

מעבדה בסטטיסטיקה 52568 - 2020-21, מטלה 2. להגשה והצגה ב-1.11

תיאור המשימה:

במשימה זו נציג את הקשרים בין הצבעה למפלגות שונות בבחירות לכנסת ה-23 (מרץ 2020) וכן את השפעת אחוז ההצבעה על ההצבעה למפלגות השונות

- א. כתבו פונקציית scatter-plot של שתי מפלגות כפי שהודגם בכיתה ונמצא בקובץ לדוגמא במודל, אבל כך **שהשטח** של כל עיגול בפלוט יהיה פרופורציוני למספר הקולות הכללי ביישוב. לצורך כך ניתן להשתמש בפרמטר s בפונקציית scatter של matplotlib. כיילו את השטח כך שהישוב עם מספר הקולות המקסימלי יקבל ערך 200.
- ב. בחרו 3 זוגות של מפלגות וציירו scatter-plot בעזרת הפונקציה שכתבתם עבור כל זוג. תארו במילים מה רואים בגרף והאם ומה זה מעיד על הקשר בין המפלגות.
- א. חשבו עבור כל זוג מפלגות מבין 8 המפלגות הגדולות ביותר את מקדם המתאם של **pearson** בין שכיחות המצביעים שלהם על פני כל הישובים כפי שהודגם בכיתה. הציגו heatmap המתאר את מטריצת מקדמי המתאם בגודל 8x8 כאשר המפלגות ממויינות על פי גדלן. הוסיפו colormap לתמונה. תארו מה קיבלתם.
- ב. הציגו שוב את heatmap אבל הפעם סדרו את המפלגות על פי נטייתן הפוליטית, מימין לשמאל (גם בשורות וגם בעמודות, באותו הסדר). אם אינכם בטוחים בסדר בין 2 מפלגות הפעילו את שיקול דעתכם והסבירו. תארו את התמונה החדשה שקיבלתם. איזה סידור אינפורמטיבי יותר?
- בסעיף זה ננסה לבדוק ולתקן את השפעת הבדלי אחוזי ההצבעה על שכיחות ההצבעה למפלגות.
 - א. תקנו עבור כל ישוב את מספר הקולות לכל מפלגה בהנחה שבכל ישוב התפלגות ההצבעה בקרב אלו שלא הצביעו או שקולותיהם היו פסולים זהה להתפלגות ההצבעה של הקולות הכשרים באותו ישוב (אבל כמובן שונה בין ישוב לישוב), כפי שנלמד בכיתה.
 - חשבו עבור כל מפלגה את שכיחות הקולות הארצית שהיתה מתקבלת עבור המקרה בו היו 100 אחוזי הצבעה ועל פי תיקון זה. הציגו ב-bar-plot זה לצד זה בצבעים שונים שכיחויות אלו מול שכיחות הקולות הארצית עבור כל מפלגה בפועל.
 - ב. חזרו על התיקון של סעיף א', אבל הפעם השתמשו בנתוני ההצבעה על פי **קלפיות** (ולא על פי ערים) בקובץ בשם **votes per ballot 2020.csv**. האם התוצאות השתנו?

הערות:

- חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים.
- השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.
- שימו לב שישנו בקובץ "ישוב" בשם "מעטפות חיצוניות" המתאר את המעטפות הכפולות ואינו מתאר ישוב מסויים. יש להסיר אותו כאשר מציגים נתונים על ישובים.
- מכיוון שקשה לזכור את אותיות ההצבעה לכל מפלגה ולשייך למפלגה, הציגו את שמות המפלגה הבאים בכל פעם שתדרשו לכך במקום באותיות: (ניתן להמיר באמצעות בניית dictionary בפייתון)

פתק	מחל	פה	ודעם	שס	ג	אמת	ל	טב
מפלגה	הליכוד	כחול לבן	הרשימה המשותפת	שס	יהדות התורה	עבודה מרץ גשר	ישראל ביתנו	ימינה

- מותר להיות יצירתיים; נסו לחשוב על שיטות אחרות לחישוב דמיון בין מפלגות ולתיקון אחוזי ההצבעה.