מבוא למדע נתונים גאו-מרחבי
2026 ה׳תשפ״ו
המחלקה לגיאוגרפיה, האוניברסיטה העברית בירושלים

<u>פרויקט מסכם</u>

מטרתו של הפרויקט המסכם היא ליישם את הטכניקות והשיטות שנלמדו בקורס, על-מנת לעבד ולנתח מסדי נתונים, המהווים חיתוך של נתוני עתק (big data) מתחומים שונים בתעשייה ובמחקר, ולהציג את התהליך ואת המסקנות ברמה אקדמית.

- תוכלו לבחור את הנתונים עליהם תתבסס העבודה מתוך אוסף הנתונים בנספח א׳, <u>או לבחור כל מקור נתונים אחר כרצונכם</u>, בתנאי שיכיל מעל 20,000 רשומות וכן מידע גיאו-מרחבי.
- למעוניינים, ניתן להגיש עבודה אחת משותפת לזוג תלמידים.
- ההגשה תכלול לפחות שני קבצים:

1. קובץ docx (MS-WORD), המכיל את סיכום העבודה, כולל הקדמה, שיטות, תוצאות ומסקנות. הצגת העבודה תכלול את מטרתה, תיאור של איסוף הנתונים ועיבודם, תוצאות ומסקנות ככל שיהיו. יש להקפיד על רמת הכתיבה והצגת הנתונים (הצגת נתונים בהירה, סגנון ותחביר תקינים וכיוב׳).
2. קובץ ipynb (jupyter notebook), המכיל את התוכנה המלאה שמעבדת את הנתונים ומייצרת את התוצאות. <u>התוכנה צריכה לרוץ במלואה</u> במחברת Jupyter. אנא הקפידו על תיעוד מפורט ונהיר של הקוד (commenting).
3. כל קבצי הגרפיקה שהופקו ע״י התוכנה.
4. במקרה שנכללו בעבודה מאגרי נתונים חיצוניים או קבצים נוספים, הנ״ל יצורפו להגשה.

- בחרו את שאלת המחקר באופן חופשי. רצוי שהשאלה לא תהיה טריוויאלית (״האם יורד גשם בחודשי החורף בישראל״), אך גם לא מורכבת מדי. **זכרו כי מטרת העבודה היא יישום השיטות שלמדתם בקורס, ולא מחקר אקדמי מעמיק**.
- מותר ואף רצוי לשלב כלים מעבר לאלו שנלמדו בקורס, כולל ספריות חיצוניות ושיטות חדשות.
- נא להימנע מאותיות בעברית בגוף התוכנה (כולל בתיעוד) או בשמות הקבצים.
- ניתן להשתמש בכלים חיצוניים דוגמת כלי AI לעבודה על הפרויקט, אך יש לציין בעבודה באילו כלים היה שימוש, עבור אילו שלבים ומה היה טיב השימוש.
- **תאריך אחרון להגשה: 28.7.2026**.

מבוא למדע נתונים גאו-מרחבי
2026 ה'תשפ"ו
המחלקה לגיאוגרפיה, האוניברסיטה העברית בירושלים

ניקוד ודרישות

<u>סיכום העבודה</u> : [50 נק']

הציון על סיכום העבודה מורכב מהחלק שיכלול את הגדרת מטרתה, נתונים ותהליכי ניקוי/עיבוד, שיטות, תוצאות, מסקנות וסיכום [15 נק'], תרשימים גרפיים מתאימים עם ביאור (caption) בגוף המסמך [25 נק'], וקבצי הגרפיקה שיוצרו במהלך העבודה [10 נק'].

סיכום העבודה יוגש בקובץ docx ואורכו לא יעלה על 6 עמודי A4, כולל גרפים וטבלאות. הסיכום צריך להיות כתוב בצורה מקצועית ותמציתית. יש להקפיד על הצגה נקייה של גרפים, <u>להוסיף ביאור לכל תרשים</u>, ולדאוג שכל המידע הרלוונטי יוצג באופן ברור.

<u>התוכנה</u> : [50 נק']

התוכנה תוגש בקובץ ipynb <u>יחיד</u> וצריכה לעמוד בדרישות הבאות :

1. על התוכנה לרוץ במלואה ממחברת jupyter. במידה ונדרשות תוספות לסביבת העבודה, יש לתאר אותן בסיכום העבודה ובגוף התוכנה בהערות. [2 נק']
2. התוכנה צריכה להיות מתועדת היטב, כך שגם קורא שאינו מצוי בתיאוריה יוכל לעקוב אחר הפעולות הנעשות בכל חלק וחלק של התוכנה. על הפעולות להתאים לשאלת המחקר ולכלול פעולות המתקשרות לניקוי הנתונים ועיבודם [5 נק']

התוכנה צריכה לכלול **לפחות** :

3. משתנה אחד מכל אחד מהסוגים הבאים : [5 נק']
   a. List
   b. Dictionary
   c. Tuple
   d. מערך numpy דו-מימדי (מטריצה)
   e. DataFrame או GeoDataFrame
4. לולאה אחת מסוג for. [2 נק']
5. בדיקת תנאי אחת (בתוך indexing, או מסוג if למשל). [2 נק']
6. שימוש בשלוש פונקציות סטטיסטיות של ספריית numpy. [3 נק']
7. שני תרשימים שיוצרו באמצעות ספריית matplotlib. [5 נק']
8. שני תרשימים שיוצרו באמצעות ספריית seaborn. לפחות אחד מהם קטגוריאלי. [5 נק']
9. פעולה מרחבית באמצעות Geopandas (חיבור מרחבי, חיתוך, אזור חיץ, חישוב מרחק וכו'). [5 נק']
10. ניתוח מרחבי באמצעות חבילת PySAL (Moran's I, LISA, ניתוח נקודות חמות). [5 נק']
11. שימוש בלמידת מכונה באמצעות ספריית SciKit learn. אין הגבלה על סוג המודל הלומד. [5 נק']
12. כתיבת נתונים מעובדים לקובץ בפורמט csv. [2 נק']
13. כתיבת נתונים מעובדים לקובץ בפורמט json. [2 נק']
14. שמירת כל הגרפים בפורמט החביב עליכם, ברזולוציה של 150dpi. [2 נק']

בהצלחה!

מבוא למדע נתונים גאו-מרחבי
2026 ה׳תשפ״ו
המחלקה לגיאוגרפיה, האוניברסיטה העברית בירושלים

נספח א׳ – אוסף הנתונים

לנוחיותכם, באתר הקורס מרוכזים מסדי נתונים מסוגים שונים, בהם תוכלו להיעזר עבור הפרויקט המסכם. אין חובה לבחור באחד או יותר ממאגרי נתונים אלה. אפשר להשתמש בכל מקור אחר לנתונים ובתנאי שיהיה בעל 20,000 רשומות או יותר, ויכיל גם מידע גיאו-מרחבי.

בנוסף, מאד מומלץ לבקר באתר https://www.naturalearthdata.com/downloads, שמציע נתונים חופשיים להורדה, כולל קבצי shp הכוללים מידע גיאוגרפי רב (גופי מים, גובה מעל פני הים, בטימטריה, תוואי נהרות, שדות תעופה, ערים ועוד נתונים רבים וטובים שכדאי לשלב).

אוסף הנתונים באתר:

1. כל עסקאות הנדל״ן בתל-אביב בשנים 2012-2017, כולל מחיר, מיקום הנכס על רשת ישראל, מחיר עסקה, קומה, שטח, ועוד. הנתונים נמצאים בקובץ יחיד בפורמט csv. אם ברצונכם להטיל את הקואורדינטות להיטל נפוץ יותר תוכלו להשתמש בספריית pyproj. רשת ישראל עם קוד epsg: 2039, ואילו WSG84 עם קוד epsg: 4326.
הצעות לעניין מחקרי: מגמות מחיר, התפלגויות מחיר כתלות במיקום, התפלגויות מרחביות של מכירות בשנים שונות וכו׳.
2. נתוני GPS של עדר פרות בצפון הארץ. הנתונים נמצאים במספר קבצי csv, אחד עבור כל פרה. ומכילים מיקום, זמן דגימה, ונתונים נוספים על איכות הסיגנל (כמה לוויינים אותרו וכיוב׳).
הצעות לעניין מחקרי: התפלגות שטחי המרעה של הפרות הבודדות ושל העדר בפרט, מהירות הפרות, שעות פעילות ומנוחה, השתייכות קבוצתית (עדר).
3. נתונים גלובאליים על כל תחנות הכוח, כולל מיקום, שייכות לאומית, בעלות פרטית, הספק וסוג הדלק. הנתונים מרוכזים בקובץ יחיד בפורמט csv.
הצעות לעניין מחקרי: מגמות בסוגי דלקים, מגמות בהספק, התפלגות מרחבית בתוך המדינה, התאמה בין סוג הדלק לבין מיקום גיאוגרפי (קרבה לגוף מים / גובה מעל פני הים)
4. נתוני נסיעה בעיר קוריאנית (באדיבות חברת "מובילאיי") מרוכזים בקבצי geojson: מהירות הרכב – מהירות ממוצעת בשעה מסוימת אזורי בנייה – שכבה קווית של אזורי בנייה, בטבלה יש עמודת זמן של זיהוי ראשון (eventTimestamp), זיהוי אחרון + 12 שעות (eventExpirationTimestamp) ומספר הפעמים שנצפה האזור בנייה הספציפי בתוך פרק הזמן הזה (eventValue) הולכי רגל ורוכבי אופניים - ממוצע שעתי למקטע דרך, לכל אחד מהם יש שדה של מספר נסיעות שבהם זוהה הולך רגל/רוכב אופניים וממוצע לזיהוי.
5. נתוני עוצמת הקליטה הסלולארית באיזור טאייפי, טאיוואן (באדיבות חברת Here), כולל הספק הסלולארי, עוצמות האות וטכנולוגיית השידור.
6. נתונים אודות קריאות במוקד 106 של עיריית ירושלים.
7. נתוני תאונות דרכים בישראל לשנים 2005-2008. כולל קובץ markers עם הקואורדינטות והמועד של כל תאונה וקובץ involved עם המידע על המעורבים בתאונה (מועד קבלת הרישיון,

קבוצת גיל, מגדר, סוג מכונית, מקום מגורים). מצורפות שתי שכבות נתונים – שכבת ערים ושכבת כבישים מOpenStreetMap.

8. כל נתוני OpenStreetMap, באמצעות החבילה osmnx.
9. שכבת המבנים הגלובלית של מיקרוסופט (נוצרה על בסיס תמונות לוויין באמצעות אלגוריתמים של למידת מכונה). ישנו מאגר נתונים מקביל עבור מדינות בארה״ב.