|  |
| --- |
| **חברי הצוות** |
| מאיה לביא 205731524, אור עזר 205781396 |

## מטלת ניתוח נתונים

1. ארגון הנתונים והעשרתם

בביצוע המטלה, ארגנו את הנתונים בדרכים שונות על מנת לחפש מגמות וגורמים המתאימים לשאלות עליהן רצינו לענות. השינויים שנעשו כללו הוספת עמודות אגרגאציה בטבלאות הגולמיות וכן איחוד טבלאות לייצור טבלאות חדשות ברמות גרעיניות שונות המתאימות לשאלותינו. מעבר לשינויים אלו, בעזרת VBA ו-Microsoft Query, יצרנו טבלאות עזר לחיפוש וייצוג התופעות. בחלק זה נציג את כלל הטבלאות עליהן מושתתות תשובותינו להמשך העבודה. נרצה לציין כי התשובות לשאלות בסעיף 2.ב לעיתים מושתתות על הטבלאות כפי שהן מופיעות בבסיס הנתונים אך לעיתים גם על מניפולציות של הנתונים מטבלאות אלו.

**שינויים על טבלאות גולמיות:**

בטבלת **defected parts**, ביצענו איחוד רשומות על פי מפתח ראשי עם 2 העמודות Box RFID ו-Part Type, כאשר הוספנו עמודה המציגה את מספר המופעים של כל מפתח ייחודי בטבלה המקורית. מספר זה מייצג את מספר החלקים הפגומים מסוג part type בתוך קופסא מסוימת. הוספנו גם עמודה שמחשבת את אחוז החלקים הפגומים מסוג part type שנמצאו בקופסא מסוימת.

**ייצור טבלאות חדשות:**

A screenshot of a social media post

Description automatically generatedלהלן תיאור של הטבלאות ששימשו אותנו לאורך הניתוח. השכבה העליונה מהווה הטבלאות המקוריות והשכבה השנייה טבלאות נוספות שיצרנו לצורך הניתוח.

* טבלה “**Production + Sales**” בנויה מאיחוד 2 טבלאות- production, sales. מטרת הטבלה היא לאחד את המידע על תהליך הייצור של כל מוצר שנמכר יחד עם התוצאה שלו בסוף (האם הוחזר) על מנת לאפשר ניתוחים של הבעיה של בעלי המפעל. רמת הגרעיניות גבוהה- הטבלה מיועדת לניתוח הנתונים.

Production

Product id

Sales

Product id

Production + Sales

Product id

* טבלה “**Employees-sales**” בנויה מאיחוד 4 טבלאות- employees, shifts, production, sales. מטרת הטבלה היא לנתח את ההשפעה של ותק העובד על אחוז ההחזרות של המוצרים שנוצרו במשמרותיו. לשם כך, ביצענו אגרגאציה על מנת לספור את מספר המוצרים שיצר העובד, מספר המוצרים מתוכם שהוחזרו, והיחס ביניהם (אחוז החזרות). רמת הגרעיניות נמוכה – הטבלה מיועדת להציג סיכום על מגמה בנתונים.

Production

Product id

shift

Sales

Product id

Returned

Production + Sales

Product id

Shift

Returned

Shifts

Shift

Employee ID

Production + Sales +Shifts

Product id

Shift

Returned

Employee ID

Employees + Sales

Product id

Shift

Returned

Employee ID

Employees

Employee ID

* טבלה “**Customer-returns**” בנויה מאיחוד 2 טבלאות- customers, sales. מטרת הטבלה היא לנתח את אחוז ההחזרות של הלקוחות. לשם כך, ביצענו אגרגאציות שספרו את מספר המוצרים שקנה הלקוח, מספר המוצרים שהחזיר, והיחס ביניהם (אחוז החזרות). לאחר ייצור טבלה זו, אגרגאציה נוספת שבוצעה בעזרתה היה השוואת שיעורי החזרה בסוגי לקוחות השונים. עבור הטבלה הכללית, רמת הגרעיניות גבוהה- הטבלה מיועדת לשימוש לניתוח הנתונים. עבור ההשוואה של סוגי לקוחות, הגרעיניות יותר נמוכה.

Customers

Customer id

Sales

Customer id

Returned

Customer Returns

Customer ID

Returned

* טבלה “**Failures + Shifts**” בנויה מאיחוד 3 טבלאות- failures, production, sales. מטרת הטבלה היא לאפשר הסתכלות על אירועי התקלות של משמרות ספציפיות ובכך לאפשר ניתוח של נסיבות והשפעות התקלות. רמת הגרעיניות גבוהה- הטבלה מיועדת לשימוש לניתוח הנתונים.

Failures

Shift

Production

Shift

Failures + Production

Product ID

Shift

Sales

Product ID

Returned

Failures + Shifts

Product ID

Shift

Returned

* טבלה “**Double Shifts**” אשר בנויה מאיחוד 3 טבלאות-Production, Sales, Shifts. מטרת הטבלה היא לבדוק האם קיים קשר בין משמרות כפולות של עובד יחיד לבין החזרות המוצרים מהמשרת השנייה מבניהם. רמת הגרעיניות נמוכה – הטבלה מיועדת להציג סיכום על מגמה בנתונים.

Shifts

Shift

Employee ID

Production

Shift

Shifts + Production

Product ID

Shift

Employee ID

Sales

Product ID

Returned

Failures + Shifts

Product ID

Shift

Employee ID

Returned

**אגרגאציות נוספות:**

בנוסף לטבלאות לעיל, יצרנו מספר טבלאות המיועדות להקל על הניתוח הסטטיסטי של מספר פרמטרים:

* טבלה ״**Time to sell: buckets״** (מתוך טבלה מקורית Sales) מכילה את זמני המדף (בסדר עולה) לפני מכירה של המוצרים ומחולק לפי דליים של עד 1000 רשומות. עבור כל דלי חישבנו את אחוזי ההחזרות מתוך המוצרים הנמצאים בו. רמת הגרעיניות נמוכה – הטבלה מיועדת להציג סיכום על מגמה בנתונים.
* טבלה ״**Energy Consumption: buckets״** (מתוך טבלה מקורית Production) מכילה את צריכת האנרגיה של מכונת ההרכבה (בסדר עולה) ומחולק לפי דליים של 977 רשומות. עבור כל דלי חישבנו את אחוזי ההחזרות מתוך המוצרים הנמצאים בו. רמת הגרעיניות נמוכה – הטבלה מיועדת להציג סיכום על מגמה בנתונים.

1. ניתוח הנתונים על פי הקווים המנחים הבאים

**קו מגמה לינארי עולה, בשילוב עם מחזוריות**

למציאת עמודת נתונים המקיימת את שני התנאים- מגמתיות לינארית יחד עם מחזוריות, חיפשנו אחר נתונים שעולים באופן רצוף אך גם כן מושפעים מתהליכים מחזוריים. Production time היא עמודה המתארת זמני סיום של ייצור המוצרים. כיוון שכל מוצר חייב להסתיים בזמן ייחודי, ברור כי הזמן יעלה לאורך הרשומות. בנוסף, זיהינו כי עמודות wait for glue ו-wait for packaging מתנהגות בצורה מחזורית. Production time כמובן צריך להיות תלוי בין היתר ב-2 עמודות אלו, ולכן ראינו לנכון לבדוק האם עמודה זו תקיים לינאריות וגם מחזוריות.

לאחר בדיקה, ראינו לפי הגרף כי אכן התופעה מתקיימת:

התאמנו קו מגמה לינארי אשר ערך מדד ה-R2 שלו הוא 99.5%, כך שהקו הלינארי מותאם במידה טובה מאוד. לאחר מכן, יצרנו את רצף הערכים שמתקיימים לפי הקו הלינארי ולבסוף, החסרנו בין הערך הרצוי (לפי הקו הלינארי) למצוי בנתונים בכדי לחשוף את המחזוריות עצמה:

אכן ניתן לראות מחזוריות מובהקת, כאשר יש מגמה יורדת אך מאוד שולית (R2=0.009).

**דעיכה או התכנסות אקספוננציאלית**

כאשר חיפשנו נתונים המתנהגים בצורה זו, התמקדנו בעמודות שערכיהן יכול להיות מאוד גדול ומאוד קטן כאשר הערך לא תלוי במדד אחר. זמן המכירה של מוצר אינו תלוי באופן ייצורו כלל ולכן ראינו לנכון לבדוק את נתונים אלו. עלולים להיות מקרים של מכירה מידית ועלולים להיות מוצרים שנמצאים במחסן לאורך זמן.

|  |  |
| --- | --- |
| *Time to Sell* | |
|  |  |
| Mean | 40106.6 |
| Standard Error | 368.3317 |
| Median | 25367.66 |
| Mode | #N/A |
| Standard Deviation | 43068.04 |
| Sample Variance | 1.85E+09 |
| Kurtosis | 3.174219 |
| Skewness | 1.701371 |
| Range | 279536.1 |
| Minimum | 0.06967 |
| Maximum | 279536.1 |
| Sum | 5.48E+08 |
| Count | 13672 |

לפי הסיכום הסטטיסטי, ראינו כי החציון נמוך מאוד ביחד לממוצע, וטווח הערכים מאוד גדול.

איחדנו את הנתונים לתוך דליים המכילים ערכים בטווחים של 1000 דקות. כלומר דלי 0 סופר את מספר המוצרים שנמכרו בין 0-1000 דקות. דלי 1 סופר את מספר המוצרים שנמכרו בין 1000-2000 דקות וכו׳.

ניתן לראות בצורה מובהקת כי הערכים מתאימים לדעיכה אקספוננציאלית.

במבחן KS לא דחינו את השערתנו שהנתונים מתפלגים אקספוננציאלית עם פרמטר ממוצע בר״מ 5%.

|  |  |  |  |
| --- | --- | --- | --- |
| KS | Dst | Dcr (alpha=5%) | תוצאה |
| H0: dist=exp; H1: else | -0.00077 | 0.004903 | אי- דחייה |

**הפרעות אקראיות**

|  |  |
| --- | --- |
| *Wait for Assembly* | |
|  |  |
| Mean | 9.611722 |
| Standard Error | 0.307361 |
| Median | 1.321185 |
| Mode | 0 |
| Standard Deviation | 35.93896 |
| Sample Variance | 1291.609 |
| Kurtosis | 30.45658 |
| Skewness | 5.289011 |
| Range | 371.1657 |
| Minimum | 0 |
| Maximum | 371.1657 |
| Sum | 131411.5 |
| Count | 13672 |

בכדי למצוא נתונים המתנהגים בצורה אקראית, חיפשנו עמודה שבה קיים הבדל משמעותי בין הממוצע לטווח הערכים.

עמודה זו קיימה את התנאי ולאחר הצגה על גרף ניתן לראות כי אכן יש תופעה של אקראיות.

**שינוי מגמה (Concept Drift) משמעותי**

לשם תופעה זו, חיפשנו היכן בנתונים קיים קפיצה משמעותית בערכים לאורך זמן של 10-100 רשומות. מצאנו כי דבר זה מתקיים בזמן תהליך ה-Packaging במשמרות של עובדת מספר 16. ראינו כי הממוצע מאוד נמוך אך עבור זמנים מסוימים המייצגים את משמשרת מספר 43 בהם היא עבדה, קיימת קפיצה משמעותית אשר נשארת לאורך כל המשמרת וחוזרת למצבה המקורי רק במשמרת הבאה.

|  |  |
| --- | --- |
| *Wait for Packaging + Packaging Time* | |
|  |  |
| Mean | 115.5525 |
| Standard Error | 16.61223 |
| Median | 184.581 |
| Standard Deviation | 98.27927 |
| Sample Variance | 9658.815 |
| Kurtosis | -2.05944 |
| Skewness | -0.03914 |
| Range | 229.4452 |
| Minimum | 2.76166 |
| Maximum | 232.2069 |
| Sum | 4044.336 |
| Count | 35 |

לפי הניתוח הסטטיסטי של רשומות 1-50, ניתן לזהות כי הממוצע נמצא בערך באמצע הקפיצה.

* 1. מהם ההסברים האפשריים להחזרת מוצרים? אילו גורמים היו עשויים להשפיע על כך? האם מדובר בגורמים הקשורים לפס הייצור, או אולי להיבטים אחרים ?
* **ותק העובד**

מצאנו כי עבור העובדים בעלי הותק הכי נמוך קיימים אחוזי החזרה גבוהים משמעותית מהעובדים בעלי ותק יותר גדול. לאחר ותק מסוים ההבדלים כבר לא היו משמעותיים והשוני מתבטא בצורה של רעשים יותר משל מגמה ברורה. דבר זה יכול להצביע לבעלים שיהיה להם משתלם להשקיע בהכשרה יותר ארוכה או השקעה בשיפור איכות העובדים הצעירים יותר.

ניתן לראות ויזואלית כי אחוזי ההחזרות של העובדים בעלי ותק נמוך (שנה ומטה) גדולים בהרבה מאחוזי ההחזרות של עובדים ותיקים יותר. על מנת להראות זאת סטטיסטית, ביצענו מבחן t עבור שני מדגמים בעלי שונות שונה. בכדי לבצע מבחן זה, היה עלינו לוודא ראשית כי ניתן להניח נורמליות של הנתונים. בעקבות מספר הנתונים הנמוך, הצלחנו לקבל את השערת האפס (נורמליות) תחת רמת מובהקות של עד 4.22%.

A screenshot of a cell phone

Description automatically generatedעבור ותק נמוך:

A screenshot of a cell phone

Description automatically generated

ועבור ותק גבוה:

תוצאות מבחן t:

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  |  |
|  |  |  |
|  | *ותק נמוך* | *ותק גבוה* |
| Mean | 0.337066765 | 0.100053803 |
| Variance | 0.000939347 | 0.000494128 |
| Observations | 7 | 13 |
| Hypothesized Mean Difference | 0 |  |
| df | 9 |  |
| t Stat | 18.06145218 |  |
| P(T<=t) one-tail | 1.11354E-08 |  |
| t Critical one-tail | 1.833112933 |  |
| P(T<=t) two-tail | 2.22708E-08 |  |
| t Critical two-tail | 2.262157163 |  |

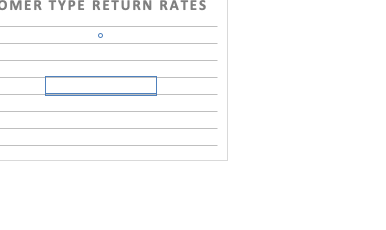
ברמת מובהקות 5%, נדחה את H0 של מבחן חד צדדי עליון ונאמר שהתוחלת של אחוזי ההחזרה של עובדים בעלי ותק נמוך יותר גבוה מתוחלת עובדים עם ותק גבוה.

* **לקוחות ספציפיים / סוג לקוח**

מצאנו כי קיימים לקוחות מסוימים אשר אחוז ההחזרות שלהם מתוך המוצרים שרכשו גבוה במיוחד. ניתן להמליץ לחקור את הנושא באופן פרטני עם הלקוח או להפחית את המכירות ללקוחות אלו באופן ספציפי.

מטרת גרף זה הוא לראות כי קיימים מספר לקוחות שמדיניות ההחזרה שלהם גבוה משאר הלקוחות.

לאחר מחקר מעמיק, גילינו כי עבור סוג לקוח industry, אחוז ההחזרות גבוה משמעותית מאשר שאר סוגי הלקוחות.

בדיקה של הנתונים בתרשים box-plot הראה כי ערך זה הינו חריג ביחס לשאר הנתונים.

בנוסף, ביצענו מבחני t על הפרש ממוצעי ההחזרות של כלל הלקוחות מהסוגים השונים אל מול industry ויצא שברמת מובהקות 5%, נדחה את השערת האפס (שתוחלת אחוזי ההחזרות של כל אחד מן סוגי הלקוחות גדול משל industry) בכל אחד מההשוואות ולכן ניתן להסיק כי תוחלת אחוזי ההחזרות גדול יותר אצל לקוחות מסוג industry משאר הלקוחות. הממצאים:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | | | | | t-Test: Two-Sample Assuming Unequal Variances | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  | *industry* | *private* |  |  |  | *industry* | *abroad* |  |  |
| Mean | 0.324 | 0.161902 |  |  | Mean | 0.324 | 0.151959 |  |  |
| Variance | 0.000368 | 0.000627 |  |  | Variance | 0.000368 | 0.000254 |  |  |
| Observations | 8 | 8 |  |  | Observations | 8 | 8 |  |  |
| Hypothesized Mean Difference | 0 |  |  |  | Hypothesized Mean Difference | 0 |  |  |  |
| df | 13 |  |  |  | df | 14 |  |  |  |
| t Stat | 14.53254 |  |  |  | t Stat | 19.51305 |  |  |  |
| P(T<=t) one-tail | 1.02E-09 |  |  |  | P(T<=t) one-tail | 7.51E-12 |  |  |  |
| t Critical one-tail | 1.770933 |  |  |  | t Critical one-tail | 1.76131 |  |  |  |
| P(T<=t) two-tail | 2.04E-09 |  |  |  | P(T<=t) two-tail | 1.5E-11 |  |  |  |
| t Critical two-tail | 2.160369 |  |  |  | t Critical two-tail | 2.144787 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| t-Test: Two-Sample Assuming Unequal Variances | | | | | t-Test: Two-Sample Assuming Unequal Variances | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  | *industry* | *student* |  |  |  | *industry* | *school* |  |  |
| Mean | 0.324 | 0.151708 |  |  | Mean | 0.324 | 0.146435 |  |  |
| Variance | 0.000368 | 0.000232 |  |  | Variance | 0.000368 | 0.000191 |  |  |
| Observations | 8 | 9 |  |  | Observations | 8 | 8 |  |  |
| Hypothesized Mean Difference | 0 |  |  |  | Hypothesized Mean Difference | 0 |  |  |  |
| df | 13 |  |  |  | df | 13 |  |  |  |
| t Stat | 20.32968 |  |  |  | t Stat | 21.24403 |  |  |  |
| P(T<=t) one-tail | 1.54E-11 |  |  |  | P(T<=t) one-tail | 8.86E-12 |  |  |  |
| t Critical one-tail | 1.770933 |  |  |  | t Critical one-tail | 1.770933 |  |  |  |
| P(T<=t) two-tail | 3.09E-11 |  |  |  | P(T<=t) two-tail | 1.77E-11 |  |  |  |
| t Critical two-tail | 2.160369 |  |  |  | t Critical two-tail | 2.160369 |  |  |  |

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | | |
|  |  |  |
|  | *industry* | *staff* |
| Mean | 0.324 | 0.145889 |
| Variance | 0.000368 | 0.000237 |
| Observations | 8 | 8 |
| Hypothesized Mean Difference | 0 |  |
| df | 13 |  |
| t Stat | 20.47149 |  |
| P(T<=t) one-tail | 1.42E-11 |  |
| t Critical one-tail | 1.770933 |  |
| P(T<=t) two-tail | 2.83E-11 |  |
| t Critical two-tail | 2.160369 |  |

נמליץ כי עבור מוצרים שמיועדים ללקוחות מסוג זה יעברו בקרת איכות נוספת לפני שליחה. כך המפעל יוכל לצמצם את ההחזרות מלקוחות אלו. נמליץ גם לפנות ללקוחות מסוג זה על מנת להבין מהם הסטנדרטים הדרושים עבורם וכך גם ללמוד איך להקטין את אחוז זה בעתיד.

עבור שאר הלקוחות, האחוזים יחסית קבועים ולכן אין צורך בשינוי באופן זה, למרות שכמובן שבדיקת איכות יכולה לתרום לכולם.

* **Failures**

רצינו לבדוק האם יש משמעות לתקלות שקורות במהלך משמרת. ראשית, עניין אותנו לדעת האם תקלה באופן כללי (ללא סיווג) משפיעה על כמות ההחזרות. לשם כך, חישבנו עבור המשמרות ללא תקלות כלל כמה החזרות היו ועבור משמרות עם תקלה כלשהי גם כן כמה החזרות היו.

ביצענו מבחן להסקה על שתי פרופורציות. כיוון שמדובר במדגם גדול (242 משמרות עם תקלות, 1316 משמרות ללא תקלות), לפי משפט הגבול המרכזי, ניתן להניח שהנתונים מתפלגים באופן נורמלי ולכן ניתן לבצע מבחן זה.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Returns | Not Returns | Total manufactured | Percent of Returns |
| Y Shifts with fialure | 853 | 3758 | 4611 | 18.50% |
| X Shifts without fialure | 1604 | 7456 | 9060 | 17.70% |
| מבחן על 2 פרופורציות | | | | |
| H0: px=py | p^ | Zst | Zcr=Z(1-alpha=93.3%) |  |
| H1: else | 0.179724 | 1.144647 | 1.5 | אי דחייה |

תוצאות המבחן הראו כי אין שוני מובהק באחוז ההחזרות עבור משמרות ללא תקלות ומשמרות עם תקלות ברמת מובהקות 6.7% (וגם עבור מובהקות נמוכה מזו).

כעת רצינו לבדוק האם ניתן להסיק מסקנה שונה עבור תקלות מסווגות- האם יש תקלות שמשפיעות על ההחזרות יותר מאחרות.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Returns | Not Returns | Total manufactured | Percent of Returns |
| X Shifts without fialure | 1604 | 7456 | 9060 | 17.70% |
| Y1 Shifts with Assembly fialure | 314 | 1546 | 1860 | 16.88% |
| Y2 Shift with glue failure | 264 | 1098 | 1362 | 19.38% |
| Y3 Shift with Package failure | 275 | 1114 | 1389 | 19.80% |
| מבחן על 2 פרופורציות | | | | |
| H0: px>=py, H1: else | p^ | Zst | Zcr=Z(1-alpha=93.3%) |  |
| assembly vs none | 0.175641 | 0.8491026 | 1.5 | אי דחייה |
| glue vs none | 0.179236 | 1.5063393 | דחייה |
| package vs none | 0.179826 | 1.8924372 | דחייה |

תוצאות המבחנים מראים כי עבור במכונת ההדבקה ומכונת האריזה, ברמת מובהקות 6.7%, נדחה את ההשערה כי אחוז ההחזרות של משמרות ללא תקלה זהה (או גדול) למשמרות עם תקלות אלו. עבור ההרכבה לא ניתן להסיק שהפרופורציות שונות ברמת מובהקות זו.

יש להדגיש כי לאחר תחנת ההרכבה מתקיימת בקרת איכות על ידי המפעיל, ולאחר מכן המוצר עובר בתחנות ללא בדיקה נוספת.

לאור התוצאות שהתקבלו, נמליץ להוסיף בקרה של המוצרים הנמצאים בפס ייצור לאחר כל תקלה ובכך ננסה להקטין את כמות ההחזרות.

* **Time to sell**

חילקנו את time to sell ל14 דליים המכילים עד 1000 רשומות. לאחר מכן, עבור כל דלי חישבנו את אחוז המוצרים שהוחזרו מתוך כלל המוצרים. מצאנו כי קיימת מגמה עולה ככל שזמן המדף גדול יותר. דבר זה מעיד על פגמים שנוצרים בזמן האחסון ונמליץ למפעל לעבוד בצורה של FIFO בכדי להקטין כמה שניתן את זמן המדף של המוצרים שנמכרים ובכך למנוע כמה שניתן את תופעה זו.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 0.009805085 | 0.009805085 | 41.9595413 | 3.03456E-05 |
| Residual | 12 | 0.002804154 | 0.00023368 |  |  |
| Total | 13 | 0.012609239 |  |  |  |

בכדי לבדוק לינאריות ביצענו מבחן F ולפי טבלת ANOVA יצא כי הקו הוא אכן לינארי בעל R2=0.77.

* **משמרות עוקבות**

רצינו להשוות בין ההחזרות של מוצרים של עובדים במשמרות רגילות אל מול משמרות כפולות (משמרות עוקבות בהן אותו העובד השגיח).

ראשית, בדקנו את המשמרות הכפולות יחד אל מול משמרות רגילות, ומצאנו כי בסך הכל, אחוז ההחזרה גדל מ-17.9% ל-18.9%: גידול של כ-5% בכמות המוצרים המוחזרים.

לאחר מכן, בדקנו האם המשמרת הראשונה מבין הכפולות מתנהגת באופן שונה ממשמרות רגילות, וגילינו שלא- 17.8% לעומת 17.9% של המשמרות הרגילות.

לבסוף, בדקנו את המשמרת השנייה בלבד אל מול שאר המשמרות, ומצאנו כי קיים הבדל יותר משמעותי- גידול מ-17.9% ל-20% החזרות: גידול של כ-12% בכמות המוצרים המוחזרים.

אחוז המשמרות הכפולות מתוך כלל המשמרות הוא רק כ-1.7% ולכן סביר להניח כי גורם זה אינו גורם מרכזי להחזרות. בניסיון לבצע מבחן סטטיסטי להשוואת פרופורציות להראות את נכונות הקשר, מצאנו כי אין מובהקות בתוצאות ולכן גורם זה אינו מוכח סטטיסטית כמשפיע על המכירות. בכל מקרה, ניתן לומר כי המשמרת הכפולה הראתה קפיצה באחוזי ההחזרות ולכן המלצתו היא להמנע ממשמרות אלו במידה וניתן, ואם התופעה נהיית יותר שכיחה, יש לבחון את הקשר מחדש כאשר גודל המדגם של המשמרת העוקבת גדול יותר.

* **צריכת אנרגיה**

כיוון שתהליך ההרכבה הוא מרכזי במפעל, רצינו לבדוק איך משפיעה צריכת החשמל של המכונה על המוצרים השונים.

חילקנו את כלל המוצרים לדליים שווי גודל וחישבנו את אחוזי ההחזרה עבור כל דלי. התוצאה:

מצאנו כי עבור מוצרים שעבורם הצריכה הייתה גדולה, אחוז ההחזרות עלה באופן חריג. בכדי לבדוק את הקשר, בדקנו עבור 2 קבוצות האם פרופורציית ההחזרות זהה- הקבוצה הראשונה הייתה עבור צריכה נמוכה (המקבץ שניתן לראות בגרף) והשנייה הייתה הדליים בעלי הנתונים החריגים.

ביצענו מבחן להסקה על שתי פרופורציות. עקב גדלי המדגם (nx=11724,ny=1954), לא ביצענו מבחן לנורמליות המדגמים כיוון שלפי משפט הגבול המרכזי, הנתונים יתפלגו בקירוב נורמלי. תוצאות המבחן:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Returns | Not Returns | Total manufactured | Percent of Returns | Average Energy Con. |
| Norm. Energy Consumption X | 1592 | 10132 | 11724 | 13.58% | 6.957282 |
| Y High Energy Consumption | 866 | 1088 | 1954 | 44.32% | 11.60036 |
| מבחן על 2 פרופורציות | | | | |  |
| H0: px<=py | p^ | Zst | Zcr=Z(1-alpha=95%) |  |  |
| H1: else | 0.179705 | 32.76673 | 1.65 | דחייה |  |

מצאנו כי ברמת מובהקות 5%, נדחה את השערת האפס ונאמר שאחוז ההחזרות עבור הקבוצה עם צריכת אנרגיה גבוהה אכן גבוה מאחוז ההחזרות עם צריכת אנרגיה סטנדרטית.

* **חלקים פגומים**

ניסינו לנתח האם אחוז פגומים בקופסא יכולה להעיד על תוצאות המוצרים שייווצרו ממנו. בכדי לבצע את הניתוח, בדקנו מהו אחוז הפגומים בקופסא כולה (בהינתן שהיא בעלת 120 חלקים) ובדקנו האם קיימת מגמה כתלות באחוז הפגומים בה. ניתן לראות כי עבור המגמה מתנהגת באופן לא צפוי, כאשר עבור אחוזי פגומים נמוכים, אחוז ההחזרות מגוון ומגיע עד 100%, ולעומת זאת באחוזי פגומים גבוהים אחוזי ההחזרות נמוכים יותר ופחות מרווחים. בחרנו שלא להתייחס לתופעה זו כמסביר לשיעור ההחזרות של המפעל.

* **זמני המתנה**

סכמנו את זמני ההמתנה שכל מוצר עבר על מנת לנתח או שיעור ההחזרות כתלות במשך ההמתנה של המוצרים. לא מצאנו כי קיים קשר ישיר ביניהם, אחוז ההחזרות נותר בטווח קטן ולכן מדד זה אינו משפיע באופן ישיר על ההחזרות.

* **כמות דבק**

בכדי לבדוק האם כמות הדבק משפיעה, חילקנו את הנתונים ל-100 דליים. עבור כל דלי חישבנו את אחוז ההחזרות מתוכו. גילינו כי אין מגמה שניתנת לזיהוי המעידה על כך שגורם זה משפיע על החזרות.