

חלק א' - הקדמה

Key Experience Indicator - KEI

- משהו שאפשר למדוד
- יש עליו הסכמה מכל השותפים
- תמיד יש מה לעשות איתו (לשפר ולתקן)

בגוגל יש KEI 120 שונים לשימוש בגוגל חיפוש. שניים מתוכם הם:

1. Long clicks - משתמש פותח לינק ונעלם או חוזר אחרי הרבה זמן
2. Short clicks - משתמש פותח לינק אבל חוזר מהר לתוצאות החיפוש

ההנחה תמיד היתה ש-Long clicks טובים יותר מ-Short clicks אבל ראיונות עם משתמשים גילו משהו אחר. הסתבר שמשותמשים שהולכים ל-Long clicks הולכים לאיבוד באתרים (כי לא כל האתרים טובים ובנויים כהלכה) לפיכך, שני הפרמטרים האלו רעים באותה מידה. התשובה של גוגל הייתה להוציא את המידע החשוב כבר לעמוד התוצאות וזה בדיוק מה שקורה כשאנחנו מחפשים פרטים על מסעדה או סרטים - שעות הפתיחה ונתוני הוויקיפדיה מוצגים בעמוד התוצאות.

ב-WeWork המוצר שונה, מדובר בחלל פיזי שמקיים בתוכו מוצרים דיגיטליים, והמטרה שלו היא לייצר אינטראקציה בין אנשים. לכן אחד מה-KEI הוא שביעות רצון לגבי דברים מאוד מסויימים (כמו עמדות עבודה) והתנהגות בתוך החלל עצמו (כמו שעות נוכחות).

סוגי מדדים

יש הרבה סוגי מדדים (פירוט במצגת), אבל הכי חשוב זה להבין את את ההבדל בין מדדי התנהגות לבין מדדי עמדות. התנהגות הרבה יותר חשוב למדוד מאשר עמדות. למה? כי בניגוד לעמדות, התנהגות לא משתנה ולא ניתן להטות אותה.

8 תרופות דמה כדי לקיים עסק ללא לב:

1. לשכור/לייצר קבוצת חדשנות (IDEO)
2. להתמקד שמה שיודעים לעשות הכי טוב - רדוקציה פוגעת באינטרס למציאת ערך חדש
3. ללמוד ולשבור מהר - האמירה היא אם שברנו, לפחות למדנו, אבל זה פשוט תהליך פוסט-רציונאלי שנועד להצדיק את התהליך השגוי. במקום זה אפשר פשוט למדוד ולהבין לאן צריך לכוון.
4. להביא תותח ממקום אחר
5. מטריקות יהירות - מסתכלים על מטריקות שלא מספרות את הסיפור האמיתי
6. אגו - היינו מאוד טובים פה, בוא נעשה גם את זה.
7. תובנות גסות - שנגזרות מ-NPS, למשל (ו-NPS הוא מדד בעייתי שלא אומר מספיק)
8. הכנסה "מרדימה" - הכנסה יוצרת אשליה שהכל בסדר

כשמדברים על מדידות, הכי חשוב לבצע מדידות שיניעו לפעולה. לצורך כך יש להבין מה הולכים לעשות עם התוצאות. במידה וזה רק "בשביל העניין" זה לא שווה, אם אין כוונה לשנות משהו זה מיותר.

חלק ב' - HEART

בגוגל רוח מאוד מודל PULSE:

1. Page views
2. Uptime (כמה אחוזים מהזמן המוצר עובד)
3. Latency (כמה זמן לוקח מהרגע שעושים פעולה ועד לרגע שהיא באמת מתרחשת)
4. Seven days active users
5. Earnings

העניין הוא שהמדד הזה לא מודד חווית משתמש. לצורך כך "המציאו" את מודל Heart:

1. Happiness
2. Engagement - תדירות/עומק שימוש. בודק האם משתמשים במוצר כפי שהתכוונו.
3. Adoption
4. Retention - לא לכל המוצר אלא לפיצרים ספציפיים. האם הפיציר שומש עלי כמשתמש? האם הוא מגביר את השימוש במוצר ואולי הוא מהווה סיבה לעזיבה?
5. Task Success - ברגע שלכל פעולה יש נקודת התחלה וסיום (כמו הוספה למועדפים) אפשר למדוד דברים שונים כמו אחוזי ההצלחה, כשלון, כמה זמן לוקח להשלים את המשימה וכו' (זה כבר מתקשר לבדיקות שמישות).

אבל, איך מחליטים מה למדוד?

מה זה אומר שהמשתמש מצליח במוצר? מה זו הצלחה (וכישלון)?

איך מגדירים הצלחה? איך מגדירים הצלחה של משתמש?

ההגדרה הזו מתחילה את הדיון.

כל מדידה צריכה להיות:

1. בעלת משמעות - אנחנו בודקים דברים משמעותיים ולא איזוטריים.
2. בעלת תוצאות ישימות - התוצאה של המדידה צריכה להניע אנשים לפעולה. אם זה רק מעניין כנראה שזה לא כזה חשוב.
3. יציבה - אובייקט המדידה צריך להיות משהו יציב. הוא חייב להיות חלק מהמהות של המוצר.

כל מדידה מורכבת משלוש:

1. Goals - כשמדברים על KEI המטרות חייבות להיות המטרות של המשתמש, לא של העסק. למשל, אלו משימות המשתמשים צריכים לעשות כדי להשיגם? מה המטרה של ה-re-design? תשובה טובה יכולה לעזור בניסוח המטרות. הכי חשוב לזכור שהגדרת המטרות צריכה להיות בהסכמה של כל המעורבים (stake holders). כולם צריכים להבין מה אנחנו מודדים, מה זה אומר ומה אפשר לעשות עם זה. אם אין הבנה והסכמה משותפת הכל יפול מהר מאוד.
2. Signals - איך הצלחה/כישלון יבואו לידי ביטוי בהתנהגות או בעמדות של המשתמש? אלו רגשות המשתמשים מביעים הצלחה/כישלון בעת שימוש במוצר? מהגדרת הסימנים אפשר לגזור מטריקות.
3. Metrics - צריך לנרמל את המספרים, זאת אומרת, לדבר באחוזים ולא במספרים מוחלטים. Log analysis הוא כלי מעולה שאפשר להוציא ממנו המון דברים טובים ומועילים, אבל צריך לא מעט עבודה כדי לחלץ משם תובנות.

לדוגמה:

- **Goal**
Help users effectively decide which book to buy.
- **Signal**
Path to purchasing a book goes directly from book page to purchase or to other book pages and not to other sources
- **Metric**
Path directness

חלק ג' - סקרים

סקר הוא לא האפשרות הכי טובה כדי לקבל תשובות לשאלות, באופן עקרוני אין על תשובות שניתנות בראיונות, הן הרבה יותר מקיפות מכאלו שמתקבלות בסקרים. לכן כעורכים סקרים צריך לאפסם אותם כדי שיביאו את המקסימום.

מה זה סקר?

סקר - "תהליך של קבלת תשובות לשאלות ממדגם אנשים לצורך מטרת השגת מספרים שניתן על מנת לקבל החלטות"

שלבי הסקר

1. מטרות - שאלות שצריך לקבל עליהן את התשובה (הכוונה היא לשאלות מחקר, לא לשאלת הסקר)
2. מדגם - אנשים שיוזמנו לסקר
3. שאלות - שאלות שעליהם יענו
4. שאלון - תכנון השאלון שעומו המשתמשים יבואו באינטראקציה

5. עבודת שטח - הרצת הסקר מהזמנה ועד ה-follow up
6. תשובות - התסכלות על התשובות וניקוי הדאטא
7. תובנות - ניתוח התוצאות וקבלת החלטות

מטרות

מה אנחנו רוצים לדעת? למה אנחנו רוצים לדעת את זה? אם זה רק כי "זה מעניין", זה לא תשובה נכונה. אלו החלטות אנחנו הולכים לקבל על בסיס התוצאות? ישנה גישת האולד-סקול ששואלת הכל על כל דבר אפשרי, אבל הדבר הנכון הוא לשאול אחת בלבד - בקונטקסט הנכון ובזמן הנכון. לאחר מכן לשפר/לתקן את מושא הסקר ואז להעלות סקר עדכני (עם שאלה אחת בלבד).

גודל מדגם

יותר חשוב לשאול את האנשים הנכונים מאשר לשאול הרבה אנשים.

איך מוצאים אנשים שישתתפו בסקר:

1. לפי הקונטקסט הנכון ובזמן הנכון (למשל, לשאול על ארוחת הצהריים מיד אחריה ולא אחרי חצי שעה כשכולם מתפזרים).
2. לשלם לחברות שימצאו.
3. ויראלי - להריץ משהו בפייסבוק ולקוות שמהו יצא מזה.

נקודה למחשבה - הסקר נשלח ל-1000 אנשים אבל רק 100 עונים. יש כאן response rate של 10%. עד כמה ה-10% מייצגים את כולם?

איך אדם מחליט האם לענות על סקר?

ה-Response triangle מגדיר שלושה קריטריונים שלפיהם אדם מחליט האם לענות על סקר:

1. כמה מאמץ זה יצרוך ממני
 2. מה יצא לי מזה
 3. האם אני סומך על מי שיעשה את הסקר (שישמור על הפרטיות, שיתייחס לתשובות וכו')
- אלו שלושת הקריטריונים שאדם מחליט האם לענות על הסקר.

שאלות

לפני שמעלים את הסקר צריך לעשות ניסוי לשאלות, לבדוק על מספר מועט של משתמשים ולראות אם הן טובות/רעות. השלב הזה יאפשר את שיפור הסקר. חשוב מאוד כי קל מאוד לעשות סקר רע. שאלות טובות מייצרות תשובות מעולות אבל מאוד קשה לכתוב ולנסח אותן.

טעויות נפוצות בניסוח שאלות

1. שאלות על העתיד (שאלות 'האם') - הטעות הכי נפוצה. הכי טוב לראיין את המשתמשים, להשקיף עליהם ולהבין את הצרכים האמיתיים שלהם.
2. שאלות על העבר - כי אנשים פשוט לא זוכרים. אנשים לא בהכרח יודעים לשחזר את החוויה שלהם בצורה מהימנה.
3. מימצוע התהליך (תהליך העליה למטוס הוא משעמם - מסכים עד לא מסכים בכלל) - על מה אתה שואל אותי? על הפעם האחרונה? על כל הפעמים? אם שואלים את המשתמשים על משהו שקרה הרבה פעמים אין להם מושג על מה הם עונים. הדבר הנכון הוא לשאול על הפעם האחרונה (וגם זה, רק אם זה קרה בזמן האחרון, גג שבוע אחרון, אחרת אנחנו פשוט לא זוכרים).
4. שתי שאלות בשאלה אחת - לא טוב. כל שאלה צריכה לעמוד בפני עצמה.
5. שאלות מובילות (כמה נמוך היה נפוליאון?) - אני אומר את התשובה שאני מצפה לה בתוך השאלה עצמה. הדבר הנכון הוא לשאול "איך היית מתאר את הגובה של נפוליאון?"
6. שתי שאלות שליליות בשאלה (האם זה הגיוני או לא הגיוני שהמוות של אלביס מעולם לא קרה?) - השאלה היא לגבי הספק, האם יש ספק לגבי המוות שלו. לכן השאלה צריכה להיות 'האם אתה מטיל ספק במוותו של אלביס?'
7. שאלות מוקשים (איפה אתה אוהב לשתות בירה?) - צריך לשאול האם אתה שותה בירה? אם כן, להציג את השאלות הרלוונטיות שנובעות מכך.
8. שאלות מוחלטות (האם אתה תמיד אוכל ארוחת בוקר?) - לאיזו תקופה השאלה מתייחסת? כל החיים? שבוע אחרון? השאלה הנכונה היא: בשבוע האחרון, כמה פעמים אכלת ארוחת בוקר?
9. זירגון מיוחד לחברה - יש לשאול בשפה של המשתמשים ולא בזו של החברה.
10. שאלות מעליבות - בן כמה אותה? עד 19, 30-19, מבוגר יותר מ-30? אם כן, אז מה, כל מי שמעל 30 הוא זקן?

השאלון עצמו

התשובות בדרך כלל מנוסחות בסולם ליקרט, אבל כמה נקודות צריכות להיות על הסולם? ההמלצה היא ל-2 או 3. או שאתה מסכים או שאתה לא מסכים. עם 3 תשובות לא נוח להוציא ממוצע. נוח יותר לעשות את זה עם 5, שכל אחד יחליט מה טוב לו. 9 ומעלה יוצרים בעיה, מה כבר ההבדל בין נקודה לנקודה? אורך השאלון משפיע על אחוז הנטישה.

הזמנה לסקר

ההזמנה צריכה להיות אישית, תגידו מי אתם. אנשים עונים כשזה בא מבן אדם ולא מכונה. תגידו מה הסיבה לסקר, למה אתם שואלים - כי אנחנו רוצים לשפר את המוצר וכו'. להעניק פרס - לפעמים הצגת הסיבה היא פרס לשכעצמו. אם יש אפשרות להוסיף תמריץ, אז כדאי. להדגיש את הנושא של הסקר להגיד מתי הסקר יסתיים/יסגר לא להגיע כמה זמן זה יקח

סקר SUS - System Usability Scale

סקר 5-1 בסולם ליקהרט שמתחקר חווית משתמש (המערכת היתה מורכבת לי/חשבתי שהמערכת קלה לשימוש/מצאתי את המערכת מסורבלת לשימוש/מצאתי את הפיצ'רים השונים במערכת ברורים ומובנים).

סקר NPS

הטענה היא שהוא מייצג נאמנות של לקוחות (עד כמה סביר להניח שתמליץ על החברה ואז סולם 1-10). הבעיה היא שזו שאלה היפוטית לגבי העתיד. היא מתעלמת מתגובות נוספות בעלות משמעות. זו שאלת התנהגות, אבל התנהגות מודדים, לא שואלים. אם זו התנהגות, בוא נמדוד אותה. היא לא actionable, אין מה לעשות עם הציון הזו ברמה האופרטיבית. מנגד, היא כן טובה ברמה השוואתית כשרוצים לבחון גרסאות שונות של אותו הפיצ'ר.

חלק ד' - טכניקות

מובהקות סטטיסטית זו לא כמות משתמשים מינימאלית שיכולה להעיד על כל המשתמשים אלא הקפדה על כך שמרווח הביטחון בין שני המבחנים לא יהיה חופפים. ככל שמסתפקים באחוז הצלחה נכון יותר (נגיד, 80% הצלחה לעומת 95%) כך המרווח קטן יותר ואז הפערים ברורים וחד-משמעיים יותר.