

## פרוייקט סיום עיבוד, ניתוח והצגת נתונים 88-6970-01

**אופן הגשה:** את הפרוייקט צריך להגיש אך ורק למייל [uziel.guy@gmail.com](mailto:uziel.guy@gmail.com) חובה לציין בכותרת את מספר הקורס. ההגשה תעשה בקובץ מחברת ג'ופיטר שתכלול את קוד הפרוייקט ואת התיעוד.

**תאריך הגשה:** 1.5.24

**מטרת העבודה:** יישום המתודות השונות שנלמדו במהלך הקורס.

**תיאור העבודה:** במהלך הקורס למדנו את החלקים השונים בעבודת הדאטא בשלב האקספלורציה והצגת המסקנות. בפרוייקט זה כל סטודנט יממש את השלבים השונים שנלמדו על דאטאסט נתון:

**דאטאסט:** הדאטאסט עוסק בגידול עגבניות שרי ובנתונים שונים המשפיעים על תפוקת העגבניות.

### **נקודות מנחות לפדרוייקט:**

1. סטטיסטיקות כלליות על הדאטאסט.
2. כמה משתנים קטגוריאליים וכמה משתנים נומריים יש?
3. תאור המשתנים הקטגוריאליים.
4. תאור המשתנים הנומריים.
5. האם יש ערכים חסרים? אם כן באלו משתנים?
6. טיפול בערכים חסרים.
7. איתור וטיפול בערכי קיצון במידה ויש.
8. נרמול ערכים נומריים.
9. חיפוש קורלציות בין משתנים בצורה ויזואלית ובצורה נומרית בדגש ליחס למשתנה tomato\_yield.
10. ויזואליזציה של ההתפלגויות של המשתנים השונים ביחד ולחוד.
11. ויזואליזציות של משתנים ביחס לtomato\_yield.

### **הערות נוספות:**

1. אתם רשאים לייבא חבילות נוספות מעבר לאלה שנלמדו בקורס (כל עוד אתם כוללים את תהליך ההתקנה בהערה).
2. בסוף כל חלק וגרף חובה לכתוב מסקנות וסיכום של התוצאות המרכזיות.
3. דגש רב ינתן לתיעוד העבודה ולמסקנות שיופקו. ולכן הקפידו לתאר בפירוט למה בחרתם בסוג מסוים של ניתוח ולא בניתוח אחר.
4. העבודה היא בזוגות.