1**. היתרונות בשימוש במידע גולמי מהלקוח ללא עיבוד או ניקוי הם :**

מהירות וגישה מיידית המידע זמין ואין צורך להשקיע זמן בתהליכי עיבוד מה שמאפשר לקבל תובנות מהירות על המידע מה שחוסך את העלויות של העיבוד ואת המשאבים, ובנוסף שמירה על שלמות המידה כאשר משתמשים במידע גולמי לא מחסרים נתונים מה שנמסר מהלקוח נשמר ללא שינוי זה חשוב עם רוצים לתת מידע מדויק וזה מוזיל את העלויות

**החסרונות** בשימוש במידע גולמי מהלקוח ללא עיבוד או ניקוי הם **:**

**מידע גולמי עשוי לכלול טעויות, נתונים חסרים או שגיאות בהזנה, דבר שיכול להוביל לתוצאות לא נכונות או לא מהימנות.** בנוסף על כך **המידע הגולמי עלול לכלול הטיות לא מכוונות הזנה לא אחידה של תאריכים, שמות או מיקומים מה שיכול לפגוע בהבנת התמונה הכוללת. ללא עיבוד או ניקוי, הנתונים יכולים להיות לא אחידים ולא מובנים, מה שמקשה על ביצוע ניתוחים מתקדמים**. המידע **עלול לכלול נתונים רגישים או פרטיים שלא עברו עיבוד להבטחת פרטיות הלקוח.**

**2. סוגי טעויות עלולות להופיע במידע גולמי שמגיע מהלקוח , וכיצד ניתן לזהות ולתקן אותן?**

**א. שגיאות כתיב והקלדה: נתונים שיכולים להכיל טעויות כתיב כמו שמות, כתובות או ערכים אחרים.**

**כדי לזהותן ניתן להשתמש באלגוריתמים של תיקון אוטומטי או להשוות את המידע עם רשומות קיימות .**

**פתרון שימוש בתוכנה לניקוי טקסטים (למשל, fuzzy matching) או קלט אוטומטי שמתקן טעויות .נתונים חסרים: לדוגמא לקוח לא ממלא שדה חובה (כגון תאריך או מקום מגוריו).**

**ב. חוסר בנתונים: לקוח לא ממלא את כל השדות**

**זיהו ניתוח של רשומות חסרות ערכים או שימוש בבדיקות עקביות (למשל, לוודא שלא חסרים ערכים בתאריך או במספר הטלפון).**

**הפתרון ניתן להנחות את הלקוח למלא את הנתונים החסרים או להשלימם בצורה אוטומטית (למשל, על ידי חיפוש אחר ערכים דומים)**

**ג. שדות לא תקינים ואחידים : תאריכים פורמטים שונים וכדומה**

**זיהוי יצירת תבנית אחידה יצירת תבניות אחידות לבדוק שהנתונים נכנסים למבנה תקני (למשל, תאריך בפורמט YYYY-MM-DD).**

**פתרון: תהליך של המרה אוטומטית לפורמט אחיד בעזרת סקריפטים או כלים לניקוי נתונים.**

**ד. נתונים לא עקביים: לקוח כתב נתונים שונים במספר רשומות כגון כתובת לא נכונה וכדומה**

**זיהוי: השוואת ערכים (כגון עיר אזור גיאוגרפי)לאחידות.**

**פתרון: תהליך של נרמול (Normalization), בו ממירים את הערכים למבנה אחיד (למשל, שם עיר או מדינה) באמצעות רשימות מקובלות.**