



Data Mining and Visualization (83676)

Final Project

Part 1

מגישים: טל רוזנצוויג (213391691), אופיר יחזקאל (214328940).

1.1 הצגת המידע – סוגי השדות והערכים:

תחילה, הצגנו את כל שדות המידע, יחד עם הטיפוס של כל שדה ומספר הרשומות (שאינן Null) בכל אחד.

1.2 הצגת המידע הסטטיסטי:

בשלב זה הצגנו לכל גרפים להתפלגויות של כל המשתנים הנומריים, גרפי עוגה למשתנים המתאימים (סטטוס והשכלה).

לאחר מכן הצגנו את ה-skewness של כל אחד מהמשתנים.

לבסוף, הצגנו ערכים סטטיסטיים מרכזיים לכל שדה בטבלה, ביניהם: ממוצע, סטיית תקן, מינימום, מקסימום וכו'.

1.3 קורלציה:

בחלק זה יצרנו מטריצת קורלציה בין כל תכונות המשתנים לצורך מציאת קשרים בין תכונות שונות. תחילה, נשים לב שכל האלכסון של המטריצה מכיל את הערך 1, זה משום שלכל תכונה יש קשר של 100% עם עצמה כמובן, ולכן תכונה זו אינה מקדמת אותנו. שנית, כפי שלמדנו מטריצת הקורלציה סימטרית (לפי על האלכסון), לכן נוכל להתמקד רק בחלק העליון שלה. בנוסף, נזכיר שכל שהערך המוחלט של תא במטריצת הקורלציה יותר גדול, כך הקשר בין 2 התכונות יותר גדול (כאשר הערך חיובי – הקשר הוא קשר ישיר וכאשר הערך שלילי – הקשר הוא קשר הפוך).

נחפש קשרים בולטים:

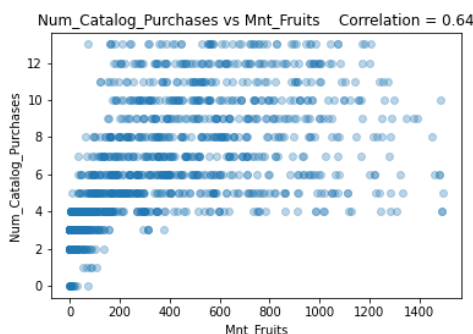
1. ניתן לראות שיש קשר הפוך גדול בין ה- $Year_Birth$ ל- Num_of_Teen (של -0.38), כלומר בין שנת הלידה של הלקוח למספר הנערים שיש בביתו. זה מאוד הגיוני, שכן ככל שהלקוח יותר מבוגר, יש יותר סיכוי שיהיו לו ילדים בגיל מבוגר יותר – נערים במקרה שלנו.
2. ניתן לראות שיש קשר ישיר משמעותי בין ה- $Income$ ל- Mnt_Fruits (של 0.55), בנוסף גם ל- Mnt_Meat (של 0.4), גם ל- Mnt_Sweet (של 0.57), ל- Mnt_Wines (של 0.41), ל- $Mnt_Gold_Products$ (של 0.41), ל- $Num_Deals_Purchases$ (של 0.58), ול- $Num_Catalog_Purchases$ (של 0.5). כלומר, יש קשר גדול בין המשכורת שמרוויח הלקוח, לכמות הפירות, הבשר, הממתקים, היינות, המוצרים העשויים מזהב והמוצרים שהיו בהנחה שהלקוח רכש. בנוסף, יש קשר גדול גם בין כמות המוצרים שהלקוח רכש באמצעות הקטלוג למשכורת שלו.
3. ניתן לראות שיש קשר הפוך גדול בין ה- $Income$ ל- Num_of_Kids (של -0.4), כלומר שככל שללקוח יש פחות ילדים, כך המשכורת שלו יותר גבוהה. הגיוני, שכן כניראה שבמצב כזה יש לו יותר זמן פנוי להשקיע בעבודה. בנוסף, יש קשר הפוך גדול בין ה- $Income$ ל- Num_Web_Visits , כלומר שככל שהלקוח מרוויח יותר, כך הוא נכנס פחות פעמים לאתר האינטרנט של החנות.
4. ניתן לראות שיש קשר ישיר גדול בין Num_of_Kids ל- Num_Web_Visits (של 0.42), כלומר ככל שללקוח יש יותר ילדים, הוא גם ביקר יותר באתר האינטרנט של החנות. הגיוני שכן, גם ככל שללקוח יש יותר ילדים, יש לו פחות זמן לנסוע ולערוך את הקניות באינטרנט וגם יכול להיות שילדיו של הלקוח עוזרים לו להשתמש האתר האינטרנט, מה שמסביר את נוכחותו הגבוהה שם.
5. ניתן לראות שיש קשר הפוך גדול בין Num_of_Kids ל- $Num_Deals_Purchases$ (של -0.5), כלומר שככל שללקוח יש יותר ילדים, כך הוא קונה פחות מוצרים בהנחה. ניתן לשער שזה כך משום שרוב המוצרים שקונים לילדים הם "דחופים" וללקוח יש פחות זמן לחכות להנחה כלשהי. בנוסף ניתן לראות שיש ל- Num_of_Kids קשר הפוך גדול גם עם $Num_Catalog_Purchases$ (של -0.5), כלומר שככל שללקוח יש יותר ילדים, כך הוא קונה פחות מוצרים באמצעות הקטלוג. ניתן להניח שזה כך, משום שככל שיש לו יותר ילדים, כל הוא יותר "עמוס" ויש לו פחות זמן להיסתכל בקטלוג. ניתן לראות גם של- Num_of_Kids יש קשר הפוך גדול עם Mnt_Fruits (של -0.49), כלומר שככל שללקוח יש יותר ילדים, הוא קונה פחות פירות, הגיוני שכן כניראה שילדים פחות אוהבים לאכול פירות.
6. ניתן לראות שיש קשר ישיר גדול בין Num_of_Teen ל- $Num_Web_purchases$ (של 0.39). כלומר, שככל שללקוח יש יותר נערים, כך הוא רוכש יותר מוצרים דרך אתר

- האינטרנט. הגיוני שכן, כיום הנערים רגילים לקנות באינטרנט וכנראה שגם יכולים לעזור להוריהם בכך.
7. ניתן לראות שיש קשר ישר משמעותי בין Mnt_Fruits ל-Num_Catalog_Purchases (של 0.64), ול-Num_Deals_purchases (של 0.62). כלומר, ששיש הרבה פירות שקונים באמצעות הקטלוג, כנראה שהחלק של הפירות בו מסודר וערוך בצורה טובה ו"מגרה" ובנוסף יש הרבה פירות שקונים בהנחה, כנראה שיש הרבה הנחות ומבצעים על הפירות.
8. ניתן לראות שיש קשר ישר משמעותי בין Mnt_Meat ל-Mnt_Wines (של 0.57), כלומר שככל שאנשים קונים יותר בשר הם גם קונים יותר יין, יכול להיות שסוג מזון זה משתלב טוב.
9. ניתן לראות שיש קשר ישר משמעותי בין Mnt_Sweet ל-Num_Deals_Purchases (של 0.75), כלומר שככל שקונים יותר ממתקים קונים גם יותר מוצרים בהנחה. כנראה שיש הנחות ומבצעים רבים על הממתקים.
10. ניתן לראות שיש קשר הפוך גדול בין Num_Deals_Purchases ל-Num_Web_Visits (של -0.51). כלומר, ככל שנכנסים יותר לאתר האינטרנט של החנות, כך פחות מוצרים בהנחה. סיבה אפשרית לכך יכולה להיות שבאתר האינטרנט יש הרבה פחות הנחות ומבצעים.
11. ניתן לראות שיש קשר בין ערך המטרה שלנו, Response, לבין Response_Campaign_3 (של 0.33). זאת לעומת שאר הקמפיינים בהם ההשפעה על ערך המטרה הייתה מועטה יותר. הסבר אפשרי לכך הוא שהקמפיין השלישי היה יותר אפקטיבי לעומת הקמפיינים האחרים. מעבר לכך ניתן להבחין בקורלציה גבוהה בין Response_Campaign_3 לבין Mnt_Fruits (0.49), Mnt_sweet (0.38) ועל Num_Deals_Purchases (0.33). הסבר אפשרי לכך הוא שהקמפיין השלישי היה מכוון לפירות ולממתקים, וכן שזה הביא לקנייה מסיבית יותר בקופונים.

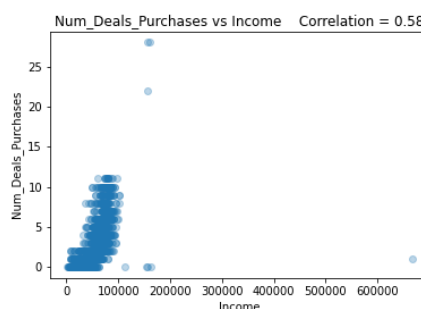
1.4 תובנות, יחסים ואפקטים של תכונות אחת על השנייה:

נתייחס לשלושה גרפים המתארים יחסים בין תכונות שונות:

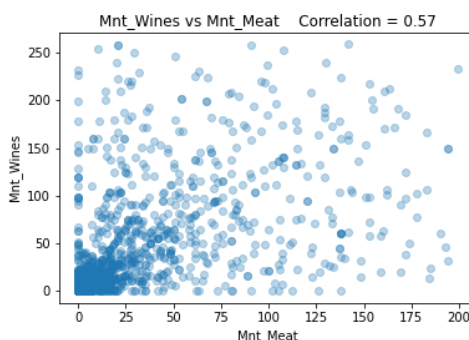
1. ניתן לראות בגרף, שבאופן יחסי, קנייה מסיבית של פירות (דו שנתית) באה ביחד עם כמות גבוהה של קניות חודשיות דרך הקטלוג.



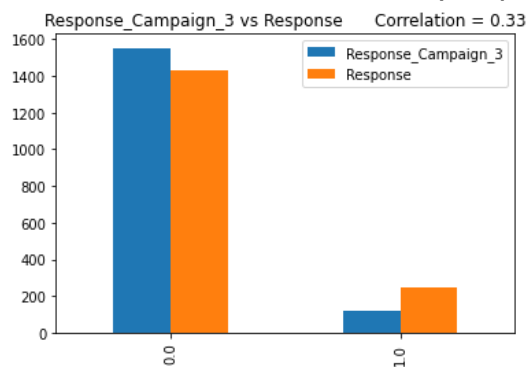
2. בגרף זה ניתן לראות שהיחס בין ההכנסה לבין מספר הקניות עם דילים, מתבטא בכך שעד גבול מסוים (בערך 100000), ככל שמרוויחים יותר, כך קונים יותר עם דילים, אך ברגע שחוצים גבול זה, הרוב לא משתמש בדילים ברכישות שלו. הסבר אפשרי לתופעה הזו, הוא שכאשר ההכנסה היא נמוכה, לא מבצעים הרבה רכישות ולכן גם הרכישות עם הדילים תהיינה נמוכות יותר, אך ככל שעולים בהכנסה, וככל שההכנסה מאפיינת את מעמד הביניים, כך השימוש בדילים עולה. מעבר לכך, כמובן שכאשר ההכנסה היא מאוד גבוהה, השימוש בדילים יפחת.



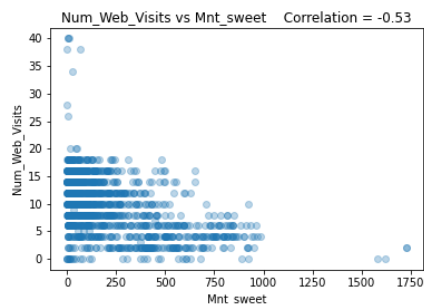
3. מגרף זה ניתן להסיק, שרוב האנשים שלא קונים בשר, גם לא קונים יין. כמו כן, נראה שיש איזשהו יחס ישר בין קניית הבשר קניית היין, דבר הגיוני, שכן בשר הולך טוב עם יין, אך עם זאת יש לא מעט אנשים שקונים יותר יין ולא קונים בשר (דוגמה אפשרית- טבעונים וצמחונים), וכן יש לא מעט אנשים שקונים בשר, אך לא קונים יין (יכול להיות שאנשים אלה לא אוהבים את תחושת העפיצות מהיין, או פשוט אנשים שלא אוהבים את טעמו החריף של האלכוהול המצוי ביין).



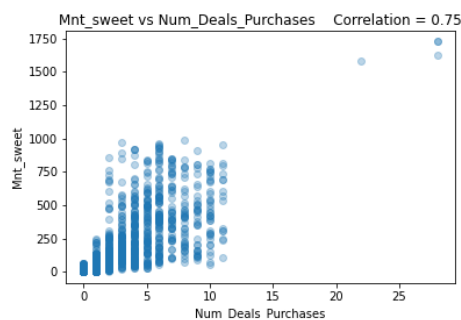
4. בגרף זה רואים שערך המטרה שלנו מתנהג בצורה כמעט זהה, כמו הערך Response_Campaign_3. ייתכן שזה נובע מהאינטנסיביות ו/או מהאפקטיביות של הקמפיין השלישי.



5. בגרף זה ניתן לראות שככל שקונים יותר באינטרנט, כך קונים פחות ממתקים. הסבר אפשרי לכך הוא שהקנייה באינטרנט גורמת לאדם להיות יותר שקול ואחראי, וכן ייתכן כי תחושת האשמה בעת קניית הממתקים גבוהה יותר כאשר קונים באינטרנט.

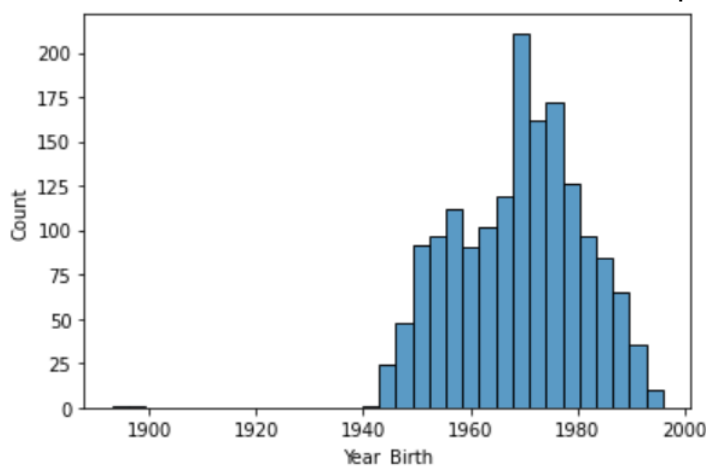


6. אפשר לראות בגרף זה, שקנייה מרובה של ממתקים מאופיינת בקנייה עם דילים. ייתכן שהסיבה לכך היא שממתקים נמכרים לרוב בדילים ואריזות גדולות.



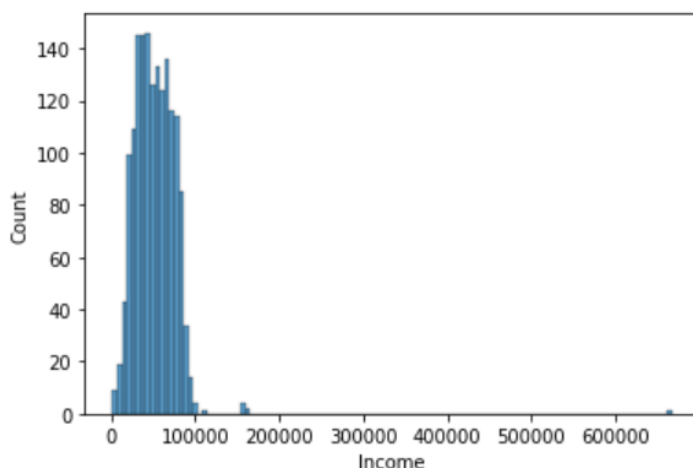
1.5-1.7 ניקוי מידע, השלמת ערכים חסרים, הפחתת מידע:

- הורדנו מה-df המקורי שלנו את השדות: ID, Revenue, Cost_Contact, Complain, שכן מדובר בקבועים שאין להם השפעה על המשתנים החסרים שאותם נרצה למלא ובשלב זה אין לנו מה ללמוד מהם. הכנסנו את המידע המעודכן לשדה data.
- הורדנו את כל הרשומות של אלו שהיו להם מעל ל-6 ערכי Null, בהנחה שחסר בהן מידע רב מידי וקשה ללמוד מהן.
- ראינו בגרפי הוירטואליזציה של סעיף 1.2 עבור השדה של שנת לידה – Year_Birth, שינם ערכים הקטנים מ-1905:



הנחנו שהנתונים עדכניים – שהשנה 2022, ולכן הורדנו את כל הרשומות, שבהן שנת הלידה היא מתחת ל-1905.

- ראינו בגרפי הוירטואליזציה של סעיף 1.2 עבור השדה שמתאר את ההכנסה – Income, שינן רשומות עם ערכים שליליים:



הנחנו שמדובר בטעות הקלדה ושיסימן המינוס נלחץ בטעות, ולכן בחרנו לתקן כל רשומה כזו לערכה המוחלט.

- שמנו לב שישנן רשומות שבהן מספר הרכישות מאתר האינטרנט (Num_Web_Purchases) גדול מ-0, בעוד שמספר הכניסות לאתר

(Num_Web_Visits) הוא 0. כמובן שזה לא הגיוני, אך מכיוון שישנו מספר רב של לקוחות כאלה, הנחנו שמדובר בבעיה באתר ובערכי ה-Num_Web_Visits והחלטנו לעדכן את הערכים הללו כך:
ראשית, נסמן:

$$mean = mean(Num_Web_Purchases - Num_Web_Visits)$$

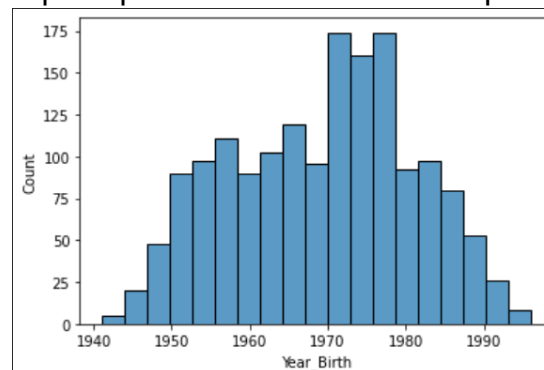
כעת:

$$Num_Web_Visits = \begin{cases} 0 & 1 \leq Num_Web_Purchases \leq mean \\ 1 & Num_Web_Purchases \geq mean + 1 \\ Num_Web_Purchases - mean & \end{cases}$$

כלומר, כאשר כמות הרכישות מהאתר היא 0, עדכנו גם את כמות הכניסות לאתר להיות 0. כאשר כמות הרכישות מהאתר היא בין 1 להפרש הממוצע בין כמות הרכישות מהאתר לכמות הכניסות אליו, עדכנו את כמות הכניסות להיות 1 וכאשר כמות הרכישות מהאתר היא גדולה שווה להפרש הממוצע בין כמות הרכישות מהאתר לכמות הכניסות אליו, עדכנו את כמות הכניסות להיות מספר הרכישות מהאתר פחות ההפרש הממוצע בין כמות הרכישות מהאתר לכמות הכניסות אליו.
לאחר מכן, חישבנו ומצאנו שההפרש הממוצע בין כמות הרכישות מהאתר לכמות הכניסות אליו הוא 3 ולמעשה מקבלים שהעדכון מתבצע כך:

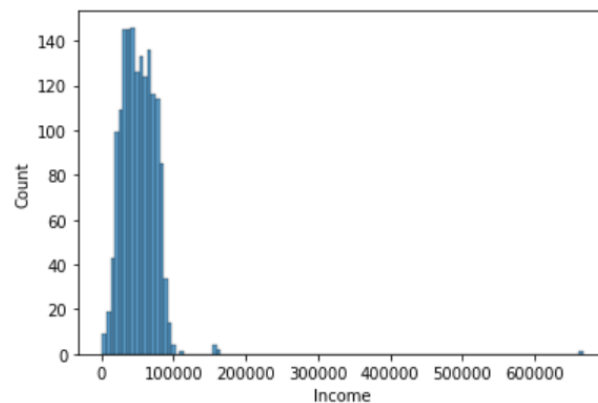
$$Num_Web_Visits = \begin{cases} 0 & 1 \leq Num_Web_Purchases \leq 3 \\ 1 & Num_Web_Purchases \geq 4 \\ Num_Web_Purchases - 3 & \end{cases}$$

- הסתכלנו בגרפי הוירטואליזציה של סעיף 1.2 עבור השדה של שנת לידה – *Year_Birth*, וראינו שההתפלגות דומה מאוד להתפלגות נורמלית, יצרנו גרף חדש לאחר הורדת הערכים שקטנים מ-1905 כפי שהסברנו קודם וקיבלנו:



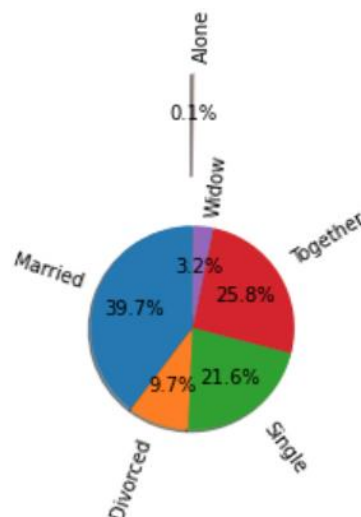
לאחר מכן, חישבנו את הממוצע וסטיית התקן, באמצעותם השלמנו את הערכים החסרים על פי התפלגות נורמלית.

- באופן דומה, הסתכלנו בגרפי הוירטואליזציה של סעיף 1.2 עבור השדה שמתאר את ההכנסה – *Income* וראינו שגם כאן ההתפלגות דומה להתפלגות נורמלית:

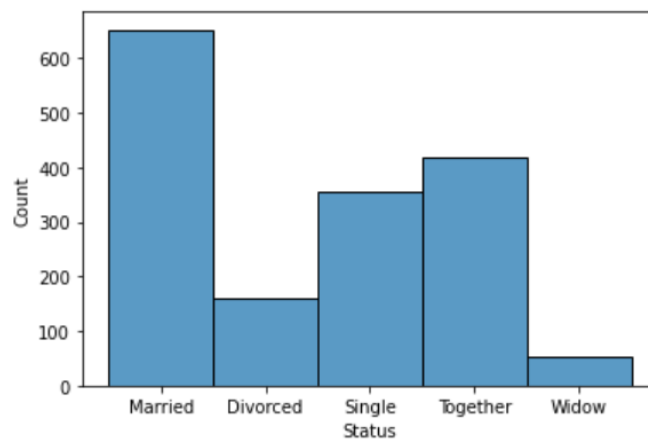


לכן, חישבנו את הממוצע וסטיית התקן ובאמצעותם השלמנו את הערכים החסרים על פי התפלגות נורמלית.

- המשכנו בהסתכלות בגרפי הוירטואליזציה של סעיף 1.2, עכשיו עבור השדה של הסטטוס – *Status*:



ניתן לראות בגרף שישנו מס' מועט של *Alone*, סיבה אפשרית לכך היא שבסטטוס הייתה הקלדה חופשית ולכן היו כאלה שכתבו דווקא את זה. מכאן, בחרנו לאחד את *Alone* יחד עם *Single*:



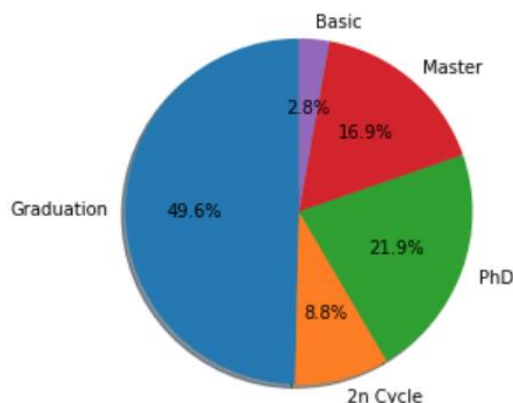
- כעת, רצינו להשלים את השדה המתאר את רמת ההשכלה – *Education*. בחרנו לעשות זאת לפי הקשר שבין רמת ההשכלה של אדם לשכר שהוא מרוויח. כדי לבדוק

שהקשר אכן קיים ועובד על פי ההיגיון (השכלה גבוהה יותר \leftarrow שכר גבוה יותר), בדקנו מהו חציון ההכנסה של כל רמת השכלה שהיא וקיבלנו:

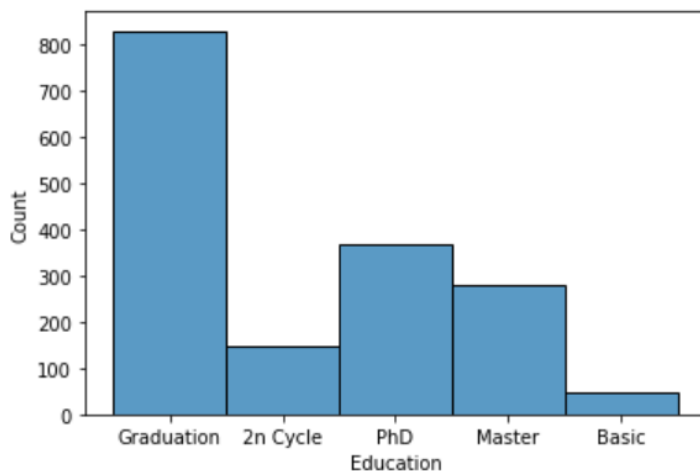
Mean of Graduation is 52768.73392181589
Mean of 2n Cycle is 48143.27272727273
Mean of PhD is 55919.15730337079
Mean of Master is 51978.173431734314
Mean of Basic is 19959.565217391304

כלומר, ניתן לראות שאכן רמת השכלה *Basic* מרוויחים הכי פחות, אחרים רמת ההשכלה של *2n Cycle*, אחר כך רמת ה-*Master*, מעליהם *Graduation* ובהכנסה הגבוהה ביותר *PhD*. סה"כ התוצאות הגיוניות ותואמות את ההיגיון, מלבד זה שרמת השכלה של תואר ראשון מרוויחים פחות מאלו שבעלי רמת השכלה של תואר שני, אך ההפרש הוא מאוד מזערי ולא מספיק משמעותי.

כעת, על פי החציון בחרנו להשלים את אלה שמרוויחים מעל \$55,000 להיות כאלה בעלי רמת השכלה של *PhD*, את אלה המרוויחים מתחת ל-\$25,000 להיות כאלה בעלי רמת השכלה של *Basic*. בשאר הערכים נעזרנו בגרף העוגה המתאר את רמת *Education* שעשינו בחלק 1.2:



ניתן לראות שמרבית האנשים בעלי רמת השכלה של תואר ראשון בצורה משמעותית יותר גדולה מאלו של תואר שני, ולכן בחרנו את אלה שמרוויחים בין \$25,000 ל-\$55,000 להיות בעלי רמת השכלה של *Graduation* (נזכיר שמדובר במשכורות שנתיות). לאחר השלמת הערכים כפי שהסברנו קיבלנו את הגרף הבא:

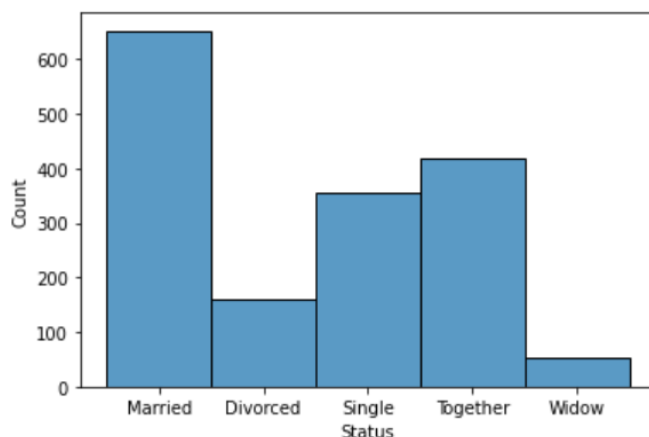


וניתן לראות שקיבלנו גרף מאוד דומה לזה של שלב 1.2 ושהשלמת הערכים לא עיוותה את תמונת המצב.

- כעת, אנו רוצים להשלים את הרשומות החסרות בשדה של הסטטוס. לשם כך בדקנו מה הגילאים של אלו שחסרה להן רשומה זו וחסרות להם גם הרשומות שמתארות את כמות הילדים ובני הנוער (בצד ימין אינדקס הרשומה ובשמאל הגיל):

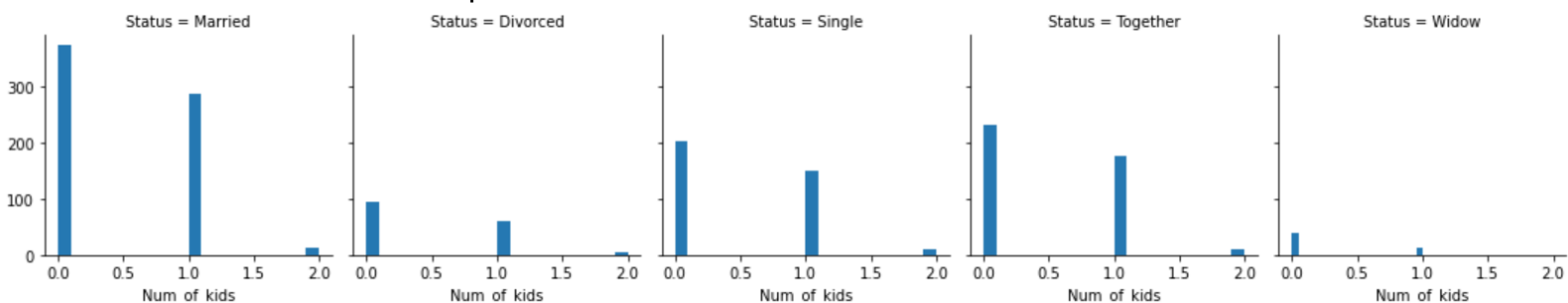
502	54.0
509	48.0
634	46.0
936	55.0
1151	52.0
1464	65.0
1502	65.0
1639	60.0

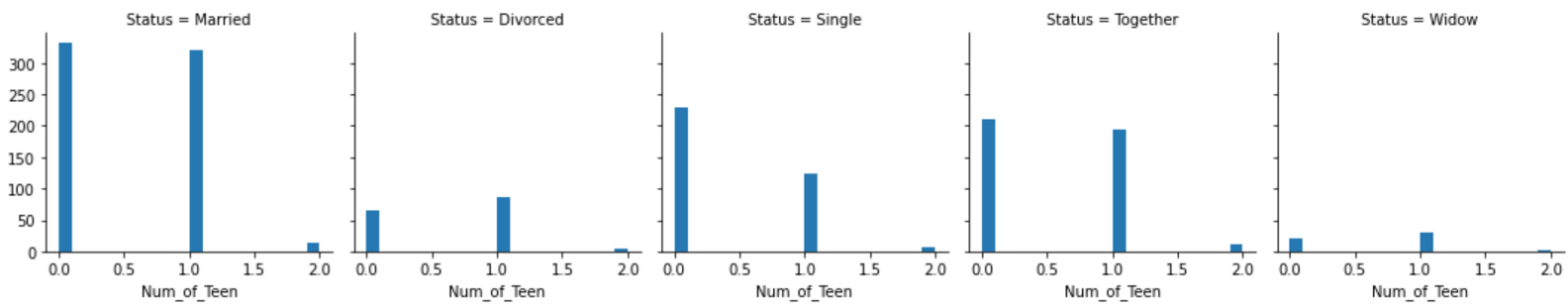
ניזכרנו בהתפלגות הסטטוס:



ובחרנו למלא את אלו שחסר לנו את הסטטוס, כמות הילדים וכמות בני הנוער שלהם בערך Married, שכן ערך זה הוא הנפוץ ביותר, שכן גם ראינו שאין לכך סתירה עם הגילאים שלהם. לאחר מכן, בחרנו את אלה שאין להם ילדים כלל ולא בני נוער להיות בעלי ערך Single, שכן זו אפשרות הגיונית ביותר לאדם ללא ילדים. את כל השאר – אלה שיש להם כמות כלשהי של ילדים או בני נוער בחרנו להיות גם Married, שכן זו האפשרות הנפוצה ביותר על פי ההתפלגות.

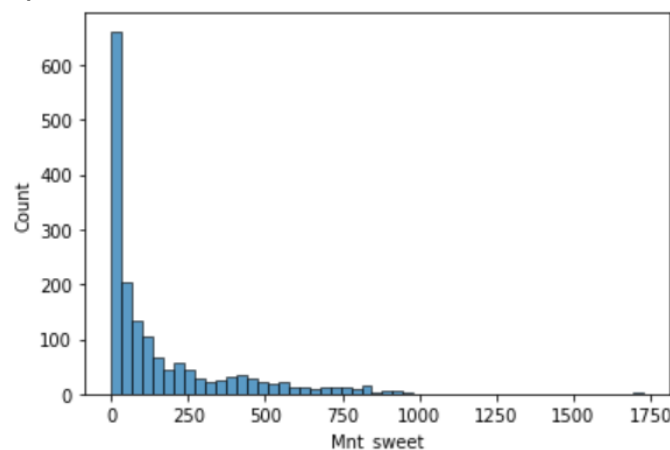
- בשלב זה רצינו להשלים את הרשומות החסרות להן ערכים בשדות כמות הילדים וכמות בני הנוער. לשם כך, בדקנו תחילה ווידאנו שאין אנשים שגם יש להם בני נוער והם גם נולדו אחרי שנת 1990 (עפ"י ההיגיון הסיכויי לכך הוא אפסי). לאחר מכן, הסתכלנו על התפלגות הילדים ובני הנוער לפי סטטוס האנשים וקיבלנו:



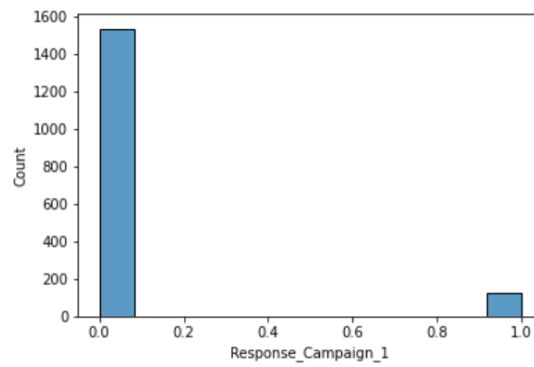


כל זאת בדומה לכמות הילדים ולכמות בני הנוער: ניתן לראות ב"מבט מהיר" שכמות האנשים שאין להם ילדים כלל כמעט זהה לכמות האנשים שיש להם ילד אחד (הסיכוי לילד אחד מעט יותר נמוך) ובנוסף ניתן לראות שכמות האנשים שיש להם שני ילדים נמוכה ביותר. לשם כך בחרנו להגריל לאנשים את כמות הילדים כך שיקבלו אפס ילדים בסיכוי 0.5, ילד אחד בסיכוי 0.47 ו-2 ילדים בסיכוי 0.03. נזכיר שוב שפעלנו בצורה זו גם כלפי כמות הילדים וגם כלפי כמות בני הנוער. לבסוף בדקנו וראינו שאכן ההתפלגויות והגרפים נשארו דומים.

- לאחר מכן, ראינו שיש לנו 14 רשומות ריקות בשדה ה-Mnt_sweet (כמות המוצרים המתוקים) ובחרנו להשלים אותן באמצעות החציון (לא מדובר בכמות גבוהה של רשומות). בדקנו וראינו שהתפלגות זו נשארה זהה לזו שמצאנו בסעיף 1.2:



- בשלב זה, אנו רוצים למלא את כמות המוצרים שנרכשו באמצעות אתר האינטרנט. לשם כך, בדומה למה שהסברנו קודם, החלטנו למלא את הערכים הללו בכמות הכניסות שהאדם נכנס לאתר ועוד ההפרש הממוצע בין כמות הרכישות מהאתר לכמות הכניסות אליו (אותו חישבנו קודם וקיבלנו שהוא שווה ל-3).
- לבסוף, נשאר לנו לעדכן את הרשומות החסרות עבור שדה ה-Response_Campaign1. הסתכלנו בהתפלגות שמצאנו עבור ערך זה בסעיף 1.2:



ניתן לראות שכמות אלו שלא הגיבו לקמפיין (ערך 0) גבוהה בהרבה מאלה שהגיבו לו (ערך 1), ולכן בחרנו לעדכן את כל הרשומות החסרות בערך הפופולארי יותר בהרבה – 0.

1.8 טרנספורמציות מידע – נורמליזציה ודיסקרטיזציה:

דיסקרטיזציה:

בחרנו לבצע דיסקרטיזציה על השדות שנת לידה והכנסה, שכן ראינו שבהכנסה יש ערכים חריגים, שרצינו בכל זאת להראות אותם ולהתייחס להשפעה שלהם. בשנת הלידה יהיה לנו מאוד נוח להסתכל על טווח של שנים. בנוסף, בשדות אלה אנו רוצים לראות את טווח הערכים ופחות משנה לנו ערך של שנה ספציפית, כזו או אחרת, או ערך של הכנסה כזו או אחרת. נציין גם שדאגנו שלא תהיה עמודה נמוכה משמעותית ביחס לעמודות אחרות. קיבלנו:

עבור שנת הלידה:

3	534
2	504
1	370
0	263

עבור ההכנסה:

0	574
3	448
1	261
2	226
4	162

נורמליזציה:

תחילה, החלטנו לאחד את שדות מספר הילדים ומספר הנערים לשדה – Children, משום שלשני הערכים יש משמעות דומה.

עבור שאר הערכים שלא ביצענו עליהם דיסקרטיזציה (והם לא נומריים): "Registration_date", "Recency", "Mnt_Fruits", "Mnt_Meat", "Mnt_sweet", "Mnt_Wines", "Mnt_Gold_Products", "Mnt_Fish", "Num_Web_Purchases", "Num_Store_Purchases", "Num_Deals_Purchases", "Num_Catalog_Purchases", "Num_Web_Visits", "Children"

ביצענו נורמליזציה, משום שעבור כל השדות הללו ראינו בגרפים הקודמים שאין ערכים חריגים בצורה קיצונית שאליהם נרצה להתייחס ומשום שאנו רוצים לשמור על השדות רציפים ועל היחס שבהן, מבלי שנצטרך לוותר על ערכים לטובת הצגה עם טווחים (כמו בדיסקרטיזציה). בנוסף, רוצים להעביר את כל השדות לאותה סקאלה, כדי שלא יהיה שדה שישפיע בצורה קיצונית יותר מהאחרים, משום שהערכים בו מאוד גבוהים.

PCA:

כאן המטרה להוריד את מימדי המידע. לצורך כך, היה עלינו להעביר את השדות – סטטוס והשכלה לערכים נומריים, המרנו על פי התמונה הבאה:

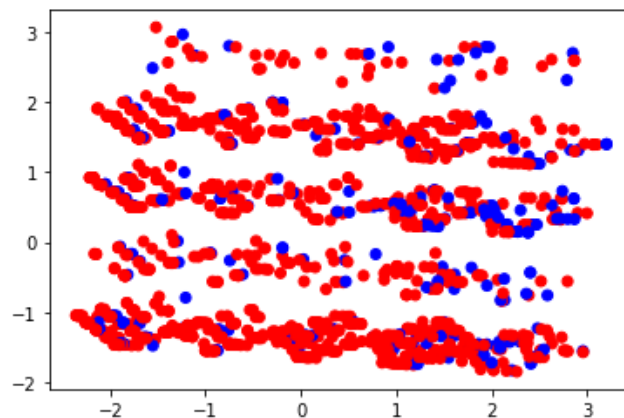
```
['Graduation' '2n Cycle' 'PhD' 'Master' 'Basic'] ---> [0, 1, 2, 3, 4]
['Married' 'Divorced' 'Single' 'Together' 'Widow'] ---> [0, 1, 2, 3, 4]
```

בשלב זה כל הערכים שלנו נומריים וכולם עברו נורמליזציה או דיסקרטיזציה, ולכן נוכל להריץ PCA.

בנוסף, ניתן לראות שסכום השוניות של טורים 1 ו-2 גדולות מ-0.5, ולכן הם מכילים את רוב המידע.

```
array([0.30403789, 0.23341864, 0.22017403, 0.13511911, 0.02191104,
       0.01236641, 0.01192092, 0.01034978, 0.00981063, 0.00743531,
       0.00573689, 0.00526694, 0.00458561, 0.00376759, 0.00231998,
       0.00220547, 0.00187258, 0.00161126, 0.00151619, 0.00147416,
       0.00129241, 0.0007295 , 0.00067606])
```

לבסוף, יצרנו גרף המתאר את 2 העמודות הראשונות, כאשר **אדום – מייצג 0 ב-Response** וכחול – מייצג 1 ב-Response.



נוכל לראות שאכן יש סדר מסויים במידע ושהוא אינו מבולגן.