

Wellbeing and lifestyle

שם הקורס: ניתוח נתונים

שם החוג: מדעים דיגיטלים להייטק

מגישים: שחק פרסר, עדי קורן וסער כהן

תיאור סט הנתונים

1. מהיכן נלקח סט הנתונים? מי אסף את הנתונים? לאיזו מטרה? האם יש לו אג'נדה בפרסום הנתונים?

סט הנתונים נלקח מהאתר Kaggle. הסקר שהנתונים מתבססים עליו נערך על ידי Authentic-Happiness.com (אתר שמוקדש לסיוע לאנשים להעריך ולשפר את חייהם באמצעות מחקרים ותובנות בתחום האושר והרווחה האישית). ו-360living.co (פלטפורמה המתמקדת בהעצמה אישית ובשיפור איכות החיים בכל היבטי החיים), הנתונים נאספו כחלק מפרויקט שמטרתו לספק תובנות על האופן שבו גורמים שונים באורח החיים משפיעים על הרווחה הכללית, כולל היבטים כמו מתח יומיומי, רשת חברתית, פעילות גופנית ודפוסי שינה. לא מצוינת אג'נדה מסחרית ברורה בפרסום הנתונים, והם מיועדים לקידום המחקר והידע בתחומים אלו.

2. מאיזו אוכלוסיה נאספו הנתונים? מתי נאספו? האם מדובר בתקופה חריגה? האם יש למועד משמעות? סט הנתונים נאסף באמצעות סקר מקוון בו השתתפו 15973 אנשים, ומתאר את אורח חייהם והתנהגותם במספר ממדים. הנתונים נאספו בין השנים 2015 ל-2021, כולל תקופת מגפת הקורונה, מה שיכול להשפיע על חלק מהממצאים, בעיקר בנוגע למתח יומיומי ושימוש ברשתות חברתיות.

3. האם הם עברו "טיפול ראשוני"? סינון? עיבוד? דיסקרטיזציה? כן, עברו עיבוד ראשוני, כולל סינון ועיבוד לפני פרסומו.

- כל נתון גולמי שנאסף עבר תהליך ניקוי להסרת שגיאות ואי דיוקים. זה כולל הסרה של נתונים חלקיים, כפולים, או שגויים.
- הנתונים עובדו וסוננו כדי להבטיח את איכותם ותקפותם. תהליך זה כלל בדיקות שונות לוודא את אמינות הנתונים ואת ההתפלגות הסטטיסטית שלהם.

- נתונים מסוימים עברו תהליך של דיסקרטיזציה (הפיכת נתונים רציפים לקטגוריות), כדי להקל על הניתוח הסטטיסטי. למשל, טווחי גילאים חולקו לקטגוריות כמו "36 to 50" ו-"51 or more".

4. האם מי שאסף את הנתונים השתמש בהם לצורך פרסום מאמר או מסמך עמדה? אם כן צרפו. לא צוין בסט הנתונים אם הוא שימש לפרסום מאמר או מסמך עמדה ספציפי. עם זאת, הנתונים זמינים לשימוש חוקרים שונים למטרות מחקר ופיתוח.

5. האם יש biases מובנים בנתונים? למשל bias Selection, bias winners? הסבירו. בנתונים שנאספו ייתכנו מספר סוגי biases:

- **Selection Bias (הטיית בחירה):** המשתתפים הם אלו שהסכימו להשתתף במחקר, ייתכן והם מייצגים אוכלוסיה עם מודעות גבוהה יותר לבריאות.
- **Volunteer Bias (הטיית מתנדבים):** המתנדבים להשתתף במחקר עשויים להיות שונים במאפיינים מסוימים מאלו שלא מתנדבים, כמו רמת השכלה או מודעות לבריאות.
- **Bias Winners (הטיית מנצחים):** במחקרים ארוכי טווח, הנבדקים ששרדו את המחקר מייצגים אוכלוסיה מסוימת ולא את כלל האוכלוסיה המקורית.

6. מי עוד כבר עשה שימוש בנתונים אלו? לאיזו מטרה? לאילו מסקנות הם הגיעו? צרפו לינק. סט הנתונים "Lifestyle and Wellbeing Data" שימש למספר פרויקטים אקדמיים וניתוחים שונים:

פרויקט אנליטיקת נתונים: נעשה שימוש בנתונים לניתוח סיום בקורס שהתמקד בזיהוי מאפייני אורח חיים שתורמים למדד מסת גוף בריא (BMI).

- **מסקנה:** הפרויקט זיהה מאפיינים של אורח חיים שתורמים לשמירה על BMI בריא.
- **לינק:** [GitHub](#)

פרויקט הבנת הקשרים בין מאפייני אורח החיים לרווחה אישית: פרויקט נוסף השתמש בנתונים כדי להבין את הקשרים בין מאפייני אורח החיים לרווחה אישית, כולל יצירת אינטראקציות גרפיות.

- **מסקנה:** הפרויקט עזר לחשוף קשרים גרפיים בין מאפייני אורח החיים לרווחה האישית.
- **לינק:** [GitHub](#)

פלטפורמת Analyst-2: השתמשה בנתונים לחקר שיפור אורח החיים והערכת איזון בין עבודה לחיים אישיים, כחלק מתמיכה במטרות הפיתוח בר קיימא של האו"ם.

- **מסקנה:** המחקר סייע בהבנת השפעות האיזון בין עבודה לחיים על הרווחה האישית ותרם להמלצות לשיפור אורח החיים.
- **לינק:** [Analyst-2](#)

7. **אילו משתנים כלולים בו ובאילו שלושה משתנים מספריים מביניהם תתמקדו בתרגיל שלכם? המשתנים בסט הנתונים כוללים מגוון רחב של משתנים הקשורים לאורח חיים, בריאות ורווחה. שלושת המשתנים הנומריים שנבחרו לתרגיל הם:**

- **DAILY_STEPS**
- **SLEEP_HOURS**
- **TIME_FOR_PASSION** (זמן לדברים שאתה אוהב)

משתנים נוספים שמתייחסים אליהם בסט הנתונים:

- FRUITS_VEGGIES
- DAILY_STRESS
- PLACES_VISITED (מקומות שביקרת בהם)
- CORE_CIRCLE (מעגל החברים הקרוב)
- SUPPORTING_OTHERS (תמיכה באחרים)
- SOCIAL_NETWORK
- ACHIEVEMENT
- DONATION
- BMI_RANGE
- TODO_COMPLETED (משימות שהושלמו)
- FLOW
- DAILY_STEPS
- LIVE_VISION
- SLEEP_HOURS
- LOST_VACATION
- DAILY_SHOUTING

- SUFFICIENT_INCOME
- PERSONAL_AWARDS
- TIME_FOR_PASSION
- WEEKLY_MEDITATION
- AGE
- GENDER
- WORK_LIFE_BALANCE_SCORE

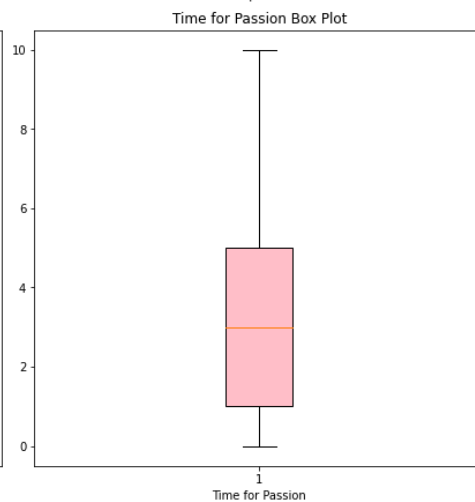
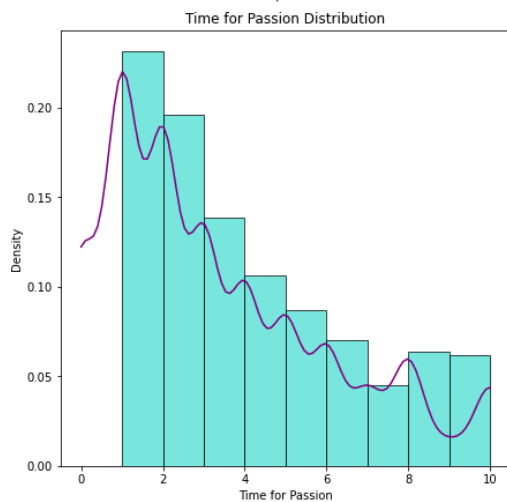
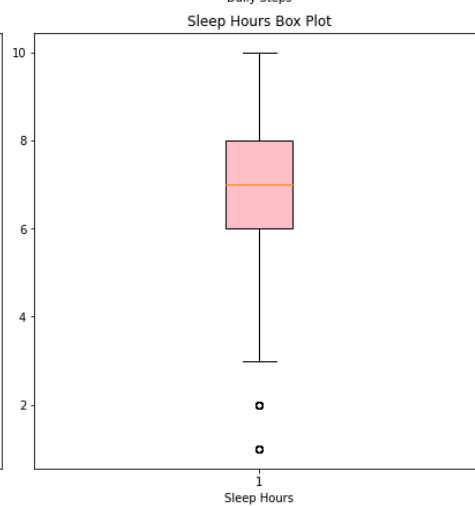
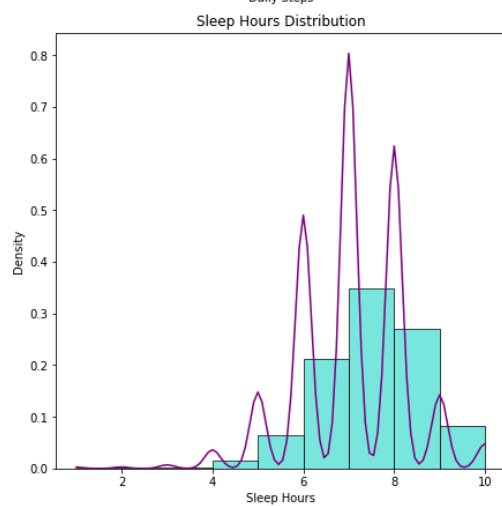
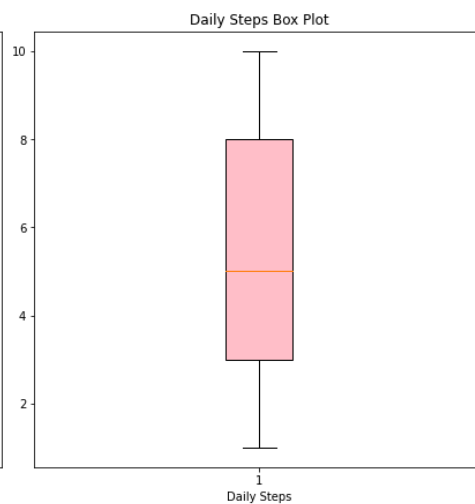
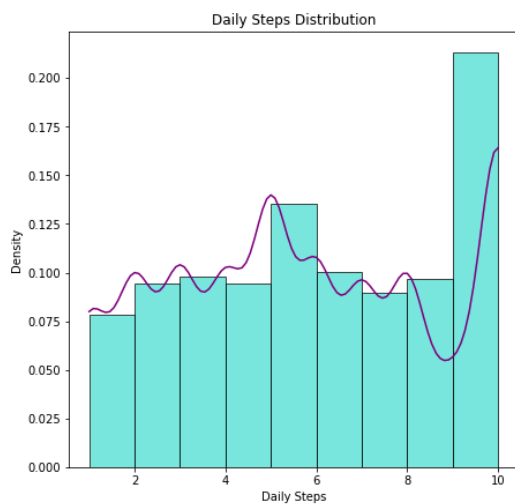
8. כמה רשומות כלולות בסט הנתונים? מה הוא סוג המשתנים? מה הם מייצגים בעולם? בסט הנתונים ישנן 15,972 רשומות. סוגי המשתנים כוללים משתנים קטגוריים ונומריים. הם מייצגים היבטים שונים של אורח חיים ורווחה אישית בעולם האמיתי.

9. הסבר קצר מדוע ניתוח של סט הנתונים מעניין מבחינה עסקית / תפעולית / ארגונית / ניהולית, כלומר אילו מסקנות ניתן יהיה לגזור מניתוח סטטיסטי שלו כדי לתמוך בקבלת החלטות כלשהי?

- ניתוח של סט הנתונים מעניין מכמה סיבות עסקיות, תפעוליות, ארגוניות וניהוליות:
- **שיפור בריאות העובדים:** ניתוח של שעות השינה וזמן לפעילות אהובה יכול לספק תובנות לגבי אורח החיים של העובדים והשפעתו על בריאותם ורווחתם. תובנות אלו יכולות לסייע בפיתוח תוכניות רווחה ובריאות שמטרתן לשפר את בריאות העובדים ולהפחית היעדרויות.
- **הגברת פרודוקטיביות:** הבנה של הקשרים בין צעדים יומיים, שעות שינה וזמן לפעילות אהובה לבין פרודוקטיביות יכולה לעזור לזהות גורמים המגבירים או מקטינים את הפרודוקטיביות של העובדים. לדוגמה, אם נמצא קשר חיובי בין פעילות גופנית לפרודוקטיביות, ניתן לעודד את העובדים לבצע יותר צעדים יומיים.
- **פיתוח מדיניות איזון חיים-עבודה:** תובנות לגבי ההתפלגות של שעות השינה והזמן לפעילות אהובה יכולות לסייע בפיתוח מדיניות שתומכת באיזון טוב יותר בין חיים לעבודה. לדוגמה, עידוד שעות שינה מספקות וזמן לפעילות אהובה יכול להוביל לשיפור באיכות החיים של העובדים, מה שיכול להתבטא בשביעות רצון גבוהה יותר וביצועים משופרים.

מסקנות אפשריות מהניתוח יכולות לכלול זיהוי הגורמים המשפיעים על רווחה ובריאות, פיתוח תוכניות לשיפור איכות החיים של העובדים, ושיפור מוצרים ושירותים בתחום הבריאות.

חלק ג'



ניתוח מאפייני ההתפלגות:

צעדים יומיים (Daily Steps) - באלפים:

- ממוצע: 5.70
- חציון: 5.0
- ערך שכיח: 10
- סטיית תקן: 2.89
- פיזור: ההתפלגות של הצעדים היומיים צמודה יחסית לממוצע ולחציון, אך הערך השכיח גבוה יותר, מה שמצביע על כך שישנם אנשים שמבצעים הרבה צעדים.

שעות שינה (Sleep Hours):

- ממוצע: 7.04
- חציון: 7.0
- ערך שכיח: 7
- סטיית תקן: 1.20
- פיזור: ההתפלגות של שעות השינה די צמודה לממוצע ולחציון, מה שמעיד על כך שרוב האנשים ישנים כ-7 שעות בלילה.

זמן לפעילות אהובה (Time for Passion):

- ממוצע: 3.33
- חציון: 3.0
- ערך שכיח: 1
- סטיית תקן: 2.73
- פיזור: ההתפלגות של הזמן לפעילות אהובה מראה פיזור רחב יותר, עם ערך שכיח נמוך, מה שמעיד על כך שישנם אנשים רבים שמקדישים מעט מאוד זמן לפעילות אהובה.

הסבר על Outliers ואנומליות בהתפלגות:

Outliers:

- **צעדים יומיים:** לא נמצאו Outliers במדד זה לפי שיטת IQR.
- **שעות שינה:** נמצאו Outliers במספר רשומות:
 - נמצאו 39 Outliers.
 - דוגמאות לערכים חריגים: שעות שינה של 1 או 2 שעות במספר מקרים.
- **זמן לפעילות אהובה:** לא נמצאו Outliers במדד זה לפי שיטת IQR.

אנומליות:

- **צעדים יומיים:** לא נמצאו אנומליות במדד זה.
- **שעות שינה:** לא נמצאו אנומליות במדד זה.
- **זמן לפעילות אהובה:** לא נמצאו אנומליות במדד זה.

הסברים נוספים:

- **הסבר להעדר Outliers ואנומליות במדד הצעדים היומיים וזמן לפעילות אהובה:**
 - הנתונים נאספו ונבדקו בצורה מדויקת, כך שאין ערכים חריגים או טעויות הזנה.
 - הנתונים מתפלגים באופן טבעי ורגיל, בלי ערכים קיצוניים במיוחד שמצביעים על חריגות.
- **הסבר להימצאות Outliers במדד שעות השינה:**
 - הערכים החריגים במדד שעות השינה מצביעים על אנשים עם דפוסי שינה קיצוניים במיוחד.
 - מדובר בקבוצה ספציפית עם מאפיינים יוצאי דופן המבדילים אותם משאר המדגם.

מאפיינים נוספים בהתפלגות:

- לא נמצאו אנומליות לפי ערכים חסרים במדדים שנבדקו, מה שמעיד על נתונים שלמים במדדים אלו.
- ניתוח סטטיסטי של המדדים האחרים מצביע על כך שהערכים נמצאים בטווחים הנורמליים ואין ערכים חריגים משמעותיים או חוסר במידע שעלול להשפיע על תוצאות הניתוח.

בסוף המסמך מצורפות כלל המסקנות שלנו בהתאם לשאלת המחקר.

מבחני השערות על התוחלת:

הנחות:

* השונות באוכלוסיה אינה ידועה

* $n = 160$, $n > 30$

לכן, ממשפט הגבול המרכזי, ממוצע המדגם מתפלג נורמאלי: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

מבחן חד-צדדי ימני על זמן לפעילות אהובה:

אנחנו משערים שהתוחלת של השעות המוקדשות לפעילות אהובה אצל אנשים הינה 3, וזאת מכיוון שהרופאים מיחסים חשיבות רבה לשילוב פעילויות אהובות ביחס ישיר לבריאותו.

$$H_0: \mu \leq 3$$

$$H_1: \mu > 3$$

מסקנה: ברמת מובהקות 0.05, אנחנו לא דוחים את השערת האפס, כלומר התוחלת של השעות המוקדשות לפעילות אהובה היא בסביבות 3 שעות. מכיוון שתוצאות המבחן הסטטיסטי t משמעותית גבוהות יותר מהערך הקריטי, ניתן להבין למה p-value שואף ל-0.

מבחן חד צדדי שמאלי על שעות שינה:

אנחנו משערים שהתוחלת של שעות השינה הינה 7, לפי המלצת הרופאים. נבצע מבחן חד צדדי שמאלי בכדי לבדוק האם התוחלת נמוכה מ 7 שעות שינה.

$$H_0: \mu \geq 7$$

$$H_1: \mu < 7$$

מסקנה: ברמת מובהקות 0.05, אנחנו לא דוחים את השערת האפס, כלומר התוחלת של שעות השינה לא גדולה מ 7.

מבחן דו-צדדי על כמות הצעדים היומית:

אנחנו משערים שהתוחלת של כמות הצעדים היומית היא 8000, שזוהי הכמות היומית המינימלית המומלצת.

כמו כן, אנחנו יודעים שזה יכול להתפלג שונה לגמרי בין אנשים- לכן נבדוק זאת במבחן דו צדדי.

$$H_0: \mu = 8,000$$

$$H_1: \mu \neq 8,000$$

מסקנה: ברמת מובהקות 0.05, אנחנו דוחים את השערת האפס, כלומר תוחלת הצעדים היומית שונה משמעותית מ-8,000.

מכיוון שתוצאות המבחן הסטטיסטי t משמעותית גבוהות יותר מהערך הקריטי, ניתן להבין למה p -value שואף ל-0.

מבחן השערות על השונות:

הנחות:

*אנו מניחים שסט הנתונים מתפלג נורמלית למרות שהוא לא, כפי שלמדנו בכיתה

חישוב השונות באוכלוסייה הכללית:

- התייחסנו לכלל הדאטה כאוכלוסייה
- השוואת השונות של מדגם קטן של שעות השינה, לשונות שלו באוכלוסייה שהוגדרה.

בחירת מדגם רנדומלי קטן:

- ניקח מדגם רנדומלי של 160 תצפיות מתוך הנתונים על שעות השינה.
- חישוב השונות במדגם הרנדומלי.

השערות:

- השערת האפס (H_0): השונות במדגם שווה לשונות באוכלוסייה הכללית (המוערכת).
- השערה חלופית (H_1): השונות במדגם שונה מהשונות באוכלוסייה הכללית (המוערכת).

ביצוע המבחן:

- שימוש במבחן חי-בריבוע (Chi-square) לבדיקת ההשערות.

תוצאות:

- אנחנו לא דוחים את השערת האפס, השונות במדגם שווה לשונות באוכלוסייה הכללית (המוערכת).

מבחן השערות על הפרופורציה:

הנחות:

$$n \cdot p_0 > 5, n \cdot q_0 > 5$$

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right) \quad \text{* עפ"י משפט הגבול המרכזי:}$$

דיסקרטיזציה של משתנה 'DAILY_STEPS':

- לקיחת מדגם רנדומאלי של 160 תצפיות
- הפרופורציה לא ידועה מראש, ונדרש לאמוד אותה
- המרת המשתנה 'DAILY_STEPS' למשתנה קטגורי על ידי חלוקתו למספר קטגוריות (למשל, ('Very Low', 'Low', 'Moderate', 'High', 'Very High')):

| | |
|-----------|--------|
| Very low | 0-1.99 |
| low | 2-4.99 |
| moderate | 5-6.99 |
| high | 7-8.99 |
| Very high | 9+ |

השערות:

- השערת האפס (H_0): הפרופורציה של אנשים עם מספר צעדים יומיים 'Moderate' או גבוה יותר היא 50% ($p = 0.50$).
- השערה חלופית (H_1): הפרופורציה של אנשים עם מספר צעדים יומיים 'Moderate' או גבוה יותר שונה מ-50% ($p \neq 0.50$).

$$H_0: P(\text{'Moderate or higher'}) \geq 0.5$$

$$H_1: P(\text{'Moderate or higher'}) < 0.5$$

חישוב הפרופורציה הנצפית:

- חישוב הפרופורציה של אנשים שיש להם מספר צעדים יומיים 'Moderate' או גבוה יותר מתוך כלל המדגם.
- אמידת הפרופורציה.
- שימוש במבחן z לפרופורציות לבדיקת ההשערה.
- חישוב הסטטיסטי z וערך ה- p .

תוצאות:

- ברמת מובהקות 0.05 מקבלים את השערת האפס, כלומר הפרופורציה של מספר הצעדים היומיים עבור קבוצות moderate or higher שווה ל-0.5.

מבחן השערות על שוויון תוחלות

- נניח ששתי התוחלות מתפלגות נורמאלית.
- נבחר משתנה קטגורי (מעל או מתחת ל 7 שעות שינה) לחלק לפיו את הרשומות. נשתמש במשתנה 'SLEEP_HOURS' שיצרנו קודם לכן.
- נבדוק את יחסי השונויות ע"י מבחן f , בכדי לדעת איזה מבחן עלינו לבצע לאחר מכן, במבחן הפרש התוחלות:

השערות:

- השערת האפס (H_0): שונויות של שתי קבוצות ה-SLEEP_HOURS שוות.
- השערה חלופית (H_1): השונויות אינן שוות
- נבחר משתנה מספרי אחר להשוואת התוחלות בין הקבוצות. נשתמש במשתנה 'DAILY_STEPS'.

השערות:

- השערת האפס (H_0): התוחלת של המשתנה 'DAILY STEPS' זהה בין שתי קבוצות ה-'SLEEP HOURS'.
- השערה חלופית (H_1): התוחלת של המשתנה 'DAILY STEPS' שונה בין שתי קבוצות ה-'SLEEP HOURS'.

ביצוע המבחן:

- נבצע מבחן f לחישוב השונויות, תוך חישוב הסטטיסטי f .
- נשתמש במבחן t לבדיקת ההשערות בין שתי קבוצות שנבחר, בהתאם לתוצאות מבחן f .

תוצאות:

- לא דחינו את השערת האפס ברמת מובהקות 0.05, התוחלת של משתנה daily steps אינה שונה משמעותית בין שתי הקבוצות הקטגוריאליות שמבדילות בין שעות השינה.

מבחן לטיב ההתאמה של שעות שינה להתפלגות נורמלית באמצעות מבחן חי-בריבוע

מטרת הניסוי היא לבדוק האם משתנה שעות שינה (SLEEP_HOURS) מתפלג אחיד- בדיד. נבצע את הבדיקה באמצעות מבחן חי-בריבוע לטיב התאמה.

נבחר מדגם של 160 תצפיות.

השערות:

- השערת האפס (H_0): המשתנה שעות שינה (SLEEP_HOURS) מתפלג אחיד בדיד $X \sim U$.
- השערה חלופית (H_1): המשתנה שעות שינה (SLEEP_HOURS) אינו מתפלג אחיד בדיד - כלומר אחרת.

חישוב פרמטרי ההתפלגות האחידה בדידה:

- התאמת התפלגות המשתנה להתפלגות אחידה בדידה.

חלוקה לקטגוריות:

- חילוק הנתונים לקטגוריות וחישוב השכיחויות הנצפות (observed) והצפויות (expected) בכל קטגוריה.

תוצאות ומסקנות:

- המבחן דוחה את השערת האפס, ומכאן ששעות השינה אינן מתפלגות אחיד בדיד.
- $p\text{-value}$ שואף ל-0, כיוון שההתפלגות אינה אחידה בדידה.

מבחן לאי תלות: ניתוח תלות סטטיסטית ודיסקרטיזציה: בין שעות שינה ו-Time for Passion

דיסקרטיזציה של המשתנים SLEEP_HOURS, TIME_FOR_PASSION:

* לא נלקח מדגם למען ביצוע במבחן זה, מכיוון שבדקים תלות בין קטגוריות שונות באוכלוסיה.

• SLEEP_HOURS:

- פחות מ-5 שעות: Less than 5 hours
- בין 5 ל-7 שעות: hours 5-7
- יותר מ-7 שעות: More than 7 hours

• TIME_FOR_PASSION:

- פחות מ-1 שעה: Less than 1 hour
- בין 1 ל-2 שעות: hours 1-2
- יותר מ-2 שעות: More than 2 hours

התפלגות מספרית של הקטגוריות:

• SLEEP_CATEGORY:

- פחות מ-5 שעות: 1365
- בין 5 ל-7 שעות: 8963
- יותר מ-7 שעות: 5644

• PASSION_CATEGORY:

- פחות מ-1 שעה: 3285
- בין 1 ל-2 שעות: 2781
- יותר מ-2 שעות: 8109

מסקנות:

• ערך ה- ρ :

- ערך ה- ρ שהתקבל קטן בהרבה מרמת המובהקות, ומכאן אנו דוחים את השערת האפס (H_0) ומסיקים שיש תלות מובהקת סטטיסטית בין שעות השינה ל-Time for Passion.

• פרשנות:

- התוצאות מראות שיש קשר מובהק בין מספר שעות השינה לבין זמן לפעילות אהובה.
- הזמן המושקע בפעילויות אהובה קשור באופן מובהק למספר שעות השינה של אנשים.

שאלת המחקר: האם בעולם המודרני אנשים שמים דגש על הבריאות שלהם ומקשיבים להמלצות הרופאים?

מסקנות- לפי המבחנים:

ניתוח מאפייני ההתפלגות:

1. **שעות שינה:** הנתונים מראים כי רוב האנשים מקפידים על שעות שינה הקרובות להמלצות הרפואיות, עם ממוצע של 7.04 שעות בלילה. זה מעיד על מודעות לבריאות ואיזון בחיי היום-יום.
2. **צעדים יומיים:** רוב האנשים אינם עומדים בהמלצה של 8,000 צעדים יומיים, עם ממוצע של 5,700 צעדים ביום. הדבר מעיד על חוסר הקפדה מספקת על פעילות גופנית יומית.
3. **זמן לפעילות אהובה:** ההתפלגות מצביעה על כך שישנם אנשים שמקדישים מעט מאוד זמן לפעילויות אהובות, עם ממוצע של 3.33 שעות ביום, אך ישנם גם אנשים שמקדישים יותר זמן מהמצופה. זה עשוי להעיד על שונות ברמת האיזון בין עבודה לחיים פרטיים.

מבחני השערות על התוחלת:

1. **זמן לפעילות אהובה:** מבחן ההשערה מצביע על כך שהתוחלת של הזמן שמוקדש לפעילויות אהובות היא בסביבות 3 שעות, ולא גבוהה יותר כפי ששוער. זה מצביע על כך שאנשים מתקשים לשמור על איזון בין עבודה לחיים פרטיים.
2. **שעות שינה:** אי דחיית השערת האפס תומכת בכך שרוב האנשים ישנים כ-7 שעות בלילה, בהתאם להמלצות הבריאות.
3. **צעדים יומיים:** המבחן מראה שהתוחלת של מספר הצעדים היומיים שונה מההמלצה של 8,000 צעדים, מה שמחזק את המסקנה שרבים אינם מקפידים על כמות הצעדים היומית המומלצת.

מבחן השערות על השונות:

1. **שעות שינה:** מבחן זה מצביע על כך ששונות שעות השינה במדגם תואמת את השונות באוכלוסייה הכללית. זה מעיד שרוב האנשים ישנים כמות דומה של שעות בלילה, מה שמראה על אחידות מסוימת בהרגלי השינה.

מבחן השערות על הפרופורציה:

1. **צעדים יומיים:** מבחן הפרופורציות מצביע על כך שלמעלה מ-50% מהאנשים מבצעים לפחות מספר צעדים "מתונים" ביום, אך רבים עדיין לא מגיעים להמלצת הצעדים היומית של 8,000 צעדים.

מבחן השערות על שוויון תוחלות:

1. **תוחלת צעדים יומיים לפי קטגוריות שעות שינה:** המבחן מצביע על כך שהתוחלת של מספר הצעדים היומיים אינה שונה באופן מובהק בין קבוצות שונות של שעות שינה, מה שמעיד על כך שאין קשר מובהק בין כמות השינה לפעילות הגופנית היומית.

מבחן לטיב ההתאמה של שעות שינה להתפלגות אחידה בדידה:

1. **שעות שינה:** המבחן דחה את השערת האפס, ומכאן ששעות השינה אינן מתפלגות באופן אחיד בדיד, אלא יש שונות מסוימת בהרגלי השינה.

ניתוח תלות סטטיסטית ודיסקרטיזציה: בין שעות שינה ו-Time for Passion:

1. **תלות בין שעות שינה לזמן לפעילות אהובה:** המבחן הראה שיש קשר מובהק בין מספר שעות השינה לבין הזמן שמוקדש לפעילויות אהובות. אנשים שישנים יותר נוטים להקדיש יותר זמן לפעילויות אהובות, מה שמצביע על איזון טוב יותר בחיים.

מסקנות כלליות:

התוצאות מראות שאנשים מודעים להמלצות הבריאותיות של רופאים, במיוחד בנוגע לשעות השינה. עם זאת, ישנה פחות הקפדה על פעילות גופנית יומית של 8,000 צעדים. כמו כן, ישנם הבדלים ברמת האיזון בין עבודה לחיים פרטיים, כאשר חלק מהאנשים מצליחים לשלב פעילויות אהובות בחיי היום-יום ואחרים מתקשים בכך.