הבדיקה תהיה מדגמית. את השאלות הבאות נבדוק בצורה מלאה ובשאר השאלות נבדוק רק קיום של תשובה וגרף (כשזה רלוונטי).

* **חלק 1 (10 נקודות) - אקספלורציה**בחלק זה מציגים סטטיסטיקות על המידע כדי להבין "איך הוא נראה". נבדוק את קיום הרכיבים הבאים:
  + הדפסת מספר רכיבי קשירות (צריך להיות 213)
  + היסטוגרמת גדלים של רכיבי קשירות (להשוות לגרף בפתרון)
* **חלק 2 (30 נקודות) - פיצול רגיל**
  + אימון על אובייקט אחד (20 נקודות) - בחלק זה מאמנים מודל על הדאטא הנתון ומשווים אותו למודל פשוט שמחזיר תמיד בתור תשובה את ממוצע ה-target על כל הדאטא. נוודא את קיום הרכיבים הבאים:
    - פיצול רנדומלי רגיל, למשל Train\_test\_split
    - אימון מודל (fit), למשל RandomForestRegressor
    - השוואה בין ה-Random Forest ל-Baseline, למשל הדפסה של MSE של שניהם על train ועל validation
  + אימון על זוגות (10 נקודות)
    - לבדוק קיום של תשובות, אמור להיות דומה לאימון על אובייקט אחד
* **חלק 3 (40 נקודות) - פיצול מיוחד 1**
  + בחלק זה מבצעים פיצול מיוחד של הדאטא (ל-train ו-validation), על פי המבנה שלו בתור גרף ולא בצורה הרנדומלית הרגילה (מפצלים ככה שבכל חלק יהיו רכיבי קשירות שלמים). נבדוק את קיום הרכיבים הבאים:
    - קיום קוד שיוצר את הפיצול (10 נקודות)
    - היסטוגרמות של ה-target או של הגודל של רכיבי הקשירות על ה-train ועל ה-validation (אם עשו השוואה של ההתפלגויות בדרך הגיונית אחרת, זה גם מתקבל).   
      (15 נקודות)
    - השוואה של ה-loss בין ה-train לבין ה-validation. תשובה פתוחה שמתייחסת להבדל ומסבירה ממה הוא נובע - התפלגות הגדלים של רכיבי הקשירות שונה בין ה-train ל-validation בגלל אופן החלוקה. (15 נקודות)
* **חלק 4 (20 נקודות) - פיצול מיוחד 2**
  + קיום התשובות, אמור להיות יחסית דומה לחלק 3.