מגישים:

קרן אור גרינברג-רוזנבאום 200748390

עומר רוזנברג 315328898

הקשר בין אקלים לרכיבה באופניים



**חלק 1: מבוא**

1. **השערה אופרציונלית:** בחודשי הקיץ והאביב, משך השימוש היומי החציוני באופניים ארוך יותר בהשוואה לחודשי החורף והסתיו בקרב מנויים.

**משתנים רלוונטיים לבדיקה: תאריך התחלה:** (Start Time and Date) — מאפשר לזהות את החודש והעונה; **עונה:** מחושבת על בסיס החודש בתאריך במשתנה (Start Time and Date) בניו יורק החלוקה היא: חורף (דצמבר, ינואר, פברואר), אביב (מרץ, אפריל, מאי), קיץ (יוני, יולי, אוגוסט), סתיו (ספטמבר, אוקטובר, נובמבר); **משך הנסיעה:** (Trip Duration (seconds)) — המדד המרכזי להשוואה. על בסיסו נחשב משך נסיעה חציוני יומי. **סוג המנוי**: (User type counts) – מאפשר להבדיל בין משתמשים קבועים שהם מנויים (member) לבין משתנים מזדמנים (casual). בחרנו לעשות סינון ולהתמקד במשתמשים הקבועים (מנויים) מתוך הנחה שיש להם דפוסי התנהגות קבועים יותר.

2. **השערה תיאורטית:** קיים קשר בין אקלים להרגלי רכיבה.

מחקרים אודות התנהגות תחבורה מבוססת אקלים מראים שתנאי מזג אויר נוחים (15-25°C) מעודדים רכיבה, שכן הם נתפסים כנוחים לפעילות פיזית, ואילו תנאים קשים כמו גשם ושלג יוצרים מחסומים פסיכולוגיים ופיזיים לרכיבה ((Nielsen et al,2013; Sabir et al, 2014). השערה זה חשובה בעינינו מאחר ואם תוכח כנכונה עבור רוכבי האופניים בניו יורק, היא עשויה לעודד השקעה ביצירת תשתיות מותאמות אקלים לרכיבה באופניים, כמו שבילים מקורים או תחנות חימום, שעשויה לעודד רכיבה גם בתנאים קשים Pucher & Buehler ,2008)). קידום שימוש באופניים תורמת לבריאות המשתמשים Weichenthal et al 2011)) על פני נסיעה בכלי ממוכן. בנוסף עידוד שימוש באופנים יתרום להפחתת זיהום אוויר בערים הנגרם מפליטות כלי רכב, שגם תורם לבריאות התושבים (Ferdinands et al 2008; Weichenthal et al 2011) . בצל גידול האוכלוסין בערים גדולות, בייחוד כמו ניו יורק, לצד השלכות משבר האקלים בגלי חום, מתן תשתיות ועדיפות לאופניים (ואפילו סובסידיה עירונית) יכולה לתת מענה על פני השקעה אדירה בהתאמת תשתיות במטרו למשל, אשר ניכר כי אינה מותאמת לגלי חום המתחזקים בעיר לפי מאמר בניו יורק פוסט מיולי 2024.

**חלק 2: השיטה**

1. קובץ הנתונים נלקח מ- Citi Bike מערכת השכרת האופניים הגדולה בניו יורק. הנתונים נאספו באמצעות מערכת ניטור המובנית בתחנות האופניים של Citi Bike ובאמצעים דיגיטליים העוקבים אחר תחילת וסיום הנסיעה, יחד עם מידע נוסף כמו סוג המשתמש והזמן שבו התבצעה הנסיעה. קובץ הנתונים מכיל מידע מפורט על נסיעות האופניים, כולל משך הנסיעה, זמן ותאריך התחלה וסיום, מיקום תחנת התחלה וסיום (כולל מזהי תחנות, קווי רוחב ואורך), סוג המשתמש (חבר במנוי שנתי או רוכב מזדמן), מידע על מגדר ושנת לידה (כאשר קיים משתמש מנוי). קובץ הנתונים הופק על ידי NYC Bike Share LLC האחראית על תפעול מערכת Citi Bike והנתונים זמינים לשימוש חופשי לציבור. הנתונים עברו עיבוד לפני הפרסום על פי הצהרת האתר, כדי להסיר סוגי נסיעות מסוימים: נסיעות שבוצעו על ידי צוות התחזוקה של המערכת כחלק מבדיקות ותפעול; נסיעות אל/מאת תחנות "ניסוי" שהיו בשימוש בעיקר בחודשים יוני ויולי 2013; נסיעות קצרות מ-60 שניות (שייתכן שהן תוצאה של ניסיונות עגינה מחדש או התחלה שגויה). קבצי הנתונים ניתנים להורדה [פה](https://s3.amazonaws.com/tripdata/index.html) עבור השנים 2013 עד 2024 כולל. עבור השנים 2017, 2016 היה חוסר עבור מספר חודשים בשנה ולכן לא נעשה בהן שימוש. בנוסף עבור 2020-2024 היה שוני באופן בו הנתונים נשמרו ולכן בחרנו להתמקד בשנים שיש אחידות בקבצים וביטוי לכל 365 ימות השנה. בנוסף בשנים 2020-2022 העולם הושפע מהקורונה (סגרים). בחודשים שבהם נרשמו יותר ממיליון נסיעות, נתוני הנסיעות פוצלו למספר קובצי CSV בתוך אותו קובץ מכווץ (ZIP). כדי לקבוע את סך כל הנסיעות בחודש מסוים, סכמנו את כל הרשומות מכל קובצי ה-CSV של אותו חודש. מדובר בקבצי ענק שהיה קשה לעבוד איתם ולכן אנחנו ערכנו עיבודים לקובץ כולל הוספת עמודות של ערכי חציון למשך נסיעה יומי (על פני שימוש בכל הרשומות היומיות), והוספת עונה לכל תצפית על בסיס החודש בו נקלטה. בנוסף עשינו סינון בין סוגי המשתמשים (מנוי לעומת מזדמן).

**2. ניתוח המשתנים והקשר להשערה**

**משתנה 1: עונה:** על בסיס תאריך התחלה (Start Time and Date)הוספנו את העונה אליה משויך היום. זהו משתנה קטגוריאלי. נסתייג ונאמר שאם היינו מסתכלים על עונה בשנה אחת העונה יכלה להיות משתנה אורדינלי מאחר ועונות יכולות להיות מסודרות בסדר כרונולוגי. אולם לקחנו עונות ממספר שנים ולכן ההבדלים לא ניתנים למדידה כמותית מדויקת (אין "מרחק" קבוע בין עונה לעונה של שנים שונות) .טווח הנתונים מכסה את כל חודשי השנה בכל שנה, ומאפשר לזהות את העונה. נסתייג ונאמר גם העונות כן נבדלות בתנאי האקלים (טמפרטורה, משקעים) ומהבחינה הזו ניתן להשוות אותן אך מידת השוני נבדלת לפי הערך (האם בודקים טמפ', גשם וכו') וזו אינה שאלת המחקר. נסתכל על השנים 2013, 2014, 2018, 2019 בהן יש ייצוג לכל 4 העונות.אין ערך שכיח מאחר ובחרנו בשנים בהן כל העונות מיוצגות בצורה שווה לאורך השנה ולאורך כל השנים הללו (25% מהתוצאות לכל עונה).

**משתנה 2: משך הנסיעה Trip Duration (seconds)** זהו משתנה מסקאלה RATIO . מאחר והשתמשנו בנתונים על פני כמה שנים כדי להתמודד עם עומס הנתונים בחרנו להשתמש בערך החציון היומי של משך הנסיעה וזאת מאחר וממוצע מושפע מאוד מקצוות קיצוניים (גם במקרה של טעויות מדידה למשל אי החזרת אופניים בזמן).  **שכיח: 603**\* לדעתנו זה חסר משמעות מאחר והמדידות הן ברמת שניה. **חציון :** 562.75 שניות **ממוצע:** 557.14 שניות **. ערכים קיצוניים:** הערך המקסימלי הוא 705 והמינימלי הוא 385.  **ערכים חסרים:** במקור היו נתונים על שנים בהן היה חוסר במידע על חודשים ללא ייצוג, ולכן בחרנו לא לכלול אותן (היה מספיק מידע משנים אחרות). **טווח:** משך הנסיעה נמדד בשניות עם טווח שנע בין 60 שניות (תצפיות קטנות מכך הוסרו מהמאגר) ועד מאות שניות.ממוצע ופיזור בזמנים לאורך העונות מוצגים בטבלה בפרק תוצאות.

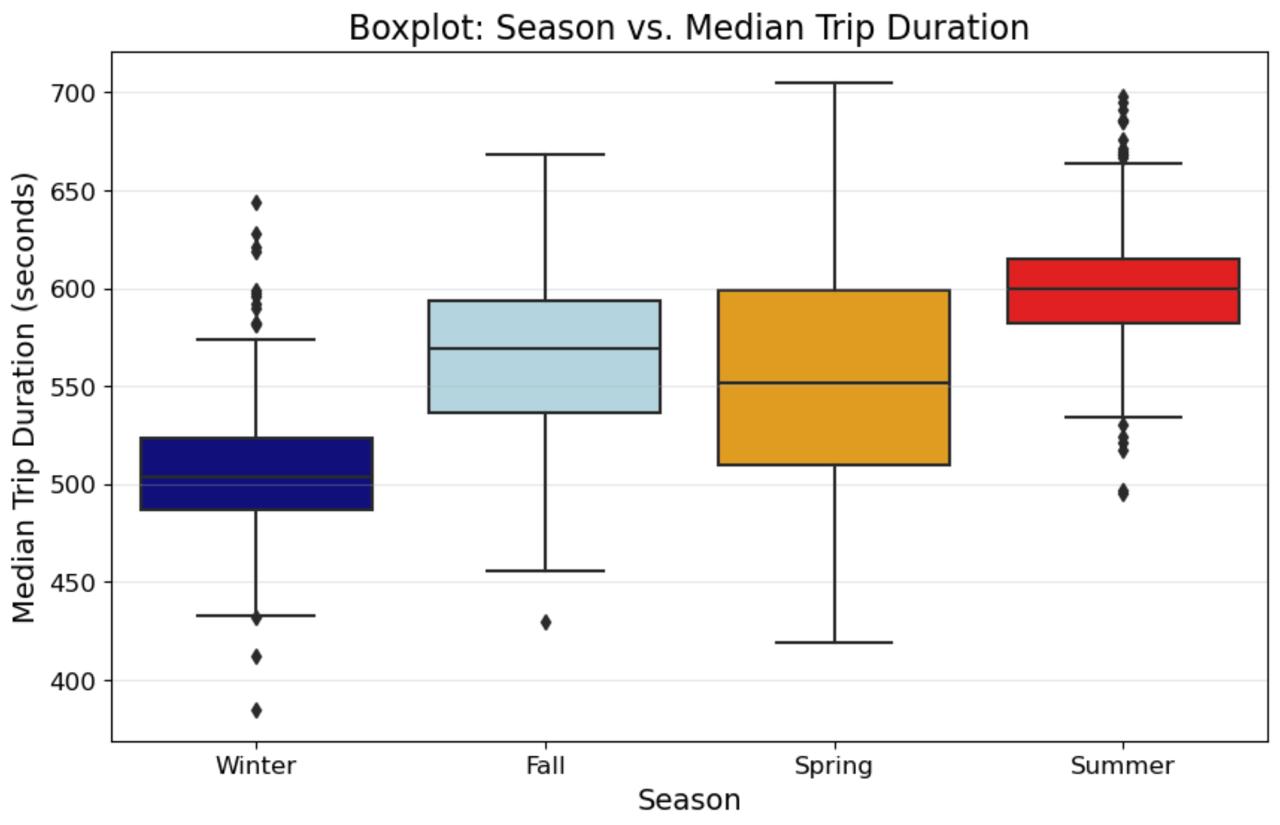
**3. תוקף ומהימנות**

שיוך "העונה" מתבסס על חותמת זמן שנחשבת מדויקת שכן נרשמת אוטומטית במערכת Citi Bike לכן המשתנה בעל **תוקף פנים** לפי הערכתנו. גודל המדגם גדול ומייצג עונות במשך מספר שנים, וניתן להניח כי ההבדלים העונתיים מבחינת נוחות הנסיעה מתקיימים לאורך השנים. גם "משך הנסיעה" נרשם אוטומטית ולכן התוקף הפנימי גבוה. המשתנה משקף את משך השימוש בפועל ואם משתמשים שוכחים להחזיר אופניים בזמן הם יכולים לערער על החיוב (אנחנו לא בטוחים שהתוצאה אכן תימחק מהרשומות ולכן השתמשנו בערך החציוני ולא הממוצע כדי למנוע הטייה של תוצאות חריגות). **תוקף הניבוי** מוצג בפרק תוצאות ומראה מובהקות שאכן נמצא קשר בין העונה למשך זמן הרכיבה. **תוקף מתכנס** נבדוק לפי הקשר בין המשתנים של משך נסיעה (חציון) יומי וכמות נסיעות יומית (שיחושב על ידינו על ידי ספירת כמות הנסיעות ביום), כששניהם בודקים את המשתנה התאורטי של הרגלי רכיבה באופניים. קיבלנו מתאם פירסון של 0.6048 שנחשב גבוה (מעל 0.5), ולכן ניתן לומר שיש תוקף מתכנס. נבדוק **תוקף מבחין** על ידי המתאם בין העונה שהיא משתנה אופרציונלי של אקלים, לבין משך הנסיעה שהיא משתנה אופרציונלי של הרגלי רכיבה. ערך מתאם ספירמן הוא 0.575 ומצביע על קשר חיובי בינוני-חזק מובהק בשל P-value: 2.064e-129 נמוך מאוד. זה אומר שככל שהעונה משתנה (למשל, מחורף לקיץ), ישנה נטייה לשינוי ברור גם במשך הנסיעה. תוצאה הזו אינה תומכת בתוקף מבחין מאחר שנמצא קשר חזק ומובהק בין שני המשתנים, ייתכן שהמשתנים אינם מייצגים תכונות שונות לחלוטין כפי שציפינו. יתכן שהאקלים (עונה) משפיע בצורה מהותית על הרגלי הרכיבה (משך הנסיעה), למשל על תשתיות הנסיעה ולכן הם אינם בלתי תלויים. נבדוק **תוקף מבנה** על ידי בדיקת הקשר בין משך נסיעה חציוני יומי לבין מספר נסיעות יומי לבין מגדר (משתנה שלישי). עבור מבחן T בין המינים עבור משך הנסיעה קיבלנו ערך T של 0.000 שמעיד שאין הבדל בכלל בין משך הנסיעה היומי החציוני בין גברים לנשים, עם-P-value של 1.000 שמציין שההבדלים בין המינים (אם קיימים) אינם מובהקים סטטיסטית, כלומר אין כל עדות לכך שמשך הנסיעה מושפע ממגדר. מבחן T בין המינים עבור מספר נסיעות נתן ערך גבוה (28.279) שמצביע על הבדל משמעותי מאוד בין גברים לנשים במספר הנסיעות היומיות עם -P-value של 0.000 שמאשר שההבדל מובהק סטטיסטית. קיבלנו ממצא סותר בעוד שמספר הנסיעות היומי מושפע ממגדר בצורה מובהקת, משך הנסיעה החציוני אינו תלוי במגדר כלל, דבר שפוגע בתוקף המבנה מאחר ושניהם משתנים אופרציונליים שבודקים הרגלי רכיבה, אבל מחזק את הבחירה שלנו במשך נסיעה חציוני מאחר ואינו מושפע ממשתנה כמו מגדר. ייתכן שמשך הנסיעה אינו משתנה רגיש מספיק כדי להבחין בין מגדרים וניתן במחקר המשך לשקול איסוף מדדים אופרציונליים נוספים (כמו סוג המסלול, מהירות נסיעה).

**מהימנות:**  **מהימנות של retest-Test** אינה ניתנת להערכה מאחר ואין מידע על בעל המנוי כמשתמש ייחודי אלא הבחנה בין תצפיות של רוכבים בעלי מנוי לבין אלו שאינם בעלי מנוי. אנו מניחים שיש אופי נסיעות שונות בין מנויים לבין למשל מבקרים ותיירים ולכן בחרנו להתמקד רק באוכלוסיית המנויים . \*במידה וקיים מידע זה שאינו נגיש לציבור אך קיים בחברה, ניתן לבדוק זאת ואז אנו מעריכים שהמשתמש כן יחזור על ההתנהגות שלו אם היא מושפעת מתנאי מזג האויר. במובן מסוים בדקנו מהימנות חוזרת של התוצאות בכך שלקחנו מדגם ממספר שנים כך שיש ייצוג לעונות ולא רק משנה אחת, וראינו שהמגמה יציבה. עבור שיוך מדידה ביום מסוים לעונה, יש מהימנות כי בכל שנה התאריך ישויך לאותה עונה, מאחר וזה שיוך קבוע. כאן יש הסתייגות כי יתכן שהימים ישתנו מבחינת נוחות הנסיעה (ימים בחורף שלא יהיו בהם גשם, או ימים נעימים יותר בקיץ) ואנו מצהירים שזו מגבלה במחקר. **מהימנות כלי המדידה גבוהה**, היות שהנתונים נאספים בצורה עקבית באמצעות מערכות דיגיטליות שמבוקרות ומכיולות. ניתן לראות שהערכים לאורך השנים ולאורך השנה גם סבירים והגיוניים. לא ניתן לחשב מדד אלפא כי המדידות נעשו בכלי אחד בלבד. אבל כן בדקנו מהימנות פנימית לנתונים על ידי חישוב מתאם בין משתני המדד (משך נסיעה חציוני) בין ימים סמוכים כדי לבדוק אם יש יציבות בערכים וקיבלנו 0.733 , מתאם חיובי חזק למדי (קרוב ל-1), שמעיד על כך שיש יציבות ברורה בין ערכי המדד בימים סמוכים. המהימנות בין שופטים גבוהה מאחר ואין פה התערבות אנושית אלא מדידת זמן נסיעה ומדידת מועד הנסיעה בהקשר העונתי על ידי אמצעים דיגיטילים. נסתייג ונאמר שתנאי מזג אויר יכולים להשתנות במהלך היום וגם בין יום ליום במהלך העונה, דבר שלא בא לידי ביטוי כאן ועלול להשפיע על מהימנות הצגת האקלים מאחר ואנו תלויים בייצוג העונה בלבד. כמו כן יתכן שמזג אויר "קשה" לאחד לא יהיה "קשה" לשני, אך זה עניין למחקר אחר.

**3. תוצאות**

בחנו את השערתנו באמצעות ניתוח סטטיסטי של אנובה שבודק האם יש הבדל מובהק סטטיסטית בין העונות מבחינת המשתנה התלוי - משך הנסיעה היומי החציוני. קיבלנו תוצאה של F-statistic: 315.275 שזה ערך גבוה שמעיד שיש שונות בין העונות, ושהעונות מסבירות חלק ניכר מהשונות בנתונים, עם P-value: 1.069e-157 נמוך מאוד דבר שאומר שהתוצאה מובהקת, ושיש הבדל מובהק סטטיסטית בין משכי הנסיעה היומיים החציוניים בעונות השונות. תוצאות ה-Boxplot מראות את ההבדלים במשכי הנסיעה החציוני היומי בחלוקה לעונות, ולפיו שמשך הנסיעה החציוני בקיץ (600) ארוך יותר באופן ברור בהשוואה לחורף (504). נראה שכאשר מזג האוויר נעים יותר, המנויים נוטים לרכוב לפרקי זמן ארוכים יותר. נציין כי לסתיו (569) משך נסיעה חציוני גבוה יותר מבאביב (509), כי בניגוד לאסוציאציה שלנו דווקא בסתיו יש טמפרטורות גבוהות יותר מבאביב ויורד פחות גשם ולכן הוא נעים יותר לרכיבה. דבר זה מתיישב למעשה עם ההשערה שלנו שבתנאי אקלים נוחים ירכבו לאורך זמן ממושך יותר. השונות של התוצאות נמוכה יותר בקיץ ובחורף שנחשבים קיצונים יותר בתנאי האקלים גם.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **q75** | **q50** | **q25** | **Std** | **Mean** | **season\_label** |
| 593.25 | 569.5 | 536.75 | 40.61 | 565.19 | Fall |
| 599 | 552 | 509.75 | 54.26 | 555.56 | Spring |
| 615 | 600 | 582 | 31.61 | 599.85 | Summer |
| 523.63 | 504 | 487 | 34.21 | 506.97 | Winter |

**4. מסקנות.**

1. ניתן להסיק שלעונה יש קשר למשך הנסיעה החציוני היומי. במונחי ההשערה התאורטית שלנו לאקלים יש קשר להרגלי הרכיבה. ראינו זאת דרך המובהקות שנצפית בערך משך הנסיעה בכל עונה, אך זה בא לידי ביטוי גם במספר הנסיעות היומיות בכל עונה, כפי שראינו בבדיקת התוקף ומציאת מתאם גבוה בין ערך זה לעונה. תוצאה זו נתמכת גם במחקרים שהראו שלתנאי אקלים השפעה על הרכיבה באופניים (Nankervis, M. 1999; Chen,et al 2022) .

2. אנו מאמינים שהתוצאות שלנו תקפות ומהימנות. ראשית הן מתיישבות עם ההיגיון התאורטי על התנהגות המשתמשים שבמזג אויר נעים ירכבו יותר ולכן לתוצאות תוקף פנים מכיוון שמבטאות את הקשר שציפינו לראות בין תנאי מזג אויר נעימים דרך עונות השנה לבין משך אורך הנסיעה. יש להן תוקף ניבוי כי איששו את ההנחה שלנו במבחן סטטיסטי אנובה שנועד לבחון הבדלים בין קבוצות (עונות) - המבחן יצא מובהק כך שזו לא מקריות. ערכי ה-Boxplot מצביעים על מגמה יציבה וברורה של משך נסיעה חציוני שונה בין עונות השנה שמצביעה על מהימנות פנימית. ההנחה שלנו התייחסה לאביב וקיץ לעומת סתיו וחורף אך לאחר בדיקה גילינו שתנאי מזג האויר בסתיו נעימים יותר ביחס לאביב ולכן תוצאות אלו מתיישבות עם ההנחה על הקשר בין תנאים נעימים לרכיבה. המתאם בין העונות למשך הנסיעה היומי החציוני עונה על ההשערה התאורטית לגבי השפעת מזג האוויר (אקלים) על הרגלי רכיבה. לתוצאה תוקף חיצוני כי הנתונים שבחרנו מתמקדים רק במשתמשים מנויים, שהם קהל מייצג לאנשים שנוסעים בקביעות על אופניים ונחשבת קבוצה יציבה יחסית מבחינת דפוסי ההתנהגות שלהם (פחות מושפעת מגורמים כמו תיירות). לכן גם סביר להניח שהתוצאות נכונות גם עבור אוכלוסיות דומות בערים אחרות, בתנאי אקלים דומים. התוקף קיים גם בשל השימוש במדידה עקבית של המשתנים (עונה, משך הנסיעה), והשימוש בנתונים יומיים חציוניים (ולא ממוצעים) שמפחית את ההשפעה של ערכים קיצוניים. יחד עם זאת נסייג ונאמר כי אקלים אינו בהכרח משתנה משפיע ישיר על הרגלי הנסיעה, ויתכן שיש משתנים מסבירים אחרים כמו השעה או היום בשבוע (יום עבודה לעומת סופ"ש), חגים, מצב התשתיות בדרכים שמושפע מהאקלים (כבישים חלקים שמובילים לנסיעות קצרות יותר ובחירה באמצעי תחבורה אחרים לנסיעות ממושכות יותר בחורף. בקיץ בגלל שחם במטרו אולי מעדיפים לנסוע באויר הצח בחוץ). יש גם תרבות הנהיגה - בניו יורק פחות רגילים לרכוב בגשם (לעומת מדינות כמו דנמרק). בנוסף יש עניין פרסונלי של תפיסת מזג אויר "קר" לעומת חם, אשר יכול להיות סוביקטיבי פר משתמש.

3. להרחבת מסקנות היינו מוסיפים אלמנטים של טמפרטורה וכמות משקעים במידה והייתה לנו גישה למקורות מידע נוספים אקלימיים כדי לדייק את ההערכה של האקלים היומי. בנוסף היינו בודקים סוג מסלול הנסיעה (נתיב אופניים מובנה או לא בנתיב ייעודי). היינו בודקים כיצד הרחבת תשתיות אופניים לאורך השנים קשורה לשימוש בהן בעונות לאורך השנים. יש לציין שהמחקר מתבסס על נתוני השכרת אופניים של חברה אחת (אף כי היא הנפוצה) וטוב היה להוסיף נתונים מחברות אחרות, או אם היה ניתן לבצע מדידות על בעלי אופניים ולא רק שוכרים (אולי דרך הטלפון, אפליקציה ייעודית).

4. עבור משך הנסיעה היינו מודדים גם מהירות הנסיעה ואורכה במספר הק"מ בנסיעה, מאחר ויתכן ונראה קשר בין זה לאקלים. למשל נסיעה יכולה להיות איטית יותר במזג אויר נעים כדי להנות מהנוף או איטית יותר בתנאי מזג אויר קשים כי אנו נזהרים יותר ברכיבה כשהכביש רטוב. בנוסף היינו מודדים את הטמפרטורה בזמן הנסיעה לדיוק תנאי האקלים בזמן הנסיעה ולאו דווקא במהלך היום (יכול להיות נקודתית קר באותה נסיעה ואח"כ נעים). היינו שומרים זיהוי משתמש כדי שנוכל לבחון דפוסי התנהגות של פרט ולא רק אוכלוסיות.

**ביבליוגרפיה**

Chen, Y., Zhang, Y., Coffman, D. M., & Mi, Z. (2022). An environmental benefit analysis of bike sharing in New York City. Cities, 121, 103475.‏

Ferdinands JM, Crawford CA, Greenwald R, Van Sickle D, Hunter E &Teague WG. (2008). Breath acidification in adolescent runners exposed to atmospheric pollution: a prospective, repeated measures observational study. Environ Health. 7;7:10. doi: 10.1186/1476-069X-7-10. PMID: 18328105; PMCID: PMC2292713.

Hannah Fierick & Matt Troutman, (2024, July 17). NYC subways melt down, stranding sweaty straphangers with delays amid 100-degree heat wave. *New York Post*. <https://nypost.com/2024/07/17/us-news/nyc-subways-melt-down-stranding-sweaty-straphangers-with-delays-amid-100-degree-heat-wave/?utm_source=chatgpt.com>

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. Transportation Research Part A: Policy and Practice, 33(6), 417-431.‏

Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., & Skov-Petersen, H. (2013). Environmental correlates of cycling: evaluating urban form and location effects based on Danish micro-data. *Transportation Research Part D: Transport and Environment*, 22, 40–44. doi:10.1016/j.trd.2013.02.017

Pucher, J., & Buehler, R. (2008). Making cycling irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 28(4), 495-528. doi:10.1080/01441640701806612.

Sabir, M., Bertolini, L., & Te Brömmelstroet, M. (2014). Weather and travel behaviour: Empirical evidence from revealed and stated preference studies. *Journal of Transport Geography*, 38, 38–47. doi:10.1016/j.jtrangeo.2014.05.003.

Weichenthal, S., Kulka, R., Dubeau, A., Martin, C., Wang, D., & Dales, R. (2011). Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environmental health perspectives*, *119*(10), 1373-1378.‏