**מדעי הנתונים – תרגיל בית 4**

**תיאור מבנה הקוד**

**מחלקת Main:**

המחלקה הראשית אשר אחראית על בניית הGUI והרצת הלולאה הראשית של התוכנית אשר מקבלת פעולות מהמשתמש ודואגת לקרוא לפונקציות המתאימות.

שיטות:

**\_\_init\_\_-**מאתחלת את ממשק המשתמש, מגדירה את כל השדות והכפתורים השונים.

**open\_file-** פונקציה אשר בוחרת נתיב לקובץ הנתונים ושומרת אותו, בודקת שהקובץ הנבחר הוא קובץ אקסל ושהוא אינו ריק.

**pre\_processing-** פונקציה אשר שולחת את קובץ הנתונים למחלקה שאחראית על הכנת הנתונים.

**cluster-** פונקציה אשר שולחת את הנתונים אחרי ניקוי והכנה למחלקה Clustering אשר אחראית לבצע את אלגוריתם k-means וליצור את הפלט הרצוי.

**input\_tests –** פונקציה שאחראית על בדיקות קלט מהמשתמש. בודקת שכל הקלט שהמשתמש הזין תקין במידה ולא שולחת הודעה מתאימה.

**מחלקת PreProcessing:**

מחלקה אשר מקבלת את הנתונים כמו שהם ואחראית על ניקוי הנתונים: השלמת ערכים חסרים, ביצוע נרמול לערכי קובץ הנתונים וקיבוץ הנתונים.

שיטות:

**\_\_init\_\_-** קליטה של הנתונים ואתחול השדות של המחלקה.

**clean –** מבצעת את השלמת הנתונים החסרים באמצעות הממוצע, נרמול שאר הנתונים וקיבוץ הנתונים לפי התכונה country ומחיקת עמודת הyear.

**מחלקת Clustering:**

מחלקה אשר אחראית על חלוקת הנתונים לאשכולות באמצעות אלגוריתם k- means.

שיטות:

**\_\_init\_\_-** קליטה של הנתונים ואתחול השדות של המחלקה.

**cluster –** הפונקציה הראשית של המחלקה אשר מפעילה את האלגוריתם K means על סט האימון, משתמשת בפונקציה חיצונית לבניית מודל k- means. מבצעת הפיכת שמות המדינות לשמם המקוצר ובאותיות גדולות על מנת לבצע את ההכנסה למפה וקוראת לשתי פונקציות נוספות שיוצרות את התרשימים המבוקשים.

**draw\_scatter\_image-** פונקציה שמייצרת תרשים scatter ושומרת בנתיב העבודה.

**draw\_map\_image-** פונקציה שמייצרת horopleth map עבור המדינות בקובץ הנתונים ושומרת בנתיב העבודה.