

שאלה 10 – מסקנות

במהלך המחקר, חקרנו את ההשפעה של אי שוויון מגדרי בתחום השכלה על מגוון משתנים סוציו אקונומיים במדינות שונות. במהלך הניתוח התחלנו בטעינה וניקוי סט נתונים של נתוני השכלה עולמיים, במהלך תהליך ניקוי הנתונים, טיפלנו בערכים חסרים על ידי החלפת הערכים החסרים עם הממוצע של אותה עמודה והתחלנו את הניתוח על סט נתונים ללא ערכים חסרים לאחר שביצענו את כל המשימות של סעיפים 1 ו 2.

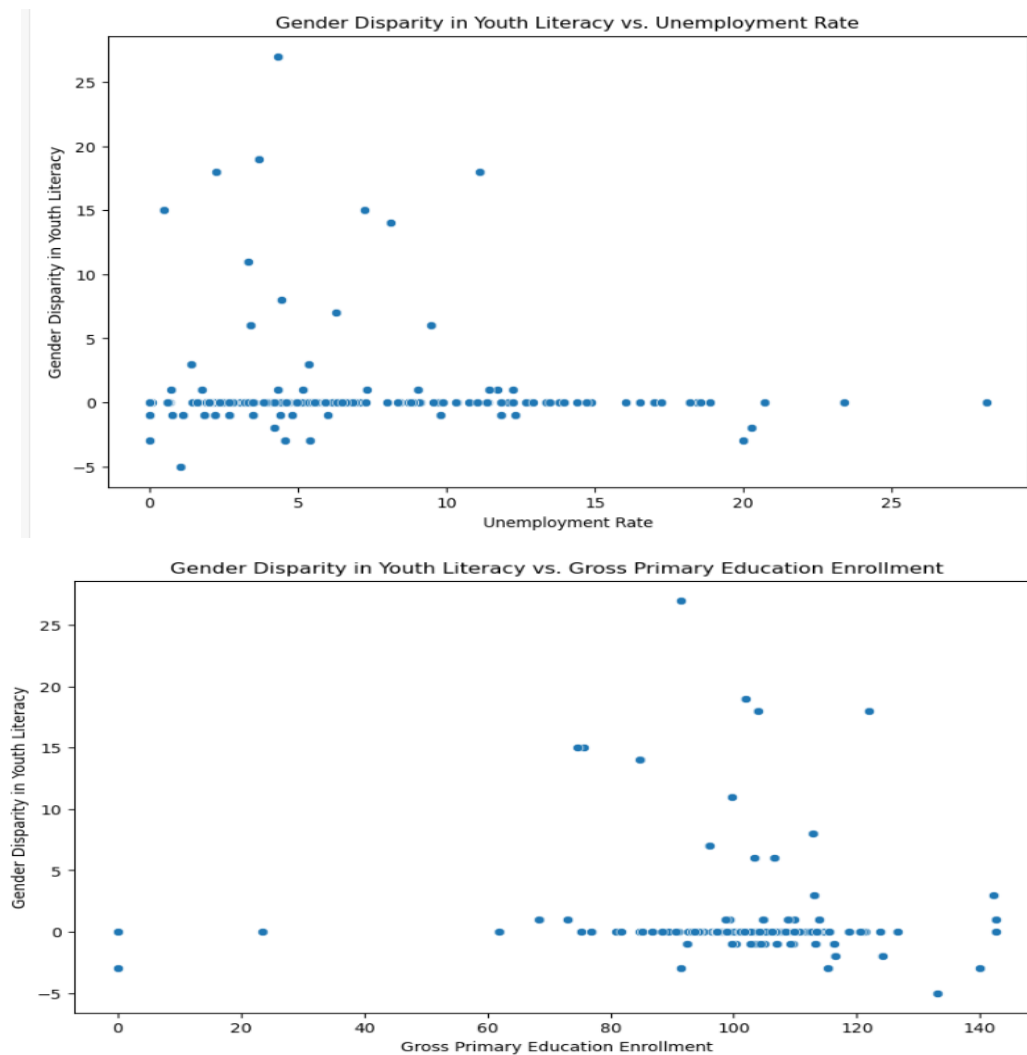
סעיף 3 –

חישוב אי שוויון מגדרי –

שמנו דגש על אי שוויון מגדרי ברמות האוריינות בקרב צעירים, וגילינו שבעוד שמדינות רבות השיגו שוויון מגדרי בתחום ההשכלה, עדיין קיים אי שוויון מגדרי במגוון אזורים. זוהי תובנה חשובה מכיוון שהיא מדגישה את האתגרים המתמשכים בהשגת שוויון מגדרי בתחום ההשכלה ברמה עולמית.

סעיף 4 – EDA

במהלך ניתוח ה EDA לא גילינו תבנית ברורה בין מדינות עם אי שוויון מגדרי גבוה באוריינות בקרב צעירים לבין רמות ההרשמה להשכלה או לרמות האבטלה במדינה. זה מצביע על כך שאי שוויון מגדרי בהשכלה לא בהכרח משפיע באופן ישיר על שני המשתנים הסוציו אקונומיים האלה. אבל, חשוב לקחת בחשבון משתנים אחרים אשר עלולים להשפיע על הקשר הזה.



סעיף 5 –

ניסוח היפותזה – היפותזה שבחרנו לבחון היא: "במדינות אשר בהן יש הבדל גדול יותר במספר הבנים והבנות שמסיימים תיכון, יהיה הבדל גדול יותר בכמה בנים ובנות יכולים לקרוא ולכתוב כאשר הם בוגרים צעירים".

רצינו לבדוק את הקשר בין סיום תיכון בקרב בנים ובנות לבין יכולות הקריאה והכתיבה שלהם, והאם משך הזמן שהם נמצאים במסגרת הבית ספרית תורמת ליכולות הקריאה והכתיבה או לא. ביצענו ניתוח קורלציות בין אי שוויון מגדרי בקרב מסיימים תיכון לבין רמות אוריינות בקרב צעירים וקיבלנו קורלציה של 0.479 וערך p-value קטן מאוד דבר אשר מצביע על קורלציה חיובית בין המשתנים. זה תומך בהיפותזה שניסחנו שמדינות עם אי שוויון מגדרי גדול ורמות מסיימי השכלת החובה הן גם בעלות אי שוויון ברמות האוריינות. זה מדגיש את החשיבות של השארת גם בנים וגם בנות בבית ספר כדי לשפר את רמות האוריינות בקרב שני המגדרים.

סעיף 6 –

ניתוח רגרסיה לינארית –

במהלך הניסיון לבצע חיזוי של רמות ההרשמה לחינוך ברמת בית ספר יסודי בהתבסס על אי שוויון מגדרי בתיכון לא הצלחנו לבנות מודל חזק על בסיס משתנים אלו בלבד וקיבלנו תוצאות R^2 נמוכות מידי ושגיאה גדולה מידי. זה מצביע על הסיבוכיות בחיזוי רמות רישום בהתבססות על אי שוויון מגדרי בלבד, ולכן צריך להתחשב במשתנים נוספים.

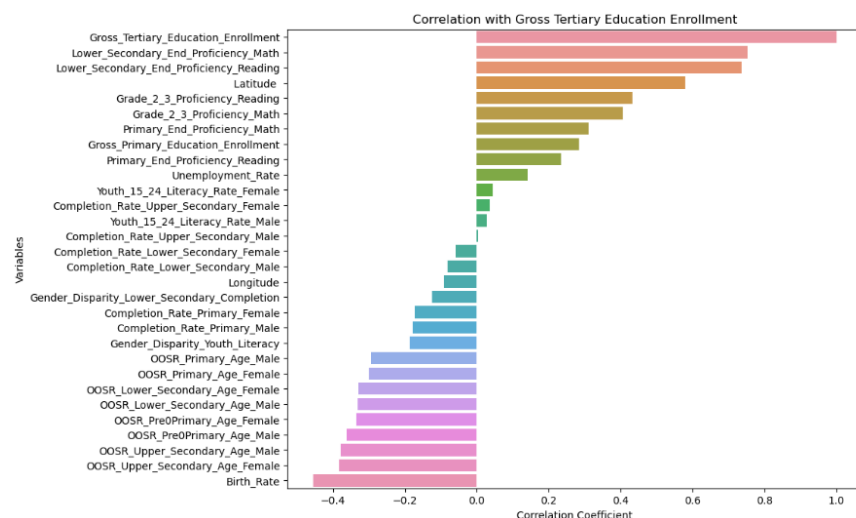
סעיף 7 –

ניתוח קורלציות – במהלך המחקר בשלב זה בדקנו מקרוב איך משתנים שונים בהשכלה קשורים אחד לשני. במיוחד, שמנו לב שיש קשר חזק בין כמות הסטודנטים שנרשמים להשכלה גבוהה (כמו אוניברסיטה) לבין ההצלחה שלהם במגוון תחומים ונושאים ולרמות ההשכלה הכלליות שלהם.

המשתנה Gross Tertiary Education Enrollment אשר מודד כמה סטודנטים נרשמו להשכלה גבוהה, הוא סימן מפתח לגבי כמה מדינה מסוימת מעריכה לימודים מתקדמים והשכלה גבוהה. במהלך הניתוח גילינו שיש קורלציה וקשר חזק בין המשתנה הזה לבין משתני השכלה אחרים, לדוגמא:

המשתנה בעל קורלציה של 0.752771 עם המשתנה שמודד כמה טובים סטודנטים במתמטיקה בתיכון. זה מצביע על כך שכל שיותר סטודנטים הולכים להשכלה גבוהה, ככה הביצועים שלהם במתמטיקה היו טובים יותר בגיל צעיר יותר.

בנוסף, הייתה קורלציה של 0.735815 עם יכולות קריאה בתיכון. לכן בדומה למתמטיקה, ככל שהסטודנטים נרשמים להשכלה גבוהה, אז גם ציוני הקריאה שלהם גודלים.



סעיף 8 – סיווג

בשלב זה, יצרנו מודל עץ החלטה ומודל יער אקראי כדי לחזות את קטגוריית קצב הילודה במדינות שונות. קודם כל הפכנו את המשתנה Birth_Rate לקטגוריאלי בעזרת Label Encoder ו qcut. מודל עץ ההחלטה סבל מהתאמת יתר מכיוון שהשיג תוצאות - Accuracy, Precision, Recall, F1-score של 1.0 מה שמראה על 100% דיוק והתאמה לנתונים. זה הגיוני עבור מודל מסוג עץ החלטה כאשר משתמשים בדאטה סט קטן יחסית ולא מסובך העץ לומד אותו בצורה מושלמת. העץ האקראי השיג דיוק של בערך 95% ולכן לא סובל מהתאמת יתר, ויתכן ויתאים גם למידע חדש בצורה טובה יותר מאשר עץ ההחלטה.

Decision Tree - Accuracy: 1.0

Decision Tree - Precision: 1.0

Decision Tree - Recall: 1.0

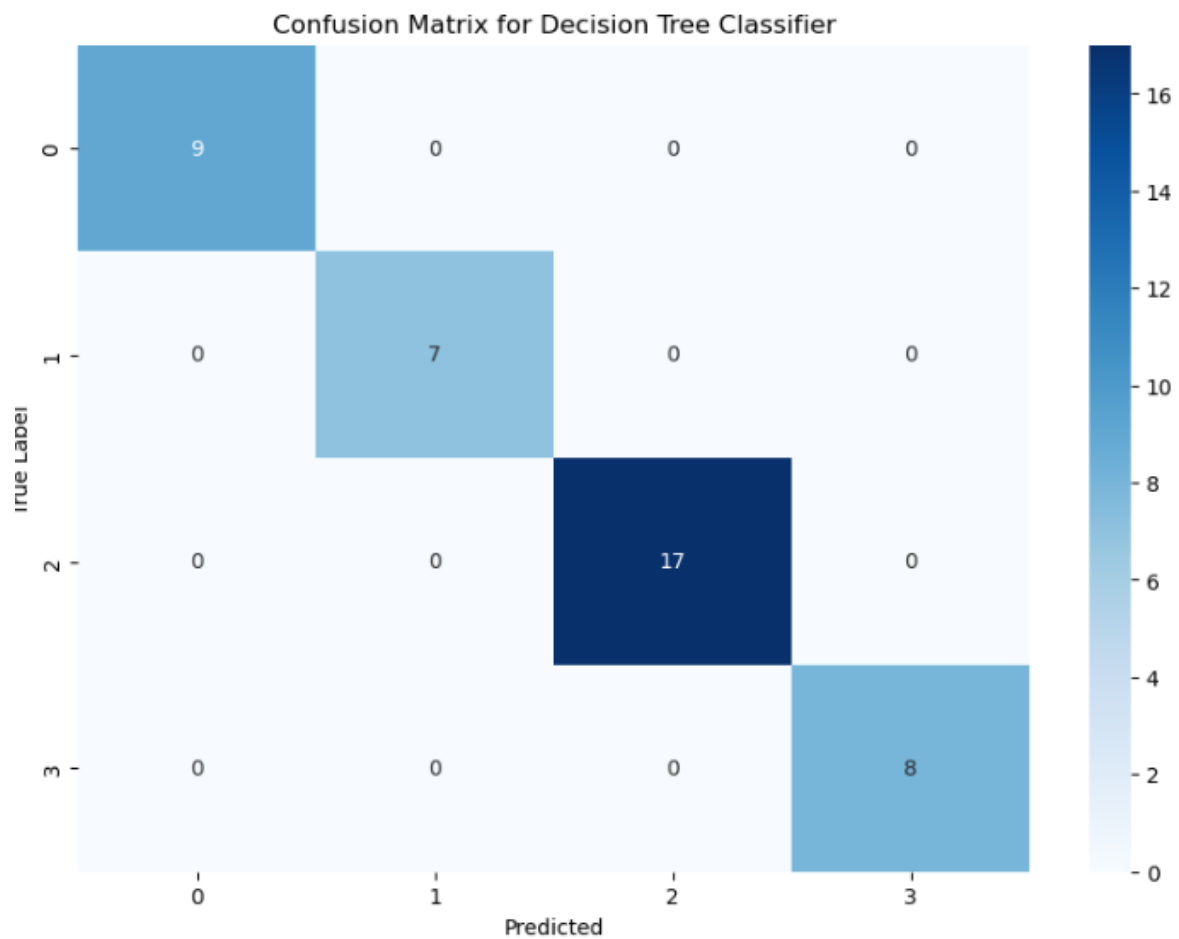
Decision Tree - F1-Score: 1.0

Random Forest - Accuracy: 0.9512195121951219

Random Forest - Precision: 0.9409722222222222

Random Forest - Recall: 0.957516339869281

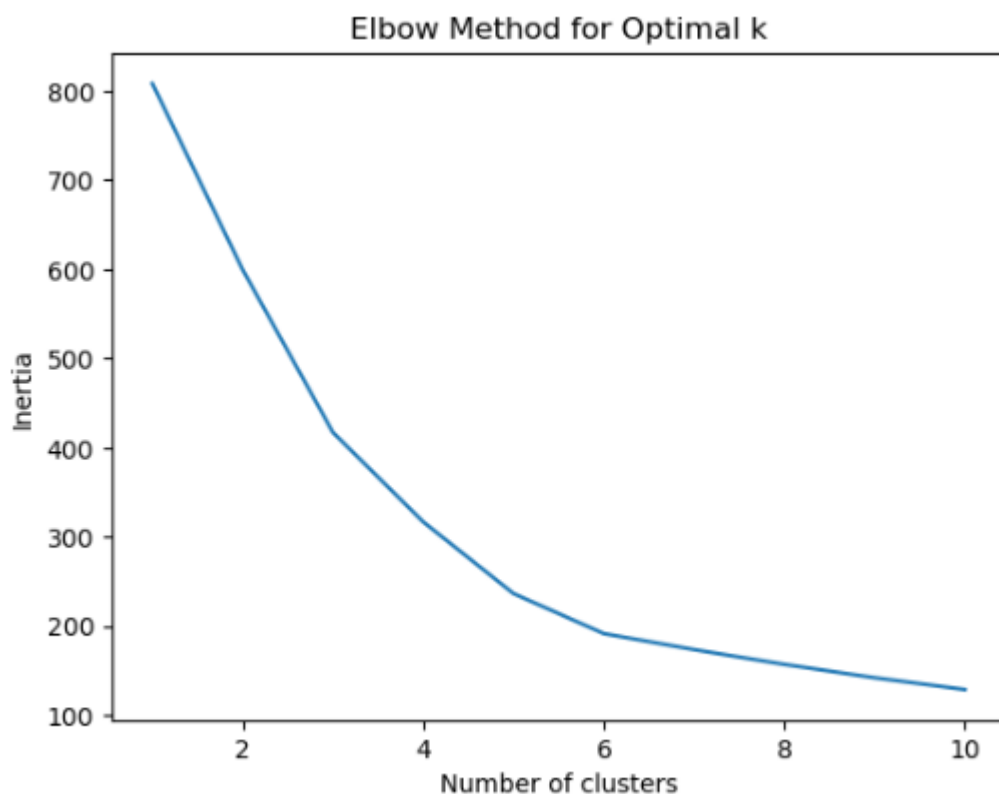
Random Forest - F1-Score: 0.9463458110516934



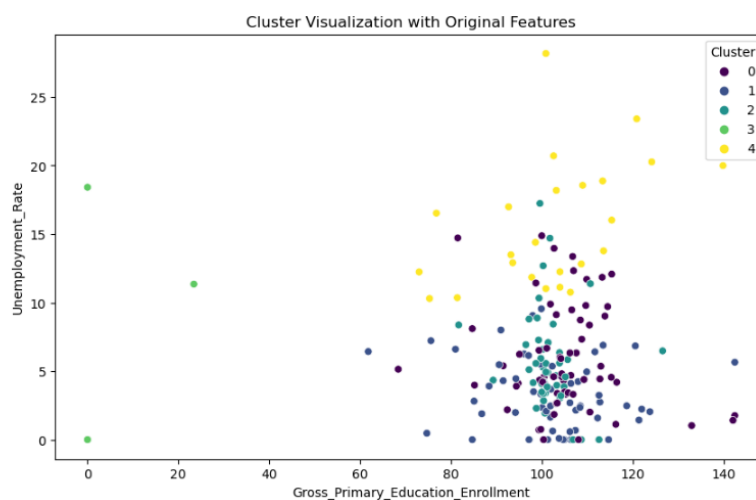
לסיכום, הסיווג שלנו הציג כי רמות השכלה ותכונות נוספות בעלות קשר חזק עם רמות ילודה בקרב מדינות שונות. הדיוק של המודלים שלנו מצביעים על כך ששיפור של ההשכלה יכול להוביל להשפעות חברתיות נרחבות, כולל מגמות דמוגרפיות, שיעור ילודה במדינה וכו'.

סעיף 9 – K-means –

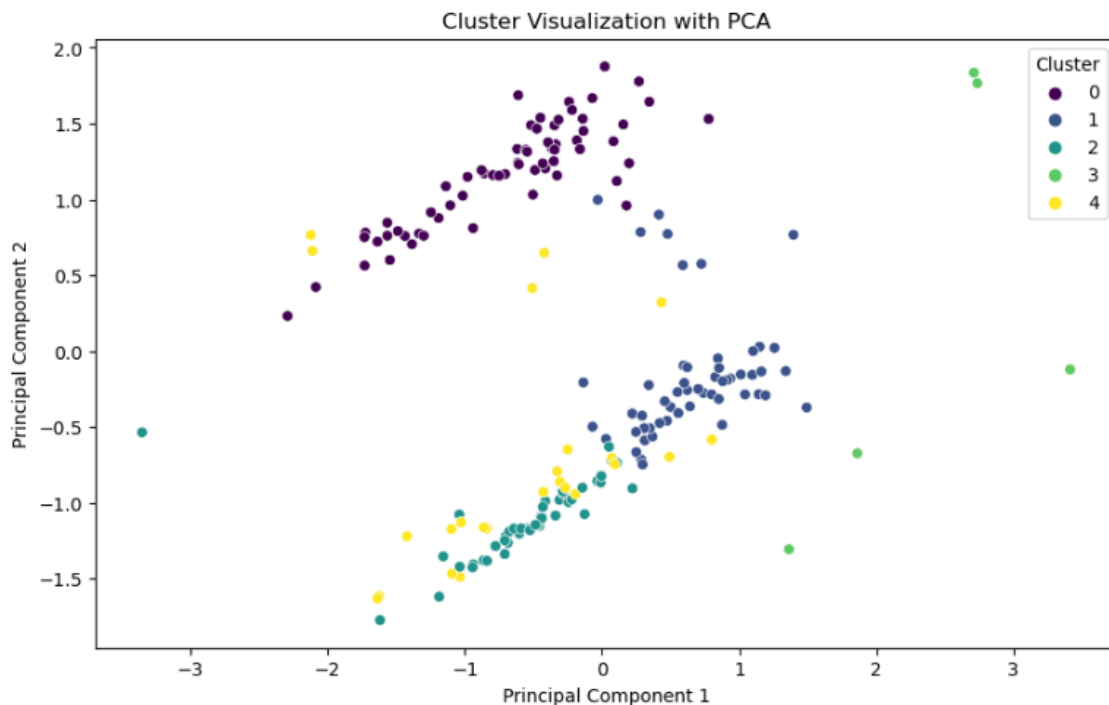
בסעיף זה ניסינו לקבץ מגוון מדינות ביחד בהתבסס על תכונות שונות כמו רמות השכלה ומצב אקונומי, והאם בנים ובנות מקבלים השכלה בצורה שווה. השתמשנו בשיטת המרפק כדי למצוא את ה K האופטימלי, שעל פי התרשים זיהינו ש $k=5$ הוא האופטימלי מכיוון שהוא נמצא בדיוק במרפק:



לאחר מכן אימנו מודל K-means על הנתונים וחילקנו אותם ל 5 קבוצות שונות



התוצאות הראו שהרבה מדינות מאוד דומות אחת לשנייה ושהקבוצות ממש חופפות, ואין הפרדה ברורה במיוחד בין קבוצה 0 לקבוצה 1 ו 2. ניסינו לבצע PCA להקטנת מימדים וקבלת הפרדה ברורה יותר ואכן התוצאות היו טובות יותר וקיבלנו קבוצות יותר ברורות:



עקיף 11 – Reflections

במהלך המחקר בדקנו איך הבדלים בחינוך בין בנים לבנות משפיעות על המצב החברתי והכלכלי במדינות שונות. על אף שלמדנו המון, המחקר שלנו הראה שהבעיות הללו הן מסובכות. גילינו שלא כל בעיה בחברה נגרמת בצורה ישירה על ידי הבדלים בין רמות ההשכלה של בנים ובנות, אבל יש הרבה משתנים שקשורים בצורה חזקה אחד לשני.

דרכים לשיפור ומגבלות:

- המגבלה העיקרית היא מידע מצומצם - שימוש במידע מגוון ורחב יותר, יכול לעזור לנו לגלות דברים נוספים ומידע חדש לגבי איך הבדלים בין מגדרים משפיעים על החברה ועל הכלכלה.
- שימוש בשיטות ניתוח מסובכות יותר כמו מודלים לא לינארית או טכניקות למידת מכונה מתוחכמות יותר יכולים לעזור לנו להבין את הקשרים המסובכים הללו.

מחקר נוסף:

- חקירת ההשפעה של התערבויות שיעשו על ידי המדינה כדי להקטין את חוסר השוויון המגדרי בהשכלה יכולה לספק תובנות לגבי איך למנוע את אי השוויון הזה.

המחקר שביצענו מדגיש את הצורך הקריטי למאמצים מתמשכים כדי להשיג שוויון מגדרי בהשכלה, ואיך המאמצים האלה יכולים לשפר את המצב הסוציו אקונומי בעולם כולו.