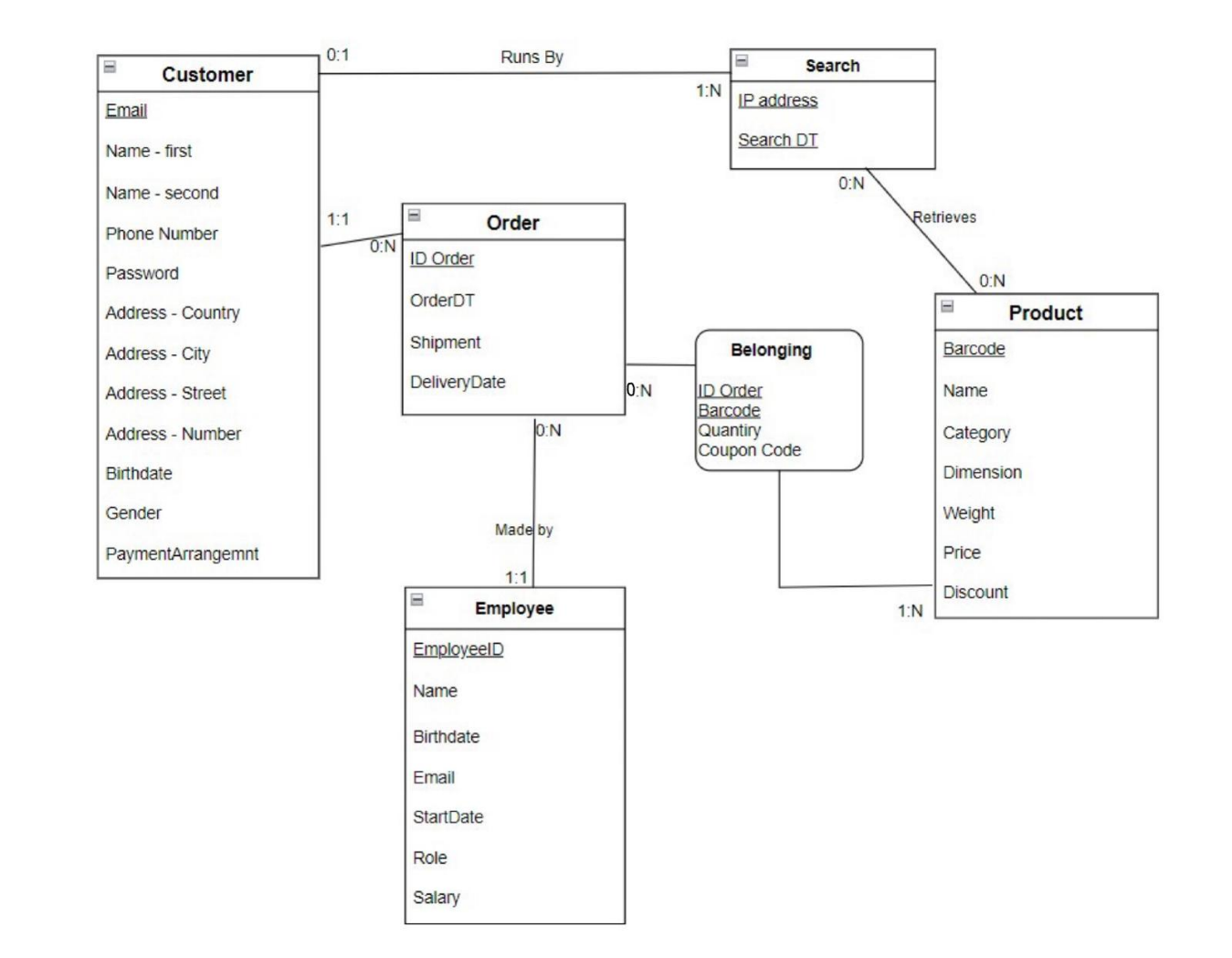
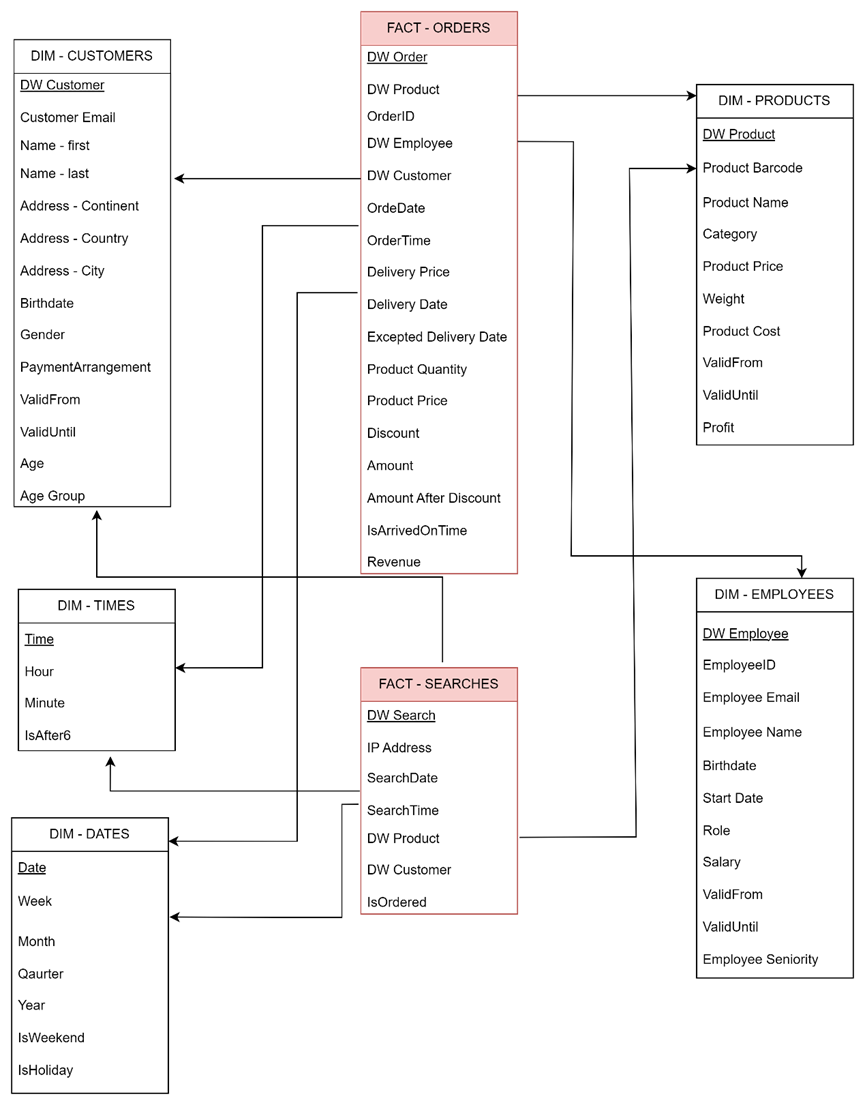
|  |  |
| --- | --- |
| **נושא הפרוייקט** | |
| kustomkitgymequipment.com | |
| יואב נחמני 208996462 |

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

**מטלה מקדימה - עדכון הנתונים**

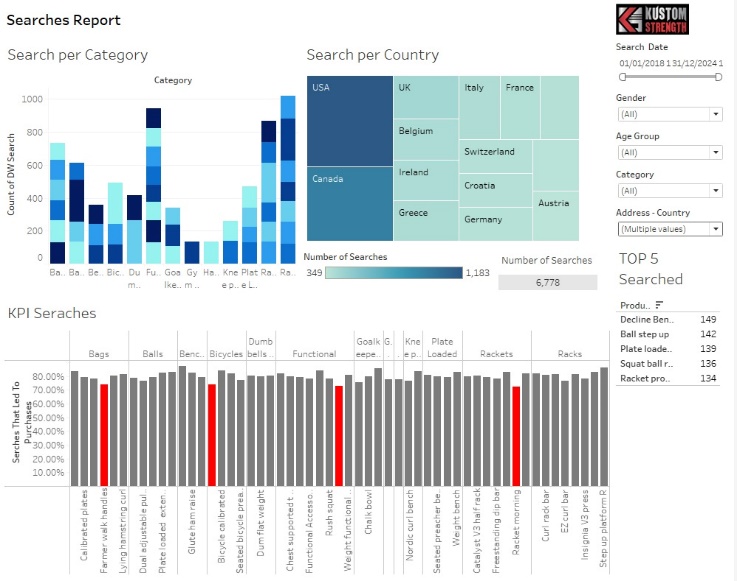
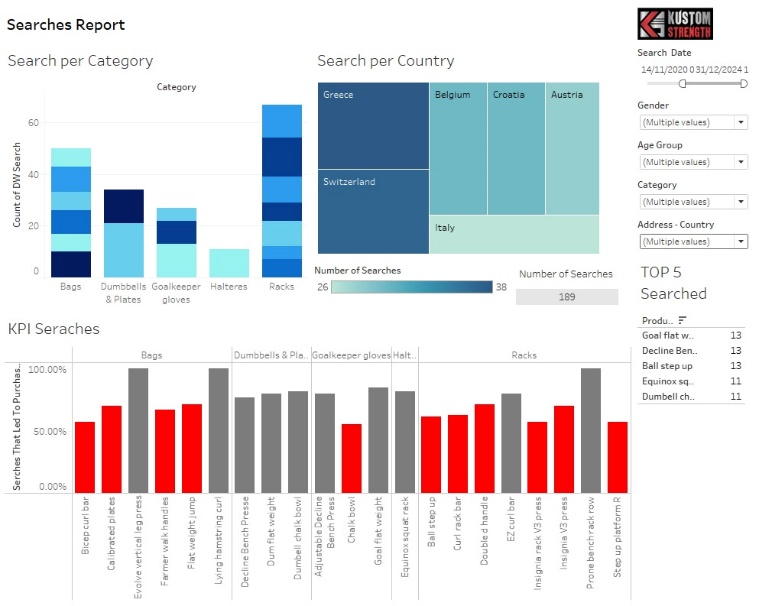
* ב-ERD שינינו את הקרדינאליות בין מוצר להזמנה ל-N:0 כדי לאפשר מצב שמוצר חדש שהתווסף למלאי עדיין לא הוזמן.
* בטבלת PRODUCTS שינינו את סולם המדידה של השדה הנחת המוצר ל-Ratio מכיוון שמדובר באחוזים.
* בסכמת הכוכב ביטלנו את קישור טבלאות המימד עם טבלאת מימד תאריכים, ועדכנו את פירוט של טבלת סכמת הכובה בהתאם.
* הוספנו את יחידות המידה של הKPI- (אחוזים), והחלפנו את מדד אחוז ההזמנות שהגיעו בזמן למדד תועלת.

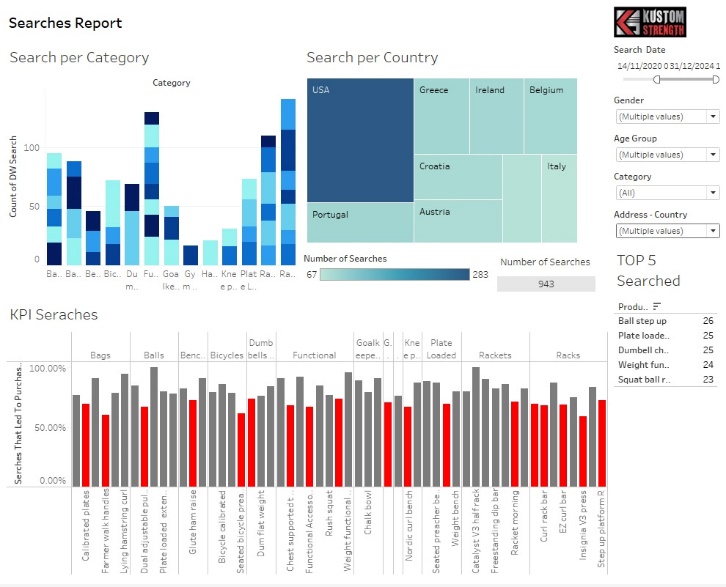
****



1. **שני דו"חות עסקיים**

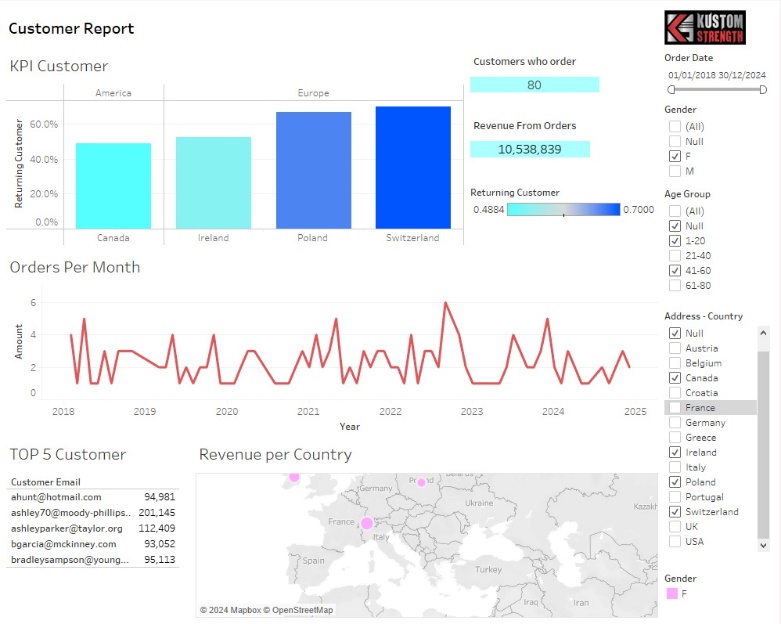
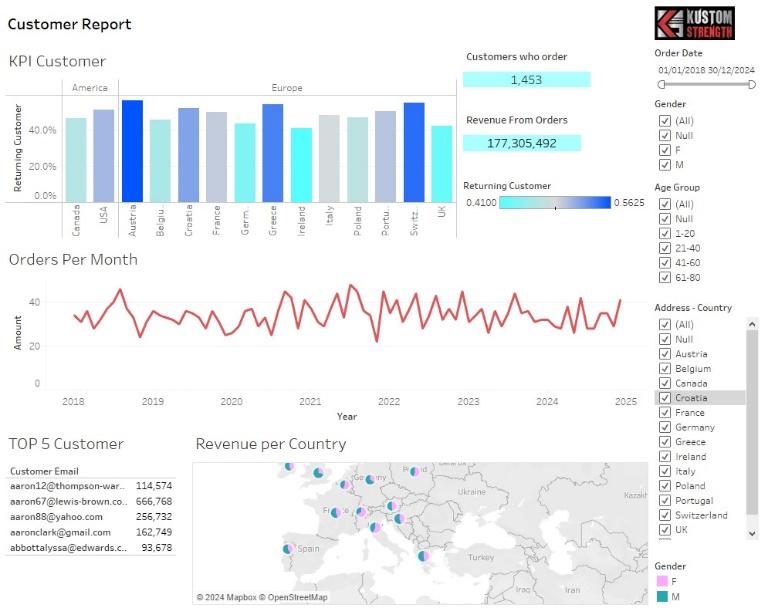
**דוח לקוחות-**

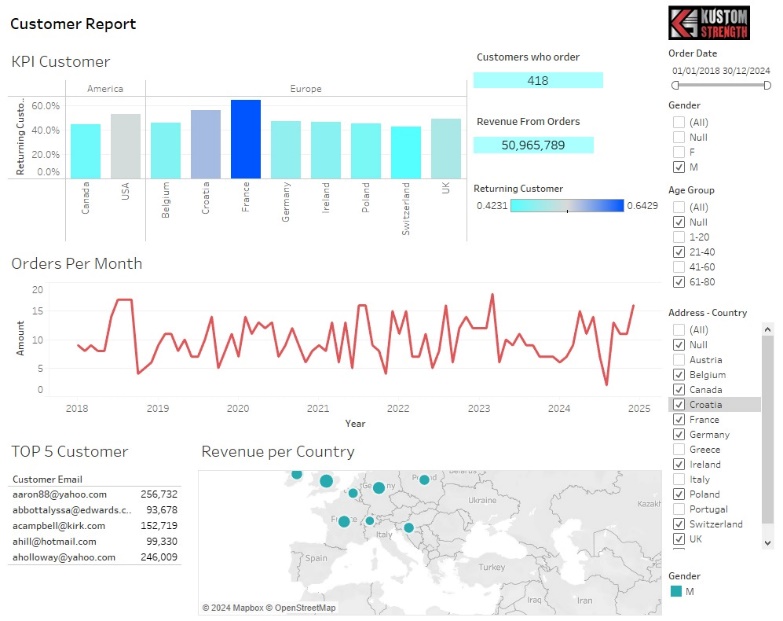
**החלטה ניהולית - טקטית:** לאילו מבין הלקוחות שביצעו הזמנה כדאי להציע הטבה בהזמנה הבאה.  
הדו"ח נועד עבור מחלקת השיווק על מנת לקבל תמונת מצב על פילוח הלקוחות במטרה למקסם את רווחיות הלקוחות.   
הדו"ח מציג: אחוז לקוחות חוזרים עבור כל מדינה (KPI), סך הכנסות מכל ההזמנות, כמות הזמנות עבור כל חודש, חמשת הלקוחות הרווחיים ביותר ורווחים עבור כל מדינה מפולח לפי גברים ונשים.  
ניתן לפלטר את התרשימים לפי: תאריך הזמנה, מין הלקוח, קבוצת גיל אליה הלקוח משתייך והמדינה בה גר הלקוח.

****

**דוח חיפושים-**

**החלטה ניהולית - אסטרטגית:** מהו התקציב אותו תשקיע החברה בפרסום מוצרים, ואילו מבין המוצרים תנסה החברה לקדם יותר. הדו"ח נועד עבור מחלקות השיווק והכספים על מנת לקבל תמונת מצב על המוצרים שיש סביבם "באז" גדול (שחיפשו אותם הרבה) לעומת אלו שפחות.  
הדו"ח מציג:KPIאחוז חיפושים שהסתיימו ברכישה עבור כל מוצר. מוצרים שפחות מ-75% מהחיפושים שלהם הובילו לרכישה מודגשים באדום על מנת לעורר את תשומת לב המנהלים. בנוסף, הדו"ח מציג גם: כמות חיפושים בכל מדינה,כמות החיפושים בפרק הזמן, חמשת המוצרים שחיפשו הכי הרבה וכמות חיפושים עבור כל קטגוריה מפולח לפי מוצרים.  
ניתן לפלטר את התרשימים לפי: תאריך הזמנה, מין הלקוח, קבוצת גיל אליה הלקוח משתייך, המדינה בה גר הלקוח וקטגוריית המוצר.

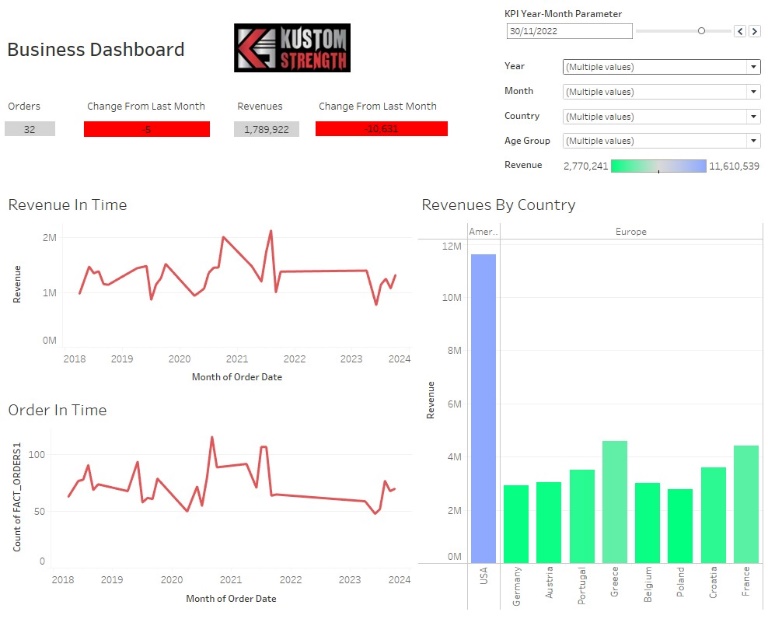
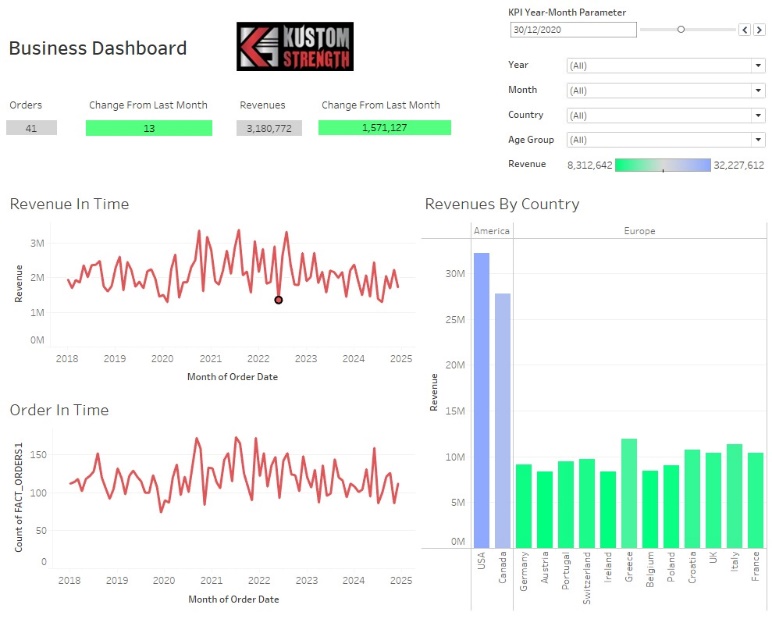


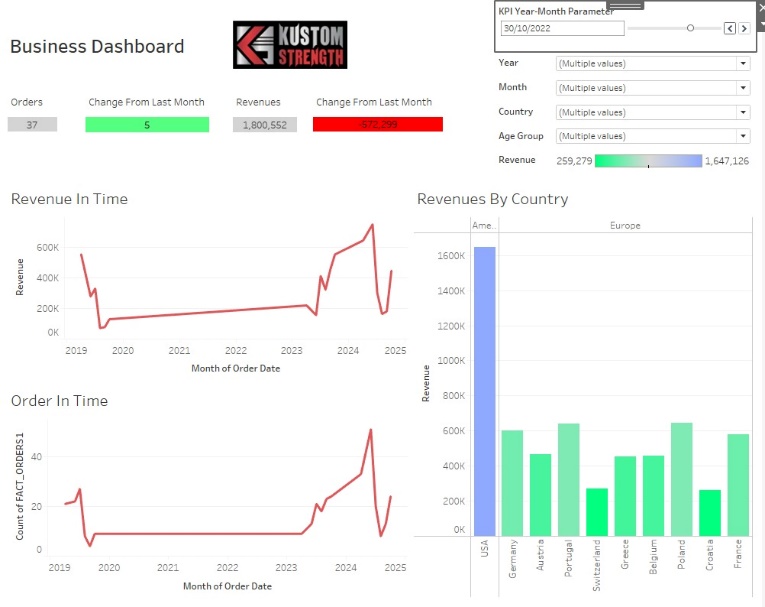


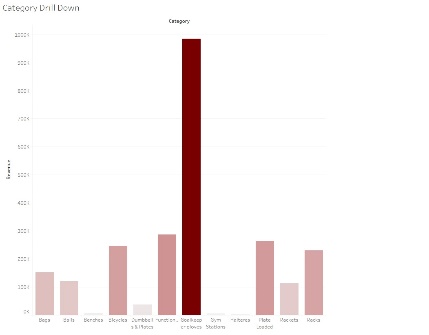
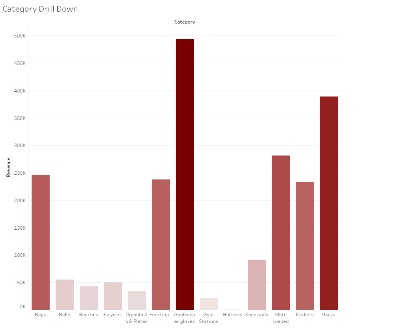
1. **לוח מחוונים:**

לוח מחוונים מיועד למנהלים בכירים. מדובר בכלי יותר דינמי המרחיב את הדו"חות הגרפיים באמצעות קישור לעולם של מדדי ביצועים ומאפשר למקבלי ההחלטות לקבל החלטות לקח שראו את התמונה הגדולה מזווית ראייה רחבה.

לוח המחוונים מכיל בחלקו העליון את כמות ההזמנות וסכום הרווחים לחודש הנבחר בפרמטר ה-KPI, ואת השינוי מהחודש הקודם. שינוי חיובי סומן בירוק, ושינוי שלילי באדום- בצורה אינטואיטיבית. בחלקו התחתון של הלוח מוצגים שני גרפים של הזמנות ורווחים לאורך הזמן, וגרף נוסף של הכנסות לפי מדינה. כמו כן, גרף הרווחים לאורך הזמן מאפשר "קידוח" ועל ידי לחיצה על חודש מסויים מוצג גרף הכנסות באותו החודש לפי קטגוריות. ניתן לפלטר את התרשימים לפי: שנים, חודשים, קטגוריית המוצר וקבוצת הגיל אליה משתייכים הלקוחות.



Drill Down :

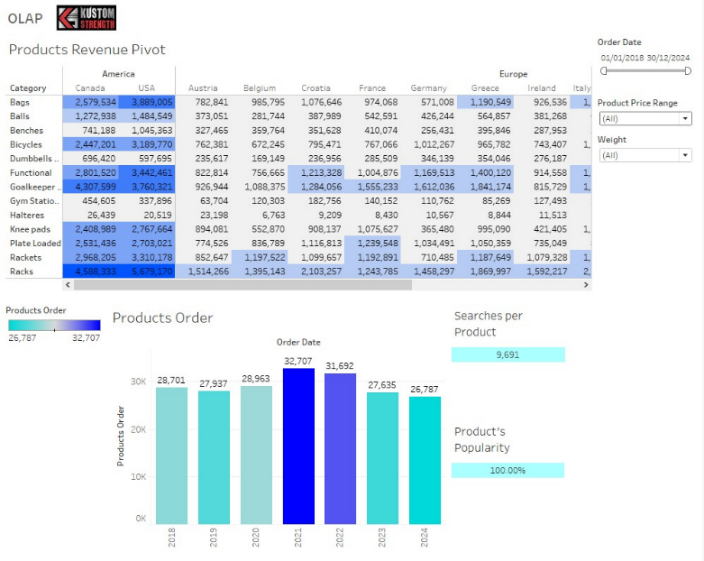


1. **כלי לתחקור אנליטי מקוון OLAP – On-Line Analytical Processing))**

כלי ה-OLAP הוא כלי תחקור גמיש ואינטראקטיבי, המאפשר לבצע תחקור מעמיק של הנתונים. ה-OLAP מיועד לאנליסטים שיש בידיהם את הכלים לתחקר בעצמם את הנתונים, לפי צרכי התחקור, בצורה דינמית וחופשית. ה-OLAP מעוצב ע"י משתמש הקצה (האנליסט לרוב) ולכן הוא מחליט את המורכבות הוויזואלית.

הכלי אותו בנינו מיועד לתחקור המוצרים שהחברה מוכרת. במרכזו נמצאת טבלת הציר, המתארת את הפילוח של הרווחים מהמוצרים לפי הקריטריונים השונים: בשורות ניתן לראות את ההיררכייה של המוצר לפי קטגוריה והמוצר עצמו, ובעמודות- ההכנסות שלו ע"פ היררכיה של יבשת – מדינה -עיר . כמו כן ניתן לראות גם גרף המתאר את פופולריות הפריט לפי כמות הפעמים שהוזמן ,וכן ייצוג טקסטואלי של מדד הביצוע הנוגע לפופולריות המוצר כאחוז מסך כל המכירות. בנוסף לכך,ניתן לראות את מספר הפעמים שחיפשו את המוצר. את כל המדדים בדו"ח ניתן לתחקר לפי משקל המוצר, טווח תאריכים שבו המוצר הוזמן וקטגוריית המחיר שלו.

ניתן לתחקר באמצעות הכלי תופעות חריגות כמו למשל אם נזהה בטבלה סוג מוצר/קטגוריית מוצרים שהרווח שלהם מהמכירות נמוך מאוד או נמוך משמעותית ביחס למוצרים האחרים בקרב פלח היבשות/מדינות/ערים הנבדקות- ונחשוד שמדובר בסוג מוצר שלא נמכר מספיק או שאינו רווחי מספיק- נוכל לבחון את כל הקריטריונים שלו ולנסות להבין מדוע, אולי תמחור המוצר אינו מתאים לערכו, אולי הוא לא מתאים לאיזור מסוים ועוד. כמו כן נוכל לבדוק עוד מדדים כגון מספר החיפושים שיעיד על התעניינות במוצר ואת מדד הפופולריות שלו שיעיד על הצלחה במכירותיו.



# יישומי BI&A וניתוח נתונים מתקדמים

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

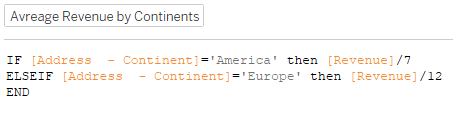
ביצענו שלושה ניתוחים סטטיסטים, לצורך הניתוחים חיברנו את תוכנת ה-R לטאבלו.

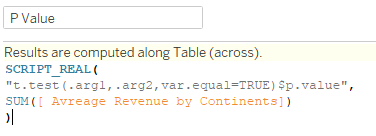
**מבחן T-test**

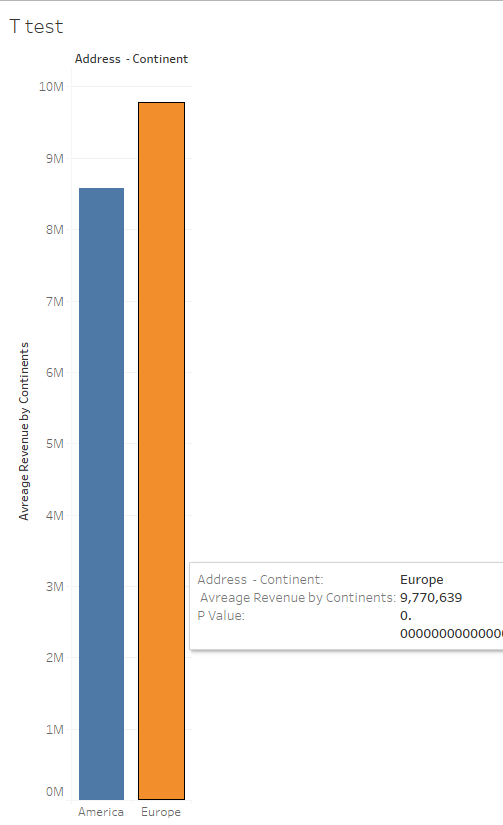
בעקבות תוצאות ה-OLAP רצינו לבדוק **האם ממוצע סך התשלומים על מוצרים לפי יבשות- שונה**. קבוצות היבשות לצורך מבחן זה הן : America, Europe.  
t-test מתאים במקרה זה כיוון שהוא עוזר לקבוע אם קיים הבדל מובהק סטטיסטי בין היבשות השונות.   
בחרנו להשוות בין יבשות, שכן כפי שראינו בOLAP, ראינו כי קיים הבדל משמעותי בין אירופה לאמריקה.

להלן השערות המבחן:

יצרנו משתנה חדש אשר מחלק את היבשות ומחשב את ממוצע ההכנסות לפי יבשת,הממוצע מחושב לפי מספר המדינות ביבשת (ביבשת אמריקה חישבנו לפי כמות הstates-).



  
באופן זה יכלנו לבצע את הבדיקה,  
 להלן חישוב ערך ה-P value:



כפי שניתן לראות הקשר הינו מובהק בר"מ 5%   
(P value < 0.05) **ולכן נדחה את השערת   
האפס ונגיד כי קיים הבדל בהתפלגות   
בין ממוצע ההכנסות בין היבשות.**

**מבחן רגרסיה ליניארית**

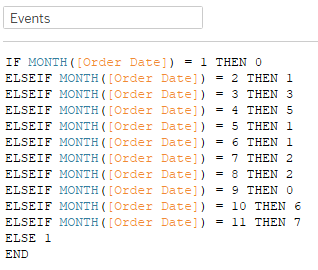
רצינו לבדוק **האם ישנו קשר בין עונות השנה (חגי ישראל ואירועים מיוחדים) בחודש לבין סך ההכנסות הכוללות ממוצרים בהזמנות.**רגרסיה ליניארית מתאימה במקרה זה מכיוון שהיא עוזרת לבחון קשר בין משתנה תלוי למשתנה בלתי תלוי ולהגיע לתובנות ומסקנות.בחרנו במבחן זה משום שרצינו לבחון האם משתנים אלו מסבירים את העלייה / ירידה בהכנסות הכוללות ממוצרים. להלן השערות המבחן:

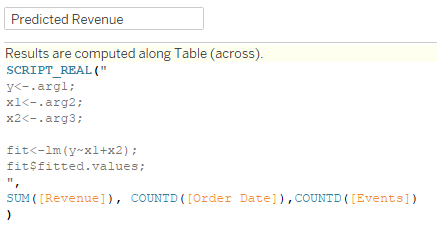
להלן משתני המודל:

Y – סך ההכנסות החזוי, משתנה מוסבר.  
1 X- סך ההכנסות ממוצרים בהזמנה, משתנה מסביר.

2X – כמות האירועים בחודש, משתנה מסביר.

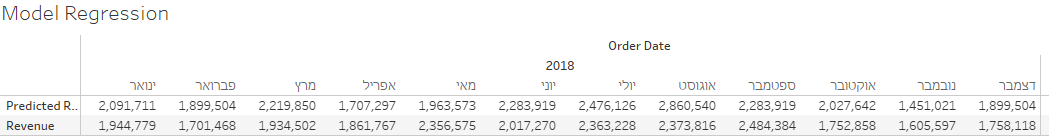
לצורך כך יצאנו שדה מחשב חדש אשר מציין את כמות האירועים לפי חודשים:



בנוסף יצרנו משתנה חדש אשר חוזה את ההכנסה הצפויה:

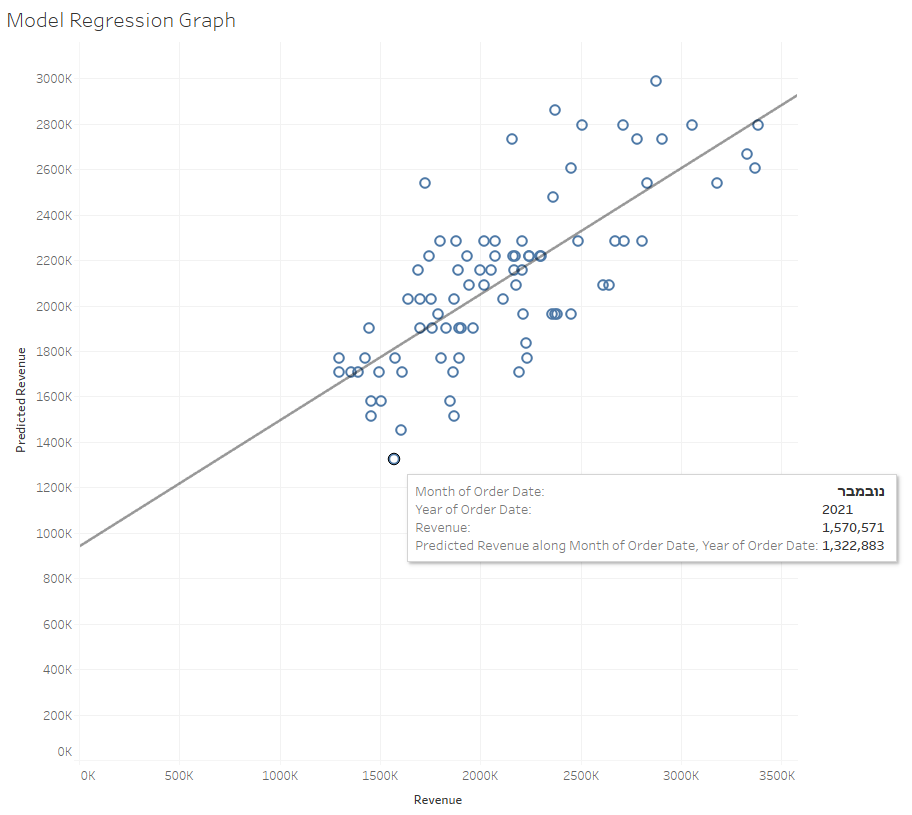
השדה בודק האם מספר ההזמנות וכמות האירועים המיוחדים בחודש מסבירים את סך ההכנסות.

הכנסנו את השדה המחושב לתוך טבלה אשר מציגה את סך ההכנסות על פי מדינות על מנת לבחון את יכולת החיזוי שלו. בחלק העליון של הרשומה מוצג הרווח האמיתי ומתחתיו מוצג הרווח החזוי.



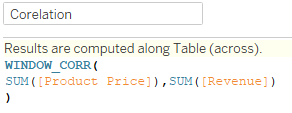
ניתן לשים לב שההפרשים אינם משמעותיים בין החיזוי לבין הערך בפועל – ומכאן ניתן ללמוד כי המשתנים הללו מסבירים את סך ההכנסות החזוי.

בדיקה נוספת שערכנו היא הכנסת הנתונים לתוך גרף. הגרף מציג את הסך ההכנסות החזוי לעומת סך ההכנסות האמיתי, עבור כל חודש בכל שנה. **ניתן לראות גם בגרף כי הפיזור אחיד ולכן יש קו רגרסיה ברור – ומכאן שהמודל מובהק.**

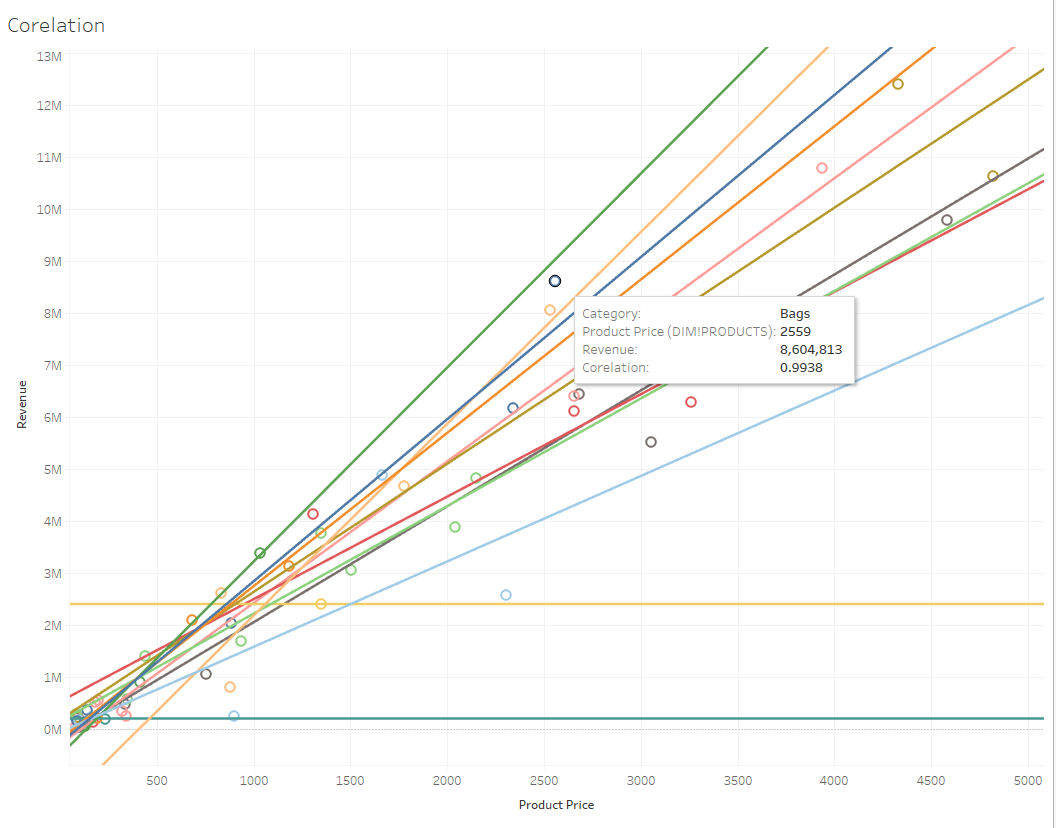


**מבחן קורלציה**

רצינו לבדוק **האם ישנו קשר בין מחיר המוצר לסך ההכנסות ממוצר**. בחרנו לבדוק זאת באמצעות קורולציה משום שזהו מדד אשר בוחן קיום קשר ואף כיוון (חיובי או שלילי) בין משתנים. באופן טבעי היינו מצפים שככול שנגדיל את מחיר המוצר כך יגדל גם סך ההכנסות ממוצר.

חישבנו את הקורלציה באופן הבא:  


הוספנו את הקורלציה כסימון על גבי גרף אשר מציג את סך ההכנסות ממוצר כתלות במחיר המוצר פר קטגוריה:



**ניתן לראות כי בכלל הקטגוריות נרשמה קורלציה גבוהה של 0.9938-0.791, כלומר קרוב ל1 ולכן יש קשר מובהק בין מחיר המוצר לסך ההכנסות ממנו**.

**יישום בינה ואנליטיקה עסקית "יוזם" (Proactive BI&A)**

בהמשך לניתוח העסקי היוזם, החלטנו לבצע קלאסיפיקציה (באמצעות עצי החלטה) של חיפושים שהפוטנציאל הגבוה שלהם להפוך להזמנות, כלומר שהמשתמש יחפש מוצר ויחליט להשלים את תהליך ההזמנה. המטרה היא למקסם את ההזמנות ולהגדיל את ההכנסות, באמצעות הצעת הטבות ללקוחות שסביר שיזמינו את המוצרים, כמו הנחות או דמי משלוח חינם, כדי לדרבן אותם להשלים את ההזמנה.

איחדנו את טבלת הלקוחות עם טבלת החיפושים. לאחר מכן כתבנו קוד ב-Python בו ביצענו Feature Extraction, נירמול נתונים ו-One-Hot Encoding על משתנים קטגוריאליים תוך התחשבות בזליגת מידע כדי למנוע Over-fitting. על הנתונים האלה הרצנו מודל עץ בעזרת ספריית sklearn כדי לראות מהו עומק העץ האידיאלי (גם כאן תוך התחשבות ב-Over-fitting) ובאמצעות 10-fold cross-validation. בעץ שקיבלנו ראינו כי במידה והלקוח חיפש עד 3 מוצרים, גילו קטן מ-54.3 שנים והוא ביצע פחות מ-13 חיפושים, אז ב-75.9% מהמקרים הוא ישלים את ההזמנה.

בחיבור ל-Tableau הרמנו שרת וירטואלי מקומי באמצעות tabpy על מנת שנוכל להריץ את הקוד Python מתוך ה-Tableau. מכיוון שאי אפשר להחזיר אובייקט של עץ החלטה אלא רק מספר, העלינו תמונה של העץ ל-sheet ב-Tableau לאחר שהרצנו את הקוד בפלטפורמה אחרת. הקוד מצורף בקובץ נפרד.

כעת נוכל באופן יזום לשלוח הצעות ללקוחות שסביר שיזמינו מוצרים, שנוכל לזהות אותם מראש ולא רק לתאר את ההזמנות שלהם בדיעבד. למעשה, הדבר נבדל מהכלים המסורתיים בכך שיכולנו לבצע סטטיסטיקה היסקית – הכללה מדגימות למטרות סיווג, שיסייעו לנו לבצע פעולות יזומות להגברת ההכנסות.

A diagram of a diagram

Description automatically generated

