הכנת נתונים עם Knime, חלק ג'

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)

## שרשור (Concatenation)

שרשור מקביל ל UNION של SQL. תצפו ב- סרטונים [1](https://youtu.be/VzH2lHbDAg0) (1:45), [2](https://youtu.be/ku6SyEZ1Pv8) (4:30) ו- [3](https://youtu.be/PYPzcQ6hgC8) (0:42). תיצרו את הטבלאות הבאות:

1. רשימה (כלומר, טבלה עם עמודה אחת) של שמות לקוחות עם 10 שנות tenure.
2. רשימת לקוחות ש- יש להם או 10 שנות tenure או שהם מתאימים לכל התנאים בסעיף 3.1.1 מחלק א' של התרגיל (ניתן להעתיק קדקודים משם). שימו לב ש- לקוח יכול להופיע ברשימה פעם אחת בלבד. (בדיקה: צריכים להיות 530 לקוחות בטבלת פלט)

## מניפולציות של נתונים (Data manipulations)

תצפו ב- סרטונים [1](https://youtu.be/mJrBXmLQ4ko) (10:40) ו- [2](https://youtu.be/_3EJ5cYdhwI) (4:56). תיצרו את הטבלה עם ה- שינוים הבאים מהטבלה המקורית:

1. עמודה בשם status שתחליף עמודת exited. הערך שבעמודת status יהיה:
   * Old client עבור לקוח עם tenure של 5 שנים ומעלה שלא עזב.
   * New client עבור לקוח עם tenure של פחות מ 5 שנים שלא עזב.
   * Former client עבור לקוח שעזב.
2. CreditScore מנורמל ע"י ניכוי של CreditScore מינימלי (לא ניתן לקבוע את המספר הזה בנוסחה) פחות אחד ובסקלה לוגריתמית שלימה. לדוגמה, CreditScore מינימלי בנתונים שלנו הוא 350. עבור לקוח עם Credit score של 400, הערך המנורמל הוא 51 (שזה 400-(350-1)). לוגריתם של 51 בבסיס 2 הוא 5.67. נזרוק את החלק האינו שלם ונכניס ערך של 5.

טיפוס העמודה (לשונית Spec - Columns) צריך להיות Integer.

רמזים:

* + ניתן להשתמש בפונקציות הנמצאות בקטגוריית constants.
  + ניתן לחשב לוגריתם בבסיס 2 של מספר כלשהו x כך: ln(x)/ln(2).

1. שמות המכילות סימן שאלה (כגון בשורה עם RowID של 10) מוחלפות בערך NULL (סימן שאלה אדום בתצוגה). [קישור לרמז](https://stackoverflow.com/q/53642200). רמז: נוסף: יש להשתמש ב- \\? כדי להתאים סימן שאלה ([קישור להסבר](https://stackoverflow.com/q/25145793)) בביטוי רגולרי.

# הרחבת ידע (אופציונלי)

למדנו על שימוש בסיסי בתוכנת Knime. הקישורים הבאים יעזרו לכם להרחיב ידע באופן עצמאי:

1. קיים קדקוד הכנה ל Time Series Analysis. תולכו ללמוד עליו [כאן](https://www.knime.com/blog/all-you-need-is-the-lag-column-node) ו[כאן](https://youtu.be/pR_7pIEqW-c).
2. [הסרטון המקושר](https://www.youtube.com/watch?v=n8HbUUc51fc) מציג שימוש ב Knime עבור כריית מידע.