בס"ד

**נתונים על הדאטהסט בעקבות EDA למציאת בקיעים וחסרים:**

ל-12% מהאנשים יש רקע רפואי.

ל-5% מהאנשים יש רקע פלילי.

ל-19% מהאנשים יש אזרחות זרה.

ל-97% מהאנשים יש ימים ללא פעילות.

לכל עובד נוצרו כל 180 הימים, כלומר אין עובד עם רשומה חסרה בתאריך מסוים.

לעובד זדוני הלייבל "זדוני" מופיעה 180 פעמים, כלומר בכל 180 הימים. אין עובד שאינו זדוני שאחד מימיו סומן כזדוני.

כמות הדפים שהודפסו במועד חריג אף פעם לא עולה על כמות הדפים הכוללת.

כמות הפקודות במועד חריג אף פעם לא עולה על כמות הפקודות הכוללת.

כמות הדפים שהודפסו בצבע אף פעם לא עולה על כמות הדפים הכוללת.

בכל הרשומות כמות ההדפסות בצבע + בשחור לבן שווה בדיוק למספר ההדפסות הכולל.

ב-99% מהרשומות הודפס יותר מדף אחד לכל פקודת הדפסה, עם ממוצע של 6.3 דפים לכל פקודת הדפסה.

ב-32% מהימים בוצעו הדפסות.

מספר הדפסות ממוצע לכל יום שהיו בו הדפסות – 31.

אחוז הימים שבהם התבצעה כניסה למשרד – 95%.

זמן שהייה ממוצע במשרד – 9.1 שעות ליום.

78% הם עובדי החברה, היתר קבלן.

35% מהעובדים בדרגת סיווג 1, 35% בדרגה 2, 13% בדרגה 3, 16% בדרגה 4.

אין שום רשומה שבה ממוצע הסיווג לבקשות צריבה היה גדול יותר מהסיווג המקסימלי.

**בעיות שהתגלו:**

97%1 מהימים של העובד הם ימי פעילות ורק 3% הם ימים ללא פעילות. כלומר מתוך 180 ימים לעובד יש רק 6 ימים ללא פעילות.

2. ב-100% מהרשומות כמות הכניסות שווה לכמות היציאות.

3. ב-0.02% מהרשומות כמות פקודות ההדפסה עולה על כמות הדפים שהודפסו

4. אחוז הרשומות שבהן כמות ההדפסות בצבע מתוך כמות ההדפסות הכוללת אינה שווה למספר שנרשם בתור אחוז ההדפסות בצבע –

Checking if color ratio matches actual color percentage...

71.47% of records have inconsistent color ratios

לצורך תיקון הבעיות נכתבו השורות הבאות בפרומפט שתפקידו לשנות את הדאטה:

1. בקוד הקודם שכתבת התברר ש97% מהימים של העובד הם ימי פעילות ורק 3% הם ימים ללא פעילות. כלומר מתוך 180 ימים לעובד יש רק 6 ימים ללא פעילות. דבר זה לא הגיוני, היות שכמעט כל העובדים אינם מועסקים יותר מ-5 ימים בשבוע, ואף אחד מהם אינו מועסק באופן רשמי ביום שישי לאחר 4 אחרי הצהריים או בשבת. בנוסף, יש גם ימי חג במהלך השנה שאינם נחשבים כימי עבודה. בנוסף כל עובד יכול לקחת מספר מסוים של ימי מחלה או ימי חופשה במהלך השנה.

2. בקוד הקודם ב-100% מהרשומות כמות הכניסות שווה לכמות היציאות. אמנם זו המציאות התקינה והרווחת, אולם שים לב שתיתכן מציאות מחשידה שבה העובד נכנס עם תג שאינו שלו ויוצא עם תג שלו וכך מספר היציאות גדול ממספר הכניסות, וכן יש מציאות מחשידה שהעובד נכנס עם תג שלו אך יוצא עם תג שאינו שלא, וכך מספר היציאות קטן ממספר הכניסות.

3. משום מה בקוד הקודם ב-0.02% מהרשומות כמות פקודות ההדפסה עולה על כמות הדפים שהודפסו. למעשה אי אפשר שפקודות ההדפסה יהיו מספר יותר גבוה מכמות הדפים שהודפסו בפועל – זה לא הגיוני! (היות שאנו סופרים רק פקודות הדפסה שלא בוטלו).

4. אחוז ההדפסות בצבע – אמור להיות מסוג int ואם זה אמור להיות שבר – תעגל. משום מה בקוד הקודם נוצר מצב שאחוז הרשומות שבהן כמות ההדפסות בצבע מתוך כמות ההדפסות הכוללת אינה שווה למספר שנרשם בתור אחוז ההדפסות בצבע, זה כן אמור להיות שווה ליחס ביניהם לאחר העיגול.