מתקשים בו. תכונות אלו הופכות אותם לאופציה מועדפת בסביבות בהן נדרש דיוק תחזיתי גבוה במיוחד. עם זאת, השימוש במודלים אלו כרוך לרוב בדרישות חישוביות גבוהות, רמות מורכבות יישומית, וצורך בהתאמה למבנה הדאטה. לכן, מדובר בגישה בעלת פוטנציאל תחזיתי גבוה, אך גם כזו המצריכה שיקול דעת בבחירתה.

9.3.4 חלופה ג - מודל בייסיאני

הגישה הבייסיאנית מציעה מסגרת הסתברותית לניתוח נתונים, הנשענת על שילוב בין ידע מוקדם לבין מידע חדש המתקבל מהדאטה. באמצעות עדכון מתמיד של ההסתברויות, ניתן להעריך את הסיכוי להתרחשות של תוצאה מסוימת תוך ניתוח מפורט של אי-הוודאות. היתרונות המרכזיים בגישה זו כוללים יכולת התמודדות עם מערכי נתונים קטנים או חלקיים, אינטגרציה של מומחיות מוקדמת כחלק מתהליך החיזוי, ומתן תחזיות עשירות המבוססות על התפלגויות הסתברותיות ולא רק על ערכים נקודתיים. בכך, היא מאפשרת גיבוש הבנה

מעמיקה יותר של הסביבה הנחקרת. השיטה דורשת תכנון מדויק של ההתפלגויות ההתחלתיות, ולעיתים גם מצריכה משאבים חישוביים נרחבים. לכן, מדובר בגישה מתקדמת הדורשת שיקול דעת תשתיתי ויישומי בהתאם לאופי הדאטה ולמטרות הניתוח.

9.4 השוואה בין חלופות

המשקלים שנקבעו לכל אחד מהקריטריונים מבטאים את תרומתם היחסית להצלחת הפרויקט, בהתאם לצרכים ולאתגרים בתהליכי השיווק של ReasonLabs.

- דיוק החיזוי (40%) הוגדר כקריטריון המרכזי, לאור החשיבות הקריטית של קבלת החלטות מדויקות בתנאים של חוסר ודאות ודאטה דליל. היכולת להעריך את הסתברות הרכישה בצורה מהימנה מהווה תנאי בסיסי להקצאת תקציבי פרסום בצורה אופטימלית.
- עלות היישום (20%) משקפת את מגבלות התקציב של מחלקת השיווק, המחייבות פתרון שאינו רק מדויק, אלא גם בר-ביצוע ברמה הכלכלית, במיוחד לאור הפצה רחבה של קמפיינים במספר פלטפורמות.
- מורכבות (15%) מתייחסת לדרישות הפיתוח, ההטמעה והתחזוקה של המודל לאורך זמן. מודלים פשוטים יותר מאפשרים הסתגלות מהירה יותר של הצוות, קיצור זמני הטמעה והפחתת תלות באנשי דאטה.
- עמידות לדאטה דליל (15%) מהווה שיקול מהותי לאור האתגר המרכזי שנבחן בפרויקט- ניתוח קמפיינים בעלי כמות חשיפות נמוכה או תיעוד חלקי. פתרונות שאינם מתמודדים היטב עם תצפיות מוגבלות, עלולים להוביל למסקנות שגויות.
- יכולת פרשנות (10%) מגלמת את הצורך בשקיפות של המודל וביכולת להסביר את תרומתם של משתנים שונים לתוצאה. קריטריון זה חשוב במיוחד כאשר תובנות המודל משמשות לקבלת החלטות גם על ידי גורמים שאינם בעלי רקע אנליטי.

קריטריון	משקל (%)	גישה בייסיאנית	למידת מכונה	רגרסיה לוגיסטית
דיוק	40%	4	5	3
מורכבות	15%	3	2	5
עלות	20%	3	3	4
יכולת פרשנות	10%	4	3	5

2	3	5	15%	עמידות לדאטה דליל
3.55	3.65	3.8	100%	ציון סופי משוקלל

9.5 קביעת החלופה האופטימלית

הערכת החלופות שבוצעה הצביעה על יתרון לחלופה הבייסיאנית, אשר קיבלה את הציון המשוקלל הגבוה ביותר בהתאם למכלול הקריטריונים שנקבעו. חוזקה העיקרי של גישה זו בא לידי ביטוי בעמידותה לנתונים דלים וביכולתה להעריך הסתברויות בצורה מבוססת, יתרון מהותי בתהליכי קבלת החלטות בתנאי אי ודאות. עם זאת, הוחלט לאמץ שילוב בין הגישה הבייסיאנית לבין חלופת למידת המכונה, אשר דורגה במקום השני אך בלטה ביכולתה להשיג רמת דיוק גבוהה במיוחד, קריטריון שהוגדר כבעל המשקל המרכזי בתהליך הבחירה. השילוב בין שתי החלופות מאפשר להפיק את המיטב מכל אחת מהן: מצד אחד, ניתוח הסתברותי רגיש לנתונים חלקיים, ומצד שני, יכולת חיזוי מדויקת הנשענת על למידת דפוסים מתוך הנתונים הקיימים. בחירה אינטגרטיבית זו משקפת תפיסה אסטרטגית שמאזנת בין חוזקות של שיטות שונות, במטרה להבטיח פתרון אופטימלי השומר על איזון בין אמינות התחזית, יעילות אנליטית ונגישות יישומית.

10. מימוש הפתרון:

הפתרון יושם באמצעות פיתוח דשבורד אינטראקטיבי מבוסס פלטפורמת Streamlit, שנועד לייעל את תהליך קבלת ההחלטות השיווקיות. הדשבורד משלב בין ניתוח הסתברותי בייסיאני לבין אלגוריתם למידת מכונה (Random Forest), ומאפשר למשתמשים להפיק תובנות הן ברמת קמפיין יחיד והן בהשוואה בין קמפיינים, תוך זיהוי שלבים קריטיים ונקודות השפעה לאורך הפאנל השיווקי. המערכת פותחה כממשק אינטואיטיבי המחולק לשלבים לוגיים: התחברות, טעינת נתונים, ניתוחים והפקת תובנות. הדגש במימוש היה על פשטות שימוש, אחידות עיצובית וזרימה טבעית בין שלבי העבודה, כך שגם משתמשים ללא רקע טכנולוגי יוכלו להפיק ערך מהכלי באופן עצמאי.

במסגרת מימוש מודלים של Machine Learning, יושם הפתרון בהתאם למתודולוגיה שהוגדרה בפרויקט. מטרת העל הייתה לפתח מודל חיזוי אמין, יציב ונגיש, אשר יספק לארגון כלי לקבלת החלטות שיווקיות מבוססות-נתונים גם בתנאים של חוסר ודאות ודלילות מידע.

יישום הפתרון באמצעות למידת מכונה:

1. חלוקת הנתונים- הדאטה חולק לסט אימון וסט בדיקה ביחס של 70%-30%. חלוקה זו נועדה לאפשר אימון המודלים על חלק מהנתונים, ולאחר מכן להעריך את ביצועיהם על נתונים חדשים שלא נחשפו בשלב האימון, ובכך למנוע בעיית overfitting.

2. איזון הנתונים- נמצא כי הנתונים אינם מאוזנים באופן טבעי, כלומר מספר הדוגמאות של חלק מהמחלקות קטן משמעותית מאחרות. מצב זה עלול להוביל לכך שהמודלים יתעלמו מהמיעוט ויעדיפו את המחלקה הדומיננטית. כדי להתמודד עם בעיה זו יושמה שיטת: SMOTE- Synthetic Minority Over-sampling Technique, היוצרת דגימות מלאכותיות חדשות של המיעוט ומאזנת את קבוצות האימון.

3. בחירת המודלים - לצורך השוואה נבחרו שלושה מודלים מרכזיים בלמידת מכונה: Logistic Regression- מודל בסיסי להבנת קווי ההפרדה בין מחלקות. Random Forest- אלגוריתם מבוסס עצי החלטה אקראיים, היוצר אננסמבל של עצים לשיפור הדיוק. Gradient Boosting- אלגוריתם המבצע חיזוק הדרגתי של טעויות עבר ומשפר את ביצועי העצים.

4. הערכת ביצועים- ביצועי המודלים הוערכו באמצעות מדדים מקובלים: Accuracy- שיעור החיזויים הנכונים הכולל. Precision- אחוז התחזיות החיוביות שהיו נכונות בפועל. Recall- אחוז הדוגמאות החיוביות שהמודל הצליח לזהות מתוך כלל הדוגמאות החיוביות. F1 Score- ממוצע הרמוני של Precision ו-Recall, המעניק איזון בין השניים. ROC-AUC- מדד המודד את יכולת המודל להבחין בין המחלקות לאורך ספים שונים.

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Logistic Regression	0.960	0.094	0.089	0.092	0.518
Random Forest	0.778	0.062	0.623	0.113	0.754
Gradient Boosting	0.867	0.076	0.435	0.129	0.714

הסבר התוצאות:

מודל	יתרונות / חסרונות	ניתוח תוצאות	מסקנה
Logistic Regression	דיוק (Accuracy) גבוה מאוד 0.96, אך Precision, Recall ו-F1 נמוכים מאוד. המודל נוטה לנבא שאין רכישה ברוב המקרים ולכן המדדים המאוזנים חלשים.	המודל מציג תוצאה מוטעית של דיוק גבוה בגלל חוסר איזון בנתונים, אך לא באמת מצליח לנבא רכישות בפועל.	לא מתאים - לא מספק יכולת חיזוי אמיתית.
Random Forset	מציג איזון טוב יחסית בין המדדים. Recall גבוה (0.623), כלומר המודל מצליח לזהות אחוז גבוה יחסית מהמקרים, של רכישה בפועל. ROC-AUC = 0.754 הטוב ביותר מכל המודלים.	בפועל המודל מצליח לתת תגובה מאוזנת למדדים האחרים ומתאים יותר לניבוי, גם Recall סביר וגם דיוק חיזוי טוב יותר מכל שאר המודלים.	המודל שנבחר - מציע את האפשרות האופטימלית בין דיוק, Recall וביצועים כלליים.
Gradient Boosting	ביצועים בינוניים - ROC-AUC טוב (0.714), אך Recall נמוך (0.435). Precision נמוך מאוד (0.076).	מצליח במידה מסוימת להבחין בין קטגוריות לא מאוזנות, אך מאבד הרבה מהיכולות מול מדדים אחרים.	חלש מול Random Forest - לא נבחר.

פיתוח הדשבורד ומבנה המערכת:

עמוד התחברות (Login Page)

תחילה יש להתחבר למערכת באמצעות עמוד התחברות, שמטרתו להבטיח גישה מבוקרת לדשבורד. ההתחברות מתבצעת באמצעות הזנת שם משתמש וסיסמה, ורק משתמשים מורשים יכולים להמשיך לשלב הבא.

עמוד העלאת נתונים (Upload Page)

לאחר ההתחברות, המשתמש מועבר לעמוד ייעודי להעלאת קובץ הנתונים. בשלב זה יש להעלות קובץ Excel המכיל את נתוני הקמפיינים, לצורך המשך תהליך העיבוד והניתוח. העמוד תומך בגרירת קובץ או בבחירה ידנית, ומוגבל לקבצים עד 200MB. לאחר טעינה מוצלחת, המערכת מזהה את שם המשתמש המחובר ומבצעת עיבוד ראשוני של הנתונים, אשר מאפשר מעבר חלק לעבודה בדשבורד עצמו. הגישה לכלי הניתוח נפתחת רק לאחר העלאת תיקינה של הקובץ.

מבנה הדשבורד:

לאחר טעינת הנתונים, המשתמש מנותב לדשבורד הראשי, הכולל ארבעה טאבים פונקציונליים, אשר מייצרים רצף עבודה אינטואיטיבי - ממבט-על, דרך השוואות בין קמפיינים, ועד ניתוח מעמיק של קמפיין יחיד:

- Tab 1 - Home: מציג מבט-על על כלל הקמפיינים הפעילים. הטאב כולל מדדים מרכזיים כגון מספר משתמשים, כמות רכישות ושיעור ההמרה. בנוסף, מוצגת טבלת ניווט המאפשרת למשתמש לבחור קמפיינים להמשך ניתוח בטאבים הבאים.
- Tab 2 - Compare Campaigns: מאפשר להשוות בין שני קמפיינים נבחרים באמצעות מודל הסתברותי בייסיאני. מוצגות התפלגויות בטא עבור שיעורי ההמרה, חישוב uplift בין הקמפיינים, והסתברות לכך שקמפיין אחד עדיף על השני.
- Tab 3 - Analyze Single Campaign: מספק ניתוח מעמיק לקמפיין בודד. הטאב כולל גרפים אינטראקטיביים המציגים את משפך ההמרה (funnel), ניתוח feature importance של מודל Random Forest, פילוחים לפי תשובות המשתמשים לשאלון, ו-user journey לאורך שלבי הפאנל השיווקי.
- Tab 4 - Campaign Insights: משמש להפקת תובנות והמלצות רוחביות על כלל הקמפיינים, באמצעות חישוב מדדים כמותיים המעריכים את היעילות התקציבית של כל קמפיין בנפרד וכן של כלל הקמפיינים כמכלול. במסגרת ניתוח זה נבחנו מדדים מרכזיים,

בהם החזר על ההשקעה (ROI), עלות רכישת לקוח (CAC), ערך חיי לקוח (LTV), ויחס LTV/CAC - המשמש כאומדן מרכזי לרווחיות הקמפיין.

Tab 1 - Home: מדדי ליבה ותצוגת קמפיינים

בראש הטאב מופיעים ארבעה מדדי ליבה מרכזיים (KPIs), אשר מספקים תמונת מצב מרוכזת על ביצועי הקמפיינים בזמן אמת: מספר הקמפיינים הפעילים כרגע, סך המבקרים שנחשפו לקמפיינים, שיעור ההמרה הכללי ומספר הרכישות שבוצעו בפועל. מתחת למדדים מופיעה טבלה המרכזת מידע בסיסי על כלל הקמפיינים: שם הקמפיין, מספר מזהה, כמות משתמשים, כמות רכישות ושיעור ההמרה. טבלה זו אינה רק מקור מידע אלא גם משמשת כמפת ניווט, שממנה ממשיך המשתמש לניתוחים מעמיקים בטאבים הבאים.

CAMPAIGNS	USERS	CONVERSION RATE	PURCHASES
12	109,998	2.74%	3,019
Active Campaigns	Unique Visitors	Overall Performance	Total Completed

Active Campaigns Summary

campaign	Campaign number	Users	Purchases	Conversion Rate
DWM-LP-Quiz-purchase-US-Broad-Web-Nov24-phone-auto-IDP	1	8779	233	2.65%
DWM-LP-Quiz-purchase-US-Broad-Web-Sep24-phone	2	3128	47	1.5%
DWM-LP-Quiz-trial-US-Broad-Web-Aug24-Check	3	3961	40	1.01%
DWM-LP-Quiz-trial-US-Broad-Web-Nov24-phone-auto-IDP-holidays	4	2776	85	3.06%
DWM-LP-Quiz-trial-US-Broad-Web-Nov24-phone-auto-IDP-ot	5	3953	77	1.95%
DWM-LP-Quiz-trial-US-Broad-Web-Oct24-phone-auto-ot	6	6929	159	2.29%
DWM-LP-Quiz-trial-US-Broad-Web-Oct24-phone-price-auto	7	5153	153	2.97%
DWM-LP-Quiz-trial-US-Broad-Web-Sep24-Phone-B	8	656	8	1.22%
DWM-LP-Quiz-trial-US-Broad-Web-Sep24-phone-auto	9	28346	386	1.36%
DWM-Quiz-purchase-Broad-Nov24-auto-IDP	10	1598	53	3.32%

איור 14- תצוגת TAB1- Home

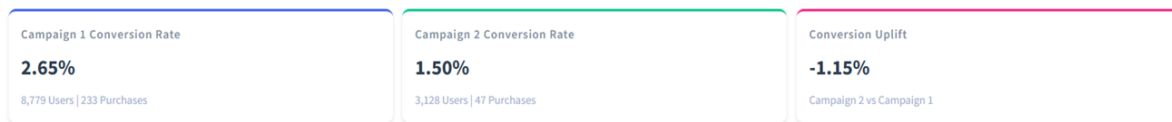
Tab 2 - Compare Campaigns: ניתוח השוואתי בין קמפיינים

בטאב Compare Campaigns מתבצע תהליך השוואה שיטתי ומעמיק בין שני קמפיינים שיווקיים, במטרה להמליץ על הקמפיין המוצלח יותר מבחינת המרות, תוך שימוש בגישה הסתברותית (Bayesian Inference). הנה הסבר של כל שלב בתהליך:

שלב 1: בחירת קמפיינים להשוואה
המשתמש בוחר שני קמפיינים מתוך רשימת הקמפיינים שמופיעה בטאב הראשי (Home). הבחירה מתבצעת לפי מזהה קמפיין (ID) שמופיע בטבלת הניווט.

שלב 2: הצגת נתוני השוואה בסיסיים

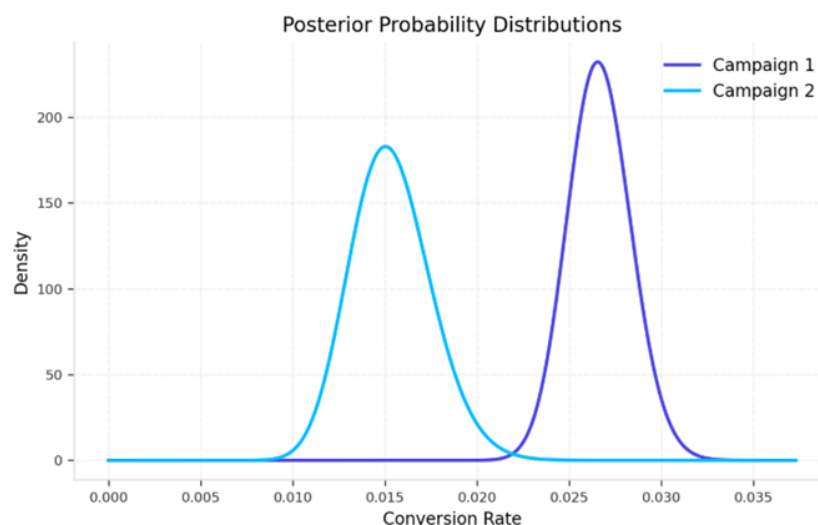
בשלב זה מוצגת השוואה נקודתית בין שני הקמפיינים שנבחרו, הכוללת את שיעור ההמרה (Conversion Rate) של כל אחד מהם, את מספר המשתמשים שנחשפו לקמפיין, את כמות הרכישות שבוצעו בפועל, וכן את מדד ה-*uplift* שמבטא את ההפרש בין שיעורי ההמרה. מטרת הצגה זו היא לספק למשתמש אינדיקציה ראשונית וברורה לגבי הביצועים של כל אחד מהקמפיינים טרם ניתוח מעמיק יותר.



איור 15- מדדי שיעור ההמרה של הקמפיינים הנבחרים

שלב 3: ניתוח הסתברותי עם התפלגות בטא

בשלב השלישי מוצג גרף הסתברותי מסוג Posterior Distributions, אשר מתאר את התפלגויות ההסתברות של שיעורי ההמרה עבור כל אחד מהקמפיינים שנבחרו. גרף זה מבוסס על חישוב בייסיאני תוך שימוש בהתפלגות בטא, והוא נועד לשקף את אי-הוודאות הקיימת בנתונים. ציר X מציג את שיעור ההמרה האפשרי (Conversion Rate), בעוד שציר Y מייצג את צפיפות ההסתברות - כלומר, עד כמה סביר שלקמפיין יש את שיעור ההמרה הנתון. הערכים בציר Y מחושבים על פי נוסחת התפלגות בטא, בהתבסס על נתוני המשתמשים והרכישות. הצגה זו מאפשרת ניתוח מעמיק יותר של ביצועי הקמפיינים, לא רק ברמה הממוצעת אלא גם מבחינת פיזור הסתברותי, ומספקת בסיס יציב להסקת מסקנות הסתברותיות.



איור 16- גרף התפלגות בייסיאנית

שלב 4: מסקנה הסתברותית

בשלב הרביעי מוצגת מסקנה הסתברותית המבוססת על ניתוח גרף ההתפלגויות: המערכת מפיקה המלצה לגבי הקמפיין בעל ההסתברות הגבוהה ביותר להוביל לשיעור המרה טוב יותר, בהתאם לחישוב הסתברותי שנעשה על סמך הנתונים שהוזנו. ככל שההתפלגויות של שני הקמפיינים נפרדות זו מזו, כך גובר הביטחון בהמלצה. לעומת זאת, כאשר קיימת חפיפה בין ההתפלגויות, המודל מצביע על רמת אי-ודאות גבוהה יותר בפערים ביניהם.

Campaign Performance Recommendation

Campaign 4

Campaign 4 has a 87.94% probability of achieving better conversion results.

איור 17-דוגמה למסקנה הסתברותית

Tab 3 - Analyze Single Campaign: ניתוח קמפיין בודד

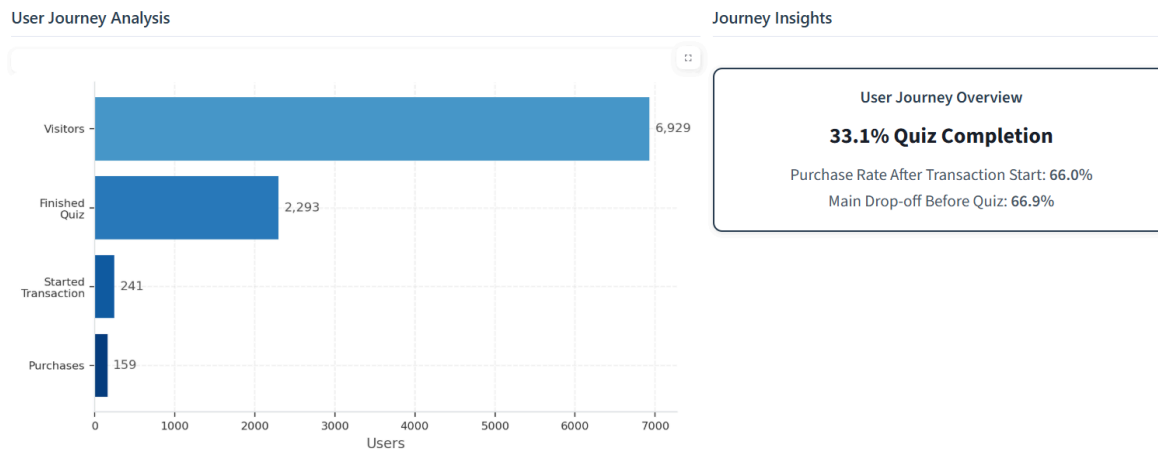
המשתמש בוחר קמפיין מתוך הטבלה המרכזית בטאב Home, ולאחר הבחירה מוצגים מדדים שמאפשרים בחינה של כל שלב בתהליך: מספר המשתמשים שהשלימו את השאלון הראשוני, כמות שהחלו תהליך רכישה, כמות הרכישות שבוצעו בפועל ושיעור ההמרה הכולל.

Finished Quiz 2,194 Completed Safety Quiz	Started Transaction 283 Checkout Started	Purchases 233 Successful Purchases	Conversion Rate 2.65% Visitor-to-Buyer Rate
--	---	---	--

איור 18- מדדי קמפיין בודד

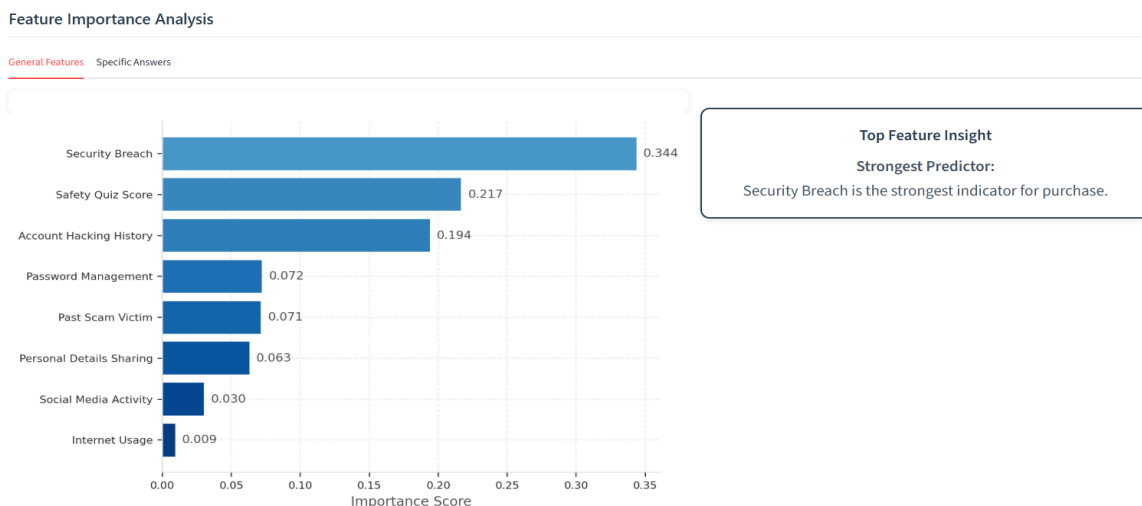
בהמשך מוצגים ארבעה ניתוחים מרכזיים:

1. User Journey Overview - תצוגה גרפית של משפך המשתמשים: כניסות, השלמת השאלון, התחלת רכישות, ורכישות בפועל. מוצג גם אחוז הנשירה המרכזי ואחוז ההשלמה.



איור 19- תצוגה גרפית של משפך המשתמשים

2. General Features - ניתוח חשיבות של תכונות כלליות באמצעות מודל Random Forest, לזיהוי משתנים המבאים רכישה. מטרתו היא להעריך את רמת המעורבות בשלב זה, ולבחון את השפעתו על שיעור ההמרה הכללי.

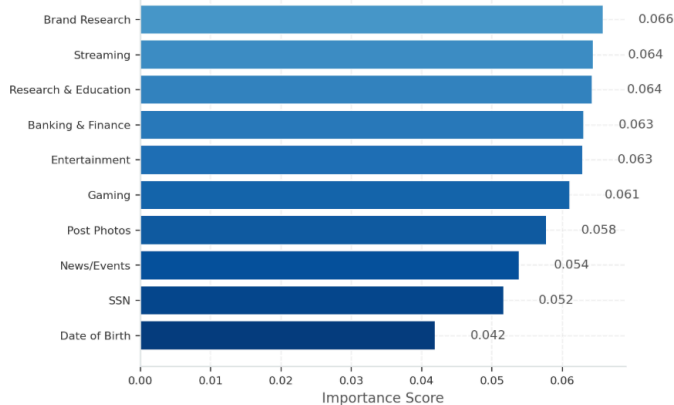


איור 20- ניתוח חשיבות משתנים

3. Specific Answers - ניתוח השפעה של תשובות ספציפיות לשאלות שנשאלו, כולל זיהוי התשובה בעלת ההשפעה הגדולה ביותר על ההמרה. התובנות המתקבלות מאפשרות שיפור משמעותי של הקריאייטיב, התאמת המסר השיווקי לפרופיל המשתמש ודיוק בטרגוט קהלים לפי תחומי עניין בעלי השפעה גבוהה.

Feature Importance Analysis

General Features Specific Answers



Top Answer Insight

Most Influential Answer:
Brand Research

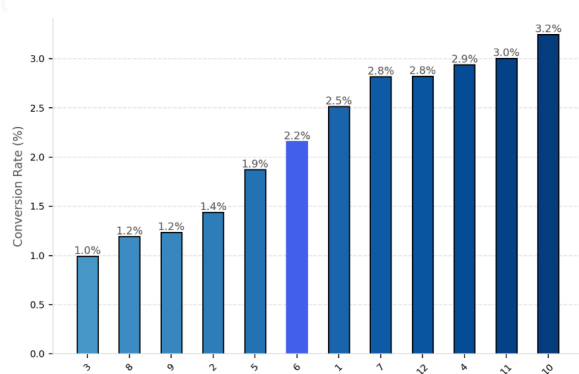
Related Question:

"What do you do on social media?"

איור 21- ניתוח חשיבות תשובות המשתמשים

4. Conversion Rate by Campaign - השוואת ביצועי המרה בין כלל הקמפיינים, כולל מיקום יחסי של הקמפיין הנבחר מבחינת אחוזי ההמרה. הקמפיין שנבחר מודגש באופן ויזואלי בגרף (הבר שלו בצבע שונה), כדי לאפשר זיהוי מיידי של ביצועיו ביחס לשאר הקמפיינים.

Conversion Rate by Campaign



Conversion Rate Insight

2.2%

Campaign 6 ranks 6 out of 12 campaigns in conversion performance.

איור 22- גרף שיעור המרה

Tab 4 - Campaign Insights: תובנות כלכליות שיווקיות

שלב 1: הזנת תקציב לכל קמפיין

בשלב זה נדרש המשתמש להזין את התקציב הכולל שהוקצה לכל קמפיין. התקציב משקף את ההשקעה הכוללת שנעשתה בקמפיין ומהווה בסיס לחישוב מדדים פיננסיים כגון עלות רכישת לקוח (CAC), החזר השקעה (ROI) וערך חיי לקוח (LTV). ההזנה מתבצעת בממשק אינטראקטיבי המאפשר הקצאה נפרדת לכל קמפיין ברשימה. עם סיום ההזנה, יש ללחוץ על כפתור "Calculate Metrics", אשר מפעיל את מנגנון החישוב ומציג את תוצאות הניתוח.

Portfolio Budget Analysis

Enter campaign budgets to compute CAC, ROI & a portfolio summary.

Budgets by Campaign

Campaign 1	0.00	-	+	Campaign 2	0.00	-	+	Campaign 3	0.00	-	+
Campaign 4	0.00	-	+	Campaign 5	0.00	-	+	Campaign 6	0.00	-	+
Campaign 7	0.00	-	+	Campaign 8	0.00	-	+	Campaign 9	0.00	-	+
Campaign 10	0.00	-	+								

Calculate Metrics

איור 23- ממשק הזנת תקציבי הקמפיינים

שלב 2: חישוב מדדים פיננסיים, קביעת סטטוס והפקת המלצות

בשלב זה מבוצע חישוב של מדדים פיננסיים מרכזיים עבור כל קמפיין, אשר משמשים כבסיס לקביעת סטטוס הביצועים והפקת המלצות להמשך הפעילות. הטבלה הבאה מציגה את הסטטוסים האפשריים ואת משמעותם:

סטטוס	קריטריון	משמעות	המלצה
No Payers	0 משלמים	לא התקבלו תשלומים כלל, אין בסיס לתובנות פיננסיות	בחינה מחדש של הכדאיות
LTV unavailable	לא ניתן לחשב יחס LTV/CAC (הכנסות = 0 או CAC לא מוגדר)	חסר מידע פיננסי מהותי	נדרש איסוף להגדרת נתונים נוספים

שקול להרחיב את הקמפיין	רווחיות גבוהה מאוד, יעילות שיווקית מרשימה	$LTV/CAC > 3.0$	Excellent
אופטימיזציה ושיפור נקודתי	הקמפיין רווחי, אך עם מקום לשיפור	$LTV/CAC \leq 3.0 \geq 1.5$	Good
בחינה אם ניתן לשפר או להפסיק	איזון תקציבי- הכנסות מכסות את ההוצאות אך ללא רווח ממשי	$LTV/CAC < 1.5 \geq 1.0$	Break-even
שקול שינוי אסטרטגי או עצירת הקמפיין	עלות גיוס לקוח גבוהה מההכנסה שהוא מניב	$LTV/CAC < 1.0$	Losing Money

Calculated Metrics

Campaign	Budget	Users	Payers	Revenue	LTV	CAC	LTV/CAC	ROI	Status	Recommendation
Campaign 1	\$8,000	8,779	192	\$15,309	\$80	\$42	1.9	91.4%	Good	Good profitability - consider increasing budget.
Campaign 2	\$1,500	3,128	43	\$1,873	\$44	\$35	1.2	24.9%	Break-even	At break-even - optimize for better ROI.
Campaign 3	\$3,000	3,961	40	\$2,786	\$70	\$75	0.9	-7.1%	Losing Money	Consider reducing budget or changing strategy.
Campaign 4	\$2,000	2,776	73	\$7,703	\$106	\$27	3.9	285.1%	Excellent	Excellent efficiency - consider scaling this campaign.
Campaign 5	\$3,000	3,953	71	\$5,046	\$71	\$42	1.7	68.2%	Good	Good profitability - consider increasing budget.

איור 24- הצגת מדדים פיננסיים של הקמפינים

שלב 3: ניתוח כולל של הקמפינים

מטרתו לספק תמונת מצב רוחבית על ביצועי המערכת השיווקית כולה, ולא רק של קמפיין בודד. המדדים המרכזיים המחושבים בשלב זה הם:

- Total Budget - סך התקציבים הכולל של כלל הקמפינים.
- Blended CAC - עלות רכישת לקוח ממוצעת, מחושבת לפי סך התקציבים חלקי סך המשלמים.
- Blended LTV/CAC - יחס הרווחיות הכולל, המשקף את האפקטיביות הפיננסית הכוללת של ההשקעה.
- Overall ROI - שיעור התשואה הכולל על ההשקעה, המבוסס על סך ההכנסות ביחס לסך התקציב.
- Profitable Campaigns - מספר הקמפינים הרווחיים (כאשר $LTV \geq CAC$) מתוך כלל הקמפינים.

שלב זה מאפשר להבין את המגמות הכלליות, לזהות הצלחות חריגות או נקודות כשל חוזרות, ולגבש החלטות אסטרטגיות כמו שימור קמפיינים אפקטיביים, הפסקת קמפיינים לא משתלמים או שינוי מדיניות התקצוב.



איור 25- הצגת מדדים משוקללים של כלל הקמפיינים

תהליך העלאת הדשבורד ל- Streamlit Cloud

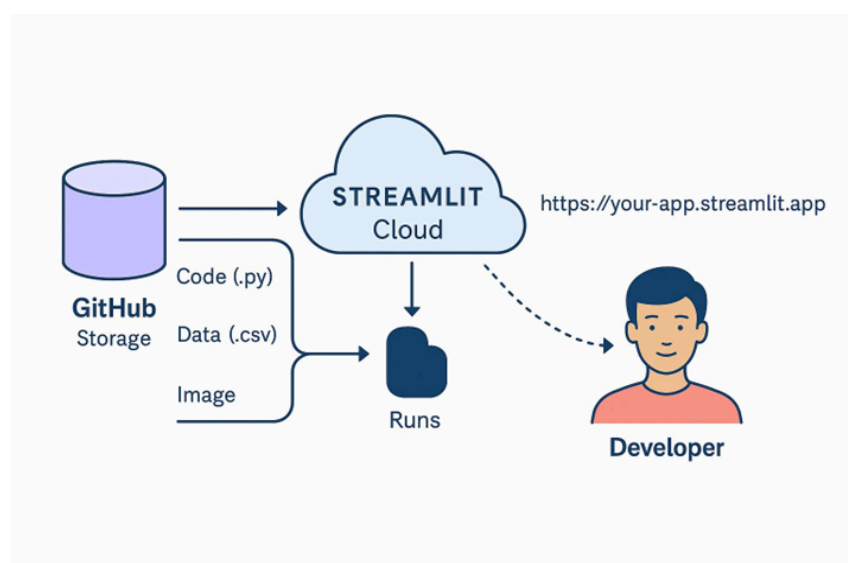
פרסום הדשבורד ב- Streamlit Cloud מבוצע באמצעות חיבור למאגר GitHub שבו שמורים קבצי הפרויקט: קוד Python, קבצי הנתונים ותכני העזר. Streamlit Cloud מושך את הקבצים מהמאגרים, מריץ את האפליקציה בסביבת ענן ומייצר כתובת URL ייחודית שדרכה ניתן לגשת לדשבורד מכל מכשיר עם חיבור לאינטרנט. השימוש בענן מאפשר זמינות מתמשכת של האפליקציה, גישה פשוטה ללא צורך בהתקנות, והפצה נוחה של הדשבורד למשתמשים, גם כאלה שאינם בעלי רקע טכנולוגי. בכך נוצר ערוץ אפקטיבי להצגת תובנות וניתוחים בזמן אמת.

קישור לדשבורד:

שם משתמש להתחברות: reasonlabs

סיסמה להתחברות: 1234

[/https://marketing-dashboard-reasonlabs.streamlit.app](https://marketing-dashboard-reasonlabs.streamlit.app)



איור 26- תהליך העלאת הדשבורד

11. הערכת הפתרון:

הערכת הפתרון בוצעה באמצעות שילוב בין ניתוח כמותי של ביצועי הקמפיינים לבין משוב איכותני מצוות השיווק בארגון, במטרה לבחון את תרומתו של הדשבורד החדש לשיפור תהליכי קבלת ההחלטות השיווקיות. הערכה זו אפשרה בחינה מקיפה של אפקטיביות הפתרון הן ברמת התוצאה העסקית והן ברמת חוויית המשתמש.

בשילובי ההטמעה הראשונים של הדשבורד, דיווח צוות השיווק על שיפור ניכר ביכולת לגבש תובנות אנליטיות על בסיס משתנים שבעבר לא נלקחו בחשבון בתהליך קבלת ההחלטות. שילובו של מודל בייסיאני לחישוב הסתברות הרכישה בכל קמפיין, לצד מודל Random Forest לזיהוי הפיצ'רים המשפיעים ביותר, יצר כלי עוצמתי שמאפשר לבחון לעומק את הגורמים המשפיעים על הצלחת כל קמפיין שיווקי.

הדשבורד שפיתחנו מציג עבור כל קמפיין את הערכת ההסתברות לרכישה, את הפיצ'רים המשמעותיים ביותר המשפיעים על ההסתברות הזו, וכן מדדים פיננסיים קריטיים כמו CAC (עלות רכישת לקוח), LTV (שווי לקוח עתידי) ו-ROI (החזר השקעה) בהתאם לתקציב שהוזן. כל אלה מוצגים באופן גרפי ברור ונוח לעיון, עם הסברים טקסטואליים המלווים את הגרפים ומסייעים בהבנתם.

מדד	לפני השימוש בדשבורד	לאחר השימוש בדשבורד	שיפור
שיעור המרה ממוצע	2.2%	4%-5%	פי 2.3
עלות רכישת לקוח (CAC) ממוצעת	500\$	370\$	ירידה של 26%

ממצאים אלו מצביעים על השפעה ממשית של הפתרון - הן בהיבט התקציבי והן ברמת האפקטיביות של הקמפיינים עצמם. צוות השיווק ציין שהמעבר לעבודה עם הדשבורד הביא לשיפור באיכות ההחלטות, ליכולת תעדוף ברורה יותר של הקצאת תקציבים, ולחסכון ניכר במשאבים.

מעבר ליעילות האנליטית, נבחנה גם חוויית השימוש הכוללת של הדשבורד בעיני המשתמשים. על בסיס המשוב מהארגון, עלה כי הכלי נתפס כנגיש, אינטואיטיבי וקל להתמצאות, גם עבור צוותים שאינם בעלי רקע טכני. המבנה הברור, השפה האחידה והאינטראקטיביות הגבוהה סייעו בהטמעה מהירה ובשימוש שוטף, והובילו להגברת הביטחון של צוות השיווק בקבלת החלטות מבוססות דאטה.

כחלק מתהליך השיפור, הושם דגש מיוחד על פרשנות ונגישות הנתונים למשתמשי הקצה. בהתאם למשוב מהארגון, שולבו בדשבורד הסברים טקסטואליים בסמוך לכל גרף ומדד מרכזי, במטרה להבהיר את המשמעות האנליטית והתפעולית של הממצאים. מהלך זה נועד לצמצם את הפער בין צוותי האנליזה לבין מקבלי ההחלטות שאינם מגיעים מרקע טכני, ומאפשר גם למשתמשים שאינם אנליסטים להבין את הדאטה, לגבש תובנות ולהגיב בזמן אמת. בנוסף, הותאם מבנה הדשבורד לזרימת העבודה הנהוגה בארגון, תוך יצירת הבחנה ברורה בין ניתוחים ברמת המאקרו עבור כל הקמפיינים לבין ניתוחים פרטניים לכל קמפיין. חלוקה זו אפשרה למשתמשים לעבור בקלות בין תצוגות רחבות לתצוגות ממוקדות, לזהות דפוסים חוזרים ולהגיב להזדמנויות בצורה יעילה. המשתמשים דיווחו כי מבנה זה שיפר באופן ניכר את בהירות הניתוחים ואת היכולת להסיק מסקנות אופרטיביות במהירות ובדיוק רב יותר.

12. דיון ומסקנות

12.1 דיון

הממצאים שהתקבלו לאורך הפרויקט מצביעים על שינוי מהותי באיכות קבלת ההחלטות השיווקיות בארגון, שינוי שנבע מהטמעה משולבת של מודלים הסתברותיים, אלגוריתמי למידת מכונה ודשבורד אינטראקטיבי תומך-החלטה. מהלך זה סימן מעבר מקבלת החלטות המבוססות על נתונים יבשים, למערכת קבלת החלטות מבוססת-דאטה, המאפשרת ניתוח בזמן אמת, זיהוי מגמות מוקדם והפחתת הסיכון בבחירה באסטרטגיות שיווקיות לא אפקטיביות. בכך, הוטמע בארגון כלי שמספק יתרון אסטרטגי, הן ברמת הדיוק האנליטי והן ביכולת להגיב במהירות לתנאי אי-ודאות בשוק.

יתרון מרכזי שבלט במהלך הפרויקט היה היכולת להתמודד עם מצבים של דלילות מידע, המאפיינים במיוחד את השלבים הראשוניים של קמפיינים דיגיטליים. שילוב המודלים ההסתברותיים יחד עם תחזיות מבוססות למידת מכונה אפשר לארגון לזהות בשלב מוקדם קמפיינים בעלי סיכויי הצלחה נמוכים, ובמקביל לחזק את ההשקעה בקמפיינים מבטיחים. בכך הושג חיסכון תקציבי ונוצר מיקוד מדויק יותר של המאמצים השיווקיים.

יישום כלי הניתוח המתקדמים תרם להבנה רחבה ועמוקה יותר של מכלול הגורמים המשפיעים על הצלחת הקמפיינים, מעבר למדד ההמרה בלבד.

באמצעות הדשבורד פותח ניתוח רב-שכבתי של הפאנל השיווקי, שאפשר לבחון את התקדמות המשתמשים לאורך שלבי הפאנל, משלב החשיפה הראשוני, דרך שאלות הסקר והאינטראקציות הביניים, ועד להחלטת הרכישה בפועל. ניתוח זה כלל גם בחינת השפעתן של תשובות ספציפיות לשאלות בפאנל, לצד הערכת חשיבות משתנים באמצעות מודל Random Forest. בכך הושגה תמונה אינטגרטיבית של הקמפיין: לא כפעולה מבודדת אלא כחלק ממערכת שיווקית מורכבת. יתרה מזאת, היכולת להציג את הנתונים בזמן אמת בדשבורד אפשרה תגובה מהירה, תיקון מהלכים תוך כדי תנועה ושדרוג ניכר במיומנויות התפעול של צוות השיווק.

עם זאת, נמצאו גם מגבלות בשימוש במודלים שנבנו. בראשן - תלות בהנחות מוקדמות ובאיכות הנתונים ההיסטוריים. כאשר הנחות אלה אינן מעודכנות או כאשר הנתונים כוללים הטיות, עשויה להיפגע מהימנות התחזיות. בנוסף, השימוש במודלים דורש השקעה תשתיתית, משאבי חישוב וליווי מקצועי, לצד הכשרה של משתמשי הקצה להבנה ולשימוש נכון בתוצרים. המעבר לתפיסת עבודה הסתברותית אינו טריוויאלי ודורש שינוי תרבותי בארגון.

מכלול הממצאים מלמד שהשילוב בין ניתוח הסתברותי, למידת מכונה ודשבורד תומך החלטה תרם לא רק לשיפור נקודתי של קמפיינים, אלא גם לחיזוק היכולת של הארגון לפעול בצורה מדויקת, מדידה ומבוססת-נתונים. הכלים שהוטמעו בפרויקט יצרו תהליך קבלת החלטות שקוף, מהיר ואפקטיבי יותר, וצפויים להוות בסיס לשיפור מתמשך גם בקמפיינים עתידיים. בכך, הפרויקט לא רק נתן מענה לצורך קיים אלא גם הציב תשתית לחשיבה שיווקית מתקדמת בארגון.

12.2 מסקנות

הפרויקט חשף את הפער שבין תובנות אינטואיטיביות לבין החלטות מבוססות-נתונים, והדגים כיצד כלים אנליטיים מתקדמים מסוגלים לגשר על פער זה באופן יישומי. עולה כי חיזוק ההסתמכות על מודלים חיזויים ומערכות תומכות-החלטה תורם לדיוק, לבקרה ולחיסכון תקציבי בעבודת השיווק.

היישום הדגיש את הצורך בגמישות תפעולית וביכולת להגיב לשינויים בזמן אמת, במיוחד בסביבה שבה קיים חוסר ודאות וחסרים נתונים מלאים. כמו כן, עלתה חשיבות ההתאמה של כלים אנליטיים למשתמש הקצה - הן מבחינת פשטות השימוש והן מבחינת פרשנות התוצרים.

לבסוף, הממצאים מעידים כי הצלחת פרויקטים מסוג זה דורשת לא רק ידע טכנולוגי, אלא גם מוכנות ארגונית לתהליך שינוי: פתיחות לחשיבה הסתברותית, הטמעת שגרות עבודה חדשות, וראייה מערכתית על כלל תהליכי השיווק.

12.3 המלצות

בהמשך לממצאי הפרויקט ולמסקנות שהוצגו, מוצעות להלן המלצות יישומיות שנועדו להבטיח מימוש מיטבי של הפתרון ותרומה ארוכת טווח לארגון. ההמלצות מתמקדות בהטמעת הדשבורד, שדרוג תהליכי השיווק, וחיזוק תהליכי קבלת ההחלטות ב-ReasonLabs.

- שילוב טכניקות מתקדמות- מומלץ להרחיב את יכולות המערכת באמצעות שילוב אלגוריתמים של למידה בלתי מונחית (Clustering), אשר יאפשרו לזהות דפוסים חבויים והתנהגויות משתמש שונים ניכרים בניתוחים מבוססי השערות בלבד. יוכל לחשוף תתי-קהלים חדשים ולסייע בגיבוש אסטרטגיות טרגוט מותאמות אישית.

גישה זו תאפשר לארגון לזהות פלחים בעלי ערך גבוה עוד בטרם ביצעו רכישה, לייעל את חלוקת התקציב בין קמפיינים שונים, ולהגביר את ההחזר על ההשקעה (ROI).

- דינמיות ועדכון בזמן אמת- כיום תהליך העבודה מחייב העלאת קובץ אקסל ידני לצורך עדכון הנתונים בדשבורד. מומלץ לפתח יכולת חיבור ישיר למקורות המידע הארגוניים (כגון CRM ERP ופלטפורמות פרסום), כך שהמערכת תוכל לשאוב נתונים באופן רציף ואוטומטי. יישום מהלך זה יאפשר עדכון מיידי של המדדים בהתאם לשינויים בשוק ובקהל היעד.
- ידע והדרכות לצוות השיווק - מומלץ לקיים סדנאות ממוקדות לצוות השיווק, שיתמקדו בהבנת עקרונות בסיסיים בעולם הבייסיאני ובלמידת מכונה, ובהסבר על אופן קריאת ההמלצות שמוצגות לצד הגרפים בדשבורד. מטרה זו היא להבטיח שהצוות לא רק ישתמש בממשק, אלא גם יוכל לפרש נכון את המשמעות של ההסתברויות, טווחי האמון וחשיבות המשתנים. בכך יוגבר הביטחון של המשתמשים בהמלצות, יפחת החשש ממונחים טכניים, והאפקטיביות של הכלי תגדל.

12.4 תרומת חברי הצוות

במהלך הפרויקט כל אחד מחברי הצוות לקח אחריות על חלקים שונים, כשכולם עבדו בשיתוף פעולה ובחלוקת משימות ברורה.

גל הייתה אחראית על כתיבת הדוח ועריכתו. בנוסף, הובילה את פיתוח הדשבורד מבחינת עיצוב ואינטגרציה של התוצרים, כדי לוודא שהמערכת תהיה נוחה ואינטואיטיבית לשימוש. איתי היה אחראי על פיתוח המודלים בפרויקט. הוא בנה את המודלים הבייסיאניים לחיזוי הסתברויות בתנאי נתונים דלים ואת מודלי הלמידה המונחית לזיהוי משתנים משפיעים. בנוסף, הוא בחן את תוצאות המודלים והסיק מהן תובנות והמלצות שהוטמעו בדשבורד. רותם דאג לקבל ולעדכן נתונים בזמן אמת, טיפל בניקוי ועיבוד ראשוני של הדאטה, והבטיח שהמודלים והדשבורד יתבססו על נתונים מלאים ואמינים. למרות החלוקה, כל אחד מאיתנו לקח חלק בכל שלבי הפרויקט - מהגדרת הבעיה, דרך פיתוח המודלים והכלים האנליטיים, ועד לניתוח הממצאים והסקת המסקנות. שיתוף הפעולה בינינו אפשר יצירת פתרון מקיף, אשר תרם למחלקת השיווק בקבלת החלטות מבוססות-נתונים ובניהול מושכל ויעיל יותר של התקציב.

13. ביבליוגרפיה:

- [1] N. Larsen, J. Stallrich, S. Sengupta, A. Deng, R. Kohavi, and N. T. Stevens. (2024). "Statistical challenges in online controlled experiments: A review of A/B testing methodology," *The American Statistician*, vol. 78, no. 2, pp. 135-149.
- [2] A. Sondhi, B. Segal, J. Snider, O. Humblet, and M. McCusker, (2021). "Bayesian additional evidence for decision making under small sample uncertainty," *BMC Medical Research Methodology*, vol. 21, art. no. 221.
- [3] R. Zimmermann and A. Auinger,(2023). "Developing a conversion rate optimization framework for digital retailers—case study," *Journal of Marketing Analytics*, vol. 11, pp. 233-243.
- [4] M. Sharma,(2023). "*Identifying factors contributing to lead conversion using machine learning to gain business insights*" ,M.Sc. thesis, School of Computing, National College of Ireland, Dublin, Ireland.
- [5] K. Pauwels and D. J. Reibstein,(2023). "The modern marketing dashboard: back to the future," *NIM Marketing Intelligence Review*, vol. 15, no. 1, pp. 10-13.

14. נספחים:נספח א: טעינת נתונים ועיבוד ראשוני

```

file_path = r"C:\\Users\\hroy2\\OneDrive\\Desktop\\my_data.xlsx"
df = pd.read_excel(file_path)

df["purchas_ind"].fillna(0, inplace=True)
df['breach_found'] = df['breach_found'].fillna(False).astype(int)

columns_to_fill = [
    "use_the_internet_for", "do_on_social_media", "enter_personal_details_online",
    "keep_your_passwords", "victim_of_online_scam", "nline_accounts_hacked",
    "safety_level_quiz_score", "breach_found", "transaction_start",
    "trial_ind", "purchas_ind", "revenue", "plan"
]

```

נספח ב: פיצול משתנים מרובי ערכים (Multi-Select)

```

columns_to_split = {
    "use_the_internet_for": range(1, 7),
    "do_on_social_media": range(1, 5),
    "enter_personal_details_online": range(1, 7)
}

for column, values in columns_to_split.items():
    df[column] = df[column].astype(str)
    for i in values:
        df[f"{column}_{i}"] = df[column].apply(lambda x: 1 if str(i) in x.split(",") else 0)

```

נספח ג: בניית משתני רוחב

```

column_groups = {
    "use_the_internet_for_answered": [f"use_the_internet_for_{i}" for i in range(1, 7)],
    "do_on_social_media_answered": [f"do_on_social_media_{i}" for i in range(1, 5)],
    # וכן הלאה...
}

for new_col, cols in column_groups.items():
    df[new_col] = (df[cols].gt(0)).any(axis=1).astype(int)

```

נספח ד: ניתוח שיעור ההמרה לפי קמפיין

```
def plot_campaign_conversion_rate(df):
    exposure = df['campaign'].value_counts().reset_index()
    purchases = df.groupby('campaign')['purchases_ind'].sum().reset_index()
    summary = exposure.merge(purchases, on='campaign')
    summary['conversion_rate'] = (summary['purchases_ind'] / summary['campaign']) * 100

    # גרף אופקי
    plt.barh(summary['campaign'], summary['conversion_rate'])
    plt.xlabel("Conversion Rate (%)")
    plt.title("Campaign Conversion Rate")
    plt.show()
```

נספח ה: קשר בין בטיחות, פרצות ורכישה

```
df['safety_level_category'] = pd.cut(df['safety_level_quiz_score'], bins=[0,20,40,60,80,100])
sns.barplot(...) # safety עבור
sns.heatmap(...) # חיבור בין שניהם
```

נספח ו: חשיבות משתנים (Random Forest)

1. ניתוח עומק - Feature Importance לפי תשובות ספציפיות

```
option_cols = [col for col in df.columns if "_" in col and not col.endswith("_answered")]
X = df[option_cols]
y = df["purchases_ind"]

model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X, y)

feat_df = pd.DataFrame({"Feature": X.columns, "Importance": model.feature_importances_})
feat_df.sort_values(by="Importance", ascending=False).head(7).plot.barh()
```

2. ניתוח רוחב- Feature Importance לפי שאלות

```

general_cols = [
    "use_the_internet_for_answered", "do_on_social_media_answered",
    "safety_level_quiz_score", "breach_found"
]
X = df[general_cols]
y = df["purchases_ind"]

model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X, y)

feat_df = pd.DataFrame({"Feature": X.columns, "Importance": model.feature_importances_})
feat_df.sort_values(by="Importance", ascending=False).head(6).plot.barh()

```

נספח ז: מודל בייסיאני להערכת יחס המרה (Beta Distribution)

```

# לכל קמפיין Beta חישוב פרמטרי
df["alpha"] = df["total_purchases"] + 1
df["beta"] = df["total_users"] - df["total_purchases"] + 1

# דגימת הסתברויות מתוך ההתפלגות הפוסטריורית
samples = np.random.beta(df.loc[0, "alpha"], df.loc[0, "beta"], size=10000)

# חישוב תחזית המרכז ותחום אמינות
mean = np.mean(samples) # תחזית הסתברותית להמרה
cred_interval = np.percentile(samples, [2.5, 97.5]) # תחום אמינות של 95%

```

נספח ח: כיוול התפלגות בייסיאנית

```

# דגימת הסתברויות מהמודל הקיים
samples = np.random.beta(alpha, beta, size=10000)

# בדיקת הכיוול: אחוז הדגימות שמתחת לשיעור ההמרה בפועל
calibration_error = np.mean(samples < observed_conversion)

# היסטורי prior אם סטיית הכיוול גבוהה מ-20%, מבצעים תיקון באמצעות
if abs(calibration_error - 0.5) > 0.2:
    alpha_prior = historical_conversion * historical_users
    beta_prior = (1 - historical_conversion) * historical_users
    alpha += alpha_prior
    beta += beta_prior

```

נספח ט: מימוש מודלים בלמידת מכונה

```
# ===== עיך X הגדרת =====
y = df1["purchas_ind"]
leakage_cols = ["revenue", "plan", "trial_ind", "transaction_start"]
X = df1.drop(columns=["purchas_ind"] + leakage_cols)

# הסרת עמודות תאריך
date_cols = X.select_dtypes(include=["datetime64[ns]", "datetime64[ns, UTC]").columns
X = X.drop(columns=date_cols, errors="ignore")

# הסרת מזהים
id_like_cols = [col for col in X.columns if "id" in col.lower()]
X = X.drop(columns=id_like_cols, errors="ignore")

# קידוד One-hot
X = pd.get_dummies(X, drop_first=True)

# ===== חלוקה Train/Test =====
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)

def evaluate_model(model, model_name):
    model.fit(X_train_res, y_train_res) # אימון על נתונים מאוזנים
    y_pred = model.predict(X_test)
    y_proba = model.predict_proba(X_test)[: , 1] if hasattr(model, "predict_proba") else None

    results = {
        "Model": model_name,
        "Accuracy": round(accuracy_score(y_test, y_pred), 3),
        "Precision": round(precision_score(y_test, y_pred), 3),
        "Recall": round(recall_score(y_test, y_pred), 3),
        "F1": round(f1_score(y_test, y_pred), 3)
    }
    if y_proba is not None:
        results["ROC-AUC"] = round(roc_auc_score(y_test, y_proba), 3)
    return results
```


הגדרת מודלים:

```
models = {
    "Logistic Regression": LogisticRegression(
        max_iter=500,
        solver="liblinear",
        class_weight="balanced"
    ),
    "Random Forest": RandomForestClassifier(
        n_estimators=200,
        max_depth=10,
        random_state=42,
        class_weight="balanced"
    ),
    "Gradient Boosting": GradientBoostingClassifier(
        n_estimators=200,
        max_depth=3,
        random_state=42
    )
}

# ===== הרצות המודלים =====
results = [
    evaluate_model(m, name)
    for name, m in models.items()
]
results_df = pd.DataFrame(results).set_index("Model")

print(results_df)
```

15. מאגר GitHub, קובץ דאטה ודשבורד אינטראקטיבי

במסגרת פרויקט זה נבנה מאגר קבצים ב-GitHub, אליו ניתן לגשת בקישור:

<https://github.com/itayhabsha/marketing-dashboard.git>

במאגר זה:

- קובץ Data.xlsx - יש להוריד למחשב המקומי על מנת להעלות אותו לתיבת ההעלאה שבדשבורד.
- קובץ Campaign_dashboard.py - מכיל את הקוד המלא של הדשבורד, הכולל את כלל הלוגיקות האנליטיות לצד עיצוב הממשק.

להפעלת הדשבורד עצמו ניתן להיכנס ישירות באמצעות הקישור:

<https://marketing-dashboard-reasonlabs.streamlit.app>

פרטי ההתחברות לדשבורד:

- שם משתמש: reasonlabs
- סיסמה: 1234