פרוייקט רגרסיה ליניארית

## Q.3

מדינה (Country): משתנה קטגורי המייצג את שם המדינה. זה מאפשר סגמנטציה וניתוח לפי מדינות, ויכול להועיל בהבנת הבדלים תרבותיים, פוליטיים וכלכליים המשפיעים על תפוצת המגיפה ( מס' המקרים המאובחנים)

תוצר לאומי גולמי לנפש (GDP per Capita): משתנה רציף המייצג את רמת החיים והעושר במדינה. מדינות עם תמ"ג גבוה יותר יכולות להשקיע יותר בבריאות הציבור ובתשתיות, מה שעשוי להשפיע על יכולתן להתמודד עם המגפה ולהקטין את מס' המקרים המאובחנים. לעומת זאת במדינות עם תמ"ג נמוך נצפה לקבל הערכת חסר במס' המקרים האמיתיים ביחס למס' המקרים המדווח של המחלה (עקב מחסור בבדיקות וחוסר במודעות למחלה) כך שגם ההשפעה ההפוכה אפשרית.

צפיפות אוכלוסין (Population density): משתנה רציף המייצג צפיפות אוכלוסייה. צפיפות גבוהה יכולה להגביר את קצב ההדבקה ואת מספר המקרים המאובחנים (עקב יכולת ההדבקה הגבוהה של הוירוס בסביבות צפופות)

צריכת אלכוהול לנפש (Total Alcohol Consumption per Capita): משתנה רציף המודד כמות צריכת האלכוהול הממוצע. מחקר שנעשה ב2021] https://publichealth.jmir.org/2021/11/e33022/[מצא קשר בין צריכת אלכוהול לבין חומרת מחלת הקורונה. לפי המחקר, יתכן כי במדינות בהן צריכת האלכוהול גבוהה יותר, תופעות חמורות יותר של המחלה יופיעו בקרב חלק מהשותים, מה שעשוי להוביל לזיהוי רב יותר של מקרים באוכלוסייה. לעומת זאת, במדינות עם צריכה נמוכה יותר, תופעות המחלה עשויות להיות פחות חמורות ולכן חלק מהמקרים עשויים שלא להיבדק בכלל. יש לציין כי זהו מחקר בודד שמצא קשר זה, ולכן יש להתייחס אליו בזהירות.

שיעור התאבדויות (Suicide Rate): משתנה רציף המייצג את שיעור ההתאבדויות ל-100,000 איש. קשור לבריאות הנפש הכללית של האוכלוסייה, עשוי להשפיע על תגובות למשברים כמו המגפה. מחקר שנעשה ב2022 מצא קורלציה חיובית בין בריאות נפשית לאשפוז חוזר וחומרת מחלת הקורונה [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9239877/] לכן בדומה למשתנה מוות מעישון ניתן להניח שתסמינים חמורים יותר יובילו לאבחון רב של האוכלוסיה.ד

יבשת (Continent): משתנה קטגורי המייצג את היבשת בה ממוקמת המדינה. בדומה למשתנה המדינה יכול להצביע על הבדלי צפיפות, תרבות ועושר, הבדלים במאפינים אלו יובילו לתגובות שונות למגפה וכך ישפיעו על קצב ההדבקה ובנוסף גם לאבחון המקרים.

מוות מעישון (Death from Smoking): משתנה רציף המודד את מספר המוות מעישון. מעישון יש השפעה שלילית על בריאות הריאות, מגדיל את חומרת המחלה, וכך מוביל לעלייה במס' המקרים המאובחנים ] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10032583/[

חוקתיות סמים לרפואה (Drug Legalization Medical): משתנה בינארי המייצג את החוקתיות של שימוש רפואי בסמים. יכול להשפיע על גישה לטיפולים. בנוסף כנראה המשתנה קורלטיבי למוות מעישון, וכמו שראינו משתנה זה אכן יכול לתרום לעלייה בכמות המקרים המאובחנים

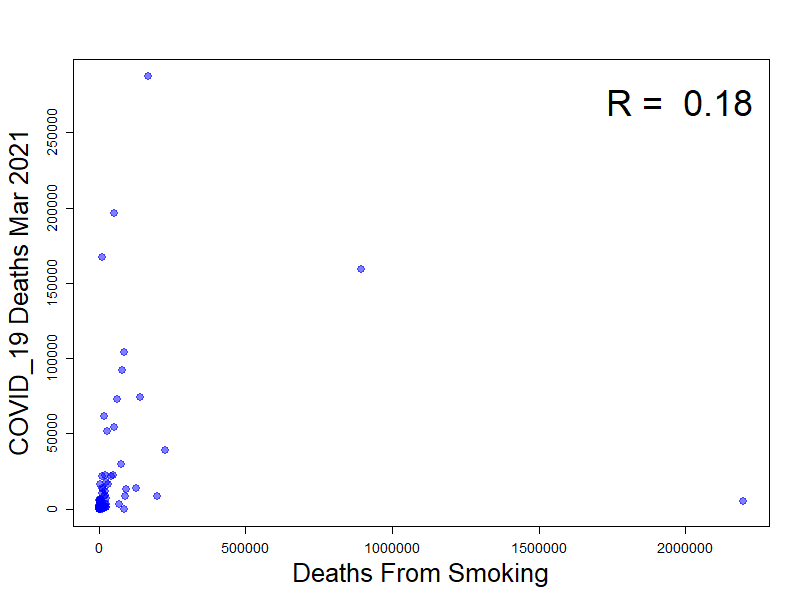
שירות צבאי חובה (Compulsory Military): משתנה בינארי המייצג האם ישנו שירות צבאי חובה. מצד אחד במדינות אלו אחוז גדול של אוכלוסייה צעירה חיה ביחד בצפיפות גבוהה, מצב שמעלה את הסיכוי להדבקה ולמקרים מאובחנים, מצד שני, מדינות אלו יכולות לתת מענה לאירוע חירום לאומי מתמשך כמו מגיפה ולהקטין מס' המקרים המאובחנים.

מקרי קוביד-19 (Covid 19 - Cases (20.3.2021)): משתנה רציף המודד את מספר המקרים נכון לתאריך המצויין. משמש כמשתנה המוסבר בניתוח זה.

מקרי מוות מקוביד-19 (Covid 19 -Death (20.3.2021)): משתנה רציף המודד את מספר המתים מקוביד-19 נכון לתאריך המצויין. ניתן להניח שמשתנה זה יהיה קורלטיבי מאוד עם המשתנה המוסבר, ככל שיותר אנשים מתים מהמחלה כך ניתן להניח שיש יותר אנשים שחולים

## Q.4

### 4.1 Deaths from smoking VS COVID-19 Death 03/2 1



ניתן לראות כי יש מתאם חיובי חלש בין מקרי מוות מעישון למקרי מוות מקורונה.קל לראות כי במדינות עם אוכלוסייה גדולה נראה מקרי מוות רבים הנובעים מעישון ובהתאמה מקרי מוות רבים מקורונה ולהפך לכן יש פה קשר נסיבתי של מאפיינים דומים על פי גודל האוכלוסיה. נוסף, כפי שהוזכר לעיל, עישון תורם להחמרת התסמינים במחלת הקורונה לכן במקרה זה, נניח סיבתיות בזהירות.

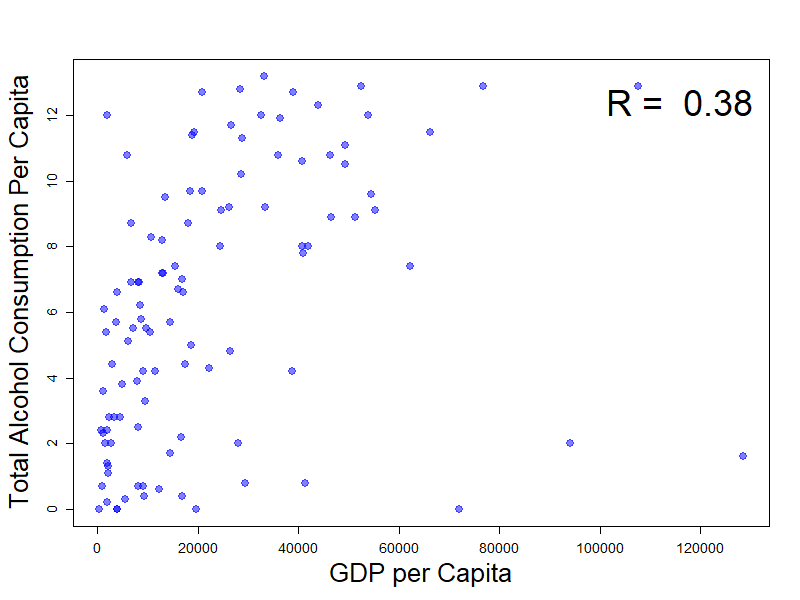
### 4.2 Population Density vs GDP Per Capita

A graph of a number of dots

Description automatically generated

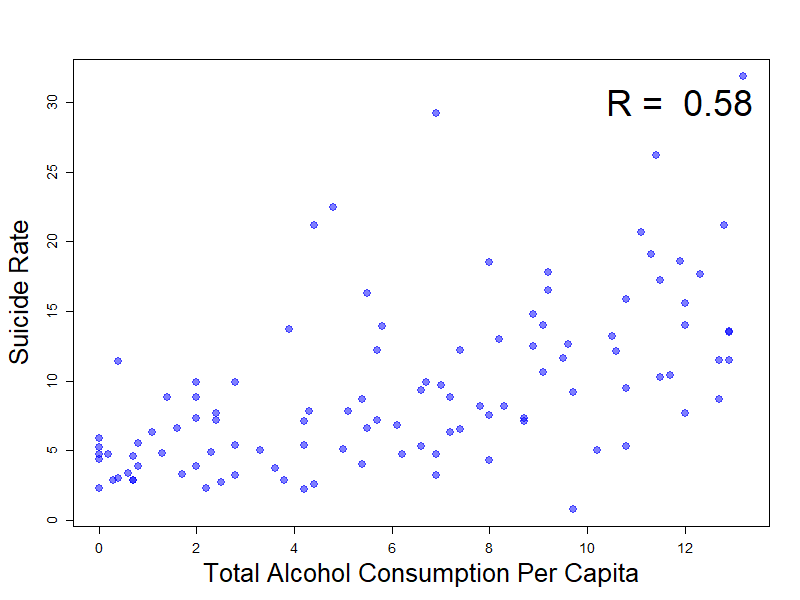
ניתן לראות כי קיים מתאם חיובי בין צפיפות האוכלוסייה ל-GDP לנפש. ניתן להניח כי מקשר זה הוא קשר קורלטיבי ואינו בהכרח מצביע על סיבתיות

### 4.3 GDP Per Capita VS Total Alcohol Consumption



ניתן לראות קשר חיובי בינוני בין ה-GDP למקרי מוות עקב אלכוהול. ניתן להניח כי אלכוהול נגיש יותר לאוכלוסייה במדינות מפותחות, בשל מחירו הגבוה, ולכן צריכתו תהיה גבוהה יותר במדינות עם GDP לנפש גבוה. בנוסף ניתן להניח הבדלים תרבותיים בצריכת האלכוהול ( למשל בתרבות המערב, צריכת אלכוהול נחשבת מקובלת, ולרוב מדינות אלו בעלות GDP גבוה)

### 4.4 Total Alcohol Consumption VS Suicide Rate



ניתן לראות כי קיים מתאם בינוני בין צריכת אלכוהול למקרי התאבדות. בהנחה שהמתאם מצביע על קשר קורלטיבי, ניתן להניח כי אלכוהול מוביל להתנהגות אימפולסיבית ושיפוט לקוי, מה שעשוי להגביר את הסיכון להתאבדות https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872355/

### 4.5 Total Alcohol Consumption VS COVID-19 Death 03/2 1

ניתן לראות מתאם נמוך בין צריכת אלכוהול למקרי מוות עקב קורונה באוכלוסיה. אף על פי שנראה קשר סיבתי נמוך, כפי שצוין לעיל, צריכת אלכוהול עלולה להחמיר את התסמינים של מחלות שונות, וכתוצאה מכך ייתכן והסיכוי למוות עולה, מה שעשוי להוביל למקרי מוות רבים יותר.

### GDP per Capita VS COVID-19 Death 03/21

### 

היינו מצפים להניח כי קיים קשר שלילי בין GDP per capita למקרי מוות, בפועל, קיבלנו מתאם נמוך בין המשתנים נוכל להניח אם משתנה מקרי המוות היו מנורמל לשיעור תמותה היינו מקבלים קורלציה חיובית בין המשתנים

## Q.5 . ניתוח תיאורי של המשתנים:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **mean** | **median** | **se** | **IQR** | **skewness** |
| **GDP\_Per\_Capita** | 22988.5 | 16064.0 | 23495.0 | 26934.5 | 1.8 |
| **Population\_Density** | 233.0 | 82.2 | 799.6 | 148.9 | 8.5 |
| **Total\_Alcohol\_Consumption\_Per\_Capita** | 6.2 | 6.2 | 4.1 | 7.2 | 0.1 |
| **Suicide\_Rate** | 9.4 | 7.8 | 6.1 | 7.7 | 1.2 |
| **Deaths\_From\_Smoking** | 50996.8 | 5898.3 | 229441.4 | 19368.9 | 8.1 |
| **COVID\_19\_Cases\_Mar\_20\_2021** | 839219.8 | 149135.0 | 2699573.8 | 472742.0 | 5.9 |
| **COVID\_19\_Deaths\_Mar\_20\_2021** | 16535.6 | 2106.0 | 42552.9 | 8515.5 | 4.1 |

תוצר לאומי גולמי לנפש (GDP\_Per\_Capita):

ניתן לראות פיזור רחב של ערכים, כאשר סטיית התקן גבוהה מעידה על שונות רחבה בין המדינות בהקשר לתוצר הלאומי הגולמי לנפש. זה משקף את ההבדלים הכלכליים הגדולים בין מדינות שונות.ע"פ צידוד גבוה ניתן להסיק כי יש מספר מדינות עם תוצר גבוה במיוחד שמשפיע על התפלגות הנתונים.

צפיפות אוכלוסיה (Population\_Density):

הפיזור הרחב (סטיית תקן גבוהה) מראה שיש הבדלים גדולים בצפיפות האוכלוסיה בין מדינות שונות. נטייה גבוהה מצביעה על כך שרוב המדינות יש להן צפיפות נמוכה יחסית, אך ישנם מדינות עם צפיפות גבוהה במיוחד. הצידוד גבוה במיוחד, ניתן להסיק אם כך שיש מס' קטן של מדינות עם צפיפות מאוד אוכלוסיה גדול מאוד בהשוואה לממוצע צפיפות האוכלוסין

צריכת אלכוהול לנפש (Total\_Alcohol\_Consumption\_Per\_Capita):

ניתן לראות כי התפלגות המשתנה הינה בקירוב נורמלית, הממוצע והחצין מלכדים, והצידוד נמוך מאוד. בנוסף במשתנה בעל פיזור צר, ניתן לראות זאת טווח הבין רבעוני ובסטיית התקן הנמוכים

שיעור ההתאבדויות (Suicide\_Rate):

ערכי החציון והממוצע די קרובים זה לזה, כך שניתן להעריך שבקירוב המשתנה מתפלג נורמלית. בנוסף ערך הצידוד חיובי, ניתן להסיק אם כך כי יש מס' מדינות אם שעור התאבדות גבוה יחסית ממוצע האוכלוסיה

שירות צבאי חובה ( משתנה קטגוריאלי) VS COVID\_19\_Cases\_Mar\_20\_2021

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Compulsory\_Military\_Service** | **mean** | **median** | **sd** | **skewness** | **IQR** |
| 0 | 927791.1 | 149513 | 3060668 | 5.611248 | 438868.5 |
| 1 | 684787.8 | 119528 | 1942581 | 4.824534 | 477656.5 |

ניתן לראות כי ממוצע וחציון המקרים גבוה יותר במדינות אשר לא קיים בהם שירות צבאי חובה, טווח הערכים וההתפלגות דומים בשתי הקבוצות ( התפלגות נוטה ימינה) בנוסף מעניין לראות את סטיית התקן הגדולה במס' אירועי קורונה במדינות ללא שירות צבאי חובה, בהשוואה למס' המקרים במדינות עם שירות צבאי חובה.

## Q. 6. ניתוח חריגים

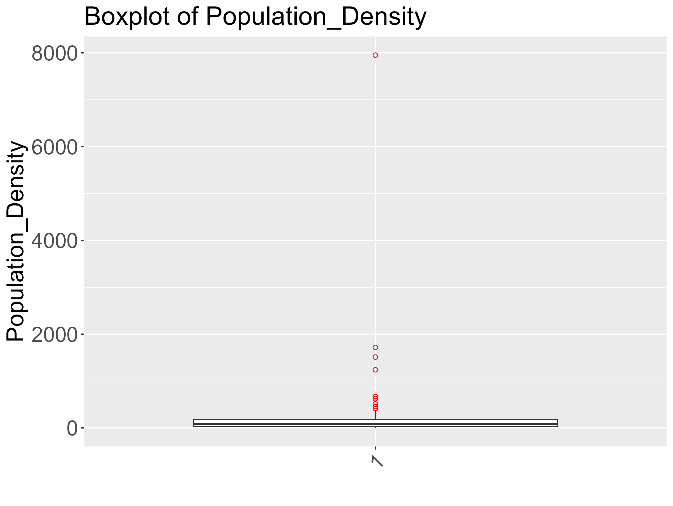
A graph with a line and a bar

Description automatically generated with medium confidence

תוצר לאומי גולמי לנפש (GDP per Capita):

נראה שיש ארבע תוצאות חריגות. בהתחשב בציפייה לפיזור רחב של תוצר לאומי גולמי לנפש, ייתכן שנראה גם תוצאות קיצוניות. במקרה זה, לא הייתי ממליץ להסיר את הערכים החריגים.

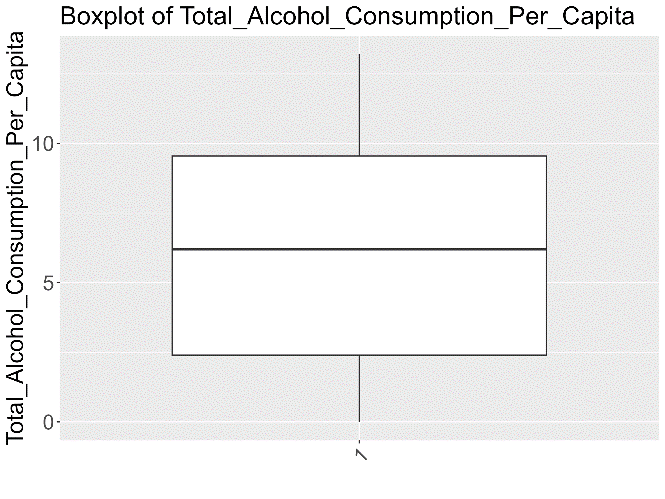
|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 97 | 128647 | 54894.75 |
| 70 | 107641 | 33888.75 |
| 99 | 94105 | 20352.75 |
| 55 | 76745 | 2992.75 |



צפיפות אוכלוסין (Population density):

יש 10 תוצאות חריגות, נחליף את אינדקס 99 לערך הmax של המשתנה עקב התוצאה החריגה. בדומה לתוצר לאומי גם פה הייתי מצפה למדינות מסויימות מאוד צפופות ולצידוד חיובי.

|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 99 | 7953 | 7552.938 |
| 74 | 1718.99 | 1318.928 |
| 76 | 1511.03 | 1110.968 |
| 9 | 1239.58 | 839.518 |
| 65 | 669.494 | 269.432 |
| 10 | 666.607 | 266.545 |
| 78 | 623.302 | 223.24 |
| 87 | 511.458 | 111.396 |
| 52 | 454.938 | 54.87595 |
| 56 | 410.527 | 10.46495 |

צריכת אלכוהול לנפש (Total Alcohol Consumption per Capita):

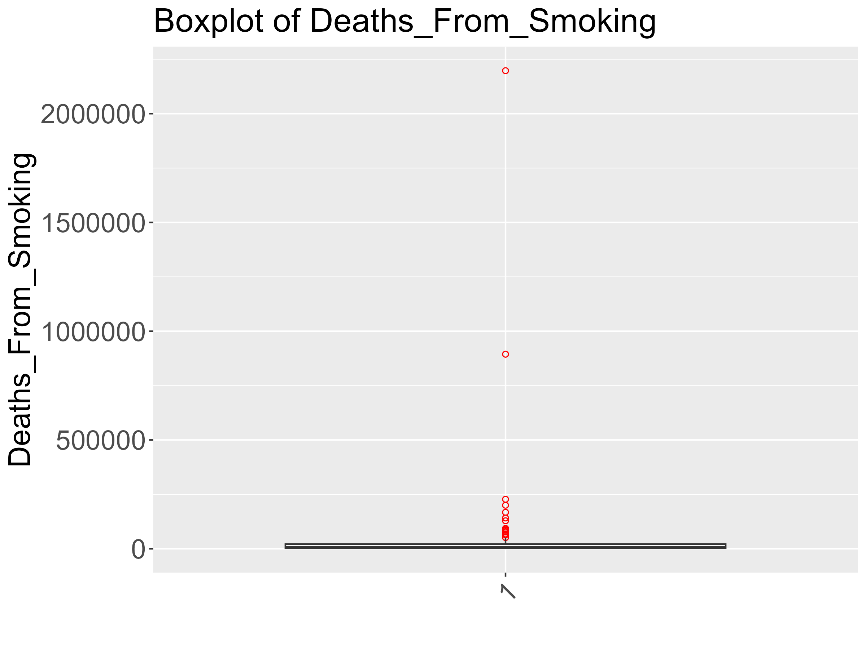
אין תוצאות חריגות

A graph with a white rectangle and black lines

Description automatically generatedשיעור התאבדויות (Suicide Rate):

יש 3 תוצאות חריגות, לא הייתי מוריד אותן עקב הסטייה הנמוכה יחסית מהתפלגות הנתונים.

|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 11 | 26.2 | 2.1 |
| 48 | 29.2 | 5.1 |
| 69 | 31.9 | 7.8 |



מוות מעישון (Death from Smoking):

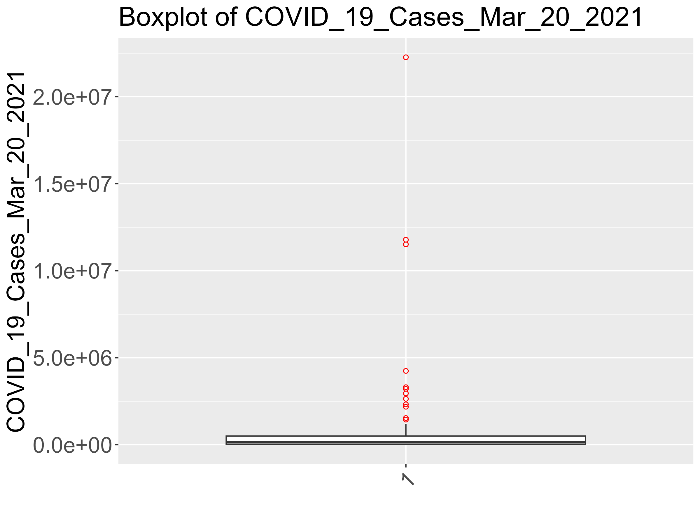
|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 27 | 2197653 | 2147072 |
| 52 | 894371.5 | 843790 |
| 53 | 227477.4 | 176895.9 |
| 59 | 200016 | 149434.5 |
| 18 | 167657.3 | 117075.8 |
| 42 | 141530.8 | 90949.31 |
| 93 | 128588.5 | 78007.01 |
| 95 | 93737.16 | 43155.67 |
| 9 | 90795.06 | 40213.57 |
| 107 | 86358.1 | 35776.61 |
| 57 | 84848.16 | 34266.67 |
| 39 | 79002.38 | 28420.89 |
| 106 | 77090.42 | 26508.93 |
| 84 | 68091.87 | 17510.38 |
| 103 | 63968.59 | 13387.1 |
| 4 | 51911.52 | 1330.029 |
| 79 | 51575.45 | 993.959 |

יש 17 תוצאות חריגות, הייתי מחליף את אינדקס 27,52 לערך ה max של המשתנה, שני הנתונים הללו חריגים בצורה קיצונית, ויכולים לאפיין מדינות מאוכלסות שיש בהן שיעור עישון גבוה, היינו מצפים לתוצאות מאוזנות אם מקרי המוות היו מנורמלות לגודל האוכלוסייה.

מקרי קוביד-19 (Covid 19 - Cases (20.3.2021)):

|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 4 | 22265753 | 21059062 |
| 18 | 11780820 | 10574129 |
| 52 | 11514331 | 10307640 |
| 39 | 4241973 | 3035282 |
| 57 | 3306711 | 2100020 |
| 103 | 3212332 | 2005641 |
| 106 | 2950603 | 1743912 |
| 42 | 2639258 | 1432567 |
| 28 | 2319293 | 1112602 |
| 79 | 2182188 | 975497 |
| 102 | 1533961 | 327270 |
| 53 | 1443853 | 237162 |
| 69 | 31.9 | 7.8 |

קיימים 12 ערכים חריגים, לא הייתי ממליץ לשנות את המשתנה המוסבר לכן נשאיר את הערכים כפי שהם. נוכל להניח כי במדינות מאוכלוסות מס' מקרי קורונה יהיו גבוהים בהשוואה למדינות עם מעט תושבים, בנוסף אנו יודעים שיש צידוד חיובי בהתפלגות האוכלוסיה לפי מדינות לכן תוצאות אלו הגיוניות



מקרי מוות מקוביד-19 (Covid 19 -Death (20.3.2021))

קיימים 19 ערכים חריגים, בדומה למס' מקרי COVID נוכל להניח כי במדינות מאוכלוסות מס' מקרי מוות כתוצאה מהקורונה יהיו גבוהים בהשוואה למדינות עם מעט תושבים, בנוסף אנו יודעים שיש צידוד חיובי בהתפלגות האוכלוסיה לפי מדינות לכן תוצאות אלו הגיוניות. משתנה זה מכיל קשר סיבתי למס' מקרי קרונה, לכן לא נשנה את ערכי הoutliers

A graph with numbers and a line

Description automatically generated

|  |  |  |
| --- | --- | --- |
| Outlier\_Index | Outlier\_Value | Deviation |
| 69 | 31.9 | 7.8 |
| 4 | 54386 | 32746.75 |
| 12 | 22624 | 984.75 |
| 18 | 287499 | 265859.8 |
| 23 | 22576 | 936.75 |
| 26 | 21988 | 348.75 |
| 28 | 61636 | 39996.75 |
| 39 | 91833 | 70193.75 |
| 42 | 74402 | 52762.75 |
| 52 | 159370 | 137730.8 |
| 53 | 39142 | 17502.75 |
| 57 | 103855 | 82215.75 |
| 79 | 196606 | 174966.8 |
| 96 | 167173 | 145533.8 |
| 98 | 21877 | 237.75 |
| 102 | 51724 | 30084.75 |
| 103 | 72910 | 51270.75 |
| 106 | 29777 | 8137.75 |