**כריית נתונים ב- R**

***מיני תרגיל מספר 6***

**שאלות:**

בתרגיל זה נתשמש במסד traindata.csv.

1. קרא את הקובץ לתוך מסד ב- R, והסר רשומות עם ערכים חסרים במשתנה income
2. צור עמודה חדשה במסד בשם "HighIncome" המקבלת את הערך 1 אם ההכנסות גבוהות מהממוצע, ו-0 אחרת.
3. פצל את המסד ל- training ו- validation
4. חזה את HighIncome בעזרת המודלים הבאים
   1. עץ החלטה
   2. רשת נוירונים
   3. רגרסיה לוגיסטית (לא למדנו!)
5. חשב את ביצועי המודלים בעזרת מדד AUC.