פרויקט סוף- ויזואליזציה של נתונים

|  |  |  |
| --- | --- | --- |
| תיאור סט הנתונים |  | סט הנתונים שבחרנו הינו סט נתונים של סקר האושר העולמי. סט הנתונים המדובר, הינו תוצרו של סקר שנערך בסוף שנת 2017 ע"י חברת Gallup שהינה חברה אמריקאית לניתוח ויעוץ ניהולי לארגונים ברחבי העולם. כמו כן, החברה נודעה בזכות סקרי דעת הקהל שמבצעת בעולם כולו. תוצאות הניתוחים שהם מבצעים, מתפרסמים ברחבי העולם מידי יום כחדשות המונעות מהנתונים עצמם. סקר זה של החברה, מסכם את ממוצע השביעות רצון מהחיים במדינות השונות בשנים 2005-2017.  המטרה של הסקר הזה היא הערכת החיים כפי שאנשים רואים אותה כאשר השאלה הנשאלת היא שכשאדם מדמיין סולם עם 10 שלבים (0 הכי נמוך ו-10 הכי גבוה) שחלקו העליון ביותר מייצג את החיים הכי טובים וחלקו התחתון מייצג את החיים הכי נוראיים שאחד יכול לחיות, היכן אותו אדם חושב שהוא עומד עכשיו ע"ג אותו סולם?  המטרה הראשונית של סקרי האושר שביצעו Gallup הייתה לספק לאנשים בשלטון הבנה ותמונה רחבה ומקיפה יותר של תחושת הרווחה במדינות שלהם ובקהילות אחרות בעולם תוך בחינת מדדים נוספים שראו כקשורים לשביעות רצון של אנשים מאיכות חייהם ומשתנים נוספים שעד לאותה נקודה חשבו שהם עצמם מה שמוביל לאושר של אדם.  את סט הנתונים הזה הורדנו מ-"github", מניתוח עבור דוח האושר העולמי שהועלה בו יש קישור לדוח ממנו ניתן להוריד את סט הנתונים בו השתמשנו.  <https://dpatl.github.io/#:~:text=The%20World%20Happiness%20Report%20is,relative%20to%20other%20country's%20scores>.  הקישור ב-github:  <http://worldhappiness.report/ed/2018/>  גלילה לתחתית המסך עד שרואים שורת טקסט לחיצה ועליה הכיתוב "Chapter 2: Online Data" אשר מובילה להורדת הנתונים.  סט הנתונים מורכב מקובץ xls אחד שבו 1,562 שורות ו-19 העמודות הבאות ואת סוג המשתנים בכל עמודה נוכל לראות בתמונה מס' 1:   1. מדינה (country)- כבר משם העמודה אנו יכולים להבין כי מדובר על דאטה טקסטואלי שיפרט עבורנו את שמות המדינות שהשתתפו בדוח. מכאן גם ניתן להבין שנוכל להתייחס למשתנה זה כמשתנה קטגוריאלי עבור המדינה בסינונים שנרצה לבצע בשלב הויזואליזציות והניתוחים הראשוניים. 2. שנה (year). 3. רמת שביעות רצון מהחיים (Life Ladder)- המדד הממוצע שנקבע בכל מדינה עבור שביעות הרצון מהחיים בה באותה שנה. ניתן להסתכל על המשתנה הזה גם כמשתנה התלוי בבחינת הקשרים בין המשתנים האחרים בסט הנתונים. 4. לוג התמ"ג לנפש (Log GDP per capita)- במקום למדוד את השינויים בתמ"ג (תוצר מקומי גולמי) עצמו בין שנה לשנה העדיפו למדוד את השינויים תחת הפעלת לוג עליהם על מנת לראות את השינויים בצורה נוחה יותר לעין. 5. תמיכה חברתית (Social support)- מדד שמטרתו לבחון האם אדם מרגיש שבעת צרה יש לו על מי לסמוך ולמי לפנות לקבלת עזרה. בסקר המקורי של Gallup התשובה הייתה בינארית (0/1) אז בסיכום של כל הנמדדים במדינה בשנה בוצע ממוצע שיצר את הנתון בסט נתונים זה. 6. תוחלת חיים בריאה משלב הלידה (Healthy life expectancy at birth)- השילוב שבין תוחלת חיים להיתכנות לחיים בריאים כבר מלידה. 7. החופש לבצע בחירות בחיים (Freedom to make life choices)- בסקר המקורי נשאלו הנבדקים האם הם מרוצים/לא מרוצים ממידת החופש שקיימת עבורם לקבל החלטות עצמאיות. גם פה התשובות שוכללו לממוצע של כל מדינה בשנה המדוברת של הסקר. 8. נדיבות (Generosity)- בסקר המקורי הנבדקים נשאלו האם הם תרמו כסף לצדקה בחודש האחרון וכמה. בסקר זה, משתנה זה הוא שארית ממוצעי התשובות של כל מדינה לפי שנה מהתמ"ג לנפש. 9. תפיסת השחיתות (Perceptions of corruption)- ממוצע התשובות בין השאלה "האם השחיתות נפוצה בכל רחבי הממשלה או לא?" והשאלה "האם השחיתות נפוצה בתוך עסקים או לא?" בין מדינות ושנים. 10. השפעה חיובית (Positive affect)- מוגדרת כממוצע של שלושה מדדי השפעה חיובית בסקר המקורי: אושר, צחוק והנאה. 11. השפעה שלילית (Negative affect)- מוגדרת כממוצע של שלושה מדדי השפעה שלילית בסקר המקורי: דאגה, עצב וכעס. 12. אמון בממשל הלאומי (Confidence in national government)- מדד שמבצע ממוצע של התשובות בכל מדינה לפי שנה לגבי השאלה "האם יש לך אמון בממשלה?". 13. איכויות דמוקרטיות (Democratic Quality)- איחוד של 2 משתנים נפרדים שבוצע להם ממוצע שבחנו "קול ואחריות" ו-"יציבות פוליטית והיעדר אלימות". 14. איכות השלטון (Delivery Quality)- מדובר על איחוד של 4 משתנים נפרדים שבוצע להם ממוצע שבחנו את השליטה על שחיתויות, שלטון החוק, איכות הרגולטורים ויעילות הממשלה. 15. סטיית התקן של דירוג שביעות הרצון מהחיים פר מדינה ושנה (Standard deviation of ladder by country-year). 16. ממוצע דירוג שביעות הרצון מהחיים פר מדינה ושנה (Standard deviation/Mean of ladder by country-year). 17. ג'יני אינדקס לפי WB (GINI index (World Bank estimate))- מדד הבוחן את המידה בה חלוקת ההכנסה (או ההוצאה עבור צריכה) בין פרטים או משקי בית בתוך משק כלשהו חורגת מלהיות חלוקה שוויונית לחלוטין. 18. ג'יני אינדקס לפי WB לפי ממוצע 2000-15 (GINI index (World Bank estimate), average 2000-15). 19. מדד ג'יני עבור הכנסת משקי בית לפי מדינה ושנה (gini of household income reported in Gallup, by wp5-year).   תמונה מס' 1: סוג המשתנים בכל עמודה |
| תהליך ניקוי וארגון הנתונים | ר | בתהליך ניקוי הנתונים עלינו לבחון הימצאות של שלושה מצבים עיקריים בסט הנתונים שלנו:   1. כפילויות בנתונים. 2. ערכים חריגים. 3. ערכים חסרים.   **כפילות בנתונים:**  מבדיקה מהירה שביצענו בעזרת הפקודה .duplicated().sum() עולה כי בסט הנתונים המדובר לא קיימות שורות נתונים כפולות.  **ערכים חריגים:**  על מנת לבחון האם קיימים לנו ערכים חריגים בנתונים בנינו פונקציה אליה נשלח את העמודות בסט הנתונים שנרצה לבחון הימצאות חריגות נתונים בתוכן. עבור כל עמודה, אל מול ערכי האחוזונים שבה נבחן האם קיימים ערכים השונים בתכלית מהשאר. כל תצפית הקיימות יא יותר ממספר חריגות מסוים תוחזר כתצפית חורגת. לבסוף הפונקציה תחזיר לנו את כל התצפיות שעלו כחריגות בסט. מהרצת הפונקציה המדוברת עולה כי לא קיימים נתונים החשודים כחריגים בסט.  **ערכים חסרים:**  את כמות הערכים החסרים בדאטה ואת מיקומם ע"ג המשתנים השונים נוכל לראות גם באמצעות פקודת .info() וגם באמצעות .isnull().sum() (תמונות מס' 2-3).  תמונה מס' 2: פלט .info() תמונה מס' 3: פלט .isnull().sum()  משתי התמונות לעיל ניתן לראות כי בסט הנתונים המדובר יש יותר משתנים עם ערכים חסרים מאשר משתנים ללא ערכים חסרים כלל. חלק מהמשתנים בעלי הערכים החסרים הם Log GDP per capita (עם 27 ערכים חסרים), Social support (עם 13 ערכים חסרים), Healthy life expectancy at birth (עם 9 ערכים חסרים) וכן הלאה בהתאם לסדר.  ישנן מס' רב של דברים להתמודד עם ערכים חסרים- להחליפם בממוצע/חציון/שכיח, להחליפם בערך מספרי שיבחר בהתאם לחישובים שיעלו כמתאימים מהתבוננות בדאטה או אפילו מחיקת התצפיות המדוברות. כמובן לפני שנבחר שיטה מתאימה לטיפול בערכים החסרים, יהיה עלינו לבחון ויזואלית ולבצע סטטיסטיקות ראשוניות של הנתונים הקיימים על מנת שאולי יסייעו לנו בבחירת מטריקה אופטימלית לטיפול בערכים החסרים. למרות זאת, כבר מהסתכלות על כותרות המשתנים וההסבר שלהם, שני מדדי הג'יני בעלי המספר הגבוהה ביותרת של ערכים חסרים (979 ו-357) יהיו קשים להשלמה ולכן ימחקו מסט הנתונים. נוסף אליהם, שלושת המדדים המדברים על סטיות תקן וממוצעים הם משתנים שהתווספו לסט הנתונים המקורי מאוחר יותר ככל הנראה ולכן גם הם ימחקו. כך שבפועל נמחק כבר עכשיו את חמשת המשתנים האחרונים ברשימה. |
| הצגת הבעיה, ניסוח שאלות, השערות וציפיות |  | **הצגת בעיות אפשריות שהנתונים יכולים לענות עליהם:** נתונים אלה מספקים מידע עשיר על מגוון הגורמים המשפיעים על אושר האוכלוסייה במדינות שונות. זה יכול להכווין את המדיניות והתכנון במדינות רבות ולפתור בעיות מרכזיות, כגון:   1. תכנון מדיניות: מנהלים מדיניים יכולים להשתמש בנתונים אלה כדי להבין אילו גורמים משפיעים באופן משמעותי על אושר האזרחים ולתכנן מדיניות שתתמקד בשיפור אותם גורמים. 2. שיפור בריאות הנפש: הנתונים יכולים לעזור לפתוח ולתמוך ביוזמות שמטרתן לשפר את בריאות הנפש של האזרחים, בהתאם לגורמים המשפיעים באופן משמעותי על אושרם. 3. קידום שוויון: בעזרת הנתונים, ניתן לאתר את הפערים בין קבוצות שונות במדינה במידת האושר, ולקבל הבנה טובה יותר של הדרכים בהן מדיניות מסוימת עשויה לקדם שוויון.   ועוד שלל בעיות, במקרה שלנו ידוע לנו שהנתונים נלקחות מדוח אשר התמקד בנושא ההגירה והאושר, ועונה על שאלות כגון: האם למהגרים חיים טובים יותר במדינות שמקבלות אותם? האם מהגרים בינלאומיים מגבירים את האושר שלהם ושל משפחותיהם על ידי הגירה? הנתונים כוללים מגוון גורמים שעשויים להשפיע על האושר, כגון תמ"ג לנפש, תמיכה חברתית, תוחלת חיים, חופש לבחור בחיים, נדיבות, תפיסות שחיתות, השפעה חיובית ושלילית, ביטחון בממשל לאומי, איכות דמוקרטית ואיכות משלוח.  להלן כמה שאלות, השערות וציפיות אפשריות המבוססות על נתונים אלה:  **שאלות:**   1. איזו מדינה יש לה את הציון הכי גבוה בשביעות רצון מהחיים, ואיזה גורמים תורמים ביותר לציון זה? 2. איך המוצר המקומי הגולמי לנפש מתממש בציון שביעות רצון מהחיים במדינות שונות? 3. איך גורמים כמו 'נדיבות' ו'תפיסת שחיתות' משפיעים על האושר? 4. איך ציון שביעות רצון מהחיים השתנה במהלך השנים במדינה מסוימת? 5. איך 'אמון בממשלה הלאומית' קשור ל'איכות דמוקרטית' ו'איכות המסירה'?   **השערות:**   1. מדינות עם מוצר מקומי גולמי לנפש גבוה יותר יהיו גם עם ציונים גבוהים יותר בשביעות רצון מהחיים. 2. רמות גבוהות של תמיכה חברתית וחופש בבחירות החיים יתממשו בצורה חיובית בציון שביעות רצון מהחיים. 3. מדינות עם תפיסת שחיתות נמוכה יהיו עם ציונים גבוהים יותר בשביעות רצון מהחיים. 4. מדינות עם 'איכות דמוקרטית' ו'איכות המסירה' גבוהות יהיו עם ציונים גבוהים יותר בשביעות רצון מהחיים.   **ציפיות:**   1. ניתן לצפות למצוא קורלציה חזקה בין גורמים כלכליים (כמו מוצר מקומי גולמי לנפש) לציון 'סולם החיים'. 2. ייתכן ויהיו הבדלים אזוריים משמעותיים בציון 'סולם החיים', שמשקפים הבדלים תרבותיים, כלכליים ופוליטיים. 3. במהלך השנים, מדינות שחוו צמיחה כלכלית או שיפור ביציבות הפוליטית עשויות להראות עלייה בציון 'סולם החיים' שלהן. 4. גורמים כמו נדיבות והשפעה חיובית עשויים להראות קורלציה חיובית אך פחותה עם ציון 'סולם החיים' לעומת גורמים כלכליים ופוליטיים. |
| univariate analysis |  | ***סטטיסטיקה תיאורית:***  *הסטטיסטיקה התיאורית בסט מתחלקת לסטטיסטיקה על נתונים מספריים וסטטיסטיקה על נתונים קטגוריאליים. מבחינת הסטטיסטיקה התיאורית של הנתונים הקטגוריאליים (מדינה) אנו יכולים לראות כי בסט קיימות לנו 164 מדינות שונות המשתתפות בדוח (תמונה מס' 4). כמובן שעלינו לזכור שכל עוד אנחנו בשלב ההתבוננות הראשוני על הנתונים נוכל להשאירם בתצורתם המקורית אך ברגע שנרצה להתחיל לבצע עליהם מבחנים סטטיסטיים כלשהם או מודלים של למידת עומק יהיה עלינו להפוך את המשתנים הטקסטואליים למספריים כאשר כל מספר יהיה תגית לקטגוריה ספציפית.*  תמונה מס' 4: סטטיסטיקה תיאורית- דאטה קטגוריאלי  *לעומת זאת, מבחינת הסטטיסטיקה התיאורית של הנתונים המספריים, אנו יכולים לראות כבר* *ממבט ראשוני את העובדות הבאות (תמונה מס' 5):*   1. *טווח השנים שהדוח מתייחס אליו הוא 2005-2017.* 2. *המדינה עם ציון מדד האושר הגבוה ביותר היא בעלת דירוג של 8.018934 והמדינה בעלת ציון מדד האושר הנמוך ביותר היא בעלת דירוג של 2.661718.* 3. *במשתנה תפיסת השחיתות (Perceptions of corruption) לפחות 75% מהתצפיות מאמינות שקיימת שחיתות כלשהי בפוליטיקה ובעסקים במדינה (אחוזון 25 עומד על 0.697359).* 4. *במשתנה השפעה חיובית (Positive affect) לפחות 75% מהתצפיות אמרו שחוו אושר, שמחה וצחוק (אחוזון 25 עומד על 0.621471).* 5. *יתרה מכך, רוב התצפיות מסרו כי לא חוו רגשות דאגה, עצב או כעס במסגרת המשתנה השפעה שלילית (Negative affect) (אחוזון 75>= 0.311515).* 6. *פחות מ-25% מהתצפיות מסרו כי אין להם מישהו קרוב להישען ולסמוך עליו בעת צרה (Social support) (<0.748304).*   *תמונה מס' 5: סטטיסטיקה תיאורית חלקית- דאטה נומרי*  ***בחינת התפלגויות ושכיחויות המשתנים:***   * ***משתנים קטגוריאליים:***   *כעת נבחן כיצד מפלגים המשתנים השונים בסט ונתחיל מהמשתנים הקטגוריאליים.*  *מדינה (country):*  *מבחינה ראשונית של שכיחות המדינות בסט* *הנתונים אנו מבינים כי המדינות מופיעות כמספר הפעמים שהן נדגמו לאורך השנים. כלומר, אם מדינה נדגמה כל שנה בטווח של 2005-2017 אז היא תופיע בסט 12 פעמים (תמונה מס' 6). מכאן אנו יכולים להבין כי לא כל המדינות נדגמו בכל טווח השנים.*        *תמונה מס' 6: גרף שכיחות המדינות*  *כבר כעת נוכל לראות כי בהמשך הדרך, כאשר נרצה לבחון קשרים מרובים בין המשתנים, ככל הנראה ניתקל בבעיה כאשר נרצה לערב את משתנה המדינה. מסיבה זו, על מנת להקל על עבודתנו ניצור משתנה חדש עבור היבשת של המדינה המדוברת ונבחן האם קיימות מגמות מולו.*    *יבשת (continent):*  *בעזרת ספריית פייטון מתאימה, יצרנו משתנה קטגורי חדש עבור יבשות במסד. כבר ממבט ראשוני על משתנה זה בתמונה מס' 7 נוכל לראות כי רוב המדינות המשתתפות בדוח משתייכות ליבשות אסיה (n=480) ואירופה (n=420) ומנגד היבשות עם הכי מעט מדינות המשויכות אליהן הן דרום אמריקה (n=120) ואוקיאניה (n=22).*  תמונה מס' 7: שכיחות היבשות במסד  *שנה (year):*  *אומנם מדובר על משתנה מספרי אך לחלוטין ניתן להתייחס אליו כמשתנה קטגורי בעל סדר בו כל שנה מהווה קטגוריה נפרדת. מתמונה מס' 8 ניתן לראות כי לפי סדר השנים השנה שהתקבלו בה הכי הרבה דוחות היא 2011 (146 מדינות) ואילו השנה עם הכי מעט דוחות היא 2005 (27 מדינות). שוב, גם גרף זה מחזק את העובדה שלא כל המדינות נדגמו בכל השנים אך ניתן לראות שמשנת 2011 מגמת הכמות יחסית יציבה. כמו כן, כפי שראינו, במסד קיימות 164 מדינות שונות ולכן אנו מבינים שגם בשנים עם כמות דגימות דומה, מדינות שונות נבדקו.*    *תמונה מס' 8: גרף שכיחות השנים*   * ***משתנים נומריים:***   *כעת נעבור למשתנים המספריים במסד הנתונים שנראה לבחון את צפיפות התפלגותם.*    *רמת שביעות הרצון מהחיים (Life Ladder):*  *מהסתכלות על אופן התפלגות המשתנה המתאר את דירוג שביעות הרצון מהחיים ניתן לראות כי ההתפלגות יחסית נורמאלית בצורתה פרט לחריגה קטנה סביב ערך 7~ בה יש עליה בצפיפות הערכים. מה שאומר שיש עלייה בכמות התצפיות שם (תמונה מס' 9).*  תמונה מס' 9: התפלגות צפיפות שביעות הרצון    *לוג התמ"ג לנפש (Log GDP per capita):*  *מהתבוננות בתמונה מס' 10 נוכל לראות את צפיפות התפלגות התוצר המקומי הגולמי לנפש במסד.* *נוכל לראות כי הנתונים מתפלגים באופן כמעט ונורמאלי, פרט לשלושה ערכי מקסימום, המקסימום הגלובלי בנק' (0.31 ,9.62) ושתי נקודות מקסימום לוקאליות אחת סביב (0.17 ,7.5) והשנייה סביב (0.2 ,10.5). גרף זה מראה לנו לפי מדד מדינה-שנה איפה קיימות הכי הרבה תצפיות בהתאם לצפיפות שלהן ומראה כי ברוב התצפיות לו"ג התמ"ג לנפש נמצא סביב בין 8.6-10.5.*  תמונה מס' 10: התפלגות צפיפות לו"ג התמ"ג לנפש  *תמיכה חברתית (Social support):*  *מבחינת צפיפות התפלגות משתנה התמיכה החברתית נוכל לראות כי היא אסימטרית שלילית (=בעלת צידוד שלילי), כלומר למשתנה ישנה נטייה לקבל ערכים גבוהים ורחוקים מהממוצע יותר מאשר ערכים נמוכים ממנו. בהתפלגות מסוג זה השכיח > חציון> ממוצע. מתמונה מס' 11 ניתן ללמוד כי ברוב המדינות לפי השנים אנשים מסרו כי מרגישים שיש להם תמיכה חברתית (0.75-1) ומס' מועט של מדינות הרגישו שלא. נוסף על כך, נוכל לראות כי ישנה קפיצה קטנה בצפיפות סביב ערכי 0.3, דבר המעיד על כמות מדינות-שנים עולה באותו ערך. פרט זה יכול להיות מקושר לחריגות בנתונים שאין להסיק מהן על מגמה כללית במשתנה אך בשל העובדה שממחינת החריגים במסד לא עלו ערכים חריגים באופן מהותי ייתכן ומדובר במדינות/ שנים ספציפיות שמגמה זו נפרדת להן כקבוצה באופן משמעותי ובהמשך נבחן זאת יותר לעומק.*  תמונה מס' 11: התפלגות צפיפות התמיכה החברתית  *תוחלת חיים בריאה משלב הלידה (Healthy life expectancy at birth):*  *מהסתכלות ראשונית של התפלגות זו נוכל לראות כי גם היא שואפת להיות נורמאלית אך היא איננה וקיימים בה מס' שיאים לוקאליים במקום נקודת מקסימום גלובאלית אחת (בדומה ללו"ג התמ"ג). נק' המקסימום הגלובאלית נמצאת ב-(0.06 ,64.27) ושתי לוקאליות נוספות נמצאות משני צדדיה כאשר זו שמימינה בעלת ערך גבוה יותר מהשמאלית לה. כלומר, ברוב התצפיות לפי מדינה-שנה תוחלת החיים הבריאה דורגה גבוהה למדי מרגע הלידה של תינוק (60-75). לעומת זאת, קיימת מקסימום לוקאלית שמאלית לפיה תוחלת החיים בלא מעט מדינות לפי שנים נמוכה יותר (סביב 51) (תמונה מס' 12). חלוקה ברורה זו לנק' מקסימום כה נפרדות בערכיהן עלולה להעיד כי בקרב מדינות מסוימות/ שנים מסוימות באופן גורף תוחלת החיים הבריאה נמוכה יותר מאשר באחרות, דבר העלול ליצור פילוג מובהק שיהיה עלינו לברר בהמשך.*  תמונה מס' 12: התפלגות צפיפות תוחלת החיים הבריאה משלב הלידה  *החופש לבצע בחירות בחיים (Freedom to make life choices):*  *ב**בחינת תמונה מס' 13 נבחין שחזרנו להתפלגות נורמאלית יותר אך עדיין אסימטרית שלילית. כלומר, ברוב המדינות לפי שנים עלה כי קיימת תחושה גבוהה שישנו החופש והשליטה העצמית של הפרט לבצע בחירות באופן חופשי בחייו (שכיח עומד על כמעט 0.8). למרות המגמה האסימטרית השלילית החזקה, קיימת משמאל לנק' המקסימום נק' פיתול שיכולה להיחשד כמקסימום לוקאלית (סביב 0.63) מה שמעיד על כמות מדינות-שנים גבוהה שמדד זה דורג אצלם נמוך יותר, כלומר קיימת שם תחושה של פחות חופש לביצוע בחירות בקרב הפרטים. אך ככלל, רוב המדינות לאורך השנים חשות כי יש לפרטים בהן חופש גדול לבחור את דרך החיים שלהן בצורה עצמאית.*  תמונה מס' 13: התפלגות צפיפות החופש לבצע בחירות בחיים  *נדיבות (Generosity):*  *בתמונה מס' 14 נוכל לראות את אופן התפלגות צפיפות דירוג מדד הנדיבות והמגמה לתרום לצדקה בכל צמד מדינה-שנה. משתנה זה מפלגת באופן אסימטרי חיובי (=בעל צידוד חיובי), כלומר למשתנה ישנה נטייה לקבל ערכים נמוכים ורחוקים מהממוצע יותר מאשר ערכים גבוהים ממנו. בהתפלגות מסוג זה השכיח < חציון< ממוצע כך שהצפיפות הגבוהה ביותר של תצפיות קיימת סביב מדד נדיבות של 0.03- את שאר כמות התצפיות מתפזרת בעיקר בערכים הגבוהים יותר. מכך נסיק כי רוב צמדי המדינה-שנה דירגו משתנה זה בערכים נמוכים-בינוניים המעידים על רצון לא גבוהה לתת צדקה. אך מגרף זה נוכל לראות שני פרטים מעניינים ביותר. הראשון הוא שישנן מדינות שמדד זה הינו שלילי עבורן, כלומר לא רק שלא תרמו אלא הם נגד או לחילופין היו רוצים לתרום אך אין ביכולתם. מהצד השני נראה שיא לוקאלי קטן סביב 0.65 המעיד על קומץ קטן יחסית של מדינות התורמות יותר מהשאר באופן שעלול להיות חריג או לחילופין מגמה נפרדת שתיבדק בהמשך.*  תמונה מס' 14: התפלגות צפיפות נדיבות הפרט    *תפיסת השחיתות (Perceptions of corruption):*  *משתנה תפיסת השחיתות בממשלה ובעסקים במדינה גם הוא בעל נטייה להתפלגות אסימטרית שלילית. רוב צמדי מדינה-שנה נמצאים בצפיפות גבוהה להפליא סביב השכיח (3.27 ,0.85) ומטה מה מעיד על מדינות רבות בהן ישנה תחושה גבוהה של שחיתות בממסד. קיימת עלייה נוספת בצפיפות סביב ערך של 0.4 המעיד על שבירת מגמת הירידה המתונה בשכיחות תחושת השחיתות עם כמות גדולה יותר של מדינות בעלות דירוג דומה סביב 0.4 (תמונה מס' 15).*  תמונה מס' 15: התפלגות צפיפות תפיסת השחיתות    *השפעה חיובית (Positive affect):*  *מדד ההשפעה החיובית, המתאר את חוויית רגשות חיוביים, כמעט ושואף להתפלגות נורמאלית פרט לעובדה במקום נק' מקסימום אחת בערך החציון יש לו שתיים כאלה. אחת בעלת צפיפות נמוכה יותר (2.6 ,0.61) ושנייה בעלת צפיפות גבוהה יותר (3.4 ,0.81). ברוב המדינות לפי שנים נראה כי קיימת תחושה יחסית גבוהה של הרגשה חיובית כאשר כמות תצפיות נמוכה ביותר דורגה מתחת ל-0.5. מכאן נסיק שבאופן גורף הרוב המוחלט של המדינות דירג מדד זה באופן גבוה התואם את הממוצע שראינו (0.7~) (תמונה מס' 16).*  תמונה מס' 16: התפלגות צפיפות ההשפעה החיובית  *השפעה שלילית (Negative affect):*  *מבחינת התפלגות מדד ההשפעה השלילית נראה כי הוא כמעט לסימטרי (נורמאלי) אך קיים צידוד חיובי יחסית קטן סביב ערך 0.5. רוב המדינות דרגו מדד זה בין 0.2-0.3, כלומר דרגו נמוך את תחושת הימצאות רגשות שליליים בחייהם כך שמגמה זו תואמת את מדד ההשפעה החיובית הגבוה שראינו קודם לכן. למרות זאת, קיימת שבירת מגמת ירידת הצפיפות סביב 0.5 מה שמעיד על יותר מדינות שנתנו דירוג גבוה יותר למדד זה. שוב, אומנם לא מצאנו ערכים חריגים בבדיקה הראשונית אך עלינו לבחון מקרוב את הפרט הזה על מנת לוודא כי אכן לא מדובר בתצפיות שעלולות להימצא כחריגות (תמונה מס' 17).*  תמונה מס' 17: התפלגות צפיפות ההשפעה השלילית    *אמון בממשל הלאומי (Confidence in national government):*  *בתמונה מס' 18 נוכל לראות התפלגות כמעט וסימטרית של אמון הפרטים בממשל הלאומי. נראה כי מחצית התצפיות נתנו דירוג גבוה למידת האמון שלהם בממשל ומחצית התצפיות נתנו דירוג נמוך יותר למידת אמונם בממשל עם ממוצע שמחזק את החלוקה הכמעט שוויונית (mean=0.48).*  תמונה מס' 18: התפלגות צפיפות אמון בממשל הלאומי  *איכויות דמוקרטיות (Democratic Quality):*  *מבחינת התפלגות צפיפות מדד האיכויות הדמוקרטיות נוכל לראות כי הנתונים לא מתפלגים בצורה אחידה או בעלת מיקוד ספציפי. עם ממוצע שלילי (0.12-) ושתי נקודות בעלות צפיפות גבוהה יותר ע"ג ההתפלגות כולה, אחת נמוכה מהממוצע והשנייה חיובית וגבוהה ממנו, נראה כי על פניו קיים רוב שחש שאיכות הדמוקרטיה במדינתם נמוכה למדי ואיננה פועלת כראוי. מצד שני קיים רוב נוסף, קטן יותר בממדיו שדירג את איכות הדמוקרטיה במדינתו גבוה יותר (1.2~). בהמשך נרצה לבחון האם חלוקה זו בדירוג איכות הדמוקרטיה אכן ברת קטלוג באופן פרטני למדינות/שנים מסוימות (תמונה מס' 19).*  תמונה מס' 19: התפלגות צפיפות איכויות דמוקרטיות  *איכות השלטון (Delivery Quality):*  *גם משתנה תחושות האנשים לגבי איכות השירותים הציבוריים והממשלתיים במדינתם לא מתפלגת באופן סימטרי. קיימות שלוש נקודות מקסימום בהתפלגות זו המהוות צפיפות נתונים גבוהה לארוך הגרף הנעות בין (-0.6)-1.8. נתון זה מראה לנו את תנודתיות דירוג מדד זה. מחצית מצמדי המדינה-שנה דירגו שלילית את המדד (median= -0.21) עם צפיפות גבוהה ביותר סביב 0.6- מה שמעיד על הסכמה יחסית רחבה בחציון זה. לעומת זאת, בקרב המחצית השנייה שנתנה דירוג גבוהה יותר למדד נראה כי התפלגות הולכת למתרחבת עם שתי נק' צפיפות גבוהה יותר, אחת סביב 0.8 והשנייה סביב 1.8. גם פה נרצה לבחון האם מגמה זו תראה באופן מובהק אל מול משתנים אחרים במסד (תמונה מס' 20).*  תמונה מס' 20: התפלגות צפיפות איכות השלטון |
| bivariate analysis |  | בשל העובדה שמשתנה המטרה במסד זה הוא רמת שביעות הרצון הממוצעת מהחיים בכל מדינה לפלי שנה, נרצה תחילה לבחון את הקשר שלו אל מול משתנים אחרים במסד.  שביעות רצון מהחיים אל מול המדינות:  על מנת לבחון את הקשר בין שני משתנים אלו נבצע boxplot כיוון שמדובר במשתנה קטגורי ומשתנה נומרי. בתמונה מס' 21 נוכל לראות את ממוצע הדירוג של כל מדינה בסדר יורד, את החציון שלה, את החלוקה לרבעים והאם קיימים ערכים חריגים. בשל גודלו של הגרף, נציג כאן רק את המדינות המדורגות הכי גבוהה והמדינות המדורגות הכי נמוך. הדבר הראשון שקופץ לעין בצורה מובהקת הוא השוני ברוחב הקופסה. בקרב המדינות במקומות הראשונים ניתן לראות קופסאות צרות למדי. כלומר, ככל שהמדינה קיבלה דירוג יותר גבוה, הציון שהיא קיבלה היה כמעט עיקבי לחלוטין בכל השנים שהדוח נלקח. לעומת זאת, בקרב המדינות שדורגו הכי נמוך, הקופסאות רחבות יותר ברוב המקרים ולכך ניתן להסיק שבשנים שבוצע בהן סקר האושר היו פערים בדירוג הניתן. חשוב לציין שכאשר בוחנים את הגרף המלא עובדה זו קופצת לעין יותר ונראה כי ככל שדירוג המדינה יורד בממוצע, כך קופסתה מתרחבת. נוסף על כך, אחוזונים 25 ו-75 מתרחבים באופן ברור לעין ככל שדירוג המדינה יורד.  תמונה מס' 21: boxplot עבור הקשר בין מדינה למדד האושר  מגרף זה, אנו יכולים להסיק כי קיים קשר יחסית חזק בין שני משתנים אלו.  שביעות רצון מהחיים מול יבשות:  תמונה מס' 22 מציגה את הציון הממוצע של "Life Ladder" לפי יבשה. ניתן לראות את ההבדלים בשביעות רצון מהחיים בין היבשות השונות, כאשר אוקיאניה (שבה נכללות מדינות כמו אוסטרליה וניו זילנד) מובילה את הרשימה, בעוד שאפריקה נמצאת במקום האחרון.  תמונה מס' 22: גרף הקשר בין היבשות למדד האושר  ממוצע שביעות רצון מהחיים מול שנים:  הגרף מראה את השינוי בממוצע שביעות הרצון מהחיים במהלך השנים. ניתן לראות כי הממוצע של שביעות הרצון מהחיים עולה ויורד באופן דינמי במהלך השנים (תמונה מס' 22).  תמונה מס' 23: גרף שנים למדד האושר    שביעון רצון מהחיים אל מול תוחלת חיים בריאה:  עוד זוג משתנים שאנו יכולים לחשוב שיהיה ביניהם קשר יחסית חזק מהיגיון בריא הן שביעות הרצון ותוחלת חיים. ואכן מהתבוננות בתמונה מס' 24 ניתן לראות את הפיזור שלהם ואת הצפיפות שלהם וניתן להבחין כי אכן קיימת מגמה חיובית וחזקה בין שני משתנים אלו. כלומר, ככל שאחד מהם עולה בערכו, כך גם השני.  תמונה מס' 24: גרף הקשר בין תוחלת חיים בריאה למדד האושר  שביעות רצון מהחיים אל מול החופש לבצע בחירות:  משתנה נוסף שנוכל לחשוד שיהיה בעל מתאם כלשהו למשתנה המטרה שלנו הוא המשתנה שבוחן את הרגשתו של הפרט לגבי החופש לו לבצע בחירות בחייו. משתנה זה לפי ההיגיון יהיה בעל קשר חזק למשתנה המטרה שלנו הרי אין זה הגיוני שאדם שמרגיש שאין לו שליטה על קבלת ההחלטות בחייו יהיה מאושר, אך מהתבוננות בתמונה מס' 25 נוכל לראות כי הקשר אכן קיים אך פחות חזק מששיערנו.  תמונה מס' 25: גרף הקשר בין החופש לקבלת החלטות למדד האושר  שביעות רצון מהחיים אל מול לוג התמ"ג לנפש:  בתמונה מס' 26 ניתן לראות שיש קשר חיובי בין שביעות הרצון מהחיים לבין התוצר המקומי הגולמי לנפש: ככל שהתוצר המקומי הגולמי לנפש במדינה גבוה יותר, יש הטיה שגם שביעות הרצון מהחיים באותה המדינה תהיה גבוהה יותר. נקודה זו מלמדת שבמדינות עם תוצר מקומי גולמי לנפש גבוה יש הטייה לראות שביעות רצון גבוהה יותר מהחיים. כמובן, ישנם גם יוצאים מהכלל, ולא כל המדינות הן באותו המגמה.  תמונה מס' 26: גרף הקשר בין לוג התמ"ג לנפש למדד האושר  שביעות רצון מהחיים מול נדיבות:  תמונה מס' 27 מראה פיזור של הנתונים, אך הקשר אינו ברור כל כך כמו בגרפים הקודמים. כלומר, אין קשר חיובי או שלילי בין שביעות הרצון מהחיים לבין הנדיבות במדינה. יש מדינות שבהן הנדיבות גבוהה ושביעות הרצון מהחיים גבוהה, אך ישנן גם מדינות בהן הנדיבות נמוכה ושביעות הרצון גבוהה, ולהפך.  תמונה מס' 27: גרף הקשר בין מדד הנדיבות למדד האושר  שביעות רצון מהחיים מול כל המדדים השונים:  את בחינת המתאם הסטטיסטי בין משתנה המטרה שלנו לשאר המשתנים במודל נבחן גם גרפי פיזור משותפים (תמונה מס' 29) וגם גרף מפת חום (תמונה מס' 28) המתארת את הקורלציה בין שביעות רצון מהחיים לבין כל המדדים השונים במאגר הנתונים בצורת המתאם עצמו במספר. מגרפים אלו נוכל לראות את מידת הקשר והפיזור שלו בין שביעות הרצון מהחיים לבין המדדים השונים. הצבעים החמים (לקראת האדום) מייצגים קורלציה חיובית, והצבעים הקרים (לקראת הכחול) מייצגים קורלציה שלילית. הערכים בגרף הם ערכי הקורלציה, כאשר ערך של 1 או -1 הוא הקורלציה המרבית האפשרית (חיובית או שלילית, בהתאמה). כך, לדוגמה, ניתן לראות שיש קורלציה חיובית בינונית בין התמיכה החברתית לבין שביעות הרצון מהחיים (ערך של כ-0.7), וזה אומר שכאשר התמיכה החברתית גבוהה, יש הטיה שגם שביעות הרצון מהחיים תהיה גבוהה.    תמונה מס' 29: פיזור הקשר בין משתני המודל  תמונה מס' 28: מפת חום עבור קורלציות בין משתני המודל למשתנה המטרה  באופן כללי יותר, בתמונה מס' 30 נוכל לראות את כלל הקורלציות בין כל המשתנים הנומריים במודל. נראה כי בקרב המשתנים המסבירים לו"ג התמ"ג לנפש ואיכות השלטון בעלי מתאם חיובי גבוה (=0.77), איכות הדמוקרטיה ואיכות השלטון בעלי מתאם חיובי גבוה (=0.86), לו"ג התמ"ג לנפש ותוחלת חיים בריאה בעלי מתאם חיובי גבוה (=0.84). מנגד, נראה כי לא קיימים קשרים שליליים בעלי מתאם חזק משמעותית (>-0.6) כאשר המתאם השלילי הגבוה ביותר הוא בין תפיסת השחיתות לבין איכות השלטון (=-0.51).  תמונה מס' 30: מפת חום עבור קורלציות בין כלל משתני המודל |
| multivariate analysis |  | כעת נסתכל על גרפים מרובי משתנים על מנת להסיק מידע נוסף על מהות הקשרים ביניהם.  תוחלת חיים, לו"ג התמ"ג ויבשת:  הגרף מצביע על מתאם חיובי בין התמ"ג לנפש של מדינה לבין תוחלת החיים הבריאה שלה בלידה. המשמעות היא שבמדינות בעלות מעמד כלכלי גבוה יותר, אנשים צפויים גם לחיות שנים רבות יותר בבריאות טובה. השפעת הכלכלה על הבריאות יכולה לנבוע מהשקעות גבוהות יותר במערכת הבריאות, תשתיות, חינוך ועוד במדינות עשירות יותר.  תוחלת החיים, לו"ג התמ"ג ושנה באפריקה מול לא אפריקה:  מדינות אפריקה, לרוב, מתקבצות בקצה התחתון של שני הצירים. משמעות הדבר היא שבדרך כלל יש להם תמ"ג נמוך יותר לנפש ותוחלת חיים בריאה נמוכה יותר בלידה בהשוואה למדינות רבות שאינן אפריקניות.  מדינות לא אפריקאיות מציגות פריסה רחבה יותר על פני הגרף, כאשר רבות מתקבצות לקצה הגבוה של שני הצירים, מה שמצביע על תמ"ג גבוה יותר ותוחלת חיים בריאה ארוכה יותר.  הגרף מאשש מגמה כללית שנראית בנתונים העולמיים: מדינות עם תוצר גבוה יותר נוטות לתוחלת חיים בריאה גבוהה יותר. עם זאת, ההבחנה בין מדינות אפריקה למדינות שאינן אפריקניות מדגישה את הפערים הכלכליים והבריאותיים בין אזורים אלו.  לו"ג התמ"ג, שביעות רצון מהחיים ויבשת:  התרשים מדמה את הקשר בין שביעות רצון מהחיים לבין לו"ג התמ"ג, לפי צבע על בסיס היבשת. ניתן לראות שקיים מתאם חיובי, כלומר מדינות שיש תוצר לנפש גבוה יותר (מדינות עשירות יותר), אזרחיהן נוטים לדווח על שביעות רצון גבוהה יותר מהחיים. ישנה מגמה כללית של מדינות עשירות יותר עם שביעות רצון גבוהה יותר מהחיים.  לו"ג התמ"ג, שביעות רצון מהחיים ומדינה (2017) Life Ladder:  התרשים מדמה את הקשר בין הכנסה לנפש בין האושר או שביעות הרצון הכללית של אנשים ברחבי מדינות שונות בשנת 2017. הצבעים בתרשים מציינים את אינדקס המדינה במערך הנתונים, כך שכל מדינה בעל צבע ייחודי. התרשים מצביעה על קשר בין הכנסה לנפש לאושר הכללי של אנשים במדינה, אם כי יש יוצאי דופן והקשר אינו מוחלט ,כגון הונג קונג ובוטסוואנה ששביעות הרצון שלהם גבוהה ביחס ללו"ג . |
| ריכוז תובנות, טיפול בדאטה והצגת הנרטיב הכללי |  | בעקבות התבוננות העומק בהתפלגות המשתנים עצמם ובקשרים שלהם עם המשתנים האחרים במסד, תחילה נבחר לטפל בערכים החסרים שעדיין קיימים בחלק מהמשתנים.  **טיפול בערכים חסרים:**  לאור העובדה שבמסד קיימות סה"כ 1,562 תצפיות ובקרב כל המשתנים בעלי הערכים החסרים מדובר על פחות מ-180 חסרים, החלטנו לא להשמיט אף משתנה בכללותו מהמסד אלא לבחון אלו דרכים להשלמת החסרים יהיו הכי מתאימות עבור כל משתנה.  כבר בתמונה מס' 6 הבוחנת את שכיחות משתנה המדינה, ראינו כי קיימות מס' מדינות שמופיעות במסד רק פעם אחת.  נציין כי את כלל הערכים החסרים השלמנו בעזרת חציון עבור פרויקט זה כיוון שעיקרו הוא הצגת הוויזואליזציות וניתוחן ופחות העבודה עם הנתונים עצמם והכנתם למודל למידה. אם זאת, עבור כל משתנה בנפרד החלטנו להסביר מראש מה הייתה יכולה להיות הדרך הנכונה להשלמת הערכים בו בהתאם להתפלגותו.  ערכים חסרים במשתנה לו"ג התמ"ג לנפש (Log GDP per capita):  כפי שראינו קודם לכן, במשתנה זה קיימים 27 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= 9.398). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 10 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, לדוגמא הקשר החזק בינו לבין משתנה תוחלת החיים הבריאה מלידה בעזרתו, דרך רגרסיה ליניארית יכולנו לחזות מה יהיה הערך המתאים ביותר עבור אותם משתנים חסרים.  ערכים חסרים במשתנה תמיכה חברתית (Social support):  כפי שראינו קודם לכן, במשתנה זה קיימים 13 ערכים חסרים. בשלב בחינת התפלגות צפיפות המשתנים, ראינו כי התמיכה החברתית מתפלגת באופן א-סימטרי עם צידוד שלילי. מסיבה זו, אנחנו רואות לנכון להשלים את הערכים החסרים במשתנה זה בערך החציון (median= 0.833) מאחר והשלמתו בממוצע עלולים להוביל להטיה חזקה יותר של הדאטה.  ערכים חסרים במשתנה תוחלת חיים בריאה משלב הלידה (Healthy life expectancy at birth):  כפי שראינו קודם לכן, במשתנה זה קיימים 9 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= 63.803). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 12 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, לדוגמא מול כל צמד מדינה-שנה על מנת לקבל החלטה משכילה יותר על אופן השלמתו. מבחינת קיבוץ המדינות לפי שנים ניתן לראות שתי מדינות שמשתנה זה כלל לא קיים עבורן (בכל הסקרים שמילאו הוא חסר)- Somaliland region ו- North Cyprusוהן משלימות באופן מלא את מס' הערכים החסרים במשתנה זה.  ערכים חסרים במשתנה החופש לבצע בחירות בחיים (Freedom to make life choices):  כפי שראינו קודם לכן, במשתנה זה קיימים 29 ערכים חסרים. בשלב בחינת התפלגות צפיפות המשתנים, ראינו כי חופש הבחירה מתפלג באופן א-סימטרי עם צידוד שלילי. מסיבה זו, אנחנו רואות לנכון להשלים את הערכים החסרים במשתנה זה בערך החציון (median= 0.74) מאחר והשלמתו בממוצע עלולים להוביל להטיה חזקה יותר של הדאטה.  ערכים חסרים במשתנה נדיבות (Generosity):  כפי שראינו קודם לכן, במשתנה זה קיימים 80 ערכים חסרים. בשלב בחינת התפלגות צפיפות המשתנים, ראינו כי מדד הנדיבות מתפלג באופן א-סימטרי עם צידוד קל חיובי. מסיבה זו, אנחנו רואות לנכון להשלים את הערכים החסרים במשתנה זה בערך החציון (median= -0.02) מאחר והשלמתו בממוצע עלולים להוביל להטיה חזקה יותר של הדאטה.  ערכים חסרים במשתנה תפיסת השחיתות (Perceptions of corruption):  כפי שראינו קודם לכן, במשתנה זה קיימים 90 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= 0.808). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 15 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, אך כפי שראינו בתמונת הקורלציות אין למשתנה זה קשרים חזקים עם אחרים.  ערכים חסרים במשתנה השפעה חיובית (Positive affect):  כפי שראינו קודם לכן, במשתנה זה קיימים 18 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= 0.717). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 16 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, אך כפי שראינו בתמונת הקורלציות אין למשתנה זה קשרים חזקים עם אחרים.  ערכים חסרים במשתנה השפעה שלילית (Negative affect):  כפי שראינו קודם לכן, במשתנה זה קיימים 12 ערכים חסרים. בשלב בחינת התפלגות צפיפות המשתנים, ראינו כי ההשפעה השלילית מתפלגת באופן כמעט באופן סימטרי אך עם מעט צידוד חיובי. מסיבה זו, אנחנו רואות לנכון להשלים את הערכים החסרים במשתנה זה בערך החציון (median= 0.25) מאחר והשלמתו בממוצע עלולים להוביל להטיה חזקה יותר של הדאטה.  ערכים חסרים במשתנה אמון בממשל הלאומי (Confidence in national government):  כפי שראינו קודם לכן, במשתנה זה קיימים 161 ערכים חסרים. בשלב בחינת התפלגות צפיפות המשתנים, ראינו כי מדד האמון בממשל הלאומי מתפלג באופן כמעט באופן סימטרי אך עם מעט צידוד חיובי. מסיבה זו, אנחנו רואות לנכון להשלים את הערכים החסרים במשתנה זה בערך החציון (median= 0.463) מאחר והשלמתו בממוצע עלולים להוביל להטיה חזקה יותר של הדאטה.  ערכים חסרים במשתנה איכויות דמוקרטיות (Democratic Quality):  כפי שראינו קודם לכן, במשתנה זה קיימים 171 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= -0.225). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 19 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, לדוגמא הקשר החזק בינו לבין משתנה איכות השלטון בעזרתו, דרך רגרסיה ליניארית יכולנו לחזות מה יהיה הערך המתאים ביותר עבור אותם משתנים חסרים.  ערכים חסרים במשתנה איכות השלטון (Delivery Quality):  כפי שראינו קודם לכן, במשתנה זה קיימים 171 ערכים חסרים. משתנה זה הושלם באמצעות מילוי הערכים החסרים בערך החציון (median= -0.21). עם זאת, אם היינו בוחנים מראש את הדרך הטובה ביותר להשלמת הערכים בו, בשל צורת התפלגות המשתנה שראינו בתמונה מס' 20 ניתן לראות כי צורת השלמת הערכים החסרים לא תהיה סטנדרטית אלא תצטרך להישען על הקשרים שיש בינו לבין המשתנים האחרים, לדוגמא הקשר החזק בינו לבין משתנה איכות הדמוקרטיה בעזרתו, דרך רגרסיה ליניארית יכולנו לחזות מה יהיה הערך המתאים ביותר עבור אותם משתנים חסרים. |