

פרויקט קורס בינה עסקית

חלק ב' – יישום המערכת

בחלק השני של הפרויקט, יישמנו את הנתונים שלנו ב-Power BI. בשלב הראשון, הקמנו את מסד הנתונים שלנו ב-SQL server בשרת, שמדמה למעשה את ה-DW שלנו. כיוון שהנתונים שלנו נלקחו מקובץ Excel, הם היו "נקיים" יחסית ולכן לא נדרש מאיתנו תהליך ETL מורכב. מחסן הנתונים ב-SQL server אפשר לנו להריץ שאילתות על הנתונים ולהבין לעומק את ההתנהגות שלהם. לאחר שעשינו ניקוי של נתונים מיותרים שהגיעו מהאקסל, הגדרנו את טבלת Fact-Event ואת כל ה-Dim, ואז יכלנו למעשה להעביר את הנתונים ל-Power BI ולקבל את הסכמה שתכננו.

בשלב השני, הגדרנו את השאלות העסקיות שנרצה לענות עליהם באמצעות הממדים והוויזואליזציות, ובחרנו כלי מתאים לכל ניתוח עבור כל שאלה. בחרנו לחלק את הניתוח לשלושה חלקים –

☒ **לשונית Revenue**

שמתמקדת בהכנסות של החברה. נעזרנו במדדים כמותיים וחישוביים ושמנו דגש על מגמתיות לאורך זמן.

☒ **לשונית User Activity**

שלמעשה מציגה את הפעילות וההתנהגות של המשתמשים כאשר מתבצעת כניסה של משתמש לאתר מסוים. הלשונית מתייחסת לאחוזי ההשמה של פרסומות ולהתרשמות של המשתמשים מהצפיות ממגוון זוויות.

☒ **לשונית ניתוח תקלות**

מתייחסת לשני סוגי תקלות שהגדרנו: "תקלות התאמת תוכן" - שלמעשה מייצג לנו Empty WF ו"תקלות דיווח מערכת" – שבעצם מסננות וממפות את סוגי התקלות שיש בדיווח של הנתונים שנרשמים במערכת. כמו לדוגמה דיווח של צפייה מעל 2 שניות (Impression) כאשר אין הזדמנות לפרסום (Inventory), ולכן אין בכלל פרסומת שניתנת לצפייה.

לשונית Revenue

בניתוח ההכנסות בחרנו לשים דגש גם על מגמתיות של הכנסה לאורך זמן, וגם על התפלגות ההכנסות לפי Publishers.

השאלות העסקיות

1. כמה הכנסה בדולרים הצלחנו להפיק?

לטובת החישוב הזה סכמנו את ההכנסה בדולרים של כל עמודת Revenuen מה-Fact-event

2. כמה כסף הכנסנו מתחילת החודש האחרון? וכמה כסף הכנסנו מתחילת השנה?

יצאנו מנקודת הנחה שנקודת הזמן הנוכחית היא התאריך האחרון במסד הנתונים שלנו – מרץ 2020. יצרנו מדד שבדוק על ידי משתנה מהו החודש האחרון בכל נקודת זמן (תוך התחשבות בעובדה שמסד הנתונים אמור להמשיך לעבוד ולהתעדכן ו"החודש האחרון" מתעדכן מחודש לחודש), וסכמנו את ההכנסות רק מחודש זה. דבר דומה עשינו עם ההכנסות מהשנה האחרונה.

3. מהי התפלגות ההכנסות לפי שנים?
לצורך תכנון עסקי ותקציבי, נרצה לדעת מהי התפלגות ההכנסות לפי שנים וכך גם לזהות טרנדים ודפוסים בהתנהגות הלקוחות והשוק. לצורך כך, הצגנו גרף עמודות.
4. מהם אחוזי הגידול והקיטון לפי חודשים?
הידע על אחוזי הגידול והקיטון מאפשר לנו לזהות טרנדים חודשיים, ואפילו עונתיות, בהתנהגות הכנסות העסק. זיהוי זה, שגורם לנו להבין אילו חודשים בשנה מובילים לצמיחה ואילו לירידה בהכנסות, מסייע להתאים אסטרטגיות שיווק במידת הצורך ולנהל חודשי צמיחה בצורה יעילה.
לטובת החישוב הזה, יצרנו מדד שכולל שני משתנים - ההכנסות מהחודש הזה וההכנסות מהחודש הקודם. המדד מחסר את סך הכנסות מהחודש הקודם מסך ההכנסות החודש הנוכחי ומחלק בסך ההכנסות מהחודש הקודם וכך למעשה מציג את אחוזי הגידול בין שני החודשים.
בחרנו להציג את המדד על ידי גרף קווי שמייצג באופן חזותי את הגידול והקיטון ברצף חודשי השנה.
5. מהו ממוצע ההכנסות למשתמש?
עניין אותנו לראות כמה בממוצע כל משתמש בודד מכניס לחברה. דבר זה לא רלוונטי לנו מבחינת שימור לקוחות, כי המשתמשים הם לא הלקוחות שלנו אלא Publishers, אבל במונחים של פעילות של משתמשים נרצה לראות כמה כל משתמש "תורם" בממוצע להכנסה שלנו בסופו של דבר, וכמה הצלחנו ל"אפטם" את הפרסום והתוכן עבור המשתמש.
6. מהו ממוצע ההכנסה לכל Event של כניסת משתמש לאתר?
גידול בממוצע כזה מצביע באופן ישיר על הצלחה במטרה המרכזית של החברה - למקסם את הפרסום בכל כניסה של משתמש לאתר.
7. מיהם עשרת המפרסמים עם נתח ההכנסה המשמעותי ביותר? ומיהם המפרסם המוביל נכון לעכשיו?
נרצה לראות מיהם המפרסמים עם נתח ההכנסה המשמעותי ביותר לחברה, וכמה כל אחד מהם מכניס לנו - דבר זה יסייע לנו להבין איפה כדאי להשקיע את עיקר המשאבים ואיזה לקוחות חשוב לשמר.
8. כמה כל מפרסם משלם לנו עבור 1000 פרסומות?
זהו למעשה מדד CPM והוא מציג עבור כל מפרסם מהו המחיר שהוא משלם לנו עבור 1000 פרסומות.

לשונית User Activity

הבנת התנהגות המשתמשים מהותית לנו על מנת למקד היכן ומהי אופטימיזצית הפרסום שאנו שואפים לעשות. כאשר נדע מהם השעות האטרקטיביות שמהוות את עיקר הגלישה באתרים, נדע באיזה שעות יש יותר כדאיות פרסום. כאשר נדע איזה קבוצות גיל גולשות וצופות יותר בפרסומות ומתי, נדע למקד את התוכן הרלוונטי עבורם. באופן יותר כללי נרצה לדעת איפה אנחנו עומדים מבחינת "הצלחות הפרסום" שלנו, מבחינת אחוזי ההשמה של פרסומות ומבחינת מימוש הזדמנויות הפרסום.

השאלות העסקיות

1. כמה משתמשים פעילים יש במערכת וכמה מתוכם גברים/נשים?
2. מהי כמות משתמשים לפי קבוצות גיל?
התפלגות קבוצות הגיל מוצגת בגרף גלגל ומציגה עבור כל קבוצת גיל משהי ממוצע הגילאים שלה. דבר זה מסייע לנו לבחור תוכן רלוונטי של פרסומות.
3. מהי התפלגות הצפיית לפי גיל ומגדר? ומהי שכבת הגיל שהכי מובילה בצפייות?

4. מהי התפלגות ההזדמנויות לפרסום לפי שעה במהלך היום?
גם מדד זה מצביע על פעילות וגלישה של משתמשים באתרים ומצביע על מגמות וטרנדים בעונתיות סביב שעות היממה.
בחרנו להציג את הנתון על ידי גרף משולה, כאשר העמודות מציגות את הImpression וקו אופקי מציג את Inventory. דבר זה מדגיש את פער או ההתאמה בין ההזדמנויות שהיו לנו לפרסם לבין ההתרשמות של המשתמשים (תצורה חזותית של מדד Fill rate).
5. האם באמת קפצה פרסומת כאשר הייתה לנו הזדמנות לפרסם? (האם נמצא תוכן מתאים למשתמש) – המדד הזה הוא המדד המשלים של Empty WF. הוא מחסיר מכל הזדמנויות הפרסום (Inventory) את הפעמים שלא הצלחנו למצוא תוכן מתאים (Empty WF) ומחלק בInventory.
6. מהם אחוזי ההתרשמות בפועל של המשתמשים מהתוכן המפורסם מתוך הזדמנויות הפרסום?
מבוטא על ידי אחוזי ההשמה – Fill Rate מחושבים על ידי סכום הImpression, (איפה ש- $Inventory > 1$) חלקי סכום הInventory.
7. מיהם publishers שיש להם הכי הרבה צפיות?
מסייע לנו להבין אצל איזה מהמפרסמים קיים התוכן הרלוונטי והמעניין יותר. בדקנו עבור ה Publisher המובילים מהם אחוזי ההשמה שלהם.
8. מיהם Publishers שאחוזי הFR שלהם גבוהים מ100%?
אחוזים גבוהים מ100% של אחוזי השמה בעצם אומרים שיש יותר מהתרשמות אחת בפרסומת אחת – מה שמצביע על צפיות חוזרות בפרסומת בודדת.
גם מדד זה מסייע בהבנה מיהם המפרסמים בעלי התוכן המעניין יותר.
9. האם הלקוח סיים לצפות באותה הפרסומת?
מדד זה מוצג כאחוזי הפרסומות שמשתמשים סיימו לצפות בהן מתוך הפרסומות שמשתמשים התרשמו מהם. (סך הcompleten איפה ש- $Impression > 1$, חלקי סך הImpression).

לשונית ניתוח תקלות

כפי שצינו, ניתוח הקלות מתחלק לשני סוגי תקלות – תקלות התאמת תוכן ותקלות דיווח במערכת.
בתקלות התאמת התוכן היה לנו חשבו להתמקד בניתוח של הגורמים הפוטנציאליים שיכולים להוביל לתקלות – אזור גיאוגרפי שעלול לעורר תקלות בעקבות תשתיות אזוריות או מכשירי הפעלה – שגם כן עשויים ליצור עיקוב בהצגת הפרסומות, בטעינה או בתמיכה.
בנוסף, מתוך הבנה שיש שעות עומס על האינטרנט או בגלישה באתרים מסוימים, ייתכן שהתקלות מתרכזות בשעות מסוימות ושיש עונתיות שחוזרת על עצמה בשעות קבועות במהלך היום – בחרנו לנתח גם את ממד הזמן בתחום זה על מנת לייעל את הטיפול ולצמצם את היקף התקלות.

השאלות העסקיות עבור תקלות התאמת תוכן

1. מהו אחוז הEWF הכולל?
למדד המחשב את סך EWF קראנו FilterEWF כיוון שהוא למעשה סוכם את EWF רק איפה ש $Empty Wf \leq Inventory$.
כיוון שהנתונים שלנו לא כולם "נקיים" ואנחנו מקבלים דיווחים שגויים, יש לנו רשומות לא הגיוניות שיש בהם אי מציאת פרסומת מתאימה כאשר אין בכלל הזדמנות פרסום – מה שלא הגיוני. ולכן הוספנו את התנאי הזה, כדי לסכום רק את הרשומות הרלוונטיות.

2. באילו אזורים גיאוגרפיים יש הכי הרבה תקלות? ומהם אחוזי התקלות בכל איזור?
את התפלגות התקלות בכל מדינה בחרנו להציג במפה, שסוכמת עבור כל מדינה את ה FilterEWF ומציגה בצורה וויזואלית היכן בועלם אנחנו צריכים לשים דגש לטיפול בתקלות אלו ואיפה יש הצטברות חריגה של תקלות.
3. באילו מדינות הנתח הגדול ביותר של התקלות?
בחרנו להציג את 15 המדינות המובילות עם אחוזי התקלות הגבוהים ביותר בטבלה אשר מכילה עיצוב מותנה שמדגיש את "רמת החומרה" שאותה מדינה נמצאת בה מבינת התקלות.
4. אילו מכשירי הפעלה מהווים את אחוז התקלות הגבוהה ביותר?
נעזרנו בגרף עוגה, המציג את אחוז התקלות עבור כל מכשיר הפעלה.
5. באילו שעות ביום יש הכי הרבה תקלות?
על ידי גרף קווי, סכמנו עבור כל שעה את סך FilterEWF. השימוש בגרף זה מדגיש את השעות שבהם יש "פיקים" של תקלות ושכדאי להתמקד בהם.
6. מהו אחוז התקלות במגמה חודשית ומהי החריגה מיעד התקלות שהוגדר?
על מנת לצמצם את אחוז התקלות באופן כללי, הגדרנו יעד שבו אנחנו מעוניינים להוריד מחודש לחודש את סך התקלות ב-10%. ביעד שהגדרנו בעצם הכפלנו את FilterEWF ב-0.9 ובעזרת מדד KPI, הצגנו את (במגמה חודשית) את העמידה או החריגה מהיעד.
7. מהו אחוז התקלות החודשי ביחס לשנה שעברה?
עבור כל חודש בשנה מסוימת, הצגנו את אחוז התקלות (FilterEWF מתוך Inventory) של שנה קודמת. (אותו נתון בדיוק שנה אחת אחורה). דבר זה מאפשר לנו להבין את מגמתיות השיפור \ הדרדור בכמות התקלות. הדגשנו את המקומות שבהם שנה שעברה היה אחוז קיצוני של תקלות.

השאלות העסקיות עבור תקלות התאמת תוכן

- איזה תקלות דיווח קיימות במערכת? ומהם אחוזי התקלות?
- ✓ Faults Complete without Impression – רשומות שבהם דווח שיש סיום של צפיה מבלי שבכלל התחילה (צפייה של שתי שניות).
 - ✓ Faults EWF without Inventory – רשומות שבהם דווח על תקלת חוסר התאמת תוכן כאשר לא הייתה לנו הזדמנות לפרסם תוכן.

סיכום

הנתונים המוצגים בדשבורד מספקים מגוון רחב של מידע לניתוח ותכנון עסקי, ניתן לראות כי החברה מצליחה לשמור על יציבות בהכנסות. הסקירה של הממוצעים השונים בגיליון Revenue כמו ממוצע הכנסה למשתמש, מספקים מידע חשוב על יכולת החברה להפיק תמורה מהשקעות פרסום ושמירה על רווח מצטבר. בנוסף, הזיהוי של עשרת המפרסמים עם נתח ההכנסה המשמעותי ביותר והמפרסם המוביל יכול לכוון את החברה לנקודות חזקות בשוק ולאפשר שיפורים עתידיים בתכנון השיווקי.

מדד ה CPM שהצגנו לכל מפרסם מספק גם כן מידע חשוב לגבי ערך הפרסום ביחס למחיר, ויכול לשמש ככלי לקבלת החלטות על השקעות ופרסום. גיליון ה Revenue מספק תמונה מקיפה של ביצועי החברה ויכול לשמש בסיס לקבלת החלטות עסקיות מושכלות.

בגיליון UserActivity רצינו להדגיש את התנהגות המשתמשים באתרים והשתמשנו בנתונים כדי למקד ולאפטם את הפרסום. בנוסף, הצגנו את הקשרים בין נתוני הפרופיל של המשתמשים (כמו גיל, מגדר וכו..) לבין ההתנהגות שלהם באתרים.

הבנה של התנהגות המשתמשים באתרים תאפשר לחברה להתאים ולשפר את הפרסום והתוכן בצורה המיטבית לפי הצורך ובכך להגדיל את ההכנסות.

בגיליון Faults רצינו להדגיש את חשיבות ניתוח התקלות במערכת כחלק מתהליך שיפור ואופטימיזציה של פעולות הפרסום והתוכן. זיהוי תקלות לפי אזור גיאוגרפי יכול להביא למיקוד והגברת נוכחות של צוות באזורים אלו לצורך פתרון מהיר של התקלות ושיפור השירות ללקוחות. הצגה של זמני העומס תעזור לייעל את הטיפול בתקלות. זיהוי ואיתור תקלות בצורה טובה יכול להבטיח חווית משתמש טובה יותר ולהגביר את יעילות הפרסומים.

לסיכום, מהנתונים שבחרנו להציג בדשבורד ה BI-ניתן לספק מידע ותובנות על פעילות המשתמשים בפלטפורמה, בעיקר בתחומי הפרסום וניתוח התקלות. באמצעות הדשבורד החברה תוכל לנתח תהליכים, לזהות מגמות ולקבל החלטות אסטרטגיות בזמן אמת.