**שמות הקבצים:**

בחירת מקורות מידע והצגתם:

**FirstCSV** –מציג ומנתח מידע על הקובץ הראשון.

**SecondCSV** – מציג ומנתח מידע על הקובץ השני.

ניקוי וארגון קובץ הנתונים:

**Outer\_JoinCSV** - מאחד טבלאות על בסיס עמודה משותפת שומר את האיחוד כקובץ , לאחר מכן מבצע פעולות ניקוי על כל DF בנפרד גם כאן שומר עבור כל אחד כקובץ (כך שיהיה מובן שהקבצים נוקו), ולאחר מכן מציג את הניתוח שנעשה עבור כל DF , לאחר מכן מנרמל את DF2 באמצעות נרמול עמודת LIFE LADDER , מוסיף את אותה עמודה ושומר את הקובץ בשם – input2\_df\_cleaned\_normalized.

הצגה ויזואלית של הנתונים + ניתוח :

**VizualizationOfData** - בקובץ הזה אני בונה 13 גפרים ייחודים כאשר כל גרף הוא סוג שונה ובעל תובנות שונות.

**Clustering\_df -** בחרתי 5 שילובים של העמודות הכי רלוונטיות עבור חלוקה לאשכולות, ולבסוף אני מבצע רגרסיה ליניארית על 2 עמודות שבחרתי.

הערות : כל הסעיפים בעבודה בוצעו כולל הבונוסים, כולל כל המודולים ואובייקטים.

פרויקט במדע הנתונים

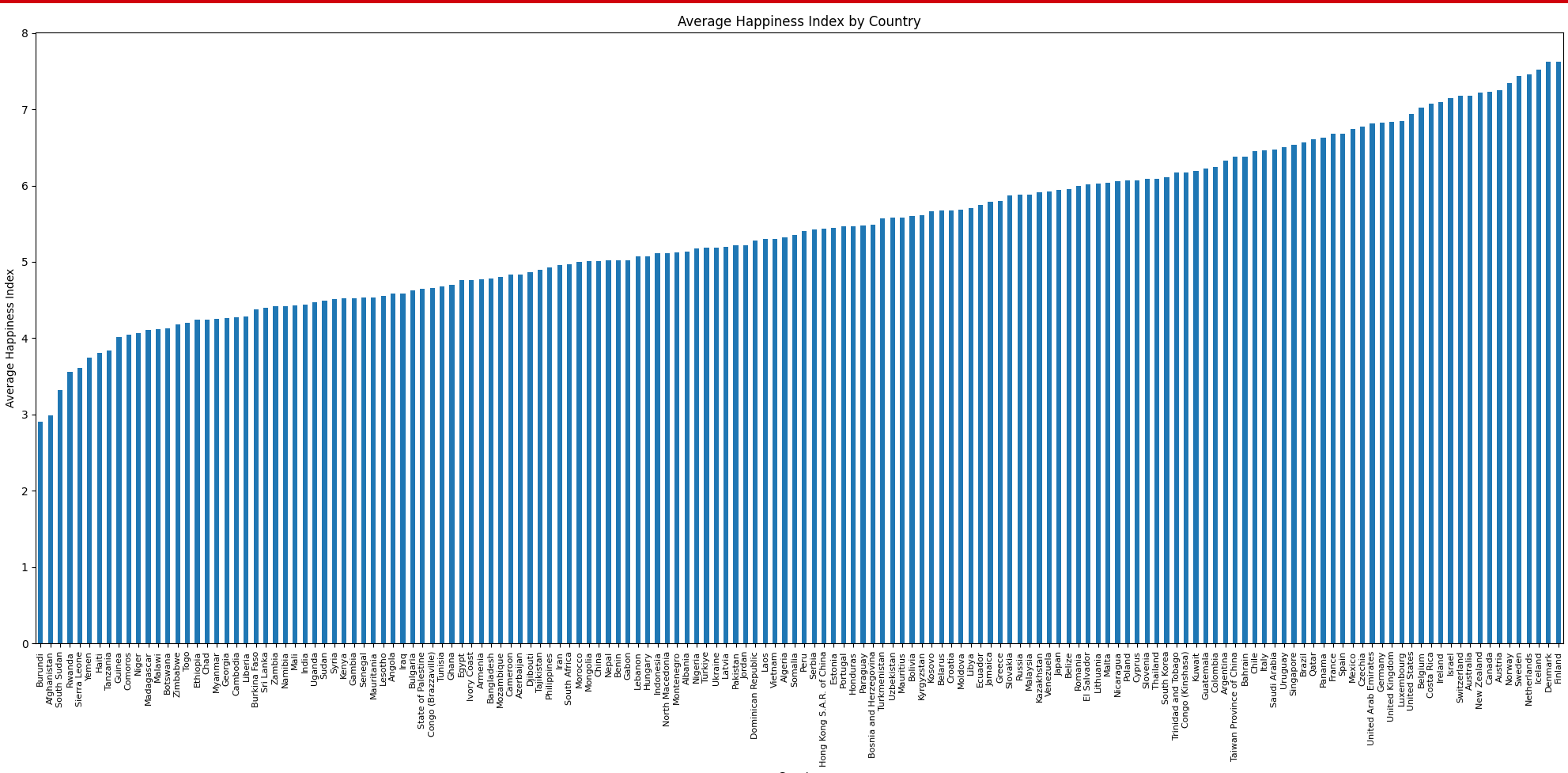
***תוכן עניינים:***

***גרפים ותובנות* ..............................................1-15**

***חלוקה לאשכולות ורגרסיה ליניארית* ................16-21**

***תובנות של INPUT1/2\_DF* .............................22-**

‏2024



גרף 1 - מציג את אינדקס האושר הממוצע עבור כל מדינה בנתונים. הגרף מסודר בסדר עולה, כך שהמדינות עם אינדקס האושר הנמוך ביותר מופיעות בצד שמאל והמדינות עם אינדקס האושר הגבוה ביותר מופיעות בצד ימין. גרף זה מאפשר זיהוי מהיר של המדינות המאושרות ביותר והמדינות הפחות מאושרות.

**המדינות המאושרות ביותר -** מדינות בצפון אירופה כמו פינלנד, דנמרק, ואיסלנד מדורגות בראש הרשימה עם אינדקסי אושר גבוהים. הדבר עשוי להעיד על איכות חיים גבוהה, תמיכה חברתית רחבה, ושירותי בריאות מתקדמים באזורים אלה.

**המדינות הפחות מאושרות -** מדינות באפריקה כמו בורונדי, צ'אד ודרום סודאן מדורגות בתחתית הרשימה עם אינדקסי אושר נמוכים מאוד. ייתכן שהדבר נובע ממלחמות, עוני, ותנאי חיים קשים במדינות אלה.

**שונות בין אזורים גיאוגרפיים –** ניתן לראות פערים משמעותיים בין מדינות ממערב אירופה למדינות באפריקה ובמזרח התיכון. הדבר מצביע על השפעות שונות כמו כלכלה, פוליטיקה, תרבות, ותנאים סוציאליים שמשפיעים על רמת האושר של האוכלוסייה.

**השפעת המצב הכלכלי והפוליטי -** מדינות עשירות עם כלכלה יציבה ושירותים ציבוריים מתקדמים נוטות להיות עם אינדקסי אושר גבוהים יותר. לעומת זאת, מדינות עם כלכלה חלשה, חוסר יציבות פוליטית וקשיים חברתיים נוטות להיות עם אינדקסי אושר נמוכים.

**זיהוי מגמות וחריגות -** הגרף מאפשר זיהוי של מדינות שחריגות מהמגמה הכללית, בין אם הן מאושרות במיוחד או פחות מאושרות, וניתן לנתח לעומק את הסיבות לכך.

תמונה שמכילה טקסט, כתב יד

התיאור נוצר באופן אוטומטי

גרף 2:

הגרף מציג את הקשר בין התמ"ג לנפש (בסקלה לוגריתמית) לבין תוחלת החיים הבריאה בלידה. כל נקודה מייצגת מדינה אחת.

**מתאם חיובי**: ניתן לראות שיש מתאם חיובי ברור בין התמ"ג לנפש לתוחלת החיים הבריאה בלידה. כלומר, ככל שהתמ"ג לנפש במדינה גבוה יותר, כך תוחלת החיים הבריאה בלידה גבוהה יותר. הקשר החזק בין התמ"ג לנפש לבין תוחלת החיים הבריאה מצביע על כך שהמדד הכלכלי משפיע במידה רבה על הבריאות הכללית של האוכלוסייה. מדינות עם תמ"ג גבוה משקיעות יותר בשירותי בריאות, תזונה, חינוך, ותשתיות, מה שמוביל לתוחלת חיים גבוהה יותר.

**פיזור הנתונים:** רוב הנתונים מצטופפים באזור התמ"ג הגבוה יותר ותוחלת החיים הגבוהה יותר, אך ישנן מדינות יוצאות דופן עם תמ"ג נמוך ותוחלת חיים גבוהה או להיפך. פיזור הנקודות בגרף מראה כי יש מדינות עם תמ"ג נמוך יחסית אך עם תוחלת חיים בריאה גבוהה יותר. זה יכול להצביע על כך שישנם גורמים נוספים המשפיעים על תוחלת החיים הבריאה, כגון מדיניות חברתית, רמת חינוך, גישה לשירותי בריאות בסיסיים, או אפילו תרבות.

**קבוצות מדינות:** ניתן לראות מגמות ברורות, כאשר מדינות עם תמ"ג לנפש גבוה מציגות תוחלת חיים בריאה גבוהה יותר, ומדינות עם תמ"ג לנפש נמוך מציגות תוחלת חיים בריאה נמוכה יותר.

**יוצאים מן הכלל:** נקודות יוצאות דופן יכולות לספק תובנות חשובות. לדוגמה, מדינות עם תמ"ג גבוה אך עם תוחלת חיים נמוכה יכולות להצביע על בעיות פנימיות כמו שחיתות, חוסר שוויון, או בעיות אחרות. מדינות עם תמ"ג נמוך ותוחלת חיים גבוהה יכולות להצביע על הצלחות בשיפור הבריאות והחיים בתנאים כלכליים קשים.

תמונה שמכילה טקסט, צילום מסך, גופן, מלבן

התיאור נוצר באופן אוטומטי

גרף 3:

הגרף מציג את ממוצע השחיתות הנתפסת עבור כל רמה של תמיכה חברתית. כל עמודה מייצגת קטגוריה של רמת תמיכה חברתית והממוצע השחיתות הנתפסת עבור אותה קטגוריה.

ניתן לראות שרמות גבוהות של תמיכה חברתית מתואמות עם שחיתות נתפסת נמוכה יותר. זה מרמז על כך שתמיכה חברתית פועלת כגורם מייצב חברתי, המקטין את רמות השחיתות.

ישנה מגמה מעניינת של עלייה בשחיתות הנתפסת ברמות בינוניות של תמיכה חברתית, מה שעשוי להצביע על כך שבמדינות עם רמות בינוניות של תמיכה חברתית, ישנם גורמים אחרים המגבירים את תפיסת השחיתות. זה יכול לכלול בעיות מערכתיות, חוסר אמון בממשלה או במוסדות אחרים.

התערבות מדינית:

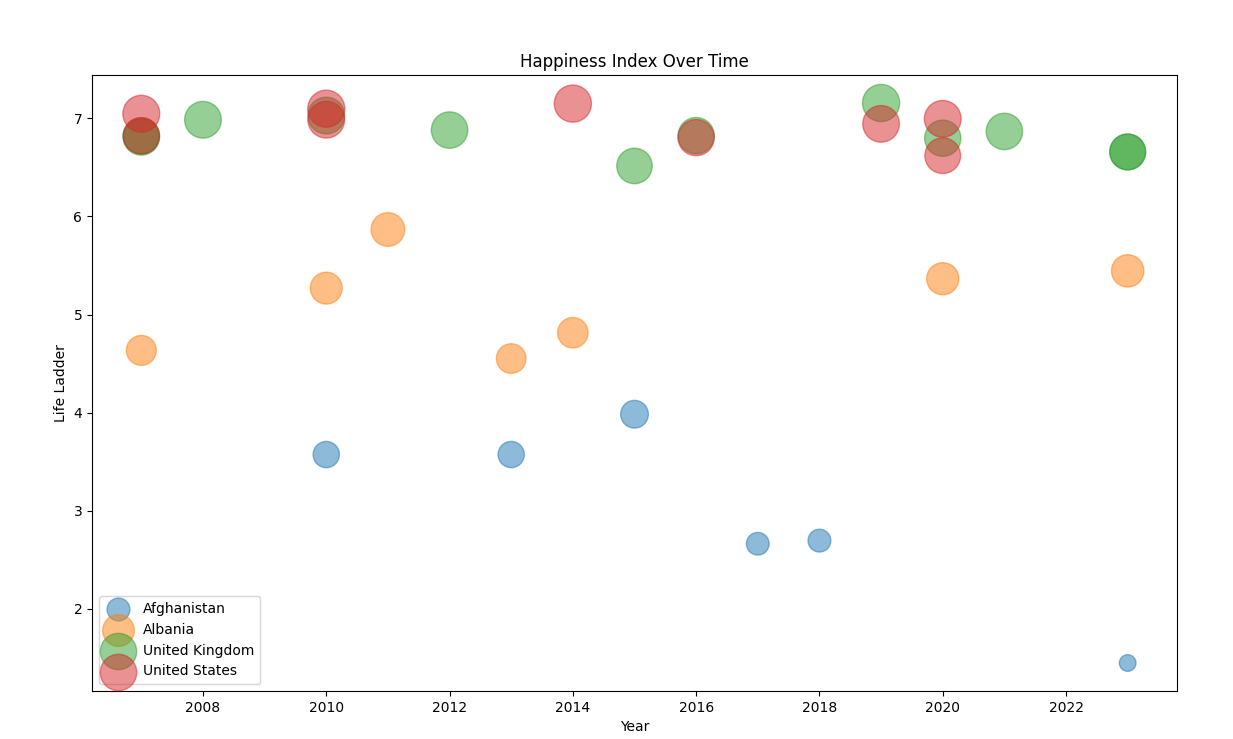
מדינות עם רמות תמיכה חברתית נמוכה או בינונית עשויות להפיק תועלת מהתערבויות מדיניות שמטרתן להגביר את התמיכה החברתית והקשרים בין האזרחים. לדוגמה, השקעה בקהילות, חיזוק ארגונים חברתיים, והגברת השקיפות והאמון במוסדות.

**השפעת תרבות ואמון:** התמיכה החברתית יכולה להיות משולבת גם בתרבות המקומית ובערכים חברתיים. מדינות עם תרבות חזקה של עזרה הדדית ותמיכה חברתית עשויות להציג רמות נמוכות יותר של שחיתות נתפסת.

**הקשר לשביעות רצון כללית:** על בסיס התובנות מהגרף, ניתן לשער כי תמיכה חברתית גבוהה יכולה לתרום לא רק להפחתת השחיתות הנתפסת, אלא גם לשיפור תחושת שביעות הרצון והאושר הכללי באוכלוסייה.

**שיפור מערכות תמיכה:** מערכות תמיכה חברתיות חזקות יכולות לכלול מערכות בריאות, חינוך, ורווחה. השקעה במערכות אלו עשויה להוביל להפחתת השחיתות הנתפסת, כיוון שהיא מחזקת את תחושת הביטחון והאמון במוסדות הציבוריים.

**השפעת הממשל:** במקרים של רמות תמיכה חברתית בינוניות עד גבוהות, ניתן לראות ירידה בשחיתות הנתפסת. זה מרמז על כך שתמיכה חברתית המנוהלת היטב על ידי הממשל יכולה להיות כלי יעיל להפחתת השחיתות.



גרף 4:

הגרף מציג את מדד האושר (Life Ladder) של מספר מדינות לאורך זמן. כל בועה מייצגת שנה מסוימת, כאשר גודל הבועה מצביע על גובה מדד האושר.

ארצות הברית והממלכה המאוחדת: שתי המדינות הללו מציגות רמות אושר גבוהות יחסית ויציבות לאורך השנים. זה מעיד על כך שמדינות מפותחות עם כלכלה חזקה ותשתיות חברתיות בריאות מצליחות לשמור על רמת אושר גבוהה ויציבה.

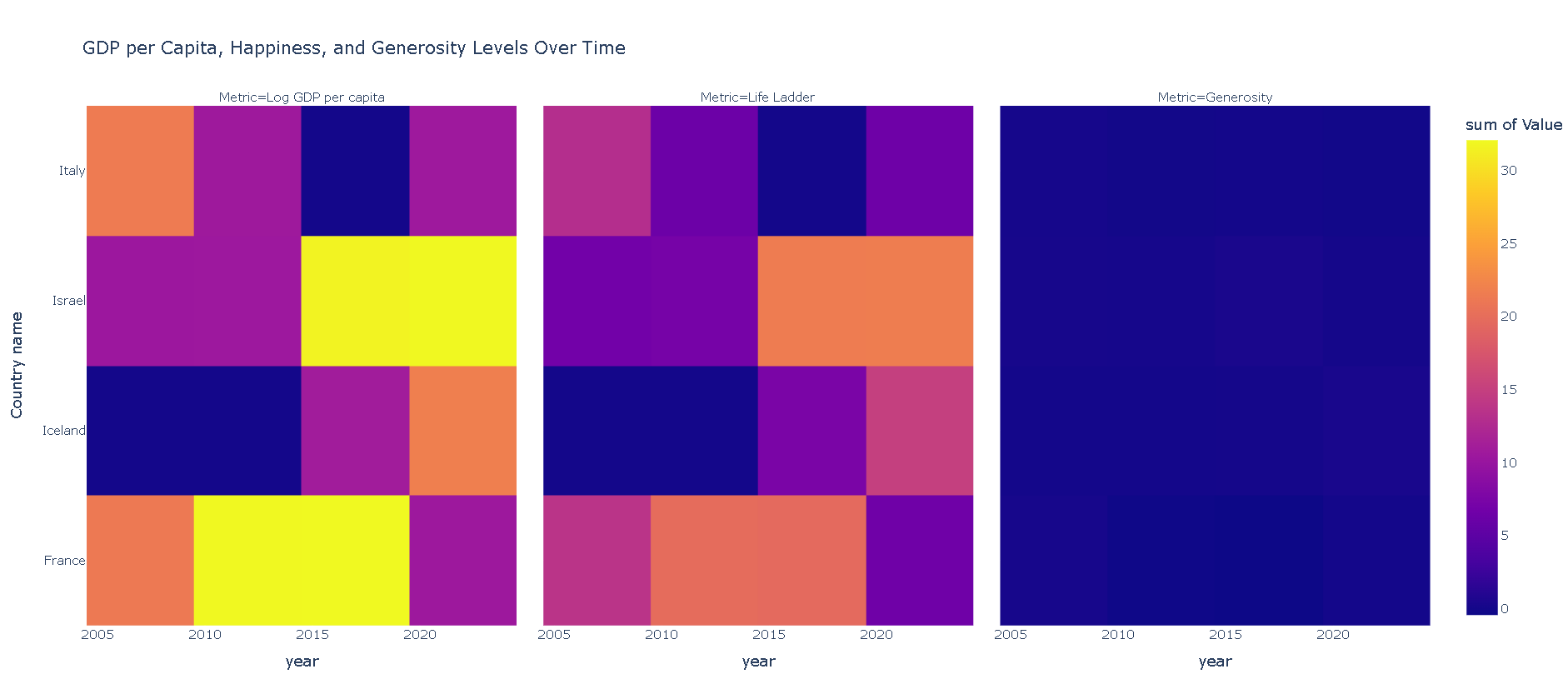
אפגניסטן: הצגת מדד אושר נמוך מאוד לאורך השנים, עם מגמת ירידה משמעותית. הדבר מצביע על השפעות של מלחמות, עוני, ותנאי חיים קשים על רמת האושר במדינה.

ניתן לראות עליות וירידות חדות במדד האושר במדינות מסוימות. למשל, שינויים פתאומיים באלבניה עשויים להצביע על אירועים פוליטיים או כלכליים משמעותיים שהתרחשו במהלך התקופה הנבדקת.

**השפעת מדיניות כלכלית וחברתית:** המדינות המציגות רמות אושר יציבות וגבוהות עשויות להוות דוגמה למחקרים על מדיניות חברתית וכלכלית מוצלחת. השקעות בחינוך, בריאות ורווחה חברתית עשויות להיות גורמים תורמים לשמירה על רמת אושר גבוהה.

**מגמות ארוכות טווח:** זיהוי מגמות ארוכות טווח במדד האושר יכול לספק תובנות חשובות על ההתפתחות החברתית והכלכלית במדינות השונות. לדוגמה, מגמת הירידה באפגניסטן לאורך זמן מעידה על התדרדרות בתנאי החיים.

**השפעות תרבותיות וחברתיות:** התרבות והחברה המקומית משפיעות גם הן על מדד האושר. מדינות עם תרבות חזקה של קהילה ותמיכה חברתית עשויות להציג רמות אושר גבוהות יותר, גם אם הן מתמודדות עם אתגרים כלכליים או פוליטיים.

****

הגרף שנוצר הוא גרף חום המציג את רמות ההכנסה לנפש (Log GDP per capita), רמות האושר (Life Ladder) ורמות הנדיבות (Generosity) לאורך זמן עבור המדינות הנבחרות. להלן תובנות חשובות שניתן להסיק מהגרף:

1. **רמות ההכנסה לנפש (Log GDP per capita):** ניתן לראות שבמדינות כמו צרפת ואיסלנד יש עלייה משמעותית ברמת ההכנסה לנפש עם השנים, במיוחד בשנים האחרונות. באיטליה ניתן לראות ירידה ברמת ההכנסה לנפש מסוימת באמצע התקופה ולאחר מכן עלייה.

2. **רמות האושר (Life Ladder):** ניתן לראות שבמדינות כמו צרפת ואיסלנד יש עלייה ברמת האושר לאורך זמן. באיטליה ובישראל ניתן לראות שינויים ברמת האושר לאורך השנים עם מגמות חיוביות ושליליות.

3**. רמות הנדיבות (Generosity):** מדד הנדיבות נשאר די קבוע ונמוך ברוב המדינות, מה שמצביע על כך שלא היה שינוי משמעותי בנדיבות לאורך השנים.

**השוואות בין מדינות**:

**צרפת:** מציגה עלייה ברורה גם ברמת ההכנסה לנפש וגם ברמת האושר לאורך השנים. הנדיבות נשארת נמוכה וקבועה לאורך הזמן.

**איטליה:** מציגה ירידה ברמת ההכנסה לנפש באמצע התקופה ולאחר מכן עלייה. רמת האושר מציגה תנודות לאורך השנים עם עליות וירידות. הנדיבות נשארת נמוכה וקבועה לאורך הזמן.

**איסלנד:** מציגה עלייה ברורה ברמת ההכנסה לנפש לאורך השנים. רמת האושר מציגה עליות וירידות לאורך הזמן. הנדיבות נשארת נמוכה וקבועה לאורך הזמן.

**ישראל:** מציגה עלייה ברמת ההכנסה לנפש לאורך השנים, במיוחד בשנים האחרונות. רמת האושר מציגה תנודות לאורך השנים עם עליות וירידות. הנדיבות נשארת נמוכה וקבועה לאורך הזמן.

**מסקנות:**

עלייה ברמת ההכנסה לנפש: רוב המדינות מציגות עלייה ברמת ההכנסה לנפש לאורך השנים, מה שמעיד על צמיחה כלכלית. **שינויים ברמת האושר**: רמת האושר מציגה תנודות לאורך השנים במדינות השונות, מה שמעיד על השפעות מגוונות של אירועים כלכליים, פוליטיים וחברתיים. קביעות ברמת הנדיבות: רמת הנדיבות נשארת נמוכה וקבועה לאורך השנים, מה שמעיד על כך שאין שינוי משמעותי בהתנהגות החברתית בתחום זה.



6

מאפשר לראות את הקשר בין אושר, תמיכה חברתית ושחיתות נתפסת.

זיהוי מדינות עם שילוב גבוה/נמוך של כל שלושת המדדים.

**קשר בין תמיכה חברתית לאושר:**

ניתן לראות שמדינות עם רמות גבוהות של תמיכה חברתית (קרוב ל-1 בציר ה-Social Support) נוטות להיות גם בעלות מדדי אושר גבוהים (קרוב ל-10 בציר ה-Life Ladder).

**שחיתות נתפסת ואושר:**

במדינות עם רמות שחיתות נתפסת נמוכות (קרוב ל-0 בציר ה-Perceptions of Corruption), נראה כי יש נטייה למדדי אושר גבוהים יותר. כלומר, במדינות בהן שחיתות נתפסת נמוכה יותר, האזרחים עשויים להיות מאושרים יותר.

**תמיכה חברתית ושחיתות נתפסת:**

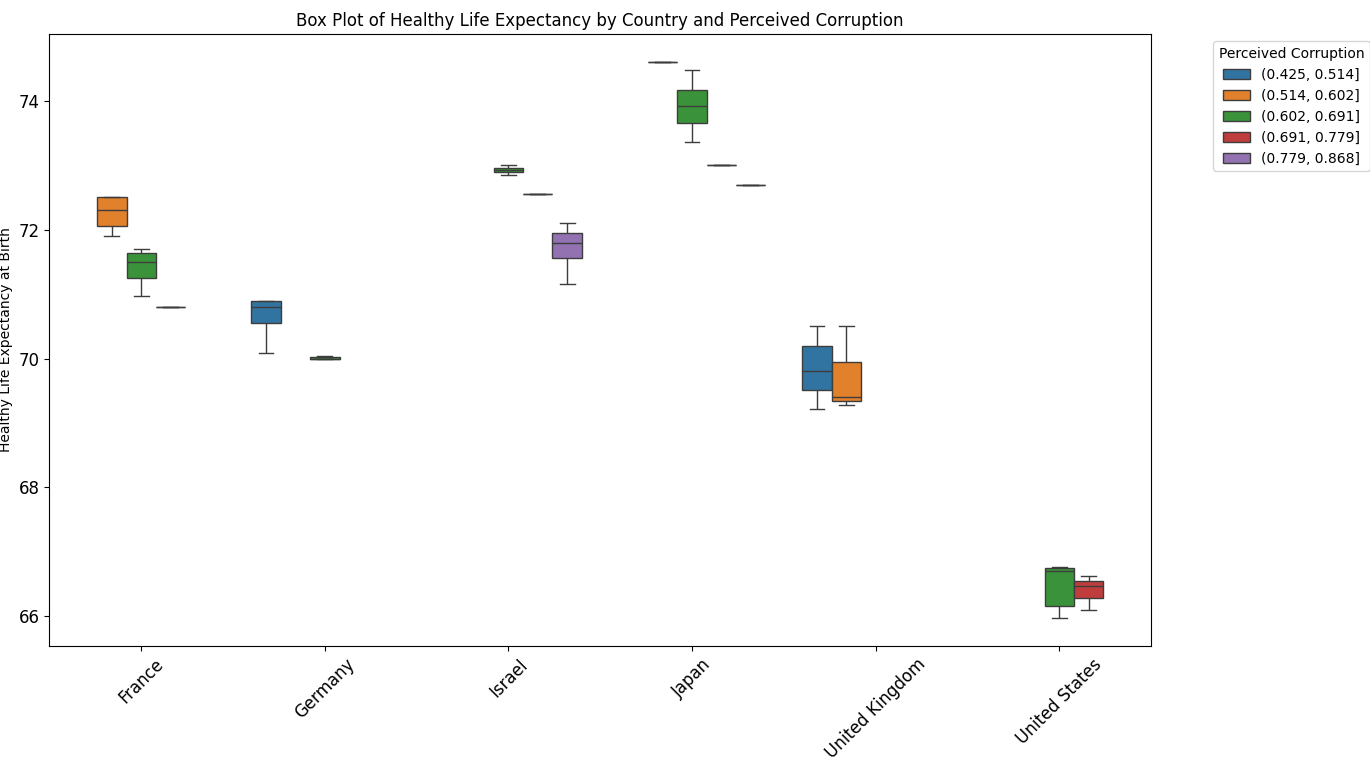
נראה שיש קשר הפוך בין תמיכה חברתית לשחיתות נתפסת: במדינות עם רמות תמיכה חברתית גבוהות, רמות השחיתות הנתפסת נמוכות יותר.

**הבדלים בין מדינות:**

הגרף מאפשר לזהות מדינות ספציפיות עם שילובים ייחודיים של אושר, תמיכה חברתית ושחיתות נתפסת. למשל, מדינות עם שחיתות גבוהה אך עם רמות אושר ותמיכה חברתית גבוהות, ומדינות עם תמיכה חברתית נמוכה אך עם רמות שחיתות ואושר נמוכות.

**פיזור רחב:**

ישנן מדינות רבות עם פיזור רחב של תמיכה חברתית, שחיתות נתפסת ואושר. זה מראה שהקשר בין המדדים אינו מוחלט וישנם גורמים נוספים המשפיעים על רמות האושר במדינה.



**גרף 7: מדדי בריאות לפי קטגוריות מדינה נבחרות והשפעת השחיתות הנתפסת (Box Plot)**

הגרף מציג את התפלגות תוחלת החיים הבריאה בלידה (Healthy Life Expectancy at Birth) במדינות נבחרות (ארצות הברית, הממלכה המאוחדת, גרמניה, צרפת, ישראל, ויפן) תוך השוואה להשפעת רמות השחיתות הנתפסת בכל מדינה. כל קופסה בגרף מייצגת את התפלגות תוחלת החיים הבריאה במדינה מסוימת, והקטגוריות של השחיתות הנתפסת מוצגות בצבעים שונים.

**הבדלים בתוחלת החיים הבריאה בין מדינות:**

הגרף מאפשר לראות בבירור את ההבדלים בתוחלת החיים הבריאה בין המדינות הנבחרות. לדוגמה, יפן מציגה תוחלת חיים בריאה גבוהה מאוד בהשוואה לארצות הברית או הממלכה המאוחדת. תובנה זו עשויה להצביע על הבדלים במערכות הבריאות, אורח החיים, והתנאים הסביבתיים בכל מדינה.

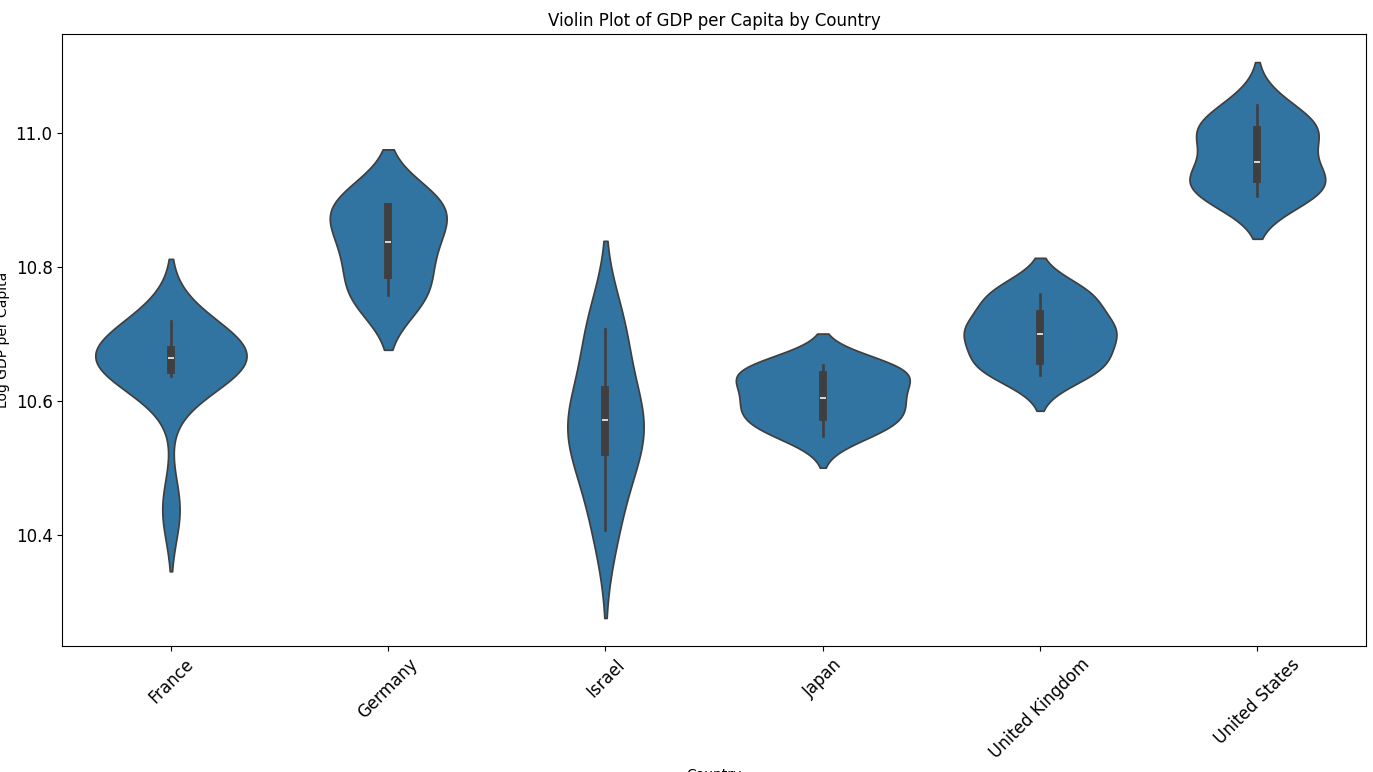
**השפעת השחיתות הנתפסת על תוחלת החיים הבריאה:**

במדינות עם רמות שחיתות נתפסת נמוכות, תוחלת החיים הבריאה נוטה להיות גבוהה יותר. ניתן לראות אם יש הבדלים מובהקים בין הקטגוריות של השחיתות הנתפסת בכל מדינה. לדוגמה, במדינה עם שחיתות נתפסת נמוכה, תוחלת החיים הבריאה עשויה להיות גבוהה יותר בהשוואה למדינה עם שחיתות נתפסת גבוהה. תובנה זו יכולה לרמוז על הקשר בין מערכת הבריאות הציבורית והשחיתות.

**שונות בתוך המדינות:**

הגרף מאפשר לראות אם יש שונות גבוהה או נמוכה בתוחלת החיים הבריאה בתוך כל מדינה. מדינות עם תוחלת חיים בריאה הומוגנית יופיעו עם קווים קצרים יותר בתיבות ההפצה (Box Plot), בעוד שמדינות עם שונות גבוהה יופיעו עם קווים ארוכים יותר. שונות נמוכה עשויה להצביע על גישה שווה לשירותי בריאות איכותיים, בעוד ששונות גבוהה יכולה להצביע על פערים באיכות הבריאות בתוך המדינה.

**ערכים יוצאי דופן (Outliers):** ניתן לזהות ערכים יוצאי דופן בתוחלת החיים הבריאה במדינות השונות. ערכים אלו עשויים להצביע על תנאים ייחודיים או חריגות באותן מדינות. לדוגמה, אם יש ערכים נמוכים במיוחד במדינה עם תוחלת חיים בריאה גבוהה, זה עשוי להצביע על אזורים מסוימים במדינה עם גישה מוגבלת לשירותי בריאות.



8

הגרף מציג את התפלגות ההכנסה לנפש (Log GDP per capita) במדינות נבחרות בצורה של "כינור" (Violin Plot). הצורה של הכינור מראה את צפיפות הנתונים, כך שניתן לראות היכן מרוכזים רוב הערכים והיכן יש פחות נתונים. לדוגמה, בישראל ניתן לראות התפלגות רחבה מאוד של הכנסה לנפש, מה שמצביע על פיזור רחב של ערכי ההכנסה.

**זיהוי פיזור ההכנסה לנפש:** במדינות עם שונות גבוהה בהכנסה לנפש, הכינור יהיה רחב יותר בחלקו העליון או התחתון, בעוד שבמדינות עם שונות נמוכה הכינור יהיה צר יותר. בגרמניה וביפן, לדוגמה, ניתן לראות שונות נמוכה יותר בהכנסה לנפש, מה שמצביע על התפלגות יותר הומוגנית של ההכנסות.

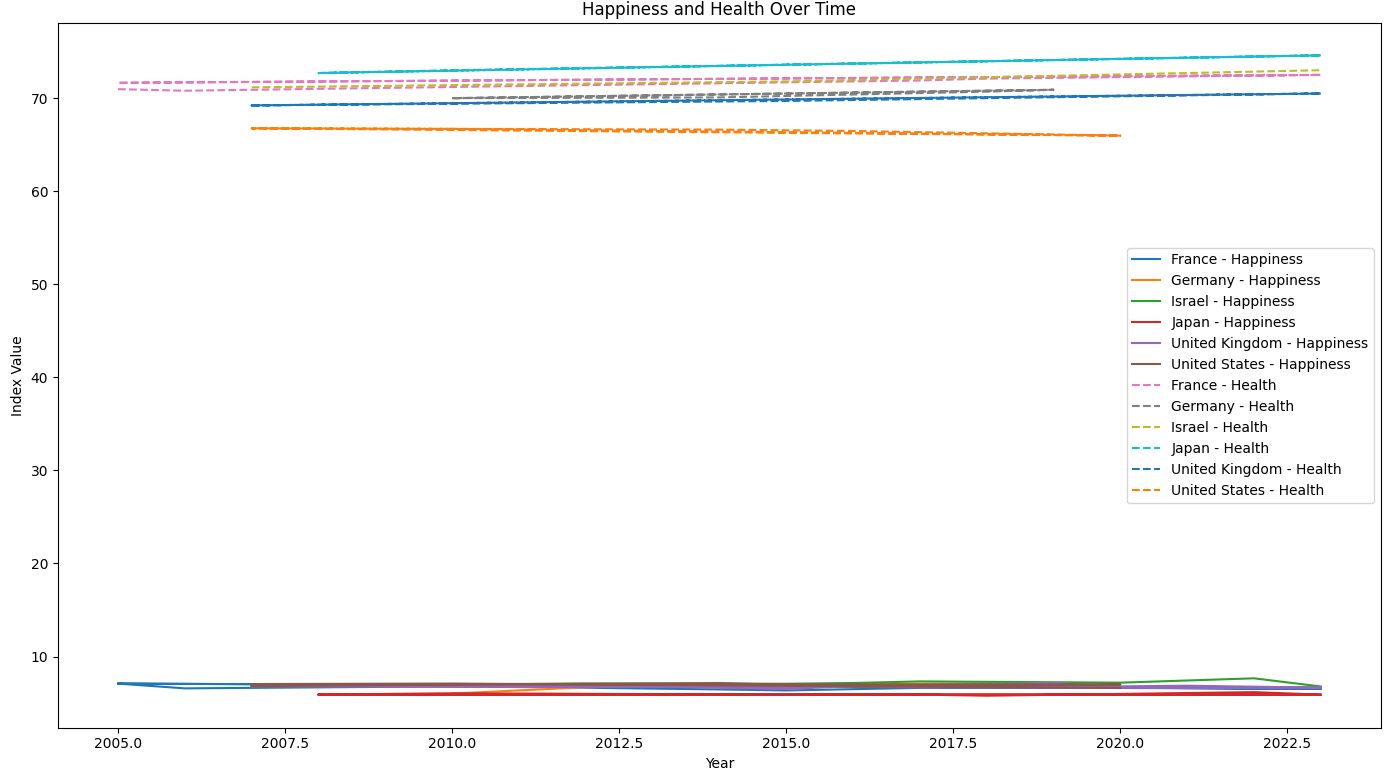
**השוואת רמות הכנסה לנפש:** הגרף מאפשר השוואה בין רמות ההכנסה לנפש במדינות הנבחרות. לדוגמה, ארצות הברית מציגה רמות הכנסה לנפש גבוהות ביותר בקרב המדינות הנבחרות, בעוד שבצרפת וישראל רמות ההכנסה לנפש נמוכות יותר יחסית. זה יכול להצביע על פערים כלכליים בין המדינות ועל רמת החיים בכל מדינה.

**השוואה בין מדינות:** הגרף מאפשר השוואה בין המדינות השונות ולראות איזה מדינות יש להן פיזור רחב יותר של הכנסה לנפש ואיזה מדינות יש להן התפלגות צרה יותר של הכנסות. פיזור רחב יותר יכול להעיד על אי-שוויון כלכלי גדול יותר בתוך המדינה, בעוד שפיזור צר יותר מצביע על הומוגניות כלכלית גבוהה יותר.

**זיהוי ערכים קיצוניים:** הגרף מאפשר לזהות ערכים קיצוניים (Outliers) בהכנסה לנפש במדינות השונות. ערכים אלו יכולים להצביע על הבדלים כלכליים משמעותיים בתוך המדינות. לדוגמה, אם יש ערכים קיצוניים גבוהים במיוחד במדינה מסוימת, זה יכול להצביע על אזורים או קבוצות אוכלוסייה עם הכנסות גבוהות מאוד.

**השלכות מדיניות:** ניתן להשתמש בתובנות מהגרף כדי להבין את ההשלכות המדיניות של פערים כלכליים ולהציע דרכים לשיפור השוויון הכלכלי. לדוגמה, מדינה עם שונות גבוהה בהכנסה לנפש עשויה לשקול מדיניות כלכלית להקטנת הפערים הכלכליים ולהגברת ההוגנות.

**השפעה על איכות החיים:** ההכנסה לנפש היא אחד המדדים המשפיעים על איכות החיים. גרף זה יכול לסייע בהבנה של הקשר בין רמת ההכנסה לאיכות החיים במדינות השונות.



9

הגרף מאפשר לראות כיצד מדדי האושר משתנים לאורך השנים במדינות הנבחרות. ניתן לזהות מגמות עלייה או ירידה באושר ולהשוות בין המדינות השונות.

**מגמות אושר לאורך זמן:** ניתן לראות שמדדי האושר (Life Ladder) ברוב המדינות שנבחרו הם נמוכים בהשוואה למדדי הבריאות (Healthy Life Expectancy at Birth).

נראה כי מדדי האושר נעים בטווח מצומצם של ערכים נמוכים (בין 5 ל-10), ויש יציבות יחסית לאורך הזמן.

**מגמות בריאות לאורך זמן:** מדדי הבריאות מציגים תוחלת חיים בריאה (Healthy Life Expectancy at Birth) הרבה יותר גבוהה ממדדי האושר, ומדדי הבריאות במדינות הנבחרות נעים בטווח רחב יותר של ערכים (בין 65 ל-75).

ברוב המדינות, מדדי הבריאות מראים עלייה קלה לאורך הזמן, מה שמעיד על שיפורים במערכות הבריאות או בתנאי החיים.

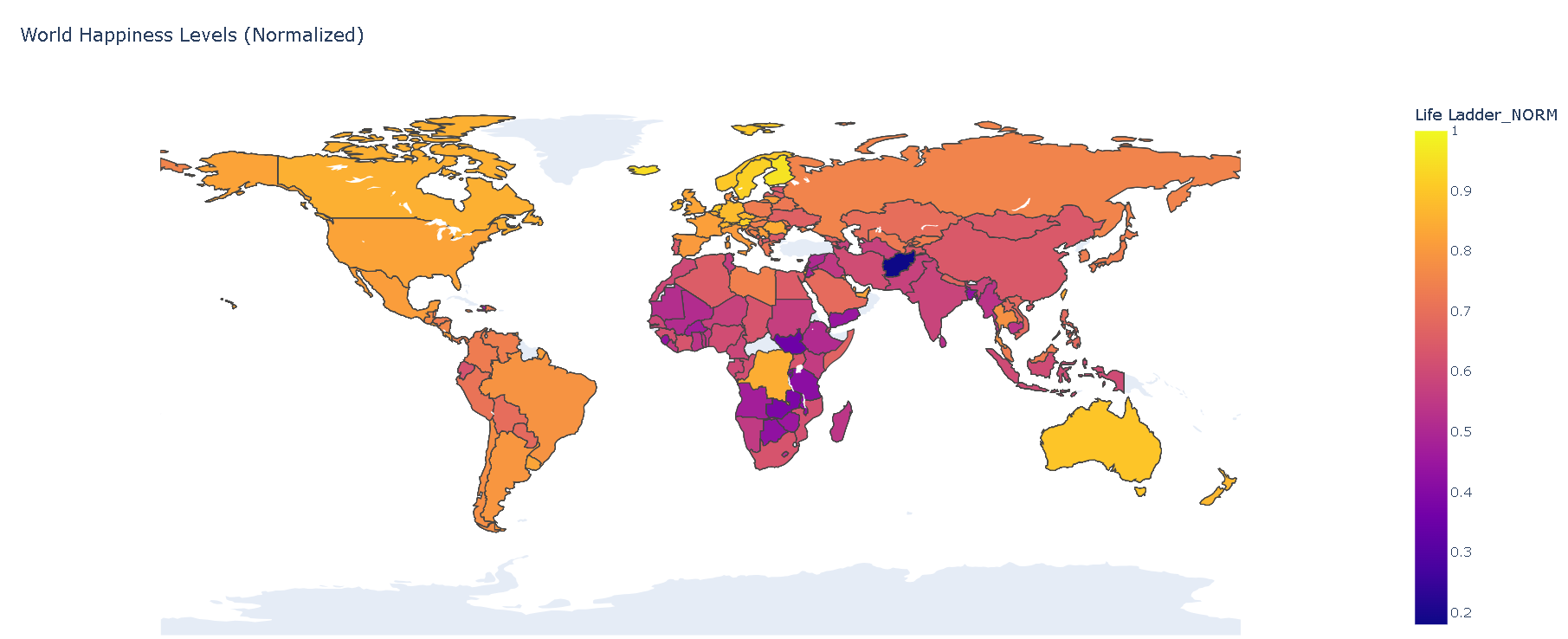
**השוואת מגמות אושר ובריאות:** ניתן לראות שמדדי הבריאות הם משמעותית גבוהים יותר ממדדי האושר בכל המדינות הנבחרות.

השוואת מגמות האושר והבריאות יכולה להצביע על כך שבריאות פיזית אינה בהכרח משפיעה בצורה ישירה על מדדי האושר באותה מידה.

**זיהוי נקודות שינוי חשובות:** ניתן לזהות נקודות בזמן שבהן יש שינוי משמעותי במדדי האושר או הבריאות. לדוגמה, ניתן לראות יציבות יחסית במדדי הבריאות לאורך הזמן עם שינויים קלים בלבד.

באשר לאושר, ישנם שינויים מועטים אך יציבות כללית במדדי האושר לאורך הזמן.

**השפעה של אירועים משמעותיים:** ניתן לבדוק את השפעתם של אירועים גלובליים או מקומיים כמו מגפות, משברים כלכליים, מלחמות או רפורמות בריאות על מדדי האושר והבריאות. לדוגמה, ניתן לראות אם יש שינויים חדים סביב תקופות של משברים כלכליים או בריאותיים.



10- הגרף הוא מפה כורופלית המציגה את רמות האושר המנורמלות (Life Ladder\_NORM) במדינות שונות ברחבי העולם. המפה מאפשרת לראות בצורה גרפית את רמות האושר המנורמלות, כאשר הצבעים מייצגים את הערכים: צבעים חמים (כתום, אדום) מצביעים על רמות אושר גבוהות, וצבעים קרים (סגול, כחול) מצביעים על רמות אושר נמוכות.

ניתן לראות בבירור כי מדינות בצפון אמריקה, מערב אירופה, ואוסטרליה מציגות רמות אושר מנורמלות גבוהות יחסית (צבעים חמים), בעוד שמדינות באפריקה ובחלקים מאסיה מציגות רמות אושר נמוכות יחסית (צבעים קרים).

**השוואה אזורית:** במדינות צפון אמריקה ואירופה המערבית יש רמות אושר גבוהות. מדינות כמו קנדה, ארה"ב, גרמניה ונורווגיה מציגות רמות אושר גבוהות יותר בהשוואה למדינות אחרות. באפריקה, במיוחד במרכז ובמזרח היבשת, ניתן לראות רמות אושר נמוכות מאוד.

**פערים בין מדינות:** פערים ברורים ברמות האושר בין מדינות שונות באותה יבשת. לדוגמה, בישראל וביפן יש רמות אושר גבוהות יחסית למדינות השכנות. במדינות כמו רוסיה וברזיל יש רמות אושר נמוכות יחסית למדינות אחרות עם הכנסה לנפש דומה.

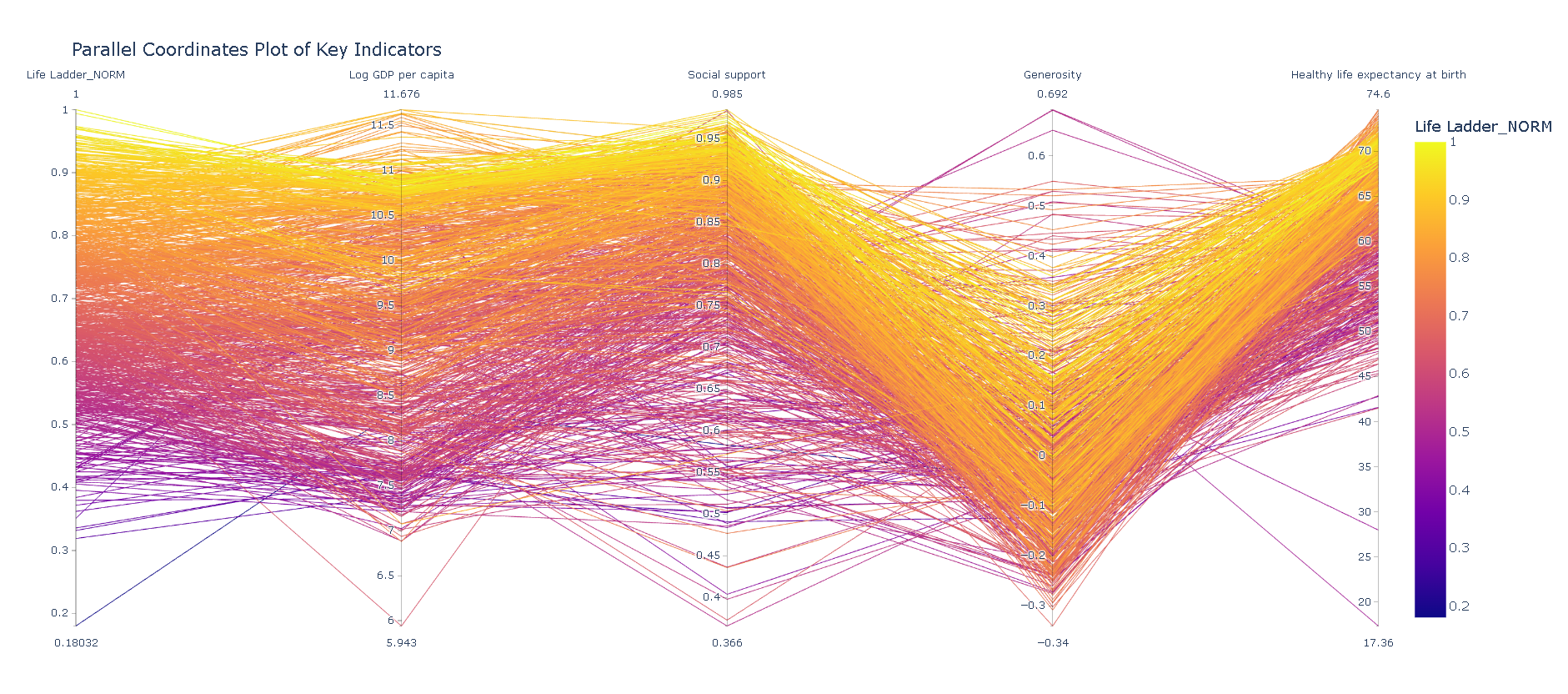
**זיהוי מדינות יוצאות דופן:** ישראל ויפן מציגות רמות אושר גבוהות יחסית למדינות בסביבתן, מה שמעיד על גורמים ייחודיים המעלים את רמות האושר במדינות אלו. מדינות כמו נורווגיה ושוודיה מציגות רמות אושר גבוהות במיוחד ביחס לשכנותיהן באירופה.

**השפעות כלכליות וחברתיות:** המפה יכולה להצביע על השפעות כלכליות וחברתיות על רמות האושר. במדינות עם כלכלה חזקה ותמיכה חברתית גבוהה, כמו קנדה, אוסטרליה ונורווגיה, רמות האושר הן גבוהות במיוחד.

במדינות עם תמיכה חברתית נמוכה וכלכלה חלשה, כמו במדינות רבות באפריקה, רמות האושר הן נמוכות.

**תובנות על אושר עולמי:** המפה מספקת תובנות רחבות על רמות האושר העולמיות ומאפשרת זיהוי של אזורים גיאוגרפיים עם רמות אושר גבוהות או נמוכות.

ניתן לראות השפעות של מגמות כלכליות, חברתיות ופוליטיות על רמות האושר במדינות השונות.



הגרף הזה הוא גרף תיאום מקבילי שמציג את היחסים בין מדדים שונים לרמות האושר המנורמלות (Life Ladder\_NORM). כל קו בגרף מייצג מדינה, והצבע של הקו מייצג את רמת האושר המנורמלת. הגרף מציג את הקשרים בין **המדדים הבאים:**

Life Ladder\_NORM (אושר מנורמל)

Log GDP per capita (תמ"ג לנפש)

Social support (תמיכה חברתית)

Generosity (נדיבות)

Healthy life expectancy at birth (תוחלת חיים בריאה בלידה)

**זיהוי מדדים חשובים:**

ניתן לראות כיצד מדדים שונים משפיעים על רמות האושר המנורמלות. לדוגמה, מדינות עם רמות אושר גבוהות (קוים בצבעים חמים) גם נוטות להיות עם ערכים גבוהים של תמיכה חברתית ותוחלת חיים בריאה בלידה.

**קשר חזק בין מדדים:**

יש קשר חזק בין תוחלת החיים הבריאה ותמיכה חברתית לבין רמות האושר. מדינות עם ערכים גבוהים במדדים אלה נוטות להיות מאושרות יותר.

**הבדלים בין מדינות:**

ניתן לראות הבדלים בין המדינות ביחס למדדים שנבחרו. מדינות עם רמות אושר נמוכות (קוים בצבעים קרים) נוטות להיות עם ערכים נמוכים של תמיכה חברתית, נדיבות ותוחלת חיים בריאה בלידה.

**זיהוי מגמות:**

הגרף מאפשר לזהות מגמות כלליות ולהבין את הדפוסים הקיימים בין המדדים השונים. לדוגמה, ניתן לראות שמדינות עם תמ"ג גבוה לנפש (Log GDP per capita) אינן בהכרח המאושרות ביותר.

**הדגשת מדדים ספציפיים:** הגרף מדגיש את החשיבות של תמיכה חברתית ותוחלת חיים בריאה ברמות האושר. המדדים האלה נראים כהשפעות משמעותיות על רמות האושר המנורמלות.

תמונה שמכילה קו, צבעוני, אומנות, צילום מסך

התיאור נוצר באופן אוטומטי

12

גרף ה-Parallel Categories מציג נתונים מרובי ממדים באמצעות קטגוריות מקבילות. כל ציר מייצג משתנה מסוים, והקווים המחברים בין הצירים מייצגים את השינויים בערכים בין המשתנים. במקרה זה, הגרף מציג את המדינות שנבחרו, קטגוריות של תמיכה חברתית וקטגוריות של שחיתות נתפסת, ומראה את הערכים המנורמלים של מדד האושר (Life Ladder\_NORM).

**השפעת התמיכה החברתית על האושר:** מדינות עם קטגוריות תמיכה חברתית גבוהה מאוד (Very High) נוטות להיות בעלות מדד אושר מנורמל גבוה יותר (Life Ladder\_NORM). תמיכה חברתית חזקה יכולה להוביל לרשת ביטחון עבור האזרחים, לספק סיוע נפשי וחברתי ולהגביר את תחושת הרווחה והאושר הכללי.

**השפעת השחיתות הנתפסת על האושר**: מדינות עם קטגוריות שחיתות נתפסת נמוכות (Low) נוטות להיות בעלות מדד אושר מנורמל גבוה יותר. שחיתות נמוכה מגבירה את תחושת הביטחון והאמון של האזרחים במוסדות המדינה, מה שיכול להשפיע באופן חיובי על תחושת האושר והרווחה הכללית.

**מדינות עם פרופיל ייחודי:** לדוגמה, ישראל מציגה תמיכה חברתית גבוהה מאוד, שחיתות נתפסת בינונית, ואושר מנורמל גבוה יחסית. ניתן להסיק שתמיכה חברתית חזקה בישראל מפצה חלקית על רמות שחיתות גבוהות יותר בהשוואה למדינות אחרות.

ארצות הברית מציגה תמיכה חברתית נמוכה יותר, שחיתות נתפסת גבוהה, ואושר מנורמל נמוך יותר. זה עשוי להצביע על קשר בין שחיתות נתפסת גבוהה ותמיכה חברתית נמוכה לתחושת אושר נמוכה.

**השוואה בין מדינות:** השוואה בין מדינות מאפשרת לראות כיצד משתנים שונים משפיעים על מדד האושר. לדוגמה, ניתן לראות שהמדינות עם תמיכה חברתית גבוהה ושחיתות נתפסת נמוכה מציגות רמות אושר גבוהות יותר, בעוד שמדינות עם שחיתות גבוהה ותמיכה חברתית נמוכה מציגות רמות אושר נמוכות יותר.

גרמניה מציגה רמות שחיתות נתפסת נמוכות ותמיכה חברתית גבוהה, מה שמוביל למדד אושר גבוה יחסית. זה מצביע על כך שהן השחיתות הנתפסת והן התמיכה החברתית משפיעות על תחושת האושר הכללית.

**מגמות כלליות:** ניתן לראות מגמות כלליות בין קטגוריות תמיכה חברתית ושחיתות נתפסת לבין מדד האושר. מדינות שמצטיינות בתמיכה חברתית ושחיתות נמוכה מראות רמות אושר גבוהות, ומדינות עם תמיכה חברתית ושחיתות גבוהות מראות רמות אושר נמוכות יותר.

התובנות מהגרף יכולות לסייע בקביעת מדיניות ציבורית ותכניות חברתיות, כדי לשפר את רמות האושר והרווחה במדינות שונות.



גרף ה-Scatter Matrix (נקרא גם גרף מטריצת פיזור) מציג את היחסים בין מספר משתנים על ידי הצגת גרפים פיזור (scatter plots) בין כל זוג משתנים. מטריצה זו מציגה גם היסטוגרמות על האלכסון הראשי של המטריצה עבור כל משתנה. גרף זה מאפשר זיהוי של קשרים אפשריים בין משתנים שונים, קורלציות וכיווני מגמה בצורה חזותית.

**קורלציה בין GDP לנורמליזציית Life Ladder**- נצפית קורלציה חיובית בין ה-Log GDP per capita לבין Life Ladder\_NORM. ניתן לראות שככל שה-GDP לנפש גבוה יותר, כך גם רמת האושר המנורמלת גבוהה יותר. הדבר מעיד על כך שהמדינות עם כלכלה חזקה נוטות להיות מאושרות יותר.

**השפעת התמיכה החברתית -** ישנה קורלציה חיובית בין Social support לבין Life Ladder\_NORM. מדינות עם רמות תמיכה חברתית גבוהות נוטות להציג רמות אושר גבוהות יותר. דבר זה מדגיש את החשיבות של התמיכה החברתית ברמת האושר של האוכלוסייה.

**השפעת הנדיבות -** הקשר בין Generosity לבין המשתנים האחרים פחות ברור, אך ניתן לראות כי במדינות עם רמות נדיבות גבוהות ישנו פיזור רחב יותר של רמות אושר ו-GDP. ייתכן שנדיבות אינה פקטור מרכזי שמשפיע ישירות על רמות האושר או הכלכלה.

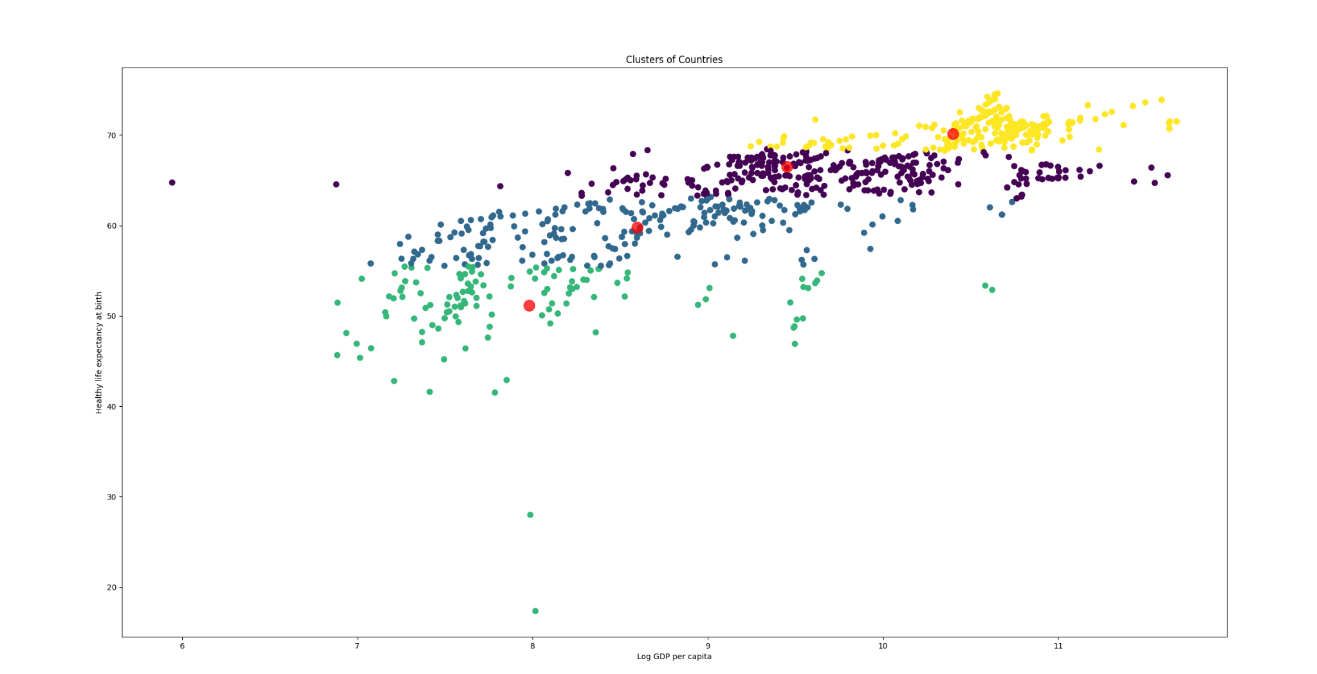
**תוחלת חיים בריאה -** נצפית קורלציה חיובית בין Healthy life expectancy at birth לבין Life Ladder\_NORM ו-GDP. ככל שתוחלת החיים הבריאה גבוהה יותר, כך גם רמת האושר וה-GDP נוטים להיות גבוהים יותר. דבר זה מראה כי בריאות טובה של האוכלוסייה מתקשרת עם אושר וכלכלה חזקה.

**מגוון גדול בין מדינות -** הגרף מציג פיזור רחב של נתונים בין מדינות שונות, מה שמדגיש את השונות הגדולה בין המדינות במשתנים הנבדקים. ניתן לראות כי כל מדינה מציגה פרופיל ייחודי של משתנים, דבר שמדגיש את הצורך בניתוח פרטני לכל מדינה.

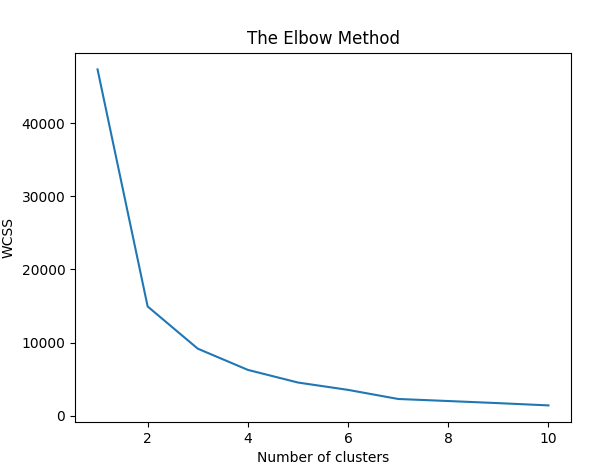
גרף ה-Scatter Matrix מאפשר ניתוח מעמיק של הנתונים ומציג את הקשרים האפשריים בין המשתנים בצורה חזותית וברורה. באמצעות גרף זה ניתן לזהות דפוסים ומגמות, ולהבין את ההשפעה של משתנים שונים על רמות האושר, הכלכלה והבריאות במדינות השונות.

חלוקה לאשכולות :

**גרף תוחלת החיים ביחס לתמ"ג:**

בחירת העמודות 'Log GDP per capita' ו'Healthy life expectancy at birth'.

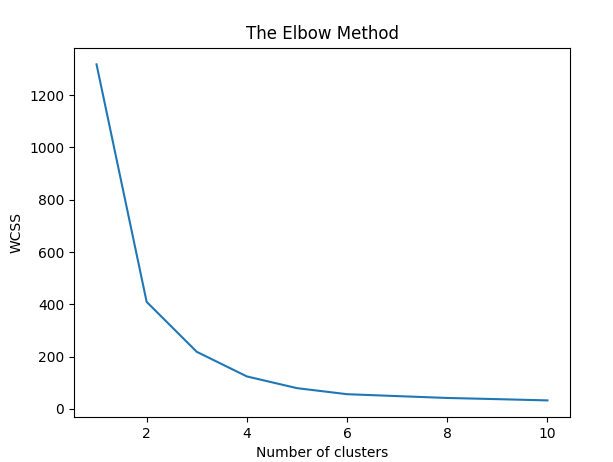
הפעלת k-means על הנתונים לזיהוי מספר אופטימלי של אשכולות וציור גרף פיזור עם אשכולות.



ניתן לראות בבירור 4 אשכולות שונים של מדינות. כל אשכול מייצג קבוצת מדינות עם רמות שונות של תוחלת חיים ותמ"ג. האשכול העליון הימני (צבוע בצהוב) מכיל מדינות עם תמ"ג גבוה ותוחלת חיים גבוהה. האשכול התחתון השמאלי (צבוע בירוק) מכיל מדינות עם תמ"ג נמוך ותוחלת חיים נמוכה. ניתן לראות שככל שהמדינה מתקדמת מבחינת תמ"ג, גם תוחלת החיים עולה. עם זאת, בתוך כל אשכול יש שונות פנימית שמשקפת את השונות בתמ"ג ותוחלת החיים בין המדינות השונות באותו אשכול. ישנם נקודות בודדות שנמצאות רחוקות מהאשכולות המרכזיים. מדינות אלו עשויות להיות יוצאות דופן בגלל גורמים ייחודיים שמשפיעים על התמ"ג או תוחלת החיים שלהן. ניתן לראות שהתוחלת החיים במדינות עם תמ"ג גבוה מתקרבת אחת לשנייה, מה שמעיד על כך שיש פחות שונות בתוחלת החיים כאשר המדינות מגיעות לרמות גבוהות של התפתחות כלכלית.

מדינות בעלות תמ"ג נמוך יחסית יכולות להשיג שיפור משמעותי בתוחלת החיים עם העלאת התמ"ג שלהן, כפי שמשתקף באשכולות התחתונים.

מדינות באשכול העליון עשויות להמשיך ולשפר את תוחלת החיים שלהן באמצעים נוספים מעבר לשיפור כלכלי ישיר, כמו מערכות בריאות מתקדמות ותמיכה חברתית גבוהה.

**מדד האושר ביחס לתמיכה חברתית:**

בחירת העמודות 'Social support' ו'Life Ladder'.

תמונה שמכילה צילום מסך, צבעוני, תרשים

התיאור נוצר באופן אוטומטיהפעלת חלוקה לאשכולות והצגת גרף פיזור עם האשכולות.

הגרף מציג את האשכולות של המדינות לפי מדד האושר (Life Ladder) ותמיכה חברתית (Social support). הנה כמה תובנות חשובות שניתן להסיק:

**מתאם חיובי בין תמיכה חברתית למדד האושר:**

ישנו מתאם חיובי ברור בין התמיכה החברתית לבין מדד האושר. ככל שרמת התמיכה החברתית במדינה גבוהה יותר, כך מדד האושר של תושביה גבוה יותר. זה מצביע על כך שתמיכה חברתית היא גורם משמעותי ברמת האושר של תושבי המדינה.

**קבוצות מדינות עם תכונות דומות:**

ניתן לזהות ארבע קבוצות מדינות עיקריות, כל אחת בצבע שונה:

אשכול תחתון (סגול): מדינות עם רמת תמיכה חברתית נמוכה יחסית ומדד אושר נמוך.

אשכול שני (ירוק): מדינות עם רמת תמיכה חברתית בינונית ומדד אושר בינוני.

אשכול שלישי (צהוב): מדינות עם רמת תמיכה חברתית גבוהה יותר ומדד אושר גבוה יותר.

אשכול עליון (כחול): מדינות עם רמת תמיכה חברתית גבוהה ומדד אושר גבוה מאוד.

**מדינות יוצאות דופן:**

ישנן מספר נקודות שנמצאות מחוץ למרכזי האשכולות, למשל מדינות עם תמיכה חברתית נמוכה ומדד אושר גבוה או להפך. מדינות אלו עשויות להיות יוצאות דופן בגלל גורמים ייחודיים, כמו מדיניות חברתית, תרבות או מצבים כלכליים ספציפיים.

**השפעה של תמיכה חברתית על אושר:**

ניתן לראות שמדינות עם תמיכה חברתית מאוד גבוהה נמצאות באופן עקבי באשכול העליון עם מדד אושר גבוה. זה מחזק את החשיבות של תמיכה חברתית ברמת האושר של התושבים.

****

**תפיסות של שחיתות ביחס למדד האושר:**

בחירת העמודות 'Perceptions of corruption' ו'Life Ladder'.

תמונה שמכילה צילום מסך, צבעוני, טקסט

התיאור נוצר באופן אוטומטיהפעלת חלוקה לאשכולות והצגת הגרף עם האשכולות.

הגרף מציג את האשכולות של המדינות לפי תפיסות השחיתות (Perceptions of corruption) ומדד האושר (Life Ladder). הנה כמה תובנות חשובות שניתן להסיק:

**מתאם שלילי בין שחיתות לאושר:**

ניתן לראות כי ככל שתפיסת השחיתות במדינה גבוהה יותר (כלומר, השחיתות נתפסת כגבוהה יותר), מדד האושר נמוך יותר. זה מצביע על כך שתפיסות השחיתות משפיעות על רמת האושר של תושבי המדינה.

**קבוצות מדינות עם תכונות דומות:**

ישנם ארבעה אשכולות עיקריים של מדינות, כל אחד בצבע שונה:

אשכול עליון (כחול): מדינות עם תפיסות שחיתות נמוכות ומדד אושר גבוה.

אשכול שני (צהוב): מדינות עם תפיסות שחיתות בינוניות ומדד אושר בינוני-גבוה.

אשכול שלישי (סגול): מדינות עם תפיסות שחיתות בינוניות-גבוהות ומדד אושר בינוני-נמוך.

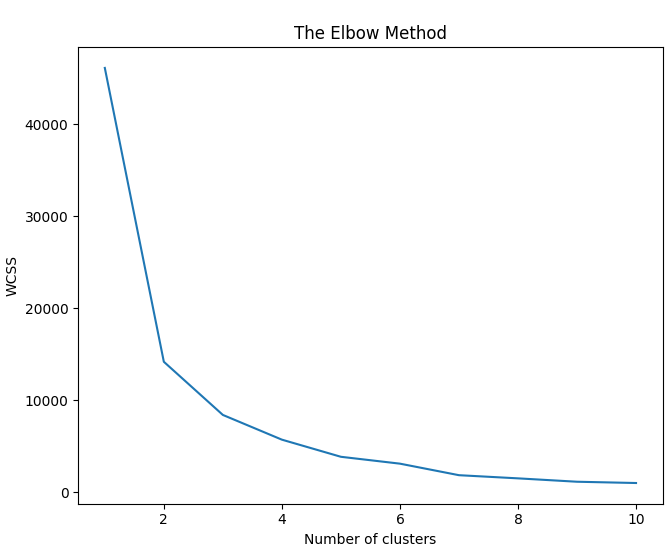
אשכול תחתון (ירוק): מדינות עם תפיסות שחיתות גבוהות ומדד אושר נמוך.

**מדינות יוצאות דופן:**

ישנן נקודות בודדות שנמצאות מחוץ למרכזי האשכולות, כגון מדינות עם תפיסות שחיתות נמוכות מאוד ומדד אושר נמוך יחסית, או להפך. מדינות אלו עשויות להיות יוצאות דופן בגלל גורמים ייחודיים, כמו מדיניות חברתית, תרבות או מצבים כלכליים ספציפיים.

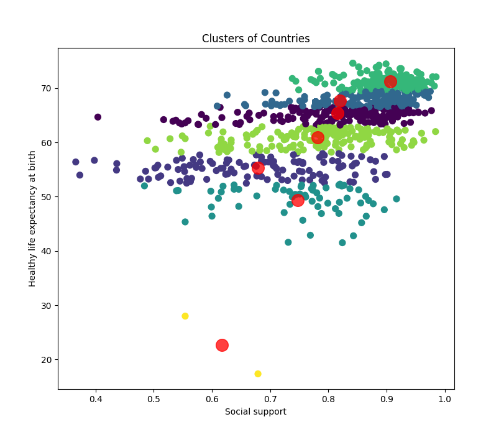
**השפעה של שחיתות על אושר:**

ניתן לראות שמדינות עם תפיסות שחיתות מאוד נמוכות נמצאות באופן עקבי באשכול העליון עם מדד אושר גבוה. זה מחזק את החשיבות של מאבק בשחיתות לשיפור רמת האושר של התושבים.

**תמיכה חברתית ותוחלת חיים:**

בחירת העמודות 'Social support' ו'Healthy life expectancy at birth'.

הפעלת חלוקה לאשכולות והצגת הגרף עם האשכולות.



הגרף מציג את האשכולות של המדינות לפי תמיכה חברתית (Social support) ותוחלת חיים בריאה בלידה (Healthy life expectancy at birth). הנה כמה תובנות חשובות שניתן להסיק:

**קשר חיובי בין תמיכה חברתית לתוחלת חיים:**

ניתן לראות מתאם חיובי בין תמיכה חברתית לתוחלת חיים בריאה בלידה. ככל שרמת התמיכה החברתית במדינה גבוהה יותר, תוחלת החיים הבריאה בלידה גם גבוהה יותר.

**קבוצות מדינות עם תכונות דומות:**

ישנם ארבעה אשכולות עיקריים של מדינות, כל אחד בצבע שונה:

אשכול עליון (כחול כהה): מדינות עם תמיכה חברתית גבוהה מאוד ותוחלת חיים גבוהה מאוד.

אשכול שני (ירוק): מדינות עם תמיכה חברתית גבוהה ותוחלת חיים גבוהה בינונית.

אשכול שלישי (סגול): מדינות עם תמיכה חברתית בינונית ותוחלת חיים בינונית.

אשכול תחתון (צהוב/אדום): מדינות עם תמיכה חברתית נמוכה ותוחלת חיים נמוכה.

**מדינות יוצאות דופן:**

ישנן נקודות בודדות שנמצאות מחוץ למרכזי האשכולות, כמו מדינות עם תמיכה חברתית נמוכה מאוד ותוחלת חיים נמוכה במיוחד. מדינות אלו עשויות להיות יוצאות דופן בגלל גורמים ייחודיים, כמו תנאים כלכליים קשים, מלחמות או מצבים בריאותיים גרועים.

השפעה של תמיכה חברתית על בריאות:

ניתן לראות שמדינות עם תמיכה חברתית גבוהה נמצאות באופן עקבי באשכול העליון עם תוחלת חיים גבוהה. זה מחזק את החשיבות של תמיכה חברתית לבריאות התושבים.

**אשכולות על פי נתונים כלכליים מרובים:**

בחירת מספר עמודות כלכליות כמו 'Log GDP per capita', 'Generosity' ועוד.

הפעלת חלוקה לאשכולות על העמודות הנבחרות והצגת גרפים מתאימים.

הגרף התלת-ממדי מציג את האשכולות של המדינות לפי שלושה משתנים כלכליים עיקריים: 'Log GDP per capita', 'Generosity', ו-'Social support'. הנה כמה תובנות חשובות שניתן להסיק:

**הבדלים בין אשכולות במדדים הכלכליים:**

ניתן לראות הבדלים ברורים בין האשכולות במדדים הכלכליים. כל אשכול מייצג קבוצה של מדינות עם תכונות כלכליות דומות.

למשל, האשכול הכחול הכהה (0) מכיל מדינות עם תמ"ג נמוך יחסית, נדיבות נמוכה ותמיכה חברתית בינונית.

לעומת זאת, האשכול הירוק (3) מכיל מדינות עם תמ"ג גבוה, נדיבות גבוהה ותמיכה חברתית גבוהה.

**מגוון רחב של מדינות בתוך כל אשכול:**

למרות שהאשכולות מוגדרים בבירור, יש שונות פנימית בתוך כל אשכול, מה שמרמז על כך שישנם גורמים נוספים שמשפיעים על המדדים הכלכליים.

**מיקומים ייחודיים של מדינות מסוימות:**

ישנם נקודות בודדות שנמצאות מחוץ למרכזי האשכולות, כמו מדינות עם נדיבות גבוהה במיוחד אך תמ"ג נמוך, או להפך. מדינות אלו עשויות להיות יוצאות דופן בגלל מדיניות כלכלית, תרבותית או חברתית ייחודית.

**הקשר בין שלושת המשתנים:**

ניתן לראות שיש מתאם חיובי בין 'Log GDP per capita' לתמיכה חברתית ולנדיבות במדינות עם תמ"ג גבוה יותר. כלומר, מדינות עם תמ"ג גבוה נוטות גם לתמיכה חברתית ונדיבות גבוהות יותר.

**רגרסיה ליניארית על** – עמודות Log GDP per capita (תמ"ג לנפש בלוגריתם) ו-Healthy life expectancy at birth (תוחלת חיים בריאה בלידה):

קשר כלכלי וחברתי ברור שמראה איך הכלכלה משפיעה על הבריאות.

הגרף מציג את הקשר בין תמ"ג לנפש בלוגריתם (Log GDP per capita) ותוחלת חיים בריאה בלידה (Healthy life expectancy at birth) יחד עם קו הרגרסיה הליניארית. הנה כמה תובנות חשובות שניתן להסיק:

**קשר חיובי בין תמ"ג לנפש לתוחלת חיים:** ישנו מתאם חיובי ברור בין Log GDP per capita לבין Healthy life expectancy at birth. ככל שתמ"ג לנפש במדינה גבוה יותר, תוחלת החיים הבריאה בלידה גם גבוהה יותר.

**כיוון קו הרגרסיה:** קו הרגרסיה מראה את הכיוון הכללי של הנתונים, שמראה שיפור בתוחלת החיים הבריאה עם עלייה בתמ"ג לנפש.

**פיזור הנתונים סביב קו הרגרסיה:** למרות שהקו מראה מגמה כללית של עלייה, ישנם פיזור נתונים רחב סביב הקו. זה מעיד על כך שלמרות הקשר החיובי, ישנם גורמים נוספים שמשפיעים על תוחלת החיים הבריאה מעבר לתמ"ג לנפש.

**מדינות יוצאות דופן:**

ישנם כמה נקודות נתונים שנמצאות רחוק מהקו, כמו מדינות עם תמ"ג נמוך ותוחלת חיים גבוהה יחסית או להפך. מדינות אלו עשויות להיות יוצאות דופן בגלל מדיניות בריאות מתקדמת, תרבות בריאותית גבוהה, או גורמים אחרים.

**תובנות מהנתונים של קובץ input1\_df.csv:**

**מידע כללי**:

* + ישנם 1029 שורות נתונים ו-5 עמודות.
  + כל העמודות מוגדרות כטיפוס אובייקט (טקסט).

**ערכים חסרים**:

* + כמות הערכים החסרים משתנה בין העמודות, עם חסרים בשיעור של כ-9% עד 11.5% בעמודות שונות.
  + העמודה עם הכי הרבה ערכים חסרים היא "Country name" עם כ-11.5% ערכים חסרים.

**סטטיסטיקה תיאורית**:

* + יש חוזרים רבים בעמודות הטקסט, עם כמה ערכים שמופיעים יותר מפעם אחת.
  + נראה שיש כמה ערכים אקראיים או שגיאות בעמודות מספריות(כמו "bmifXR7bqV" בעמודת מחיר), שצריך לנקות.

**ניתוח של ערכים קיצוניים ושגיאות**:

* + יש נתונים שלא נראים תקינים בעמודת המחירים והעובדים, כמו טקסט במקום מספרים. זה דורש תיקון לפני כל ניתוח נוסף.
  + הדפסת שורות עם נתונים חסרים עשויה להצביע על דפוסים או קשרים בין הנתונים החסרים לעמודות אחרות.

**תובנות מנתוני קובץ input2\_df.csv:**

**מידע כללי:**

* הקובץ מכיל 2363 שורות ו-11 עמודות.
* ישנן עמודות בעלות נתונים מספריים וטקסטואליים, כאשר שתי עמודות הן מסוג float64 והשאר הן object.

**ערכים חסרים:**

* ישנם ערכים חסרים בכל העמודות, עם שיעורים שונים מ-9% עד כ-21.7%.
* "Generosity" ו-"Life Ladder" הן העמודות עם הכי הרבה ערכים חסרים, מה שמחייב טיפול מיוחד בניתוחים שמבוצעים על הנתונים האלה.

**איכות וניקוי נתונים:**

* ישנם ערכים שאינם נראים רגילים או תקינים בנתונים, כגון "Error" ו"Unknown", אשר יכולים להצביע על נתונים שגויים או חסרים שלא זוהו כך באופן אוטומטי. יש לבדוק ולנקות את הנתונים מאלו לפני ניתוח נוסף.

**סטטיסטיקה תיאורית:**

* ישנם ערכים חוזרים בעמודות הטקסטואליות וערכים נדירים שחוזרים בעמודות המספריות.
* הנתונים דורשים ניתוח מעמיק יותר לזיהוי טרנדים והבנת הקשרים בין המשתנים.