

המכללה האקדמית תל אביב יפו
בית הספר לכלכלה וניהול
החוג למערכות מידע

274077, Big Data

ד"ר גלית גורדוני

תשע"ח, סמסטר א', מועד א', 28/01/2018

--	--	--	--	--	--	--	--	--

מספר זהות

משך הבחינה: 3 שעות

ללא חומר עזר

בבחינה 20 שאלות, יש לענות על כולן. משקל השאלות זהה - 5 נקודות לשאלה.
בכל שאלה יש לסמן את התשובה הנכונה (אחת בלבד), באופן ברור, בדף התשובות בלבד. אם סימנת יותר מתשובה אחת - השאלה תיפסל!
יתקבלו תשובות שנכתבו על גבי דף התשובות בלבד. לא ייבדקו תשובות שלא ייכתבו בדף המיועד לכך.

סימון תשובות בטופס הבחינה לא יהווה בסיס לערעור.

יש להחזיר את דף התשובות ביחד עם טופס הבחינה, אחרת הבחינה לא תיבדק.

**נא לשים לב כי דף התשובות מוכן לטופס בחינה של 6 תשובות. בבחינה זו יש 5 תשובות לכל היותר.
יש להקפיד למלא את התשובות במקום הנכון!**

בהצלחה!

לפניך נתונים לחמש השאלות הבאות:

כיום, בזכות אמצעים טכנולוגיים כגון כלי מיפוי (GIS) ומודלים המבוססים על ביג דאטה, ניתן לזהות אילו שינויים חלים לאורך זמן בכמות ובתפוצה של מיני בעלי חיים בעולם. במסגרת פיתוח מערכת שתאפשר להעריך באופן אוטומטי את דרגת הסיכון להכחדה של מיני בעלי חיים שונים, התבקשת להתייחס למספר בעיות אפשריות באיכות הנתונים.

שאלה מספר 1:

הבעיה הראשונה מתייחסת לקשיים בקבוצת נתונים (aggregate) והשוואה בין מקורות מידע שונים. לדוגמא, במאגר גנטי הכולל מאגרים של חיות בר ממדינות שונות, קיימות שתי בעיות עיקריות בקבוצת נתונים מרמת המיקרו של בעל חיים יחיד לרמת המקרו – כמות חיות בר מסוגים שונים ברמת המדינה:

- בעלי חיים בעלי אותם מאפיינים גנטיים, מופיעים בקבצים של מדינות שונות, כשייכים למינים שונים של בעלי חיים. לדוגמא, קיפוד מצוי המסווג במדינות מסוימות כקיפוד מצוי ובמדינות אחרות כקיפוד חולות
 - בקבצים של מדינות שונות, קיפודים בעלי מאפיינים שונים, מופיעים כשייכים לאותו מין של בעל חיים. לדוגמא, קיפוד מצוי, קיפוד חולות וקיפוד מדבר המסווגים כולם כקיפוד מצוי
- קובץ העבודה שלך הוא קובץ הכולל מיזוג של 30 קבצי מדינות שונות. בכל שורה בקובץ מופיע בעל חיים. בטורים מופיעים משתנים, כמו שם המדינה בה הוא חי ומאפייני הרצף הגנטי שלו לפי בדיקת הדם שבוצעה.
- מהו מרכיב הטעות בביג דאטה המופיע בתיאור בעיות בקבוצת הנתונים של מיני חיות בר?

- א. טעות בתאים הנובעת מטעות אפיון (cell errors-specification error).
- ב. טעות בטורים הנובעת מכפילות (column error-duplication error).
- ג. טעות בשורות הנובעת מסלקטיביות (raw error-selectivity).
- ד. טעות בטורים הנובעת מטעות בתיג (column error-erroneous labeling).
- ה. טעות בתאים הנובעת מנתונים חסרים (cell errors-missing data).

שאלה מספר 2:

התבקשת לחשב את אחוז הקיפודים מסוגים שונים (קיפוד מצוי, קיפוד חולות, קיפוד מדבר וכדומה) מתוך כלל הקיפודים המופיעים במאגר, על פני 30 מדינות. מה צפויות להיות השלכות הטעות בביג דאטה על ממצאי ניתוח תיאורי (descriptive analytics) לבחינת המספר הכולל של קיפודים בכל המדינות והשיעור היחסי של סוג קיפוד ספציפי מכלל הקיפודים?

- א. הערך המסכם של מספר הקיפודים הכולל בכל המדינות יהיה מדויק. הערך המסכם של כל סוג קיפוד יהיה מוטא, כלפי מטה או כלפי מעלה. לכן, ניתן יהיה להציג מגמות מדויקות של מספר הקיפודים הכללי ובכל אחת מהמדינות בנפרד. לא ניתן יהיה להציג מידע מדויק על השיעור היחסי של סוג קיפוד ספציפי מכלל הקיפודים.
- ב. הטעות המתוארת בביג דאטה לא תתבטא בממצאי הניתוח התיאורי מכיוון שמטרת הניתוח רק לתאר את הנתונים הקיימים. לכן, ניתן יהיה לחשב במדויק את הכמות הכללית של קיפודים ובהשוואה בין מדינות לפי סוג הקיפוד.
- ג. הנתונים חסרי ערך ולא ניתן להשתמש בהם לתיאור הכמות הכללית של קיפודים בכל המדינות יחד וגם לא לתיאור הבדלים, בין מדינות, בכמות הקיפודים.
- ד. ממצאי הניתוח התיאורי יאפשרו להסיק על הגורמים להבדלים בין המדינות השונות בכמות הקיפודים הכוללת ולחשב את סיכויי ההכחדה של מינים שונים.
- ה. מכיוון שמאגר הנתונים כולל כמות עצומה של נתונים, הטעויות בביג דאטה לא ישפיעו על הערכים המסכמים שיחושבו מהנתונים המופיעים במאגר.

שאלה מספר 3:

איזו מבין ההמלצות הבאות, יש לאמץ בעת הנגשה לציבור של מאגרי מידע הכוללים מידע גנטי של חיות בר במדינות שונות?

- א. מומלץ להעלות למאגרים הפתוחים נתונים שהם לא בפורמט pdf מכיוון שאלו הם נתונים לא-מובנים.
- ב. מומלץ במאגרים פתוחים להעלות רק נתונים בפורמט pdf ותמונות, מכיוון שאלו נתונים שאורגנו במבנה התואם את ההגדרות במאגר הנתונים.
- ג. מומלץ במאגרים פתוחים להעלות רק נתונים בפורמט pdf כדי שניתן יהיה לצמצם טעויות בטורים שמקורן בהגדרות לא מדויקות.
- ד. מומלץ במאגרים פתוחים להעלות רק נתונים בפורמט pdf כדי שניתן יהיה לצמצם טעויות בשורות שמקורן בסלקטיביות.
- ה. מכיוון שהמאגרים פתוחים לכל, אין חשיבות לפורמט הקבצים המועלים למאגר, כל עוד המטא-דאטה שלהם מפורטת.

שאלה מספר 4:

כיצד ניתן למזג בין מאגרי המידע הגנטיים של חיות בר ממדינות שונות?

- א. ניתן למזג ברמת המיקרו – כלומר, בשורות יופיעו חיות הבר שעליהן יש נתונים. המיזוג יכלול הוספת מקרים. המשתנה הממזג הנו מדינה.
- ב. ניתן למזג רק ברמת המקרו – כלומר, בשורות יופיע מינים שונים של חיות בר והנתונים בתאים יהיו מקובצים.
- ג. שימוש במספר הגדרות שונות לאותו מין של חיית בר יאפשר מיזוג קבצים מכיוון שהוא מאפשר לצמצם טעויות בזיהוי יחידת הניתוח.
- ד. שימוש בהגדרות שונות לאותה חיית בר יקל על השוואת מגמות בין קבצים שנצרכו במדינות שונות.
- ה. ניתן למזג רק ברמת המיקרו – כלומר, בשורות יופיעו המדינות השונות והמיזוג יכלול הוספת משתנים.

שאלה מספר 5:

נניח וקיימת אפשרות להוסיף נתונים, עבור כל אחת מחיות הבר המופיעות בקובץ, מתוך מאגרי מידע חיצוניים. לדוגמא, רמת זיהום האוויר באזור הגיאוגרפי שבו נמצאה החיה. מהו המשתנה הממזג?

- א. שם האזור הגיאוגרפי שבו נמצאה החיה.
- ב. המספר הסידורי של החיה המאפשר לזהות ממי נלקחה דגימת הדם.
- ג. המין שאליו משתייכת חיית הבר.
- ד. כמות חיות הבר בכל אזור גיאוגרפי.
- ה. הסימנים הגנטיים של החיה כפי שזוהו בבדיקת הדם.

לפניך נתונים לחמש השאלות הבאות:

אחת המגמות המאפיינות את השימוש בביג דאטה, הנה של בנקים העושים שימוש ב- predictive analytics כדי לקבל החלטות בנוגע למתן הלוואת משכנתא ללקוחות ספציפיים. לדוגמא, השוואה בין הלקוח ללקוחות אחרים, בעלי מאפיינים דומים, כדי לחזות האם אותו לקוח יצליח לעמוד בהחזרי המשכנתא.

שאלה מספר 6:

בבנק נעשה שימוש בנתונים פנימיים על פעילות עבר של הלקוחות ומאפייניהם לשם קבלת החלטות על מאפייני הלוואת משכנתא שתאושר ללקוחות ספציפיים. המודל כלל השוואה בין הלקוח שביקש הלוואה לבין לקוחות אחרים בעלי מאפיינים דומים כמו למשל, גובה המשכורת החודשית, מצב משפחתי, גודל המשפחה והיסטוריה של מידת העמידה לאורך זמן של הלקוחות בהתחייבויות האשראי שלהם. המודל שימש לחיזוי הסיכוי שהלקוח יעמוד בהחזרי המשכנתא.

במחלקת המחקר התלבטו לגבי כדאיות הוספת נתונים חיצוניים למודל החיזוי, כדוגמת מידע על תחביבים, השכלה, ניסיון בשוק העבודה ותחומי עניין של הלקוח מתוך הפרופיל של הלקוח ברשת החברתית. מהו המידע שאמור להופיע בשורות בקובץ הנתונים?

- א. הלקוח הוא יחידת הניתוח ולכן בכל שורה אמור להופיע לקוח אחד.
- ב. בשורות אמור להופיע מידע על התנהגות עבר של הלקוח. בכל שורה אמור להיות מאפיין אחר כמו ממוצע סכום הרכישות החודשי.
- ג. בשורות אמור להופיע פרופיל הלקוח ברשת החברתית. בכל שורה אמור להיות מאפיין אחר כמו למשל, תחביבים, השכלה, ניסיון בשוק העבודה ותחומי עניין של הלקוח.
- ד. הסיכוי שהלקוח יעמוד בהחזרי המשכנתא הוא יחידת הניתוח ולכן בכל שורה יופיע סכום המשיכה החודשי הממוצע של הלקוח מחשבון הבנק שלו.
- ה. בשורות יופיעו רק לקוחות שקיבלו הלוואות בעבר. הניתוח מתייחס למידע היסטורי ולכן בכל שורה אמור להופיע לקוח שקיבל בעבר הלוואה.

שאלה מספר 7:

מהו המידע הנדרש כדי למזג בין הנתונים הקיימים בחברה לנתונים חיצוניים מהרשת החברתית?

- א. בנתונים הקיימים ובנתונים החיצוניים אמור להופיע לפחות משתנה זהה אחד שיאפשר לזהות את הלקוח הספציפי.
- ב. בנתונים הקיימים ובנתונים החיצוניים אמור להופיע מידע משלים שאינו חופף.
- ג. בנתונים הקיימים ובנתונים החיצוניים אמור להופיע מידע מפורט וזהה על התנהגות הלקוחות שיאפשר להשוות את מידת הדיוק של הנתונים בין מקורות המידע השונים.
- ד. בנתונים הקיימים ובנתונים החיצוניים אמור להופיע מידע ברמת המקור.
- ה. אף תשובה לא נכונה.

שאלה מספר 8:

המתנגדים להוספת נתונים חיצוניים למאגר הנתונים הפנימי של החברה טענו שהעובדה שרק לחלק מהלקוחות יש פרופיל ברשתות חברתיות תגרום לפגיעה באיכות הנתונים. מהו מרכיב הטעות בביג דאטה שבו מתמקדת הביקורת?

- א. טעות בתאים הנובעת מנתונים חסרים (cell errors-missing data).
- ב. טעות בטורים הנובעת מטעות בתיוג (column error-erroneous labeling).
- ג. טעות בתאים הנובעת מטעות אפיון (cell errors-specification error).
- ד. טעות בשורות הנובעת מכפילות (raw error-duplication error).
- ה. טעות בשורות הנובעת מסלקטיביות (raw error-selectivity).

שאלה מספר 9:

ביקורת נוספת התייחסה לכך שגם אם ללקוחות יש חשבון פעיל ברשתות חברתיות יתכן והמידע המופיע בו לא עדכני או לא אמין. מהו מרכיב הטעות בביג דאטה שבו מתמקדת הביקורת?

- א. טעות בתאים הנובעת מטעות בתוכן (cell errors-content error data).
- ב. טעות בתאים הנובעת מטעות אפיון (cell errors-specification error).
- ג. טעות בתאים הנובעת מנתונים חסרים (cell errors-missing data).
- ד. טעות בטורים הנובעת מטעות בתיוג (column error-erroneous labeling).
- ה. טעות בשורות הנובעת מסלקטיביות (raw error-selectivity).

שאלה מספר 10:

לו הייתם חלק מצוות המחקר, האם הייתם ממליצים לבנק לבצע סקר בקרב הלקוחות, שיתייחס למידת הביטחון שיש להם ביכולת ההחזר העתידית שלהם את תשלומי המשכנתא, ולכלול את המידע שיתקבל מהלקוחות במודל החיזוי?

- א. אף תשובה לא נכונה.
- ב. כן, מכיוון שהוספת נתוני הסקר תאפשר לצמצם טעות בתאים הנובעת מנתונים חסרים (cell errors-missing data).
- ג. כן, מכיוון שהוספת נתוני הסקר תאפשר לצמצם בטורים טעות בתיוג (column error-erroneous labeling).
- ד. כן, מכיוון שתשובות לסקר משקפות התנהגויות בזמן אמת ומעודכנות יותר מהנתונים הפנימיים של החברה.
- ה. נתוני סקרים המבוססים על דיווח עצמי של משיבים לא מתאימים למודל חיזוי אלא רק לניתוח תיאורי.

לפניך נתונים לחמש השאלות הבאות:

בספטמבר 2017 פורסם מסמך מדיניות של המשטרה הבריטית המתמקד בשימושים האפשריים בביג דאטה. במסמך הוגדרו 4 מטרות עיקריות לשימוש בביג דאטה:

- א. חיזוי אזורי פשיעה לשם זיהוי אזורים בהם יש סיכוי גבוה להתרחשות אירועי פשיעה. החיזוי יאפשר הקצאת משאבים יעילה לטיפול בפשיעה
- ב. חיזוי הסיכוי של יחידים להיפגע מפשיעה לשם זיהוי יחידים בסיכון גבוה להיפגע, חישוב הסיכוי של יחידים להפוך לנעדרים ולהפוך לקורבנות פשיעה
- ג. ניתוח נתונים הנאספים במכשירי מעקב אלקטרוניים, כמו זיהוי אוטומטי של לוחיות רישוי ומצלמות מעקב
- ד. ניתוח נתונים ממקורות פתוחים לציבור (open-source data), כגון רשתות חברתיות, לשם הבנה מעמיקה יותר של בעיות פשיעה ספציפיות ופיתוח שיטות למניעתן

שאלה מספר 11:

אחת הבעיות המרכזיות לשימוש בדאטה לשם השגת מטרות המשטרה הנה שהנתונים מנוהלים במספר רב של מאגרי מידע נפרדים ולא תואמים. על מנת להפיק ערך מהמידע הקיים בידי המשטרה נדרש ליצור מאגר מידע לאומי הממזג את הנתונים, ממאגרים שונים, למאגר של ביג דאטה. לאיזה שלב בתהליך היצירה והניתוח של ביג דאטה מתייחסת המגבלה המתוארת בדו"ח?

- א. שלב השינוי (transform) בתהליך הטיפול בנתונים (ETL).
- ב. שלב ההיווצרות (generate step).
- ג. שלב ההרחבה (extract) בתהליך הטיפול בנתונים (ETL).
- ד. שלב הטעינה (load) בתהליך הטיפול בנתונים (ETL).
- ה. שלב הסינון - הדגימה (filtering-sampling/reduction).

שאלה מספר 12:

למשטרה הבריטית יש כמות עצומה של נתוני פשיעה היסטוריים. במידע ההיסטורי נעשה שימוש לחיזוי הסיכוי להתרחשות אירוע פשיעה ברובע ספציפי. החיזוי נועד לאפשר ניהול יעיל של כוחות הסיור המשטריים והקצאת פטרולים בהתאם לרמות הסיכון השונות המחושבות לישוברים שונים ברחבי המדינה. מהי יחידת הניתוח?

- א. ישוב.
- ב. הסיכון להיפגעות מפשע ברמת היחיד.
- ג. אירוע פשיעה.
- ד. מספר השוטרים המוצבים בכל אזור בבריטניה.
- ה. אף תשובה לא נכונה.

שאלה מספר 13:

אחת ההמלצות בדו"ח המחקר כללה יצירת מטא-דאטה אחידה שתקל על הוספת נתונים למאגרי המידע של המשטרה. ההמלצה התמקדה ביצירת מילון סטנדרטי של מושגי מפתח שיקל על הכנסת נתונים חדשים למאגר וגם על שליפת המידע ממנו.

איזו סוג טעות אמורה למנוע יצירת המטא-דאטה האחידה?

- א. תיוג שגוי או לא מדויק של הנתונים המופיעים בטור (column error-erroneous labeling).
- ב. טעות בשורות הנובעת מסלקטיביות (raw error-selectivity).
- ג. טעות בשורות הנובעת מכפילויות (duplications).
- ד. טעות בשורות הנובעת מהוספות מוטעות (erroneous inclusions).
- ה. טעות בשורות הנובעת מהשמטות (omissions).

שאלה מספר 14:

המידע ממקורות פתוחים לציבור הרחב כמו למשל, נתוני טוויטר ופייסבוק, שהמשטרה הבריטית עושה בו שימוש לניבוי פשיעה, הנו ברמת המיזו. המשטרה ממזגת מידע מהרשתות החברתיות עם קבצים משטרתיים המתעדים אירועי פשיעה. מהי יחידת הניתוח בקובץ הממזג את שני מקורות המידע?

- א. אירוע פשיעה.
- ב. הרשת החברתית (למשל, פייסבוק וטוויטר).
- ג. חשודים שהמשטרה עצרה בעבר.
- ד. אנשים שהגישו בעבר תלונה במשטרה.
- ה. זמן התרחשות אירוע פשיעה.

שאלה מספר 15:

אחד האתגרים המרכזיים בקישור בין מאגרי המידע השונים של המשטרה הנו להצליב מידע על חשוד ספציפי, הקיים במאגרים נפרדים. מהו הקושי בהצלבת נתונים משלושת המאגרים הבאים:

- מאגר דנ"א הכולל דגימות דם שנלקחו מחשודים שנעצרו על ידי המשטרה
 - מאגר תמונות פנים של חשודים
 - מאגר הכולל סרטים ממצלמות אבטחה המפוזרות ברחובות ערים
- א. הקושי טמון בכך שהנתונים לא-מובנים. כדי למזג את המידע יש להפוך את הנתונים למובנים.
 - ב. הקושי טמון בבחירת הפלטפורמה לאחסון הכמות העצומה של הנתונים.
 - ג. הקושי טמון בכך שבכל מאגר מופיעות הרבה מאוד יחידות ניתוח אך מעט מאוד משתנים.
 - ד. הקושי טמון בכך שבכל מאגר מופיעים הרבה מאוד משתנים אך מעט מאוד יחידות ניתוח.
 - ה. אף תשובה לא נכונה.

לפניך נתונים לחמש השאלות הבאות:

ב- 23 ביוני 2016 התקיים בבריטניה משאל עם על החברות באיחוד האירופי. סקרים שבוצעו בקרב אוכלוסיית המצביעים בבריטניה נכשלו בחיזוי המגמות בדעת הקהל ובתוצאות המשאל. לאחר שהתקיים המשאל בוצע מחקר שהתמקד באפשרויות הגלומות בשימוש בנתוני ביג דאטה.

שאלה מספר 16:

לאחר משאל העם בוצע מחקר שעשה שימוש בביג דאטה. נאספו מטוויטר כ-29 מיליוני ציוצים בנושא הברקזיט. החוקרים חישובו את מספר הציוצים שבהם הופיעה המילה "לעזוב" ואת מספר הציוצים שבהם הופיעה המילה "להישאר".

במחקר נמצא שבמשך כל חמשת החודשים שקדמו למשאל העם מספר הציוצים אשר כלל את המילה "לעזוב" היה כמעט כפול ממספר הציוצים שכלל את המילה "להישאר".

הטענות הבאות מתייחסות לממצא שהתקבל במחקר. מהי הטענה הנכונה?

- א. הממצא הוא תוצר של ניתוח תיאורי (descriptive analytics).
- ב. הממצא הוא תוצר של מודל חיזוי (predictive analytics).
- ג. הממצא הוא תוצר של מודל סיבתי (prescriptive analytics) שבחן את השפעת הקמפיין התקשורתי על דעות המצביעים.
- ד. גולשים בטוויטר לא מייצגים את כלל אוכלוסיית הבוחרים ולכן השימוש בטוויטר לחיזוי של מגמות בדרך כלל שגוי.
- ה. אף תשובה לא נכונה.

שאלה מספר 17:

במחקר נבחנו ציוצים (tweet) וציוצים חזרה (re-tweet). מהי רמת הקבוצ (אגרגציה) של הנתונים?

- א. יחידת הניתוח היא ציוץ וכל הציוצים חזרה המתייחסים אליו. רמת הקבוצ היא מיזו.
- ב. יחידת הניתוח היא הגולש המעלה את הציוץ או את הציוץ חזרה לטוויטר. רמת הקבוצ היא מיקרו.
- ג. יחידת הניתוח היא המיקום הגיאוגרפי שממנו הגיע הציוץ. רמת הקבוצ היא מיקרו.
- ד. יחידת הניתוח היא התאריך שבו הועלה הציוץ לטוויטר. רמת הקבוצ היא מיקרו.
- ה. יחידת הניתוח היא אות (סיגנל) מהטלפון החכם המציין את מיקום הציוץ. רמת הקבוצ היא מיקרו.

שאלה מספר 18:

הסקרים שבוצעו לפני משאל העם והמחקר שהתבסס על ביג דאטה התמקדו בכוונת ההצבעה במשאל הברקזיט ובנושאים המעסיקים את הציבור הבריטי בהקשר של הברקזיט. מה ההבדל בין סקר לבין מחקר העושה שימוש בהתנהגות גולשים ברשתות חברתיות?

- א. כל התשובות נכונות.
- ב. במחקר העושה שימוש בנתוני ביג דאטה המגמות בדעת הקהל והנושאים המעסיקים את הגולשים, בהקשר של הברקזיט, עולים מתוך השיח הקיים בטוויטר.
- ג. במחקר העושה שימוש בנתוני ביג דאטה קיימות טעויות רבות בתהליך היווצרות הנתונים.
- ד. הגולשים המציצים או מציצים חזרה בטוויטר לא מייצגים את האוכלוסייה הכללית של בעלי זכות ההצבעה.
- ה. בסקר, הנושאים הקשורים לברקזיט כמו למשל המצב הכלכלי והגירה, מוגדרים מראש על ידי החוקרים.

שאלה מספר 19:

במחקר הביג דאטה נאספו נתוני מיקום שאפשרו למפות את המגמות בדעת הקהל לפי אזורים גיאוגרפיים. האם ניתן לשלב בין נתוני סקרים בקרב בעלי זכות הצבעה בבריטניה לבין הנתונים ממחקר הביג דאטה על ציורים בבריטניה?

- א. לנתוני הסקרים, ברמת המיקרו, ניתן להוסיף נתונים, ברמת המקרו, מטוויטר.
- ב. הסקרים מבוססים על מדגם של בעלי זכות הצבעה והמידע מטוויטר מתייחס רק לאוכלוסייה של גולשים. האוכלוסיות לא חופפות ולכן לא ניתן למזג את הנתונים למאגר מידע אחד.
- ג. הסקרים מבוססים על מדגם מייצג בעוד בטוויטר האוכלוסייה סלקטיבית. לכן, לא ניתן למזג בין הנתונים.
- ד. הסקרים מוגבלים לאזור גיאוגרפי מסוים בעוד טוויטר לא מוגבל למיקום ספציפי של איסוף נתונים. לכן, לא ניתן לשלב בין הנתונים.
- ה. אף תשובה לא נכונה.

שאלה מספר 20:

בחודשים שלפני המשאל רוב הציורים בנושא הברקזיט היו מבריטניה וארצות הברית. בימים שלפני הברקזיט, ציורים בנושא, הופיעו בתדירות גבוהה מאוד במדינות שונות בעולם ובמגוון שפות. האם ניתן למזג את הנתונים ממדינות, שהן לא בריטניה, עם נתוני הסקרים בבריטניה?

- א. לא ניתן למזג. מכיוון שלא קיים משתנה ממזג שיאפשר לצרף את הציורים ממדינות מחוץ לבריטניה לנתוני סקרים בבריטניה.
- ב. ניתן למזג. יש נתוני מיקום, כך שאין בעיה למזג בין הקבצים.
- ג. ניתן למזג. המיזוג יכול הוספת משתנים (add variables) מטוויטר לקובץ נתוני הסקרים.
- ד. ניתן למזג. המיזוג יכול רק הוספת מקרים (add cases) מטוויטר לקובץ נתוני הסקרים.
- ה. ניתן למזג. המיזוג יכול הוספת מקרים (add cases) ומשתנים (add variables) מנתוני טוויטר לקובץ נתוני הסקרים.

--- סוף המבחן ---