

השפעת הצלחה בלימודים על הזוגיות

מגישה : אביטל אקרמן 208933606

הקדמה

את הדאטא סט לקחתי מ Kaggle, זהו מידע שנאסף על סטודנטים בתואר בהנדסה והוראת המדעים ב-2019 מאוניברסיטה בארצות הברית.

קישור לדאטא :

<https://www.kaggle.com/datasets/csafrt2/higher-education-students-performance-evaluation>

בתוך הדאטא סט הזה יש 145 סטודנטים (שורות), ו-331 משתנים (עמודות). המשתנים הם קטגוריים, ומגיעים מקודדים.

שאלות העבודה:

- האם זוגיות מושפעת מהשקעה בלימודים? – שאלה זו רלוונטית מכיוון שלימודים דורשים הרבה משאבים מנטליים וזמן רב, במיוחד למי שרוצה להצטיין. לכן מעניין לדעת האם סטודנטים שמשקיעים שעות רבות ומגיעים להישגים טובים בלימודים, מצליחים לתחזק זוגיות בצורה פחותה מסטודנטים שמשקיעים פחות. בנוסף לכך, רוב הסטודנטים הם בגיל הפריחה של זוגיות, וזה עניין שמעניין אותם רבות. לא תמיד זה קל לשלב לימודים וחופש/תחזוקה של זוגיות. לכן רציתי לראות אם יש קשר בין כל מיני פרמטרים שונים המראים על השקעה בלימודים והמצב הזוגי של הסטודנטים והאם ניתן לחזות מצב זוגי לפי פרמטרים אלו.
- האם יש השפעה של מגדר הסטודנט למצב הזוגי שלו? – שאלה זו מעניינת מכיוון שהדאטא סט הוא של סטודנטים להנדסה עם רוב גברי, ורציתי לדעת אם יש השפעה של המגדר על הזוגיות.
- האם מצב המגורים של הסטודנט משפיע על המצב הזוגי? – שאלה זו מעניינת מכיוון שמצב מגורים יכול להשפיע על הרצון והיכולת להבין בן/בת זוג למקום המגורים. לדוגמה סביר להניח שסטודנטים שגרים עם ההורים ירגישו פחות בנוח להביא את הדייטים שלהם לבית בגלל חוסר פרטיות, לעומת סטודנטים שגרים לבד למשל.
- האם דרך ההגעה לקמפוס משפיעה על המצב הזוגי? – שאלה זו מעניינת מכיוון שאנשים שגרים רחוק מהקמפוס יכולים להרגיש מנותקים מהסביבה החברתית העיקרית של הקמפוס, דבר שיכול להשפיע על היכרות בקמפוס ועל זוגיות.

חקירת הדאטא

תמונה 1 – הדאטא סט

	studentid	age	gender	hs_type	scholarship	work	activity	partner	salary	transport	...	prep_study	prep_exam	notes	listens	likes_discuss	classroom	cuml_gpa	exp_gpa	course id	grade
0	STUDENT1	22-25	Male	Other	50%	Yes	No	No	\$135-200	Bus	...	Alone	Closest date to the exam	Always	Sometimes	Never	Useful	<2.00	<2.00	1	60-64
1	STUDENT2	22-25	Male	Other	50%	Yes	No	No	\$135-200	Bus	...	Alone	Closest date to the exam	Always	Sometimes	Always	Useful	2.00-2.49	2.50-2.99	1	60-64
2	STUDENT3	22-25	Male	State	50%	No	No	No	\$201-270	Other	...	Alone	Closest date to the exam	Sometimes	Sometimes	Never	Not useful	2.00-2.49	2.00-2.49	1	60-64
3	STUDENT4	18-21	Female	Private	50%	Yes	No	Yes	\$201-270	Bus	...	Alone	Regularly during the semester	Always	Sometimes	Sometimes	Not useful	2.50-2.99	2.00-2.49	1	60-64
4	STUDENT5	22-25	Male	Private	50%	No	No	Yes	\$271-340	Bus	...	With friends	Closest date to the exam	Sometimes	Sometimes	Sometimes	Not useful	2.00-2.49	2.00-2.49	1	60-64

הדאטא סט הגיע מקודד למספרים לפי מפתח שהוצג ב Kaggle. המרתי את הקטגוריות המספריות לקטגוריות שמיות. לאחר מכן רציתי לשנות את הייצוג של הציונים (שהוצגו כאותיות לועזיות) לייצוג של טווח ציונים מספריים, כדי שיהיה נוח יותר ואינטואיטיבי יותר להסתכל. ניתן לראות שיש כאן ערך לא מוסבר כל כך של משכורות. לא נרשם סדר הגודל של המשכורת (דולרים בודדים או אלפי דולרים), האם היא משכורת שבועית/יומית/שנתית והאם זוהי משכורת בזמן התואר או משכורת משוערת לאחר התואר (תמונה 1). המשכורת הממוצעת לבוגר תואר ראשון בהנדסה בארצות הברית ללא ניסיון היא 91,000 דולר בשנה. בדאטא סט, הערכים גבוהים בהרבה מערך זה, ולכן לא יכולים להיות ערכים משוערים באלפי דולרים בשנה. אלו יכולים להיות ערכים ליום עבודה לאחר התואר, או לשבוע/חודש עבודה בעבודה מזדמנת במהלך התואר, אך אין דרך לדעת. בנוסף, הורדתי את עמודת "studentid" והפכתי אותה לאינדקס.

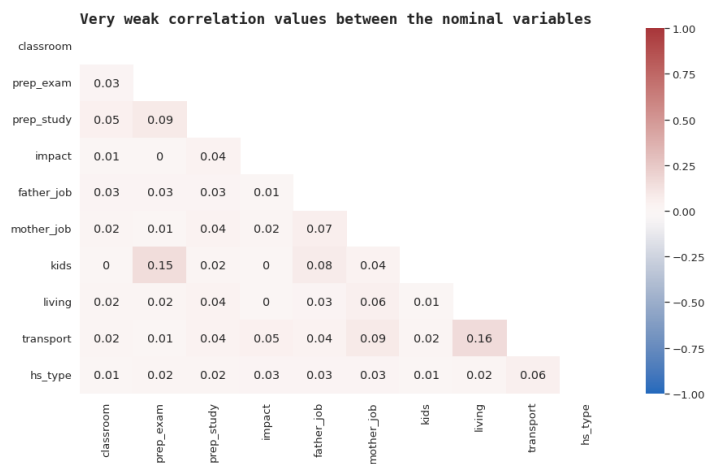
הדאטא סט הזה היה נקי, ללא ערכים חסרים או עמודות כפולות.

בדיקת קורלציה בין המשתנים

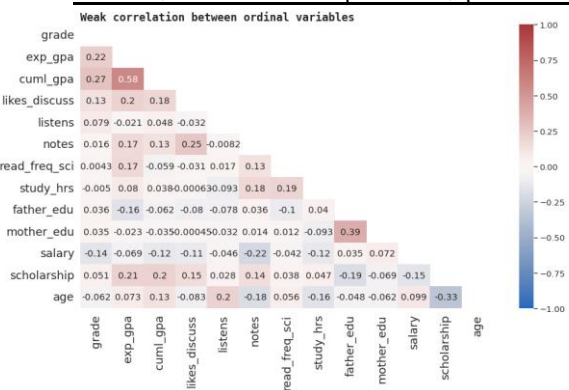
בדאטא יש 3 סוגי משתנים קטגוריים שלכל אחד יש מבחן סטטיסטי שונה הבודק את הקורלציה שלהם. חילקתי את המשתנים לפי סוג המשתנה שלהם ויישמתי את השיטה המתאימה:

- משתנים נומינליים - cramer's V test / חי בריבוע. בין המשתנים הנומינליים לא נראתה קורלציה גבוהה (תמונה 2).
- משתנים בינאריים – קורלציית פירסון. לא נראתה קורלציה גבוהה בין רוב המשתנים הבינאריים. הקורלציה המשמעותית ביותר הייתה קורלציה שלילית (-0.36) בין מין ונוכחות בשיעורים (תמונה 3)
- משתנים אורדינליים - Kendall method. לא נראתה קורלציה גבוהה בין רוב המשתנים האורדינליים. הייתה קורלציה גבוהה (0.58) בין ממוצע הציונים בסמסטר הקודם (cuml_gpa), לממוצע הציונים המוערך בסמסטר הקרוב (exp_gpa). בנוסף, הייתה קורלציה גבוהה בין השכלת האם להשכלת האב (0.39) (תמונה 4)

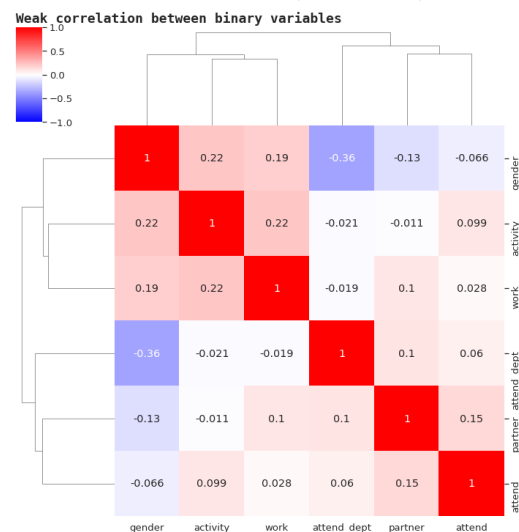
תמונה 2 - קורלציה בין משתנים נומינליים



תמונה 4- קורלציה בין המשתנים האורדינליים



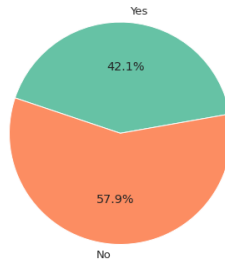
תמונה 3 - קורלציה בין המשתנים הבינאריים



רציתי לבדוק את התפלגות הסטודנטים הנמצאים בזוגיות ולא נמצאים בזוגיות בדאטא. ניתן לראות שיש אחוז גדול יותר של סטודנטים שלא נמצאים בזוגיות מאשר סטודנטים הנמצאים בזוגיות (תמונה 5).

תמונה 5- התפלגות זוגיות בדאטא

Distribution of partner variable in our data



ניתוח המידע

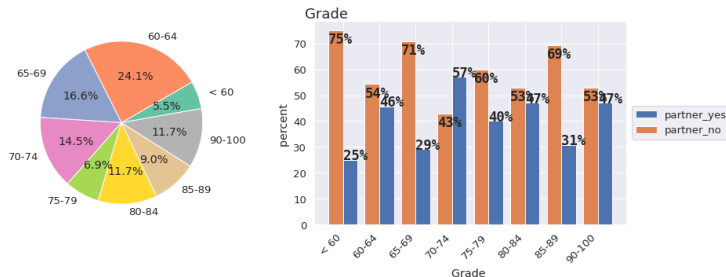
לקחתי כמה קטגוריות שעניינו אותי והסתכלתי על ההתפלגות של כל הקטגוריות בתוך כל משתנה בדאטא ועל אחוז הסטודנטים המתחזקים זוגיות בכל קטגוריה במשתנה. לאחר מכן חישבתי P-value לכל קטגוריה ב two tail test כאשר השערת האפס הייתה שלא נראה שוני במצב זוגי בכל קטגוריה. את חישוב ה P-value ביצעתי באמצעות שיטת פרמוטציה כאשר הערך הסטטיסטי שהשתמשתי בו הוא ההפרש בין כמות הסטודנטים שנמצאים בזוגיות ללא נמצאים בזוגיות בערך מוחלט בכל קטגוריה. ביצעתי Bonferroni Correction לערך ה α (שבתחילה היה 0.05) ע"י חלוקה במספר הקטגוריות.

מהו הקשר בין השקעה בלימודים למצב הזוגי?

לאחר שהסתכלתי על הדאטא, החלטתי להסתכל על פרמטרים שמשפיעים על השקעת הסטודנטים בלימודים תחת ההנחה שלסטודנטים שמשקיעים זמן רב בלימודים לא יהיה פנאי לתחזוק מערכת יחסים. הסתכלתי על הפרמטרים: ציון סופי בקורס (grade), ממוצע ציונים בסמסטר הקודם (cuml_gpa), שעות לימוד שבועיות (study_hrs), נוכחות בשיעורים (attendance) והשקעה לפני מבחן האמצע (prep_exam).

מה הקשר בין ציונים למצב הזוגי?

תמונה 6- התפלגות הציונים לפי מצב זוגי

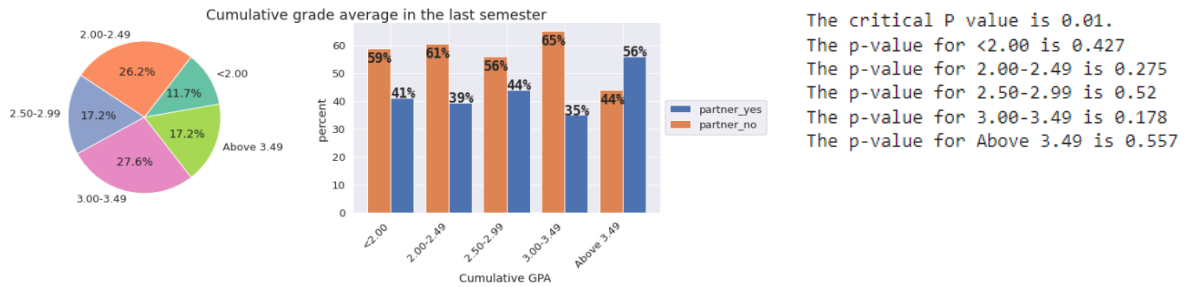


The critical P value is 0.00625.
 The p-value for 65-69 is 0.008.
 The p-value for 80-84 is 0.693.
 The p-value for 70-74 is 0.74.
 The p-value for 85-89 is 0.326.
 The p-value for 60-64 is 0.216.
 The p-value for 75-79 is 0.764.
 The p-value for 90-100 is 0.671.
 The p-value for < 60 is 0.352.

ניתן לראות שברוב קטגוריות הציונים אחוז גבוה יותר של סטודנטים שלא בזוגיות (תמונה 6). בנוסף ניתן לראות שההבדל המשמעותי ביותר הוא בציונים מתחת ל 60, בהם יש 75% של סטודנטים שלא נמצאים בזוגיות, ובטווח של 65-69 ישנם 71% מהסטודנטים שלא נמצאים בזוגיות. לפי ערכי ה P-value שחישבתי (תמונה 6), ניתן לראות שבאף אחת מקטגוריות הציונים אין הבדל משמעותי מבחינה סטטיסטית. מכאן שלפי הדאטא לציונים אין השפעה משמעותית על המצב הזוגי.

מה הקשר בין ממוצע הציונים בסמסטר הקודם למצב הזוגי?

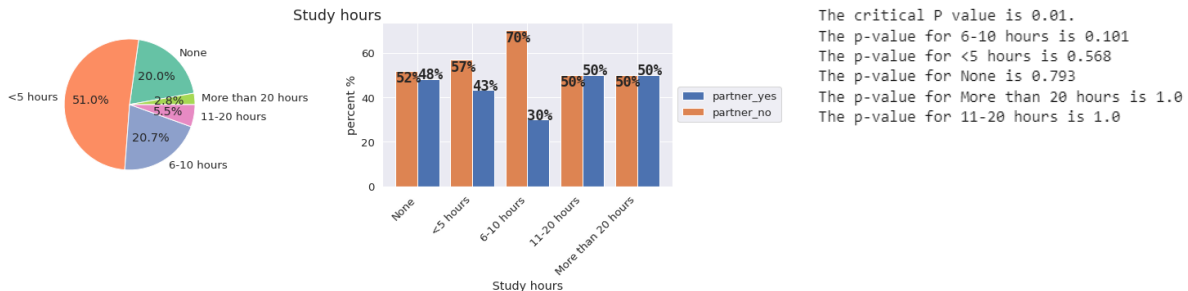
תמונה 7- ממוצע הציונים בסמסטר הקודם לפי מצב זוגי



אפשר לראות שברוב ממוצעי הציונים יש אחוז סטודנטים גבוה יותר שלא נמצאים בזוגיות, ורק בממוצע הגבוה ביותר (מעל 3.49) יש אחוז סטודנטים גבוה יותר שנמצאים בזוגיות (תמונה 7). לפי ערכי ה P-value שחישבתי (תמונה 7), ניתן לראות שבאף אחת מקטגוריות ממוצע הציונים אין הבדל משמעותי מבחינה סטטיסטית. מכאן שגם לממוצע הציונים בסמסטר הקודם אין השפעה משמעותית על המצב הזוגי.

מה הקשר בין שעות לימוד שבועיות למצב הזוגי?

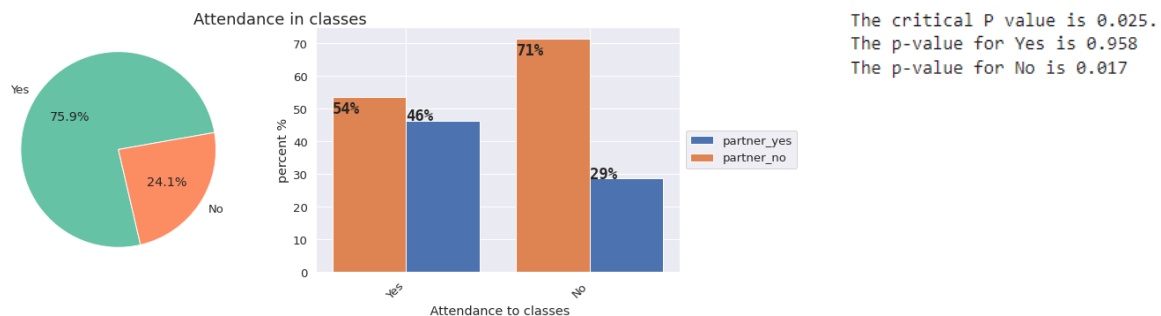
תמונה 8- שעות לימוד שבועיות לפי מצב זוגי



אפשר לראות שברוב טווחי שעות הלימוד המצב הזוגי של הסטודנטים די דומה, ורק בטווח של 6-10 שעות יש אחוז סטודנטים גבוה יותר שלא נמצאים בזוגיות (תמונה 8). לפי ערכי ה P-value שחישבתי (תמונה 8), גם כאן ניתן לראות שבאף אחת מקטגוריות שעות הלימוד השבועיות אין הבדל משמעותי מבחינה סטטיסטית. מכאן שגם לפרמטר זה אין השפעה משמעותית על המצב הזוגי.

מה הקשר בין נוכחות בשיעורים למצב הזוגי?

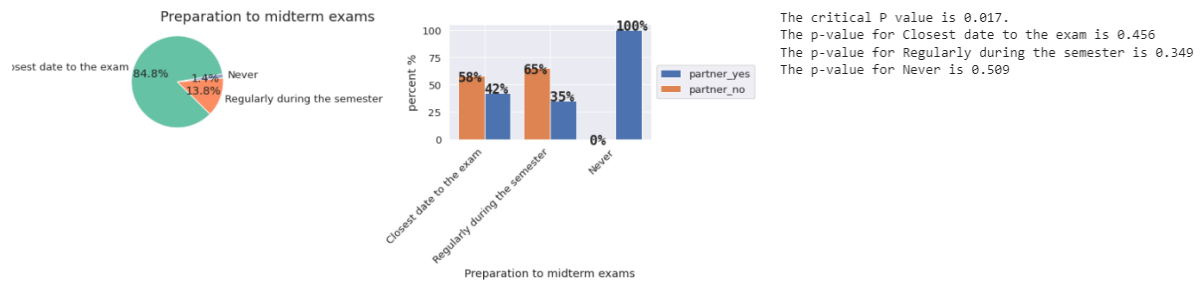
תמונה 9- נוכחות בשיעורים לפי מצב זוגי



אפשר לראות שסטודנטים שנוכחים בשיעורים הם בעלי מצב זוגי דומה, לעומת סטודנטים שלא נוכחים, שרובם ללא זוגיות (71%) (תמונה 9). לפי ערכי ה P-value שחישבתי (תמונה 9), ניתן לראות שיש השפעה סטטיסטית לחוסר נוכחות בשיעורים על המצב הזוגי, ויש קשר בין הפרמטרים.

מה הקשר בין אופן הלמידה למבחן אמצע למצב הזוגי?

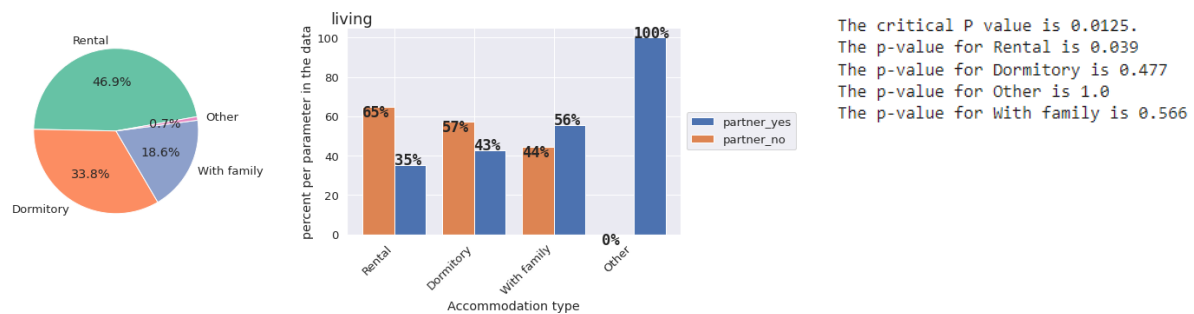
תמונה 10- אופן הלמידה למבחן אמצע לפי מצב זוגי



אפשר לראות שסטודנטים שלא לומדים למבחן אף פעם כולם נמצאים בזוגיות, אך לפי ההתפלגות הכללית של הקטגוריה הזו, הם מהווים אחוז קטן (1.4%) מכלל הסטודנטים. עבור סטודנטים הלומדים קרוב יותר למבחן האמצע, המצב הזוגי שלהם דומה, וסטודנטים שלומדים לאורך הסמסטר (תמונה 10). לפי ערכי ה P-value ניתן לראות שאין משמעות סטטיסטית לפרמטר זה.

מה הקשר בין סידור המגורים למצב הזוגי?

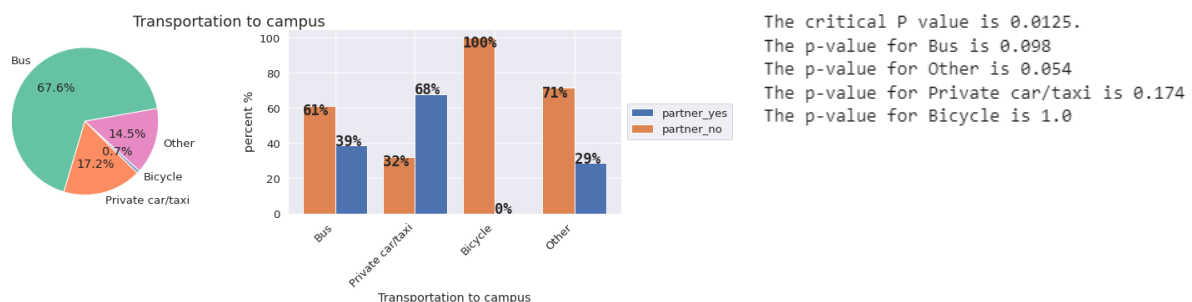
תמונה 11- סידור מגורים לפי מצב זוגי



אפשר לראות שסטודנטים בעלי סידור מגורים "אחר" (Other) כולם נמצאים בזוגיות, אך הם מהווים אחוז מזערי (0.7%) מכלל הסטודנטים. רוב הסטודנטים שגרים בשכירות לא נמצאים בזוגיות, ובשאר הקטגוריות, מצבם הזוגי דומה (תמונה 11). לפי ערכי ה P-value שחישבתי (תמונה 11), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין דרך ההגעה לקמפוס למצב הזוגי?

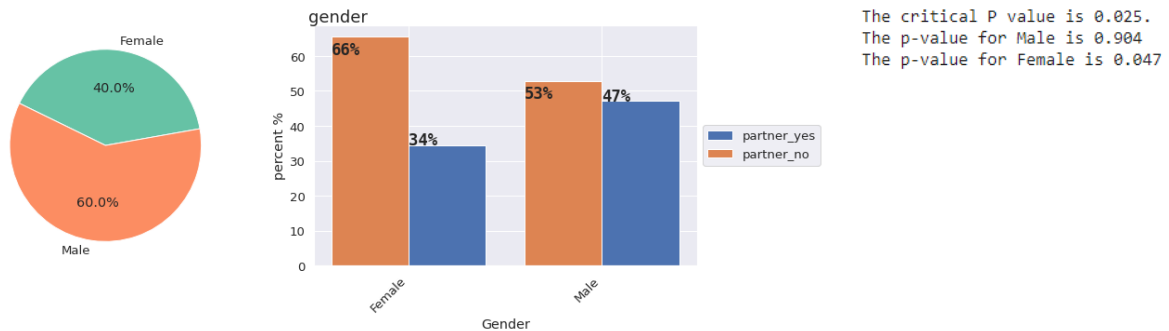
תמונה 12- דרך ההגעה לקמפוס לפי מצב זוגי



אפשר לראות שסטודנטים המגיעים לקמפוס באופניים כולם לא נמצאים בזוגיות, אך הם מהווים אחוז מזערי (0.7%) מכלל הסטודנטים. רוב הסטודנטים שמגיעים ברכב פרטי או מונית נמצאים בזוגיות, בעוד שבשאר הקטגוריות רוב הסטודנטים לא נמצאים בזוגיות. (תמונה 12). לפי ערכי ה P-value שחישבתי (תמונה 12), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין מגדר למצב הזוגי?

תמונה 13- מגדר לפי מצב זוגי



ניתן לראות שההתפלגות של זכר ונקבה היא התפלגות אופיינית לפקולטות להנדסה ומדעים (רוב גברי). בנוסף, אפשר לראות שרוב הנשים לא נמצאות בזוגיות, בעוד שגברים במצב זוגי דומה. (תמונה 13). לפי ערכי ה P-value שחישבתי (תמונה 13), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

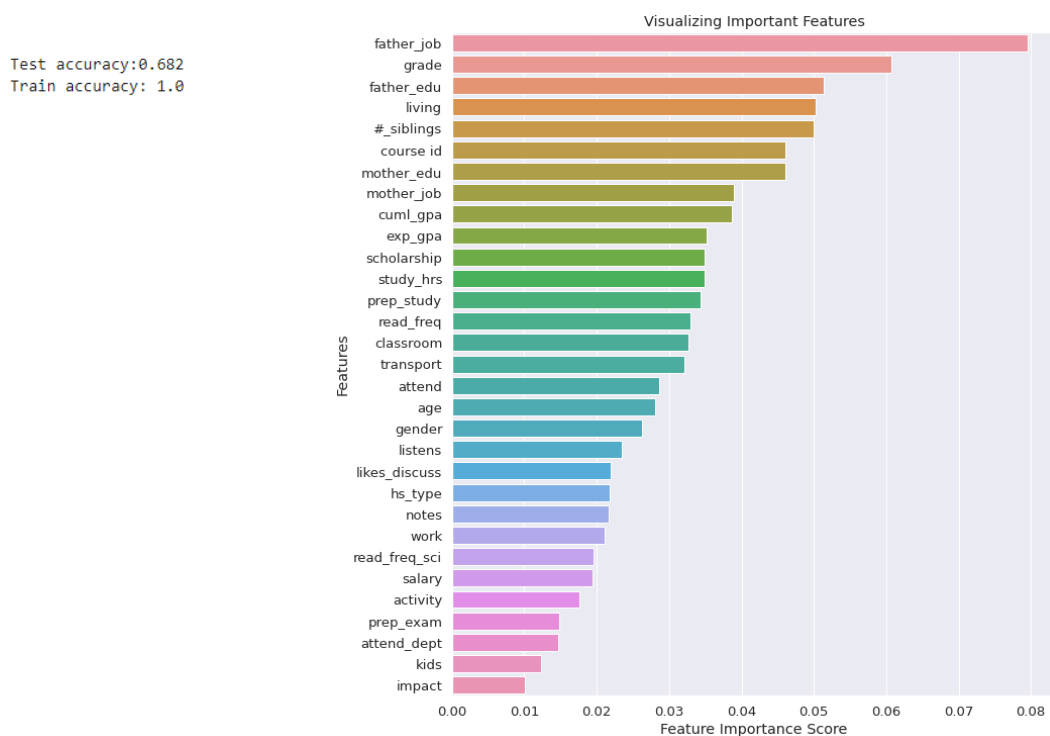
מסקנות

ברוב הקטגוריות שחשבתי שתהיה השפעה על המצב הזוגי, לא נראתה השפעה משמעותית מבחינה סטטיסטית מלבד נוכחות בשיעורים, בה נראה קשר בין חוסר נוכחות בשיעורים לבין חוסר זוגיות.

חיזוי המצב הזוגי לפי כלל הפרמטרים

לאחר שהסתכלתי על התפלגות המצב הזוגי בכל אחת מהקטגוריות וחישבתי את ערכי ה P-value, ראיתי שלרוב הפרמטרים אין השפעה על המצב הזוגי. תכננתי להשתמש ב Random Forest כדי לחזות את המצב הזוגי לפי הפרמטרים שבדקתי, אבל בגלל ההשפעה הנמוכה שלהם על המצב הזוגי, אני צופה שהחיזוי לא ייתן אחוזי הצלחה גבוהים ויתרחש overfitting לדאטא. בכל זאת, החלטתי לנסות לחזות את המצב הזוגי לפי כל הפרמטרים של הדאטא (ולא רק הפרמטרים שחשבתי שיהיו רלוונטיים) ולנסות ולראות את החשיבות שלהם בחיזוי.

תמונה 14- חשיבות הפרמטרים בחיזוי Random Forest



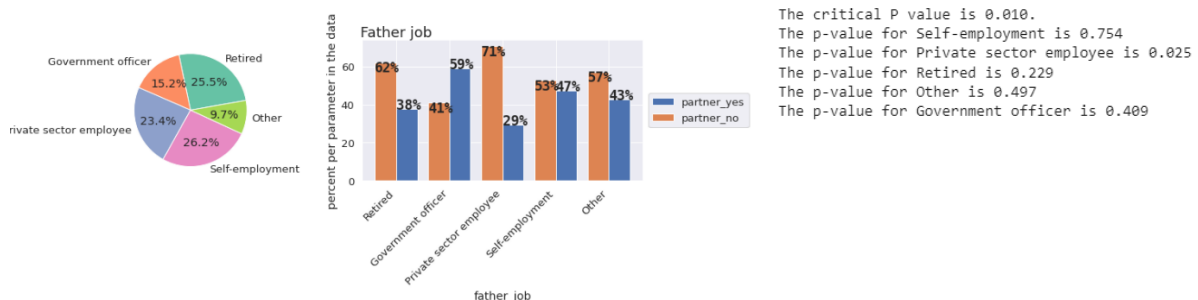
ניתן לראות שעבור סט האימון התקבל דיוק של 100% (כצפוי בגלל overfitting), ועבור סט ה test התקבל דיוק של 68%. למרות אחוז זה, ניתן לראות שלכל הפרמטרים אין אחוז חשיבות גבוה (לכל היותר 8%) ולכן אין לפרמטרים בדאטא סט יכולת חיזוי טובה למצב הזוגי. אם נסתכל על הפרמטרים עם אחוז החשיבות הגבוה ביותר מבין אלו (מעל 3.5%), רובם הם לא הפרמטרים שציפיתי שישפיעו על המצב הזוגי. (תמונה 14)

ניתוח הפרמטרים בעלי חשיבות גבוהה יחסית לפי ה Random Forest

רציתי לנתח את הפרמטרים החדשים באותה צורה כמו הפרמטרים הקודמים. לקחתי את הפרמטרים בעלי חשיבות גבוהה מ 3.5% שהם: עבודת האב (father_job), השכלת האב (father_edu), מספר אחים (#_siblings), מספר קורס (course_id), השכלת האם (mother_edu), עבודת האם (mother_job). את שאר הפרמטרים בעלי החשיבות הגבוהה (grade, living, cuml_gpa), ניתחתי כבר קודם.

מה הקשר בין עבודת האב למצב הזוגי?

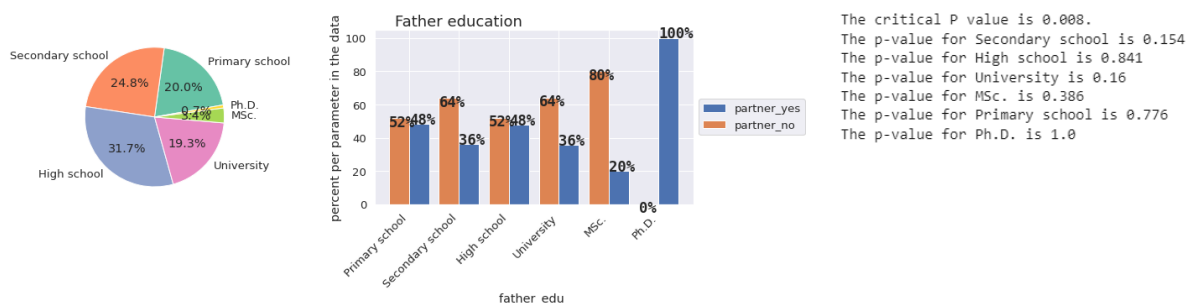
תמונה 15- עבודת האב לפי מצב זוגי



אפשר לראות שברוב עבודות האב יש אחוז גבוה יותר של סטודנטים בזוגיות, מלבד עובד מדינה, שבו יש רוב של סטודנטים שנמצאים בזוגיות (59%) (תמונה 15). לפי ערכי ה P-value שחישבתי (תמונה 15), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין השכלת האב למצב הזוגי?

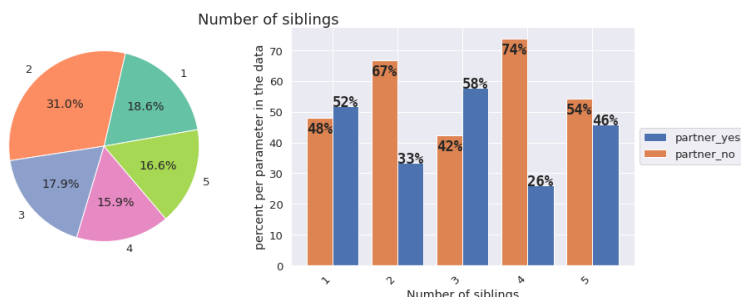
תמונה 16- השכלת האב לפי מצב זוגי



אפשר לראות שברוב השכלות האב יש אחוז גבוה יותר של סטודנטים לא בזוגיות, מלבד Ph.D. ש 100% מהסטודנטים שלאביהם יש השכלה זו נמצאים בזוגיות, אך הם אחוז מזערי (0.7%) מכלל הסטודנטים. (תמונה 16). לפי ערכי ה P-value שחישבתי (תמונה 16), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין מספר האחים למצב הזוגי?

תמונה 17- מספר האחים לפי מצב זוגי

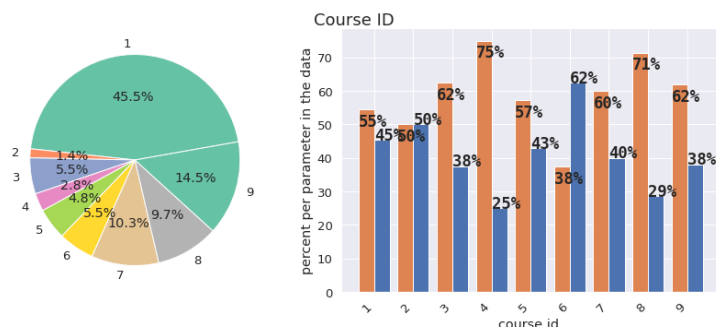


The critical P value is 0.010.
 The p-value for 3 is 0.443
 The p-value for 2 is 0.048
 The p-value for 5 is 0.642
 The p-value for 1 is 0.775
 The p-value for 4 is 0.026

אפשר לראות שברוב כמויות האחים יש אחוז גבוה יותר של סטודנטים לא בזוגיות, מלבד 3 אחים ואח, שם יש אחוז גבוה יותר של סטודנטים בזוגיות. (תמונה 17). לפי ערכי ה P-value שחישבתי (תמונה 17), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין מספר הקורס למצב הזוגי?

תמונה 18- מספר קורס לפי מצב זוגי

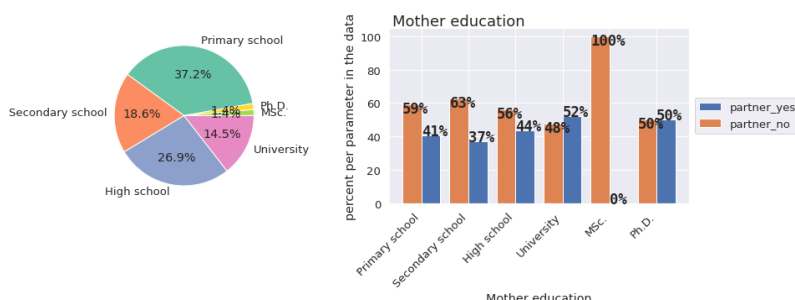


The critical P value is 0.006.
 The p-value for 4 is 0.634
 The p-value for 1 is 0.698
 The p-value for 8 is 0.084
 The p-value for 3 is 0.757
 The p-value for 5 is 0.468
 The p-value for 6 is 0.746
 The p-value for 7 is 0.645
 The p-value for 9 is 0.28
 The p-value for 2 is 1.0

אפשר לראות שברוב הקורסים יש אחוז גבוה יותר של סטודנטים שלא נמצאים בזוגיות, מלבד קורס 6, בו יש רוב של סטודנטים הנמצאים בזוגיות (62%). ניתן לראות גם ש 45.5% מהסטודנטים נמצאים בקורס 1, ובו המצב הזוגי די דומה (תמונה 18). לפי ערכי ה P-value שחישבתי (תמונה 18), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין השכלת האם למצב הזוגי?

תמונה 19- השכלת האם לפי מצב זוגי

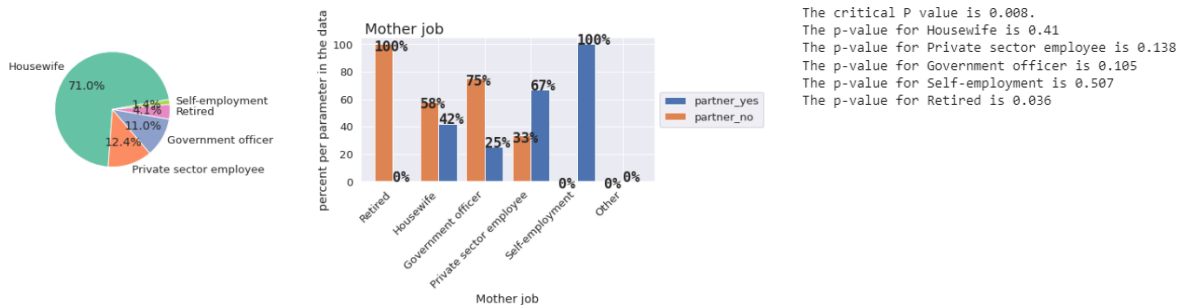


The critical P value is 0.008.
 The p-value for Primary school is 0.352
 The p-value for Secondary school is 0.192
 The p-value for High school is 0.545
 The p-value for University is 0.712
 The p-value for MSc. is 0.499
 The p-value for Ph.D. is 1.0

אפשר לראות שברוב השכלות האם יש אחוז גבוה יותר של סטודנטים לא בזוגיות, מלבד Ph.D והשכלה אוניברסיטאית שבהם יש אחוז דומה במצב הזוגי. ב MSc ו Ph.D מהווים אחוז מזערי מכלל הסטודנטים (1.4%) (תמונה 19). לפי ערכי ה P-value שחישבתי (תמונה 19), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מה הקשר בין עבודת האם למצב הזוגי?

תמונה 20- עבודת האם לפי מצב זוגי



אפשר לראות שברוב עבודות האם יש אחוז גבוה יותר של סטודנטים לא בזוגיות, מלבד עובדות הסקטור הפרטי ועצמאיות, כאשר הסטודנטים שהאימהות שלהם עצמאיות כולם נמצאים בזוגיות, אך מהווים אחוז קטן מכלל הסטודנטים (1.4%). מצב דומה קורה עבור סטודנטים שהאימהות שלהם פרשו, עם 4.1% מכלל הסטודנטים. אציין שאין סטודנטים שהאימהות שלהן עובדות בעבודה אחרת שלא צוינה. (תמונה 20). לפי ערכי ה P-value שחישבתי (תמונה 20), ניתן לראות שאין השפעה סטטיסטית של פרמטר זה על המצב הזוגי.

מסקנות

ניתן לראות שגם בפרמטרים בעלי החשיבות הגבוהה ביותר לפי ה Random Forest, אין חשיבות סטטיסטית להבדל בין הסטודנטים שנמצאים בזוגיות לאלה שלא. דבר זה מתיישב טוב עם התוצאה של ה Random Forest לחשיבות הפרמטרים (תמונה 14), בה ניתן לראות שלאף פרמטר לא הייתה חשיבות משמעותית. לכן אפשר להניח שהחיזוי שה Random Forest ביצע היה ספציפי לדאטא סט הקטן, ולא ניתן להסיק ממנו מסקנות משמעותיות לגבי הקשר בין הפרמטרים השונים למצב הזוגי של הסטודנטים. בהתחלה רציתי לבדוק את כוח החיזוי של כמה משתנים בודדים על המצב הזוגי של הסטודנטים, אבל בגלל החשיבות הנמוכה של הפרמטרים וחוסר החשיבות הסטטיסטית שלהם, אני לא רואה טעם לעשות זאת, כי כל חיזוי ייתן תוצאות מאוד ספציפיות לפי הדאטא סט המוגבל עצמו, תוצאות שלא תהיה להן משמעות רבה.

סיכום

לרוב הפרמטרים שבדקתי, לאף אחת מהקטגוריות שלהם לא הייתה משמעות סטטיסטית על ההפרש בין מספר הסטודנטים שנמצאים בזוגיות לבין מספר הסטודנטים שלא נמצאים בזוגיות בערך מוחלט. הפרמטר היחיד שהייתה לו משמעות היה נוכחות בשיעורים, כאשר חוסר נוכחות בשיעורים היה בקורלציה גבוהה עם סטודנטים שלא נמצאים בזוגיות. ניסיתי גם להפעיל Random Forest על כל הדאטא סט, וראיתי שהחשיבות של כל פרמטר היא מאוד נמוכה (8% ומטה). לאחר מכן רציתי להסתכל על חשיבות סטטיסטית של הפרמטרים עם אחוז חשיבות הכי גבוה מבין הפרמטרים ב Random Forest, אך לאף אחד מהם לא הייתה חשיבות סטטיסטית ביחס למצב הזוגי של הסטודנטים. לכן ניתן להניח שמהדאטא סט הזה לא ניתן לחזות מצב זוגי. כנראה כי הוא קטן מדי וספציפי מדי.

רעיונות להמשך

כדי לבצע חיזוי טוב של השפעת פרמטרים שונים של סטודנטים על המצב הזוגי, צריך לבצע מחקר גדול יותר עם מספר רב יותר של סטודנטים מפקולטות ואף אוניברסיטאות שונות. אם היה ברשותי מספר גדול של דוגמאות, הייתי יכולה לבצע חיזוי באמצעות רשת נוירונים שיכולה ללמוד לבד מה הפרמטרים החשובים, ואולי גם לבצע חיזוי טוב יותר למצב הזוגי עבור סטודנטים.