

הממדים של איכות הנתונים

מחבר: איציק סיון

ניתן לבדוק איכות נתונים באמצעות 7 מדדים:

- **דיוק Accuracy**
 - מספר התווים שנשמרו לשם משפחה. אם לא נשמרו מספיק תווים ייתכן ויתבצע קיצוץ בשם.
 - איות נכון של שמות, לדוגמא שם משפחה עם ואתיות סופיות באמצע
 - מספר הספרות באחוז הריבית-בפעולות בין בנקאיות מדובר בסכומים מאד גבוהים. אם באחוז הריבית יש רק 2 ספרות אחרי הנקודה העשרונית במקום 3/4, ההבדל עלול להיות מאד משמעותי.
 - בדיווח על תאונה נשמר רק התאריך. במקרה כזה הנתון לא מדויק כיוון שגם השעה חיונית
- **שלמות Completeness**
 - אם בפרטי לקוח פוטנציאלי חסר הדואר האלקטרוני, המשמעות היא חוסר אפשרות לשיווק בערוץ זה
 - אם בפרטים האישיים לא נרשם המגדר, זה עלול לגרום לבזבז משאבים במסעי שיווק למוצרים לא רלוונטיים
 - הפתרון: הקפדה ברורה בין מאפייני חובה ומאפיינים שאינם חובה.
- **התאמה לתקנים Conformity**
 - יש תקן בינלאומי לקודים של מדינות ISO-3166-1 חוסר הקפדה על שימוש בתקן עלול לגרום לבעיות הסבה של מידע בין מערכות שונות
 - יש תקן בינלאומי לסוגי מטבעות ISO 4217. חוסר הקפדה על שימוש בתקן זה עלול לגרום לבעיות משמעותיות במערכות בהן יש העברת כספים
 - יש תקן בינלאומי לשפות ISO. 639-1 חוסר הקפדה על שימוש בתקן עלול לגרום לבעיות במערכות הפעלה של מוצרים בשימוש בינלאומי
 - יש תקן ישראלי של הלשכה המרכזית לסטטיסטיקה לענפים כלכליים במשק. חוסר תאימות לתקן זה בארגונים בהם נדרש מידע על פעילות בתחומים שונים עלול לגרום לבעיות.
 - לא ייתכן שבאותו ארגון יהיו שתי מערכות שונות עם קודים שונים לסניפים של הארגון
 - הפתרון: התאמה לתקנים בינלאומיים, ארציים וארגוניים
- **רמת עדכניות Currency**
 - האם לנתונים יש משמעות נכון להיום
 - אם בין המוצרים שהחברה משווקת יש מוצרים שלא משווקים יותר אבל שכחו לעדכן את הסטטוס שלהם
 - אם החברה מחזיקה פרטים על לקוחות שחלקם כבר הלכו לעולמם
- **עקביות Consistency**
 - עקביות בערכים למאפיינים זהים: בישות עובד הערכים האפשריים למגדר הם F/M בישות לקוח הערכים האפשריים הם ז/נ
 - עקביות בטרנזקציות, לדוגמא: מבצעים העברת כספים מחשבון לחשבון. החשבון שלך חוייב אבל הצד השני לא קיבל את הסכום שהועבר
 - עקביות בשמות למאפיינים זהים
 - עקביות בגיבויים
- **כפילות Duplication**
 - תוכן כפול בישות ספציפית - בדרך כלל עקב בחירת מפתח בעיתי ובדיקות סבירות לקויות.

• **קישוריות אמינה Integrity**

- בעיות בקשרים בין תת ישויות. לדוגמא: לעובד נרשמו מספר קורסים מה קורה אם הקורס בוטל פיסית מישות הקורסים?
- ללקוח יש מספר הזמנות והמערכת אפשרה לבטל את הלקוח?

- דרגה או תפקיד של עובד , ללא תאריך שינוי הדרגה/התפקיד
- עדכון כתובת ללא ציון תאריך עדכון אחרון
- עדכון מחיר של מוצר ללא ציון תאריך השינוי
- עדכון קבועי מערכת כגון מע"מ, או אחוז הנחה לאזרחים ותיקים ללא ציון תאריך העדכון
- הצגת פרטים על ספר ללא ציון תאריך פרסום הספר

בקיצור, אחת הבדיקות החשובות הן בתהליך הבניה והן בסקר מודל הנתונים צריכה להיות בדיקה מקיפה של הצורך בתאריכים/תקופות בכל תתי הישויות כולל טבלאות ה Lut