|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **מס' קבוצה** | **נושא הפרויקט** | | | **תאריך הגשה** |
| 17 | Giella.com | | | 10/03/2024 |
| **שמות ומספרי ת"ז של המגישים** | | | | |
| אפיק דדון (206747776) | | קארין כנעני (209350537) | עוז גיל-און (316468586) | |

**חלק ב' – יישומי בינה ואנליטיקה עסקית**

# מטלה מקדימה - עדכון הנתונים

תיקון מקור הנתונים

**המודל הטבלאי המתוקן:**

* **CUSTOMERS** (Email, First Name, Last Name, Gender, Age, Phone, Country, City, First Order)
* **SEARCHES** (Search DT, IP Address, Content, Country, City, Email (CUSTOMERS))
* **PRODUCTS** (ProductID, Color, Category, Price, Formula, Shimmer, Aroma, Additives, Customize, Quantity)
* **REVIEWS** (ReviewID, Stars, Review Date, Email (CUSTOMERS), ProductID (PRODUCTS))
* **ORDERS** (OrderID, Order Date, Shipping Method, Shipping Date, Country, City, Payment Method, Discount, {Search DT, IP Address} (SEARCHES), Email (CUSTOMERS))
* **PRODUCT\_ORDER** (Order ID (ORDERS), ProductID (PRODUCTS), Quantity)

השינויים שבוצעו:

**CUSTOMERS**: מחיקת שדה Birthdate

**PRODUCTS**: מחיקת שדה Discount (החלטנו שבמקום לתת הנחה על מוצר מסוים, ההנחה תינתן בפרק זמן מסוים על כלל ההזמנות המבוצעות באתר. לכן ההנחה היא לא פר מוצר אלא פר הזמנה, והשדה עבר לטבלת הזמנות).

**ORDERS**: הוספת השדה Discount, שכן ההנחה היא על כלל ההזמנה ולא לפי מוצר.

**PRODUCT\_ORDER**: הוספת הטבלה במקור הנתונים באקסל

תיקון לקובץ המארגן את הנתונים בסכמת כוכב

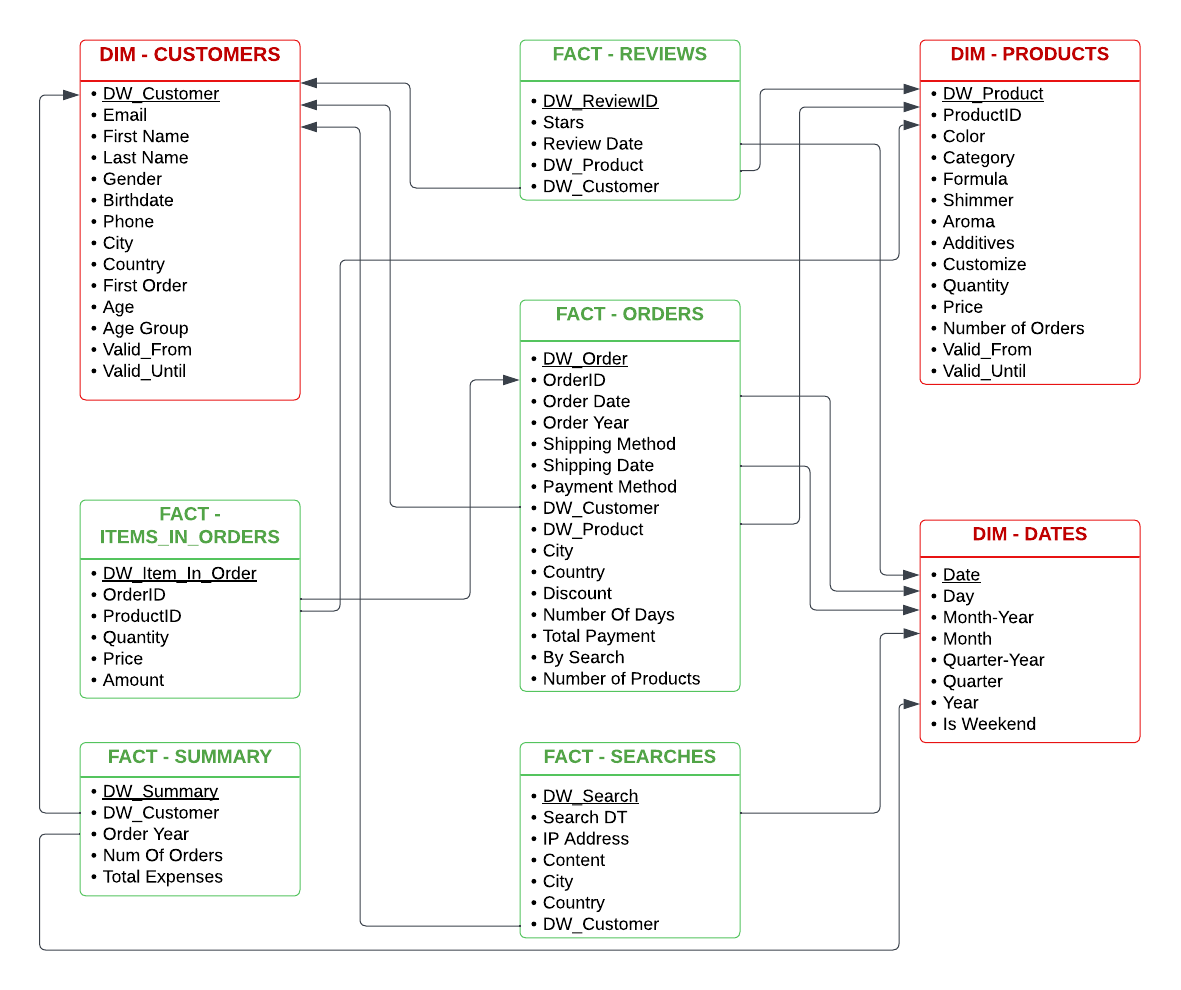
**CUSTOMERS**: נוסף השדה Country

**PRODUCTS**: נמחקו השדות Discount ו- Actual price (ההנחה היא כבר לא פר מוצר אלא פר הזמנה). נוסף השדה Number of Orders המציין את מספר ההזמנות שכללו את המוצר.

**CITIES**: הטבלה נמחקה

**SEARCHES**: נוסף השדה Country

**ORDERS**: נוספו השדות Country ו- Discount. החישוב של השדה Total Payment עודכן כך שיתחשב בהנחה. נוסף השדה By Search הנותן אינדיקציה האם ההזמנה בוצעה כתוצאה מחיפוש. נוסף השדה Numberof Products המציין את מספר המוצרים שנכללו בהזמנה.



**1. דו"חות עסקיים**

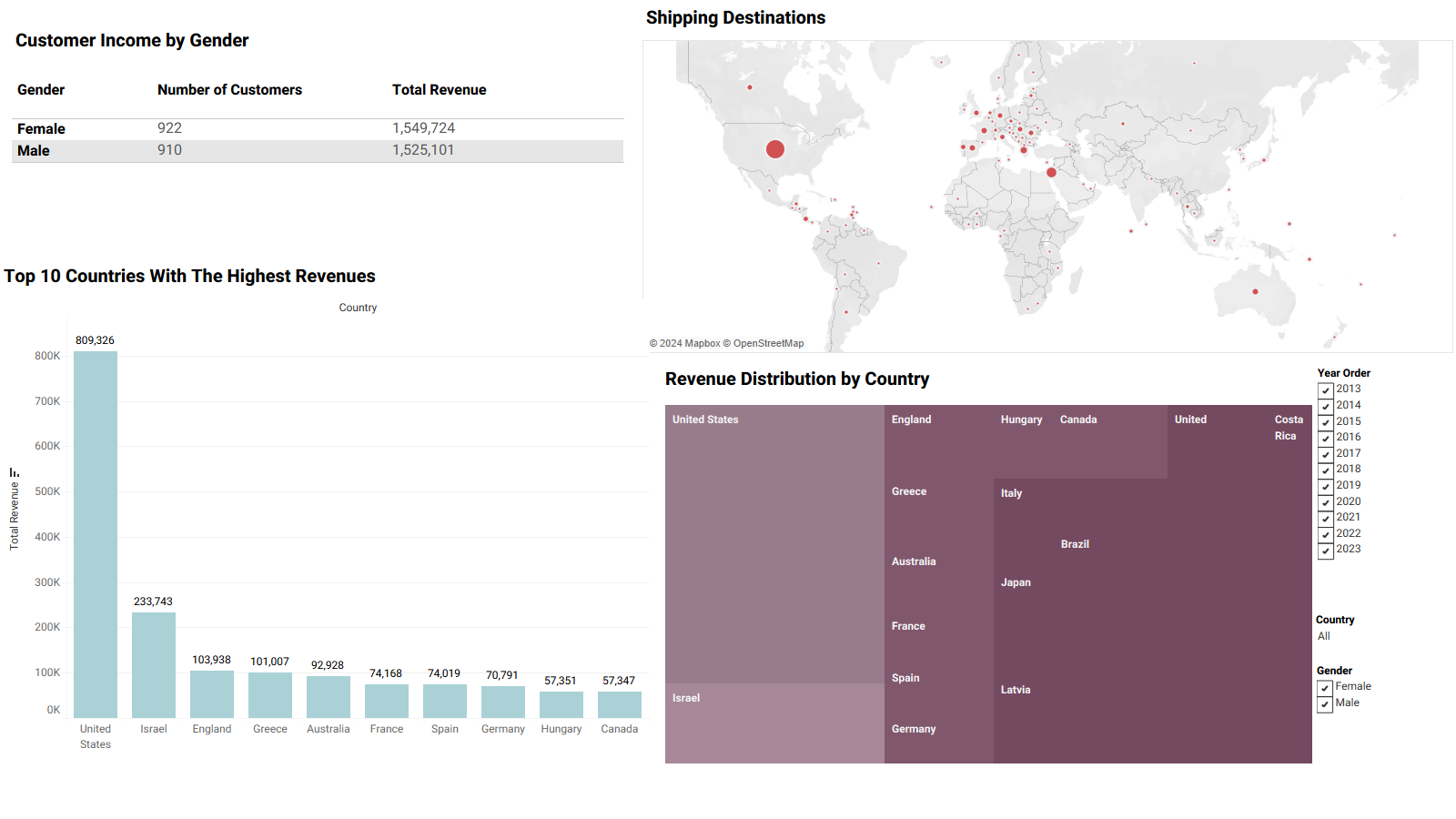
דו"ח ניהולי – אסטרטגי

הדוח מיועד לאנשי ההנהלה הבכירה והכספים ומתעסק בהחלטה האם להרחיב או לצמצם את נקודות המשלוח ליעד מסוים בעולם. הדוח משקף את ה-KPI המתאר את יחס התרומה של יעד מסוים להכנסות החברה:

*במטרה להשוות בין תרומתם של יעדים שונים להכנסות החברה, הדוח מציג מפת עולם בה מתוארות מדינות שונות המהוות יעדי משלוח. על המפה מסומנות נקודות בגדלים שונים, גודל הנקודה כגודל התרומה היחסית של היעד להכנסת החברה. בנוסף, תרשים ה-* Tree-Map *מתאר את המדינות השונות אליהן מבוצעים משלוחים. בלחיצה על הריבוע של המדינה ניתן לראות את ההכנסה הכוללת מאותה מדינה. תרשים העמודות מציג את עשרת המדינות עם ההכנסה הגבוהה ביותר לחברה.*

*כאמצעי טקסטואלי, בנינו טבלה המציגה עבור כל מגדר את כמות הלקוחות וההכנסה הכוללת.*

*הדוח מאפשר סינון כולל על פי שנה, על פי מדינה ועל פי מגדר. בבחירת מסנן יעודכנו האמצעים הוויזואליים והטקסטואליים בדו"ח, מלבד הגרף המתאר את עשרת המדינות בעלות ההכנסה הגבוהה ביותר, אותו לא ניתן לסנן על פי מדינה.*

**

דו"ח ניהולי – טקטי

הדוח מיועד לאנשי השיווק, המכירות והכספים ומתעסק בהחלטה האם לתת הנחה לזמן מוגבל על כלל ההזמנות המבוצעות באתר. הדוח משקף את ה-KPI המתאר את יעילות ההנחה:

ה-KPI בא לידי ביטוי בתרשים עמודות שמציג עבור כל שנה את כמות הרכישות שבוצעו בתקופת הנחה יחד עם כמות הרכישות שבוצעו שלא בתקופת הנחה.

במטרה לבחון את היחס בין גודל ההנחה לבין כמות הרכישות, הדוח מציג תרשים פיזור המתאר את כמות ההזמנות שבוצעו עבור כל גודל הנחה, מה שמאפשר למידה על מגמות רכישה על פי אחוזי הנחה.

כאמצעי טקסטואלי, על מנת לתאר את ההשפעה של מתן ההנחות, הדוח מציג טבלה המתארת עבור כל קטגורית מוצר את כמות הרכישות הכוללת שבוצעה ממנו, וכמה אחוז מתוך הרכישות בוצעו בתקופת הנחה. בשורה האחרונה של הטבלה ניתן לראות סיכום של כלל הנתונים (עבור כלל המוצרים).

לצורך ניתוח נוח ויעיל, הדוח מציג את כמות הרכישות הכוללת שבוצעה.

*הדוח מאפשר סינון כולל על פי קטגורית מוצר, על פי טווח תאריכי הזמנה ועל פי מדינה. בבחירת מסנן יעודכנו כלל האמצעים הוויזואליים והטקסטואליים בדו"ח.*

תמונה שמכילה טקסט, צילום מסך, תרשים, קו

התיאור נוצר באופן אוטומטי

**2. לוח מחוונים (Digital Dashboard)**

לוח המחוונים שלנו מותאם למנהלים ברמה האסטרטגית, כולל מנכ"לים ואנשי מקצוע בתחום הפיננסים. הלוח מתמקד בביצועים מרכזיים כמו שביעות רצון לקוחות, שימור ומדדי מכירות. לוח המחוונים מספק תמונת מצב של נתונים חיוניים כגון: דירוג לקוחות, שיעורי שימור, מוצרים שנמכרו, כמות ממוצעת להזמנה, הכנסה חודשית ושנתית, יחס הזמנות לחיפוש ומכירות לפי קטגוריות. על ידי מעקב אחר שביעות רצון לקוחות ושיעורי שימור, לוח המחוונים מאפשר חידוד מתמשך של המוצר ויוזמות שיווק מותאמות. חלוקה של מכירות לפי קטגוריה, יחד עם המסננים הרלוונטיים, מחזקת אסטרטגיות רלוונטיות המגיבות להעדפות המגוונות של הלקוחות. יתר על כן, ההדמיה של מגמות הכנסות מסייעת בקביעת יעדים ריאליים ותכנון אסטרטגי.

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטיהמסננים - קטגוריה, שנה של תאריך ההזמנה, מגדר וקבוצת גיל, מאפשרים למשתמשים להתעמק בפלחים ספציפיים לצורך ניתוחים ממוקדים. לדוגמה, בחינת מכירות לפי קטגוריות עוזרת להקצות משאבים בצורה יעילה, בעוד שמסננים לפי קבוצת גיל חושפים תובנות לגבי התנהגות לקוחות. בכדי לבחון היבטים לעומק יותר ישנה אפשרות קידוח עבור גרף ההכנסות לפי שנה - כאשר לוחצים על שנה ספציפית, נפתחת טבלה שמפרטת באילו מדינות נרכשו הכי הרבה מוצרים, לאחר מכן באילו ערים, ולבסוף מפולח לפי קטגוריה.

**3. כלי לתחקור אנליטי מקוון OLAP – On-Line Analytical Processing))**

כלי ה-OLAP שבנינו נותן מענה על היבטים של פילוח לקוחות ועוזר בתחקור אנליטי שמתמקד בשאלה כיצד פלחי לקוחות שונים (בהתבסס על מדינה, עיר וקבוצות גיל) תורמים לכמות ההזמנות הכוללת ולסכום ההזמנה הממוצע.

תחילה, טבלת ה-PIVOT משמשת כטבלה מקיפה, החושפת את יחסי הגומלין המורכבים בין מוצרים ולקוחות על פני נתונים דמוגרפיים מגוונים. הראש הטופס

טבלה מאפשרת להשוות בקלות דפוסי רכישה בין מדינות ובין קטגוריות מוצרים שונות, ומאפשרת תובנות לגבי העדפות אזוריות. הערכים בטבלה המציינים את מספר הרכישות מסווגים לפי צבעים כאשר כל מי שמעל 100 מסומן בירוק, כל מי שבין 10 ל-100 מסומן בכתום, וכל מה שמתחת ל-10 מסומן באדום.

כדי להבין את הקשר בין המחיר הכולל לכמות ההזמנות עבור כל קטגוריות מוצרים השתמשנו בתרשים בועות כאשר כל בועה מתארת קטגוריה שונה. באמצעות התרשים ניתן לזהות מוצרים עם נפחי מכירות גבוהים. ראינו כי יש פער מאוד גדול בין הקטגוריה שמניבה את ההכנסה הגדולה ביותר לבין זו שמניבה את ההכנסה הקטנה ביותר.

רצינו לבדוק האם תופעה זו קורית באופן גורף או שמשתנה לאורך השנים ולכן השתמשנו בתרשים קו שמתאר עבור כל קטגורית מוצר את ההכנסות ממנו לאורך השנים. מהגרף עולה כי יש תנודות רבות לאורך השנים בהכנסות מהקטגוריה שמניבה את ההכנסה הגבוהה ביותר, לעומת הקטגוריה שההכנסה ממנה היא הנמוכה ביותר, הנשארת יחסית יציבה לאורך השנים.

כדי לחזק את הממצאים, רצינו לבדוק מה ההבדלים במחיר ההזמנה הממוצע עבור כל קטגוריה. במקרה זה עלה מן הגרף שאין הבדל מהותי בין הקטגוריות השונות.

כדי לבדוק את ההשפעה של גורמים נוספים על הממצאים, יצרנו מספר אפשרויות סינון: סינון על פי קטגוריית מוצרים כדי להתמקד בפלחי מוצרים ספציפיים ולנתח את ההכנסות מהם באופן ממוקד יותר, וכן סינון על פי מדינה וקבוצות גיל כדי לפלח את הנתונים לפי נתונים דמוגרפיים שונים של לקוחות ולהבין את התנהגות הרכישה שלהם. בנוסף, ישנו מסנן לבחירת תקופות זמן ספציפיות (למשל חודשים, רבעונים, שנים) לניתוח מגמות ודפוסים לאורך זמן.

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטיכדי לבדוק האם המגמות העולות מן התרשימים השונים אכן נכונות, ביצענו בהמשך העבודה (חלק 4) מבחנים סטטיסטיים לבחינת ההשערות.

# 4. יישומי BI&A וניתוח נתונים מתקדמים

**יישום א': תחקור אנליטי מאשש מלווה בניתוח השערות סטטיסטי**

T TEST

בניתוח זה, בדקנו האם יש הבדל משמעותי בהכנסה הכוללת בין שתי קבוצות גיל: קבוצת הגיל הצעירה (עד גיל 50) וקבוצת הגיל המבוגרת יותר (גיל 50 ומעלה).

השערות:

השערת אפס (H0): סך ההכנסה זהה עבור שתי קבוצות הגיל.

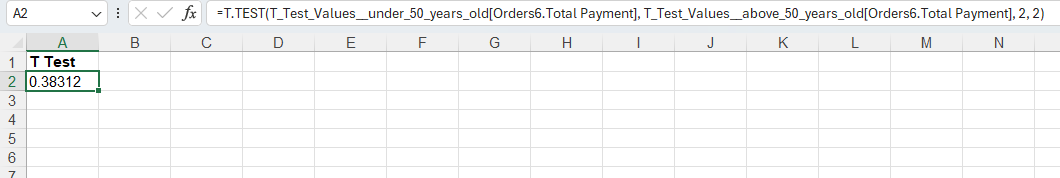
השערה חלופית (H1): קיים הבדל בסך ההכנסה בין שתי קבוצות הגיל.

ה-T-Test נבחר כמבחן הסטטיסטי המתאים לניתוח זה בשל יעילותו בהשוואת האמצעים של שתי קבוצות עצמאיות. מבחן זה מאפשר לקבוע אם הבדלים כלשהם שנצפו בהכנסה הכוללת בין שתי קבוצות הגיל הם מובהקים סטטיסטית, תוך התחשבות בשונות בתוך כל קבוצה.

לאחר ביצוע מבחן ה-T באמצעות אקסל, השגנו ערך p-value של 0.38312. ערך זה מציין את ההסתברות לצפות בהפרש שהושג בהכנסה הכוללת בין שתי קבוצות הגיל אם השערת האפס הייתה נכונה.

עם ערך p-value גדול מרמת המובהקות המקובלת של 0.05, איננו מצליחים לדחות את השערת האפס. לכן, אנו מסיקים כי אין הבדל משמעותי בהכנסה הכוללת בין קבוצת הגיל הצעירה לקבוצת הגיל המבוגרת.

תובנה זו יכולה לספק מידע על תהליכי קבלת החלטות עסקיים הקשורים לאסטרטגיות שיווק, הצעות מוצרים ומאמצי מיקוד לקוחות.



קורלציה

בניתוח זה רצינו לקבוע האם יש קשר בין כמות המוצרים שהוזמנו לבין סכום התשלום הכולל.

השערות:

השערת האפס (H0): אין קשר בין כמות המוצרים שהוזמנו לבין סכום התשלום הכולל.

השערה חלופית (H1): קיים קשר בין כמות המוצרים שהוזמנו לבין סכום התשלום הכולל.

טכניקה סטטיסטית: קורלציה (מתאם)

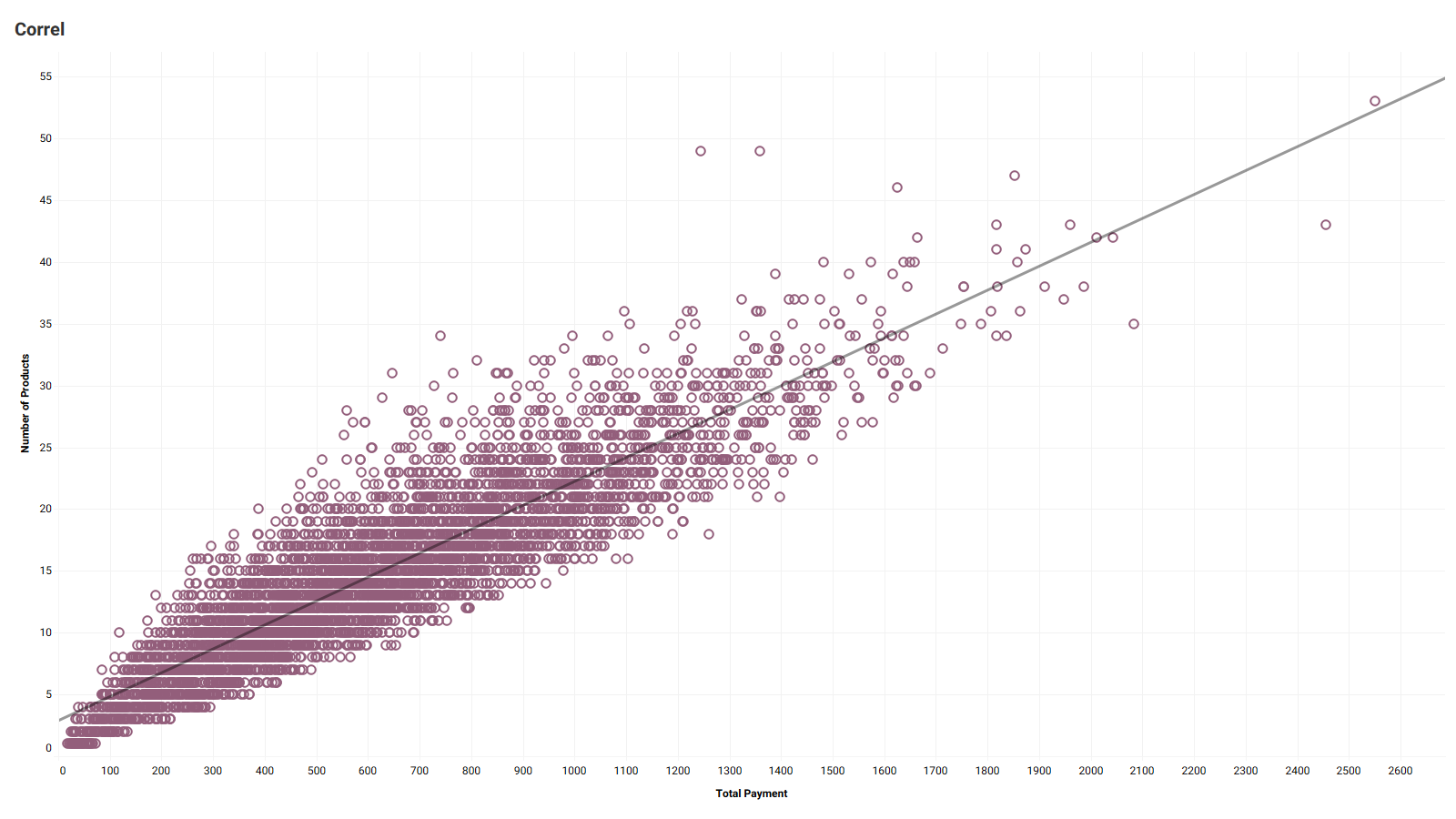
ניתוח הקורלציה נבחר כמבחן הסטטיסטי המתאים למקרה זה זו בשל יכולתו להעריך את מידת הקשר הליניארי בין שני משתנים רציפים. על ידי חישוב מקדם המתאם, נוכל לקבוע את החוזק והכיוון של הקשר בין כמות המוצרים שהוזמנו לבין סכום התשלום הכולל. מקדם מתאם קרוב ל-1 מצביע על קשר חיובי חזק, בעוד שמקדם קרוב ל- 1- מצביע על קשר שלילי חזק.

לאחר ביצוע המבחן באמצעות אקסל, קיבלנו מקדם מתאם של 0.897168933. ערך זה מצביע על קשר חיובי חזק בין כמות המוצרים שהוזמנו לבין סכום התשלום הכולל. לכן נדחה את השערת האפס. במילים אחרות, ככל שכמות המוצרים שהוזמנו עולה, גם סכום התשלום הכולל נוטה לעלות. ממצא זה נתמך עוד יותר על ידי הגרף, מתאר מגמה לינארית חיובית בין שני המשתנים.

תובנה זו יכולה לספק מידע על תהליכי קבלת החלטות עסקיים הקשורים לניהול מלאי, אסטרטגיות תמחור וחיזוי הכנסות.

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטי



חי בריבוע

בניתוח זה שאפנו לבחון האם יש קשר בין לקוחות שקיבלו את ההזמנה תוך 3 ימים לבין דירוג גבוה של המוצר.

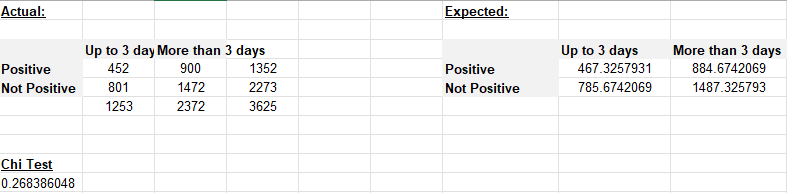
השערות:

השערת אפס :(H0) אין קשר בין לקוחות שקיבלו את ההזמנה תוך 3 ימים לבין דירוג גבוה של המוצר.

השערה חלופית :(H1) קיים קשר בין לקוחות שקיבלו את ההזמנה תוך 3 ימים לבין דירוג גבוה של המוצר.

טכניקה סטטיסטית: מבחן Chi-Squared

מבחן Chi-Squared נבחר כמבחן הסטטיסטי המתאים לחקירה זו בשל התאמתו לניתוח הקשר בין שני משתנים קטגוריים. מבחן Chi-Squared מאפשר לקבוע אם קשרים שנצפו הם מובהקים סטטיסטית. במקרה זה נבחן האם קיים קשר מובהק בין המשתנים "זמן אספקת הזמנה (תוך 3 ימים)" ו"דירוג מוצר" (דירוג גבוה של 4 או 5 כוכבים).

לאחר ביצוע מבחן Chi-Squared באמצעות ,Excel התקבל ערך סטטיסטי למבחן של 0.268. ערך זה מייצג את רמת המובהקות של מערכת היחסים בין לקוחות שקיבלו את ההזמנות שלהם תוך 3 ימים ודירוג גבוה של המוצר. עם ערך p-value גדול מרמת המובהקות המקובלת של 0.05, איננו מצליחים לדחות את השערת האפס. לכן, אנו מסיקים כי אין קשר בין לקוחות שמקבלים את ההזמנות שלהם תוך 3 ימים לבין מתן דירוג גבוה למוצר. תובנה זו יכולה לספק אסטרטגיות עסקיות הקשורות לאספקת הזמנות ולשביעות רצון לקוחות.