הכנת נתונים עם Knime, חלק ב'

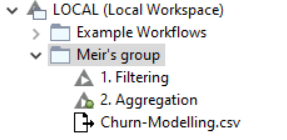
נמשיך ללמוד טכניקות הכנת נתונים לאנליזה דרך למידה מסרטונים תוך כדי ביצוע תרגילים קצרים ושיומש ב- [Knime](https://www.knime.com).

# הכנת נתונים לאנליזה (המשך)

## צבירה (Aggregation)

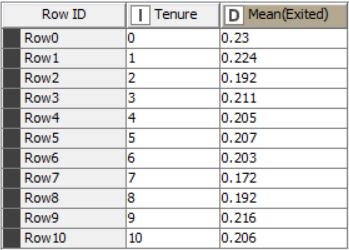
למדתם על צבירה בקורס בסיסי נתונים. במסגרת עבודה עם בסיסי נתונים במודל היחסים בעזרת שפת שאילתות SQL, מבצעים צבירה בעזרת שימוש בפונקציות צבירה בקטע Select (או Having) והוספת קטע Group By לפי הצורך. הרעיון של צבירה מוסבר בסרטון [1](https://youtu.be/bDwF-TOMtWw) (1:47). בעזרת ה- סרטונים [2](https://youtu.be/JQ-OWMt48ew) (7:09) ו- [3](https://youtu.be/i2LIKPOgNoM) (3:33), תיצרו את הטבלאות הבאות:

\*מומלץ לייצר בשלב זה Workflow (פרויקט) חדש כדי לשמור על סדר (וכן לכל נושא) , למשל כך:



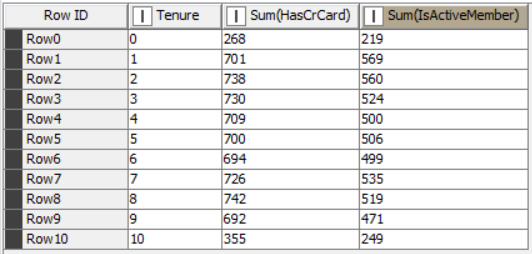
1. עמודות: tenure, הסתברות של עזיבה עבור ה- tenure ההוא. (הבהרה: שימו לב שעזיבה מקבלת ערך 1 כאשר יש עזיבה ואפס כשלא, ולכן הסתברות העזיבה עבור tenure מסוים שווה לממוצע של ערכים בעמודת exit עבור אותו tenure).

**פלט לבדיקה**:



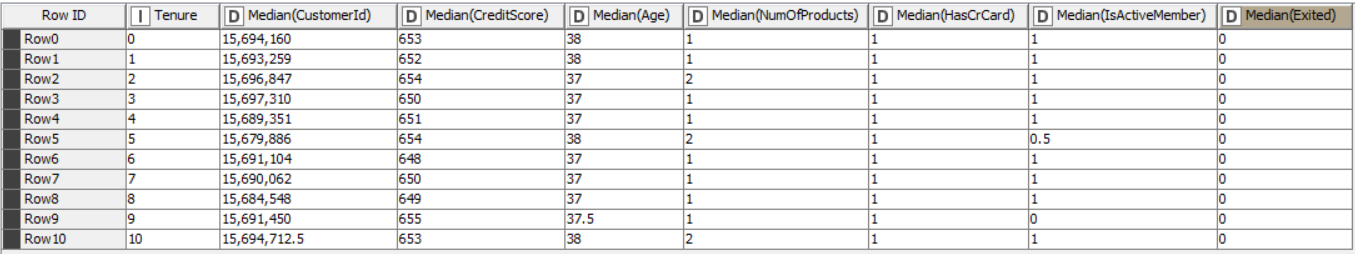
1. עמודות: tenure, עמודה מחושבת עבור כל עמודה בטבלה המקורית ששמה מתחיל ב- Has או Is (חובה להשתמש ב- regular expressions). הערך המחושב הוא כמות הלקוחות עם 1 בעמודה בטבלה המקורית עבור ה- tenure ההוא.

**פלט לבדיקה**:



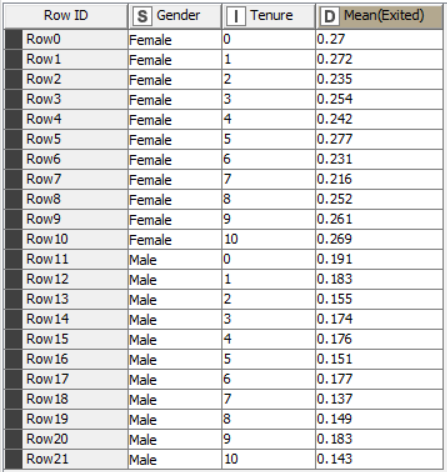
1. עמודות: tenure, עמודה מחושבת עבור כל עמודה עם טיפוס מספר שלם בטבלה המקורית. הערך המחושב הוא חציון הערכים בעמודה בטבלה המקורית עבור ה- tenure ההוא.

**פלט לבדיקה**:



1. עמודות: tenure, gender, הסתברות של עזיבה עבור tenure ו- gender ההם.

**פלט לבדיקה**:



## Pivoting

Pivoting – הרחבת הרעיון של צבירה. בצבירה, כל שורה מקבילה לקבוצת שורות עם ערכים זהים בעמודות מסוימות (group columns). ב- pivoting שמים ערכים שונים של עמודות מסוימות (pivot columns) בתור כותרות העמודות. עבור כל שורה ועמודה, הערך ב- תא בתוך הטבלה מחושב בעזרת פונקציית צבירה. תבצעו בעזרת ה- סרטונים [1](https://youtu.be/zo9YdH9kgKQ) (6:43) ו- [2](https://youtu.be/WXt7iiu9c98) (4:37):

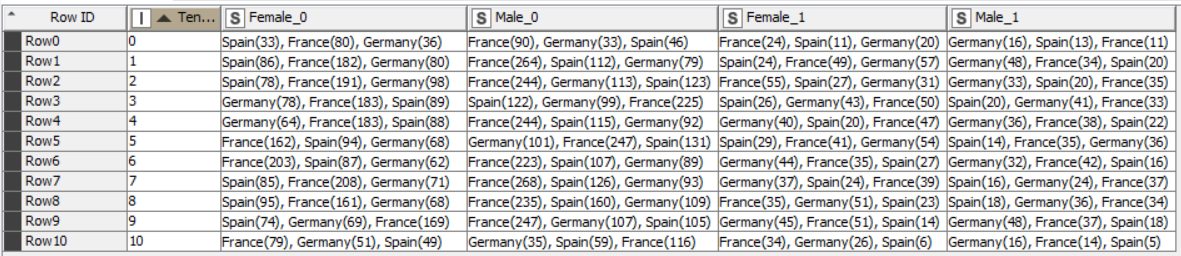
1. תיצרו את הטבלה הבאה (כולל שמות העמודות) המראה עבור כל מספר שנים בעמודת tenure וכל מגדר, את הסתברות העזיבה. שימו לב ש- יש לבחור ב- Pivot Table מהתפריט כדי לצפות בטבלה ולא באופציה האחרונה.

**פלט לבדיקה**:

### 

1. תיצרו את הטבלה הבאה המראה עבור כל מספר שנים בעמודת tenure וכל מגדר וסטטוס עזיבה, את מספר הלקוחות מכל מדינה.

**פלט לבדיקה**:



להרחבת ידע: [3](https://youtu.be/1Dilumi6X2I) (3:44).

## חיבור (Join)

למדתם על Join בקורס בסיסי נתונים. סרטון [1](https://youtu.be/6BigLM6vbhs) (3:27) מזכיר על Join ועל הסוגים השונים שלו: Inner Join ו- Left/Right/Full Outer Join. תצפו ב- סרטון [2](https://youtu.be/X8q14s2kjIU) (5:24). תיצרו טבלה ש- יש בה עמודות:

1. Surname
2. Gender
3. Age
4. Exited
5. Prob. Exited – הסתברות עזיבה עבור מגדר וגיל של הלקוח

רמז: ניתן ליצור קדקוד(י) עזר.

**פלט לבדיקה** (תחילת הטבלה, ממוין לפי surname):

