**40205**

**BI PROJECT**

**מרצה:** אור פרץ

**שמות הסטודנטים:**

ליאור חזן 316417278

אביב אמינוף 208253526

חן אלפסי 209042670

סלע בר 208437582

לירן מילס 318595808

1. מטרת הפרויקט הינה חקירת עולם הספורט והקשר שלו אל בני אדם בהתמקד על כמות אימונים, כמות שעות, רכישת ציוד ומוצרי ספורט מגוונים, קבוצות גיל ותחומי עניין.

שאלת מחקר מס' 1- מהו הקשר בין גיל האנשים לבין תדירות האימונים.  
KPI – אם עוצמת הקשר מעל 0.7, יוגדר כהצלחה. KPI זה הוא מדיד מבחינה כמותית, כמו כן הוא ספציפי לפרמטרים מסוימים בדאטה סט, עוצמת קשר מעל 0.7 זה משהו בר השגה, מדד זה נותן לנו אינפורמציה רלוונטית לשאלת המחקר, לכן KPI זה מקיים את מודל הsmart.

KPI – אם תדירות האימונים בקרב המתאמנים בגילאים עד 18 גדולה מתדירות האימונים בקרב המתאמנים מעל גיל 18, נגדיר שיש קשר בין תדירות האימונים לבין הגיל. KPI זה מדיד כי ניתן לבצע השוואה בין קבוצות הגילאים, הוא ספציפי ורלוונטי לשאלת המחקר, כמובן הוא בר השגה ולכן הוא מקיים את מודל SMART.

KPI – נחשב כמות אימונים ממוצעת בשבוע לכל קבוצת גילאים ונשווה את התוצאות על מנת להסיק את הקשר בין גיל האנשים לתדירות האימונים. KPI זה הוא מדיד (ממוצע מספר האימונים בשבוע), בר השגה (אפשר לחשב אותו בקלות) רלוונטי וספציפי לשאלת המחקר ולכן מקיים את מודל smart.

שאלת מחקר מס' 2- האם יש קשר בין תדירות האימונים לרכישת ציוד FITNESS.

מטרה עסקית – עליית כמות מכירות מוצרי FITNESS.  
KPI – אם לפחות 60% מהמתאמנים, שתדירות האימונים שלהם לפחות 3 פעמים בשבוע, ירכשו ציוד FITNESS נגדיר שכמות אימונים גבוהה מובילה למוטיבציית רכישת ציוד FITNESS. KPI זה הוא מדיד ונותן אינפורמציה רלוונטית לשאלת המחקר, הוא רלוונטי וספציפי לשאלת המחקר, כמו כן הוא בר השגה מכיוון שניתן לחשב אותו מתמטית באמצעות מדדים סטטיסטיים ולכן הוא מקיים את מודל smart.

KPI – נחלק את כלל האוכלוסייה המתאמנת ל2 קבוצות- קבוצה אחת מתאמנת בין אפס לשני אימונים בשבוע, והקבוצה השנייה מתאמנת 3 פעמים ומעלה בשבוע. נחשב את יחס רכישות ציוד FITNESS בכל קבוצה ונערוך השוואה שעל פיה נדע אם יש קשר בין תדירות האימונים לבין רכישת הציוד. KPI זה הוא אפשרי למדידה, ספציפי ורלוונטי לשאלת המחקר ונותן אינפורמציה רלוונטית שעשויה להבהיר את התשובה עבור שאלת המחקר כמו כן נוכל לחשב זאת באמצעות חישובים מתמטיים בסיסיים ( יחס) ולכן מקיים את מודל הsmart.

KPI – נסכום את כמות הרכישות של ציוד FITNESS בקרב קבוצת תדירות אימונים בשבוע נכפיל באחוז הקבוצה מכלל האוכלוסייה ונסיק לפי זה באיזה קבוצת תדירות אימונים רוכשים הכי הרבה ובאיזה הכי מעט. KPI זה ניתן למדידה באמצעות חישובים מתמטיים בסיסיים שלמדנו שעשויים לתת אינפורמציה רלוונטית עבור טיב הקשר בין תדירות האימונים לרכישת ציוד FITNESS, בנוסף חישוב זה הוא בר השגה ויכול לתת ערך מספרי ספציפי לשאלת המחקר, לכן מקיים את מודל smart.

1. הגדרת Data Warehouse –
   1. בחרנו בסכמת Star מכיוון שבמקרה שיש קישור של יחיד ליחיד בין טבלאות כמו במקרה שלנו, הסכמה המתאימה ביותר היא סכמת Star, היא מאפשרת קישור ישיר פשוט בין שתי טבלאות על ידי שדה מפתח הקיים בשתי הטבלאות.
   2. הDatawarehouse של סכמת הכוכב המתאימה לדאטה סט שלנו מורכבת במרכז מטבלת הfitness analysis הכוללת את כל השדות, ממנה יש שתי טבלאות המחוברות בקשר יחיד ליחיד לטבלת fitness analysis באמצעות PK של ID.

הטבלאות המקושרות הן Personal Details וTrainer Details.

כאשר המשתמש ירצה לשלוף נתונים על פרטי המתאמן הוא יצטרך לשלוף באמצעות שאילתה שכוללת join, ואילו בעזרת סכמת הstar הוא יוכל לפנות לטבלה ייחודית (PersonalDetails) וספציפית וממנה לשלוף את הנתונים המבוקשים ללא Join.

Use Case Diagram עבור שני מקרים של הוצאת פלט דרך שאילתה ללקוח

תמונה שמכילה טקסט, תרשים, צילום מסך, עיגול

התיאור נוצר באופן אוטומטי

1. הגדרה ומימוש ETL –
   1. **Extract** - שלב חילוץ הנתונים כולל שליפת מערך הכושר ממקורותיו. מקורות אלו יכולים לכלול מסדי נתונים, גיליונות אלקטרוניים, סקרים או כל מאגר נתונים אחר שבו מאוחסנים נתוני הכושר. תהליך המיצוי אוסף את הנתונים הגולמיים הקשורים לתכונות הכושר, כגון שם, מין, גיל, העדפות פעילות גופנית, רמת כושר, הרגלי אכילה וגורמי מוטיבציה.  
      במקרה שלנו, הדאטה סט נאסף מאתר "קאגל", בצורת גיליון אקסל אלקטרוני שמכיל נתונים גולמיים הקשורים לעולם הכושר.
   2. **Transform** -לאחר חילוץ הנתונים, הם עוברים שינויים שונים כדי להבטיח את האיכות, העקביות והתאימות שלהם. השינויים הספציפיים שלנו למערך הכושר כוללים:

ניקוי ואימות: הסרת רשומות לא רלוונטיות או חלקיות, תיקון שגיאות או חוסר עקביות בנתונים ואימות שלמות הנתונים.

סטנדרטיזציה: הבטחת פורמטים ויחידות עקביים עבור שדות כגון גיל או זמן.

גזירת משתנים חדשים: חישוב מדדים נוספים על סמך נתונים קיימים.

טרנספורמציות אלו עוזרות להכין את הנתונים להמשך ניתוח ודיווח.

**Load**: הנתונים שעברו טרנספורמציה נטענים לאחר מכן אל היעד, שיכול להיות מחסן נתונים, מסד נתונים או כלי אנליטי. שלב הטעינה כולל:

הגדרת פורמט היעד: יצירה או עדכון של המבנה של מסד הנתונים של היעד או מחסן הנתונים כדי להתאים לנתוני הכושר שעברו שינוי. זה כרוך בהגדרת הטבלאות, העמודות וסוגי הנתונים המתאימים.

מיפוי: מיפוי שדות נתוני הכושר שהשתנו לתכונות המתאימות בפורמט היעד. לדוגמה, מיפוי השדה "שם" לעמודה "שם" בטבלת היעד.

טעינת הנתונים: ביצוע פעולת הטעינה בפועל, כגון הכנסת הנתונים שעברו טרנספורמציה למסד הנתונים היעד או למחסן הנתונים.

על ידי ביצוע תהליך ETL זה המותאם למערך הכושר המסופק, ניתן לחלץ את הנתונים, לשנות אותם על ידי יישום ניקוי, סטנדרטיזציה ולאחר מכן לטעון ליעד יעד להמשך ניתוח, דיווח או קבלת החלטות הקשורות לכושר. ותובנות הקשורות לתרגיל.

1. **SQL –**

SELECT age, often, CORR(often, AVG(often) OVER (PARTITION BY age)) as corr

FROM fitness ,Trainerdetails;

ORDER BY corr;  
 **שאילתא אשר מציגה את הקורלציה בין המשתנים גיל ותדירות אימונים, ומוציאה כפלט את הגיל התדירות ואת הקורלציה ומסדרת לפי קורלציה בסדר יורד.**  
  
SELECT COUNT(often) FILTER (WHERE often = 'Never' AND equipment = 'yes') OVER () as countNotEx

FROM fitness;  
  
ש**אילתא אשר מוציאה כפלט את כמות האנשים אשר אינם מתאמנים אך למרות זאת רכשו ציוד.**  
  
SELECT COUNT(often) FILTER (WHERE often = '3 to 4 times in a week' OR often = 'every day' AND equipment = 'yes') OVER () / COUNT(often) FILTER (WHERE equipment = 'yes') OVER () as totalCustomers

FROM fitness;  
  
**שאילתא אשר מבצעת סכימה של מספר האנשים אשר רכשו ציוד במידה והם מתאמנים בין 3-4 אימונים או כל יום ומחלקת במספר הרכישות הכולל על מנת לבדוק האם יש עמידה במדד אשר הוצב בתרגיל 1**  
SELECT RANK() OVER () as "rank order", COUNT(\*) FILTER (WHERE "Do you exercise \_\_\_\_\_\_\_\_\_\_\_ ?" = 'alone' AND "Have you ever purchased a fitness equipment?" = 'yes') OVER () as "Bought alone", COUNT(\*) FILTER (WHERE "Do you exercise \_\_\_\_\_\_\_\_\_\_\_ ?" = 'with a group' AND "Have you ever purchased a fitness equipment?" = 'yes') OVER () as "Bought for group"

FROM fitness;

**שאילתא אשר סוכמת את מספר האנשים שרכשו ציוד ומתאמנים לבד ומספר האנשים שרכשו ציוד ומתאמנים בקבוצה ומציגה את הסכומים של כל אחת מהקבוצות כרכש לבד ורכש עבור קבוצה**  
SELECT COUNT(\*) FILTER (WHERE trainerDetails.gender = 'female' AND fitness."Would you say you eat a healthy balanced diet?" != 'Yes' AND fitness."What prevents you from eating a healthy balanced diet, If any?" = 'Temptation and cravings') OVER () as mustHaveSweets

FROM trainerDetails;

JOIN fitness f;

ON trainerDetails.id = fitness.id;

**שאילתא הסוכמת את מספר הבנות אשר נופלות בדיאטה בגלל דחף למתוק.**

SELECT "Age",COUNT(\*) FILTER (WHERE fitness."What motivates you to exercise?" = 'I want to lose weight' AND fitness.equipment = 'yes') OVER (PARTITION BY "Age") as "bought equipment for losing weight"

FROM(SELECT CASE WHEN trainersDetails."Your age" BETWEEN '15 to 18' THEN '15 till 18'

WHEN trainersDetails."Your age" BETWEEN '19 to 25' THEN '19 till 25'

WHEN trainersDetails."Your age" BETWEEN '26 to 30' THEN '26 till 30'

WHEN trainersDetails."Your age" BETWEEN '31 to 40' THEN '31 till 40'

WHEN trainersDetails."Your age" BETWEEN '40 and above' THEN '40+'  
END as "Age"

FROM trainersDetails

JOIN fitness

ON trainersDetails.id = f.id) as age  
 **שאילת שמראה לפי התפלגות גילאים את כמו הרכישות של ציוד עבור אנשים אשר מתאמנים בשביל להוריד משקל**

1. המסקנות שלנו מניתוח הנתונים הם שספורט הינו תחום מאוד חשוב בחיי האדם, לשמירה על הבריאות, לחץ דם, מצב נפשי מאוזן ושמחת חיים. מצאנו כי אנשים בני 19-25 מתאמנים יותר משאר קבוצות הגיל, ובפרט באימונים אינטנסיביים, מצאנו שקבוצות הגיל הגבוהות יותר מוצאות עצמן בענפי ספורט רגועים יותר עם דופק נמוך. מצאנו כי יש קשר בין דיאטה לבין עיסוק ספורט, וכי נשים "נופלות" בדיאטה יותר מאשר גברים. לסיכום, יש קשר בין רכישת ציוד ספורט לבין עיסוק בספורט, ככל שמתעסקים יותר בספורט, נוצרת אהבה לתחום ובכך הרצון לקנות ציוד רלוונטי עולה.