

תכנות הנדסי בשפת פייתון – תרגיל בית 6

מתרגל אחראי: אלעד סופר

הנחיות כלליות:

- מועד אחרון להגשה: כמפורסם בתיבת ההגשה ב-Moodle.
- מטרת התרגיל הינה לתרגל כתיבה, ולרכוש מיומנות בכלים שנלמדו עד כה בכיתה.
- קראו את העבודה מתחילתה ועד סופה לפני שאתם מתחילים לפתור אותה. ודאו שאתם מבינים את כל המשימות.
- רמת הקושי של המשימות אינה אחידה.
- את התרגיל יש לפתור לבד!
- בתיבת ההגשה במערכת ה-VPL ישנו קובץ שלד לכתיבת הקודם שלכם. עבור כל משימה כתבו הפתרון שלכם במקום המתאים למשימה. **בעבודה זאת סופק לכם קובץ שלד בודד: ex6.py**
- יש לצרף למשימה קובץ pdf המכיל את הגרפים והתשובות כפי שהתבקשתם, ובנוסף את קובץ הפייתון המלא בפורמט הבא:
`Ex6_<id>.py` עם תעודת הזהות שלכם במקום הרלוונטי לדוגמא: `ex6_123456789.py`.
- אין למחוק שום קטע קוד הנמצא בשלד. עליכם רק להוסיף את הפתרון שלכם בתוכו.
- אין לשנות את שם/שמות ה-Requested Files.
- השאלות יבדקו באופן אוטומטי. הפלט שעליכם להחזיר בכל תרגיל צריך להיות בדיוק כפי שנדרש. כמו כן, באופן אקראי יבדקו גם עבודות באופן ידני.
- כאשר תבוצע בדיקה ידנית, תתבצע גם בדיקת Readability - שימו לב שאתם משתמשים בשמות משתנים אינפורמטיביים וכותבים הערות בכל סעיף.
- בדיקה עצמית: כדי לוודא את נכונותן ואת עמידותן של הפונקציות לקלטים שונים, בכל שאלה הריצו אותן עם מגוון קלטים: אלה שמופיעים בדוגמאות וקלטים נוספים עליהם חשבתם. וודאו כי הפלט נכון. הבדיקה תתבצע על מגוון דוגמאות ולא בהכרח אלה שיינתנו פה.
- ניתן להשתמש בחומר הנלמד עד לפרסום העבודה ורק בחומר הזה.
- אין להשתמש בחבילות או מודולים חיצוניים (כדוגמת math) למעט מקרים שבהם צוין אחרת במפורש.
- במידה ולא צוין אחרת, **אין** להניח את נכונות הקלט.
- משקל כל שאלה הוא זהה.
- במידה ולא עניתם על סעיף מסוים, נא מלאו את הפונקציה/שיטה בכל מקרה על מנת שהקוד שלכם יצליח לעבוד.

במטלה זו נקרא את קובץ ההצבעה בבחירות לכנסת ה-24 (מרץ 2021) ונציג את דפוסיה ההצבעה בישובים.

לפני ביצוע המשימה וודאו כי

1. התקנתם את המודולים pandas, matplotlib, numpy.

2. הורידו את קובץ הנתונים `votes_per_city.csv`, והגדירו את הנתוב אליו למשתנה `"data_path"`.

בשלב הנתון התבצעה טעינה של אובייקט DataFrame מקובץ הנתונים votes_per_city.csv, אשר מכיל את כלל נתוני ההצבעות, שימו לב כי הטבלה כוללת מידע אשר לא רלוונטי להצבעות. עמודות כגון: סמל, ועדה, "בזב" וכו'.

הטבלה המקורית נטענה לאובייקט בשם "df_2021_raw".

כך שכל שורה מייצגת את הנתונים של יישוב/עיר וכל עמודה מייצגת קטגוריה כלשהי כך שהאיבר $M[i, j]$ מייצג את כמות ההצבעות של תושבי היישוב ה- i עבור הקטגוריה ה- j .

לנוחות, טבלה מקוצרת נטענה לאובייקט בשם `df_2021`. בטבלה זאת הוסרו מהטבלה העמודות הלא רלוונטיות כך שהטבלה תכיל אך ורק את עמודת היישובים ואת העמודות אשר מייצגות **מפלגות בלבד**.

- ניתן לראות תצוגה של חמשת השורות הראשונות של אובייקט ה"df_2021_raw" מטה.
- בכל סעיף בו התבקשתם להציג היסטוגרמה או גרף כלשהו, צרפו את הגרף לpdf עם כותרת תחת השאלה המתאימה.

[illegible]

- ניתן לראות תצוגה מקוצרת של חמשת השורות הראשונות של האובייקט "df_2021" מטה.

שם ישוב	אמת	ב	ג	דעם	ז	ח	ט	י	שס	ת
אביבים	47	0	0	3	0	...	0	0	1	8
אדירים	1	7	4	0	0	...	0	0	61	6
אנשי	72	7	0	3	0	...	0	0	0	15
אחווה	14	4	0	0	0	...	0	0	2	8
איבים	0	4	0	0	0	...	0	0	1	0

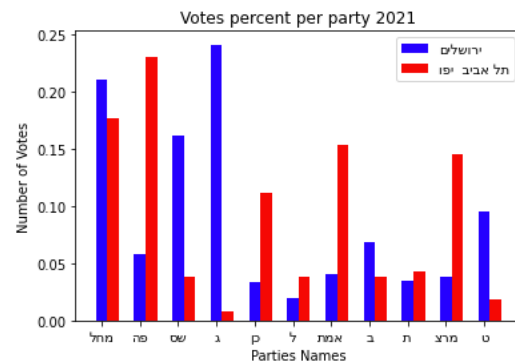
- שימו לב שהטבלאות נטענו כמשתנים גלובליים (ברמת המודול), ולכן יש לכם גישה מלאה אליהם מכלל הפונקציות.

1. מצאו את העמודות המציינות את מספר הקולות הפסולים ואת מספר כלל המצביעים בכל ישוב והציגו היסטוגרמה של **אחוז** הקולות הפסולים בישוב על פני כל הישובים.

על מנת ליצור את ציר הא, השתמשו ב-50 תאים שווי מרחק בין 0 לבין אחוז הקולות הפסולים המקסימלי. בנוסף, כתבו פונקצייה אשר מוצאת את שלושת הישובים עם **אחוז** הקולות הפסולים המקסימלי והדפיסו את שמם ואת האחוז.

- יש לצרף לpdf את ההיסטוגרמה ואת שמות היישובים ואת אחוז הקולות הפסולים בכל יישוב.

2. כתבו פונקציה אשר מוצאת את שתי הערים "הבינוניות" בעלות מספר המצביעים הגבוה ביותר. עיר בינונית מוגדרת כעיר בה מספר המצביעים הינו בטווח של 5,000-50,000.
- לאחר מכן הפונקציה תציג bar-plot של התפלגות ההצבעות עבור אותן זוג ערים כך שאך ורק המפלגות שעברו אחוז החסימה (מעל 3.25% הצבעה כלל ארצית) יוצגו בגרף.
- bar plot יוצג באופן הבא:
 - ציר הא יכיל את שמות המפלגות וציר הע יכיל את אחוז ההצבעה בעיר.



3. עבור שני ווקטורי התפלגות הצבעות המתארים את היישוב הנ ו היישוב הא בהתאמה,

$$P^j = [p_1^j, \dots, p_n^j], P^k = [p_1^k, \dots, p_n^k]$$

נגדיר את האיבר הכללי של ווקטור התפלגות באופן הבא:

$$p_i^j = \frac{\text{voters in the } j_{th} \text{ city to the } i_{th} \text{ party}}{\text{All the voters in the } j_{th} \text{ city}}$$

- נבחין כי המונה מייצג את ההצבעות ביישוב כלשהו למפלגה ה, והמכנה את סה"כ הקולות באותו היישוב.

נגדיר את מדד המרחק הריבועי, MSE, בין שני ווקטורי הסתברות להיות:

$$MSE(P^j, P^k) = \sum_{i=1}^n (p_i^j - p_i^k)^2$$

- (א) מצאו את שני הישובים המייצג ביותר את אחוז ההצבעה בכלל ישראל - כלומר הישובים שהתפלגות ההצבעה בהם (על פני כל המפלגות, כולל אלו שלא עברו את אחוז החסימה) היא הקרובה ביותר להתפלגות ההצבעה הארצית עבור מדד זה (ממזער את המרחק הריבועי מהתפלגות ההצבעה הארצית) והציגו bar-plot של הישוב ביחד עם התפלגות ההצבעה הארצית
- (ב) חזרו על א' עבור שני הישובים שהתפלגות ההצבעה בהם היא הרחוקה ביותר מהתפלגות ההצבעה הארצית.

הערות מיוחדות:

- עצבו את הגרפים כנדרש, משמע, תנו כותרת לצירים וכותרת כללית לגרף. תנו כותרות לעמודות במידת הצורך. הוסיפו legend המתאר כל קו במידת הצורך. שימו לב לגבולות הצירים. השתמשו בצבעים וכו' כדי להדגיש הבדלים.
- הגישו את קובץ הפייתון יחד עם pdf בתיבת ההגשה במודל.