A black background with white text

Description automatically generated

**מפענחים את הרשת: ניתוח דינמי של שיתוף אופניים בשיקגו**

עבודת סיכום קורס

שם הקורס: אלגוריתמים ברשתות

מספר הקורס: 350028

שנה: תשפ"ד סמסטר: ב'

שם המרצה: ד"ר אבנר פריאל

שמות התלמידים: נתן יצחק פור, כאלד כלף, לירז ליזי כהן

מספרי תעודות זהות: 211412986, 207966607, 318865854

תאריך הגשה: 12/08/2024

# הלינקים המבוקשים

ל-GitHub של הפרויקט לחץ [כאן](https://github.com/NatiIPReact/NetworkAlgorithms)

ל-GitHub המכיל את הדאטה לחץ [כאן](https://github.com/NatiIPReact/DivvyTripsData)

למצגת הפרויקט לחץ [כאן](https://www.canva.com/design/DAGIZnZdOH4/MD-c2Afwwl-b_BNosiSxLA/edit?utm_content=DAGIZnZdOH4&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

# מבוא

אופניים שיתופיים הוא שירות המאפשר לשכור אופניים לרכיבה זולה ונוחה בעיר. האופניים מפוזרות בתחנות ייעודיות ברחבי העיר, וניתן לשכור אותם, בתשלום, ולהחזיר אותם בכל תחנה אחרת. בשנים האחרונות חברות המציעות שכירות אופניים בערים מרכזיות בעולם הפכו ליותר ויותר פופולריות, כאשר ב-2021 היו בערך כ-1900 מערכות כאלו ברחבי העולם. במאמר הקרוב ניסינו להשתמש בשיטות ניתוח רשתות על מנת ללמוד דברים חדשים על רשת האופניים הציבורית שבשיקגו. בשיטה אנו מנתחים באופן דינמי ולאורך זמן את רשת האופניים ואת הקשרים בין התחנות, ובכך ננסה ללמוד על הרשת כולה ולהפיק מסקנות ותוצאות שלא היו ידועות עד כה, שבעזרתן אולי נוכל לעזור למקבלי ההחלטות והחברות המסחריות להחליט על הצעדים הבאים שלהם בעולם הקשור ברשת האופניים הציבורית.

כלומר, מטרת המחקר שלנו היא לייעל ולשפר את רשת האופניים, ובכך נשדרג את איכות חווית הנוסעים ונגדיל את רווחי החברה.

# סיכום

כדי להבין טוב יותר את המאפיינים של מערכות שיתוף אופניים ולנסות להגיע למסקנות חדשות עליהן, השתמשנו בשיטות ניתוח רשתות כדי לנתח את הקשר בין תחנות בתוך מערכת שיתוף האופניים. ראשית, השתמשנו בנתונים של החברה “Divvy”, המפעילה שירות שכירות אופניים בשיקגו, ובספריה NetworkX על מנת לבנות רשת אופניים ציבורית. שנית, ניתחנו את הרשת, בנינו קהילות, והשווינו מדדי מרכזיות, צפיפות, ועוד, בין טווחי זמן שונים כדי להבין את היחסים בין התחנות של מערכת האופניים הציבורית. התוצאות הראו שניתן להשתמש בנתונים מהשטח, כגון כמה קרובה התחנה למקומות מרכזיים בעיר כמו תחנות רכבת, אטרקציות, ועוד, בשביל לדעת אם התחנה מרכזית או לא. יתרה מזאת, למדנו שהמצב האקונומי באזור התחנה משפיע על נסיעות הנוסעים בה. ראינו שהמשתמשים בדרך כלל נוסעים בין תחנות מאותו מצב סוציואקונומי, וכשזה לא קורה הם נוטים לנסוע לשכונות עשירות. למדנו גם שהמנויים לשירות לא בהכרח מנצלים את זוגות התחנות האפשריים וגם לא בהכרח נוסעים יותר מלקוחות מזדמנים, על אף שהם בדרך כלל משלמים יותר כסף. לבסוף, השתמשנו בתוצאות ובמסקנות שקיבלנו ובנינו בעזרתן מודל Random Forest העוזר לחזות את ערך ה-Betweenness של הצומת בהינתן כמה הצומת קרוב למקומות מרכזיים בעיר. ובנוסף, בנינו מודל Deep Learning, המשתמש ברשת נוירונים שעוזר לבחור היכן עלינו למקם תחנות חדשות, בהינתן קרבתן הרצויה למקומות מרכזיים בעיר, וההכנסה החציונית המיועדת באזור התחנה. הכנו גם אפליקציה ידידותית למשתמש המשתמשת במודל זה ומראה את תוצאותיו על מפת העיר שיקגו.

# סקירת ספרות (עבודות קשורות)

מערכות האופניים השיתופיים הפכו למשומשים יותר בשנים האחרונות, למרות הפריחה הכלכלית במדינות המערב והחזקת הרכבים הפרטיים של אנשים. ככל שמספר הצרכנים גדל, כך גם רשת האופניים שבעיר גדלה, והרצון לחקור אותה, לצרכים כלכליים, שיפור חיי ובטיחות התושבים, הנדסה נכונה של העיר, ועוד. ניתוח רשתות (להלן, NA) הוא כלי מחקר רב עוצמה שיכול לספק תובנות חדשות לגבי התנהגות הרשת ולתת מסקנות על רשת האופניים והתנהגות הנוסעים. גישת NA מתמקדת בחקר קשרים ואינטרקציות בין יחידים בקבוצה, תוך התייחסות למבנה ולמאפיינים של הרשת החברתית שנוצרת, עם דגש לחשיבות הצמתים שבה. בכלליות, NA מבוסס על העובדה שהחיים החברתיים נוצרים על-ידי מערכות יחסים ודפוסים שלהם. למחקר רשתי יש שני סוגים של פלטים, ויזואלי ומתמטי. הפלט הויזואלי הוא עיבוד של דיאגרמת הרשת המציגה את הצמתים והקשרים ביניהם ובעזרתו ניתן לענות במהירות על שאלות לגבי מי מחובר למי ברשת. הפלט המתמטי הם מדדים המורכבים לרוב ממדדי מרכזיות של צומת, המזהים את חשיבות הצומת לתפקוד הכולל של הרשת, וניתוח קבוצות על-פי מדדים.[[1]](#footnote-1) לדוגמה, ניתן לחקור איך הכנסה משפיעה על המסלולים שבהם נוסעים המשתמשים. כל מסלול מהווה חיבור בין שתי תחנות, תחנת המקור ותחנת היעד, ומכיל תכונות כמו מספר הנסיעות בין שתי התחנות ומדדי הכנסה נוספים. וכך, אוסף של מסלולי אופניים יכולים להיות מוצגים בעזרת רשת, שאותה מנתחים בעזרת NA. מחקר שפורסם ב-sirs באוגוסט 2023 בדק את זה על רשת האופניים של החברה “Sitycleta” בלאס פאלמאס, ספרד. המחקר מצא כי לתחנות שנמצאות בשכונות עם הכנסה גבוהה יש סיכוי גבוה יותר להיות מחוברות ביניהן בעוד שלתחנות הנמצאות בשכונות עם הכנסה נמוכה יש סיכוי נמוך יותר להיות מחוברות לתחנות אחרות. למעשה, הסיכוי שקיימים מסלולים המחברים בין שתי תחנות בשכונות עשירות הוא גבוה יותר מפי 2 ממסלולים המחברים בין תחנות בשכונות עניות. ובנוסף, סיכוי לקיום מסלול המחבר בין שכונה עשירה לעניה גבוה ביותר מ-50% ממסלול המחבר בין שתי שכונות עניות. כלומר, ככל שיש במסלול יותר תחנות בשכונות עשירות, כך הסיכוי לקיום המסלול גבוה יותר.[[2]](#footnote-2) מחקר נוסף שהתבסס על NA שבדק את רשת האופניים הציבורית בעיר נאנג'ינג שבסין, ניסה להבין את השימוש של הנוסעים במערכות האופניים הללו ומצא שהנסיעות הכי נפוצות היו בין אזורי מגורים לתחנות הרכבת התחתית ולסופרמרקטים, בעוד שתחנות הממוקמות ליד תחנות הרכבת היו צמתים מרכזיים וחשובים לרשת האופניים הציבוריים. החוקרים שמו לב שהאופניים שירתו את הנוסעים הן לנסיעות קצרות והן לארוכות, וששימוש האופניים בתחנות קשורות גם לשימוש בתחנות הסמוכות להן.[[3]](#footnote-3) ניתוחי רשת התחבורה יכולים לעזור בפיתוח רשתות חדשות והוספתן לערים בהן לא קיימת מערכת שכירות אופניים. ניתוח דינמי של רשתות אלו יכול לעזור ולתת סיבות למקבלי ההחלטות להחליט היכן להוסיף תחנות חדשות, במיוחד ברשתות חדשות וקטנות. למעשה, ניתוח הרשת צריך להיות תנאי מוקדם לפני הוספות ושינויים לרשת, משום שכך נוכל לבדוק איך זה ישפיע על הרשת וחיי התושבים בעיר, עוד לפני שהתחנה נבנתה, ובכך למנוע טעויות קריטיות בעתיד.[[4]](#footnote-4) לסיכום, ניתוח רשתות הוא כלי עשיר היכול להפיק תובנות חדשות במהירות וביעילות. גישה המחקר הזו אמינה לאורך זמן, ויכולה לעזור לרשויות לבנות ולשפר רשתות תחבורה מתקדמות ומורכבות שישפרו את חיי התושבים בעיר.

# מתודולוגיה

השתמשנו ב-Dataset על נסיעות אופניים שיתופיים של החברה Divvy בשיקגו, שבאילינוי, ארצות הברית. את ה-Dataset מצאנו באתר Google Dataset וניתן להורידו מהאתר data.gov של ממשלת ארה"ב [בקישור](https://catalog.data.gov/dataset/divvy-trips). הדאטה מכיל 650 אלף רשומות המתפרסות על פני 370 ימים, והוא מכיל 18 עמודות הכוללות מידע על הנסיעה (מיקומים, זמנים), האופניים, והמשתמש. אנחנו הוספנו לדאטה המקורי נתונים נוספים שיעזרו לנו לענות על שאלות המחקר שעליהם נסביר בהמשך. השתמשנו בספרייה NetworkX המשמשת לחקירת רשתות על מנת לבנות מהדאטה שהבאנו חמש רשתות מכוונות (אחת לכל רבעון ואחת הכוללת את כל הדאטה ביחד), ולמענה על שאלות המחקר שכתבנו. בנינו רשת מכוונת ובשביל לבנות את הרשת בתוכנה, השתמשנו בשדות FROM STATION NAME (להלן, FSN) כ-source של הקשת – שהוא התחנה ממנה יצאה הנסיעה (שם נלקחו האופניים), ו-TO STATION NAME (להלן, TSN) כ-target של הקשת שהוא התחנה בה הסתיימה הנסיעה (לשם הוחזרו האופניים). כלומר, התחנות הן צמתי הרשת והקשתות הן הנסיעות שיוצאות מ-FSN ל-TSN, ולכן הקשתות מכוונות, כלומר זו רשת מכוונת. ומכאן, הרשת בנויה ומוכנה לניתוח, שיוסבר בהמשך.

# שאלות המחקר

במחקר זה נבדוק האם תחנות הקרובות למקומות מרכזיים בעיר הן תחנות מרכזיות ברשת, נבדוק האם ואיך המצב הסוציואקונומי של השכונה בה ממוקמת התחנה משפיע על נסיעות המשתמשים, ונבדוק האם בקהילות עם יחס גדול יותר של מנויים יש בהתאמה יותר נסיעות והאם הם "מנצלים" את זוגות התחנות באופן טוב יותר.

# הכנת הנתונים

קובץ הנתונים המקורי גדול מאוד ושוקל כמעט 5GB (אי אפשר לפתוח את כולו ב-Excel אפילו!) ולכן הדבר הראשון שעשינו הוא להוציא את הנסיעות שאנו משתמשים בהן לקובץ נפרד, וכך גודל הקובץ ירד ל-100MB ונשארנו עם 650 אלף הרשומות שאנו צריכים. מכאן, לא נותר הרבה לעשות על הדאטה המקורי חוץ מלהוסיף נתונים משום שהדאטה מסודר ואין בו ערכים ריקים (חוץ מעמודות לא נחוצות – שאותן מחקנו). לצורך מענה על שאלות המחקר הוספנו שש עמודות:

SourceArea, TargetArea – עמודות אלו שומרות את השכונה/אזור בעיר שיקגו בה תחנת המוצא/יעד נמצאת בהתאמה. בשביל להשיג את המידע הזה השתמשנו בנקודות הציון על התחנות וב-API של Google Maps. את נקודות הציון שלחנו ל-API הנ"ל וקיבלנו בחזרה את אזור התחנה בשיקגו. בעזרת ה-API בנינו מילון הממפה תחנה לאזור שלה ובעזרתו עשינו מיפוי של התחנות בכל נסיעה לאזורן.

SourceMedianIncome, TargetMedianIncome – עמודות אלו שומרות את השכר השנתי החציוני באזור תחנות המוצא/יעד בהתאמה. השתמשנו באזורים שמצאנו בשתי העמודות הקודמות ובנינו מילון הממפה כל אזור לשכר החציוני בו. בעזרת המילון עשינו מיפוי של האזורים בכל נסיעה לשכר החציוני בהן. את המידע הנחוץ מצאנו בחיפוש ידני בגוגל ובאתרים כמו Point2Homes.

SourceProximityIndex, TargetProximityIndex – אלו העמודות המייצגות את מדד הקרבה בתחנות המוצא/יעד. מדד הקרבה הוא מדד שאנחנו הגדרנו ובנינו והוא מייצג את המרחק בין התחנות למקומות מרכזיים בעיר. על מנת לחשב אותו קודם הגדרנו את המקומות בעיר שנחשבים חשובים (לפי חיפוש שעשינו בגוגל), ואז חישבנו את המרחק בין התחנות לכל המקומות החשובים שהגדרנו בעזרת נקודות הציון של המקומות והתחנות. מדד הקרבה של כל צומת הוא ערך המרחק המקסימלי בינה לבין ארבעת המקומות שהגדרנו כחשובים. בנוסף, השתמשנו ב-Min-Max Normalization למדד על מנת לקבל ערך בין 0 ל-1. בנינו מילון הממפה בין תחנות לערך מדד הקרבה שלהן ובעזרתו מיפינו תחנות לערכן בכל נסיעה.

מצורף קובץ "PreProcessing.ipynb" המכיל את הקוד להכנת הנתונים כולל הסברים מפורטים בתוכו.

# מהלך הפרויקט והערכת התוצאות

הדבר הראשון שעשינו לאחר בניית הרשתות שהוזכרו לעיל הוא לבדוק את הרשת מול גרפים רנדומיים שיצרנו בעזרת Watts-Strogatz ו-Erdos-Renyi. מצאנו שהרשת שלנו Small-World משום שה-Clustering Coefficient היה גבוה בהרבה מהגרפים הרנדומיים, וה-Average Path Length היה נמוך יותר. מבדיקה שעשינו עם התפלגות הדרגות, ראינו שההתפלגות היא לא Power law ולכן קבענו שהרשת היא לא Scale Free.

לפני שבדקנו את השערותינו, קודם בנינו והדפסנו מפת חום של התחנות בשיקגו שיעזרו לנו לקבל תחושה של הדאטה.

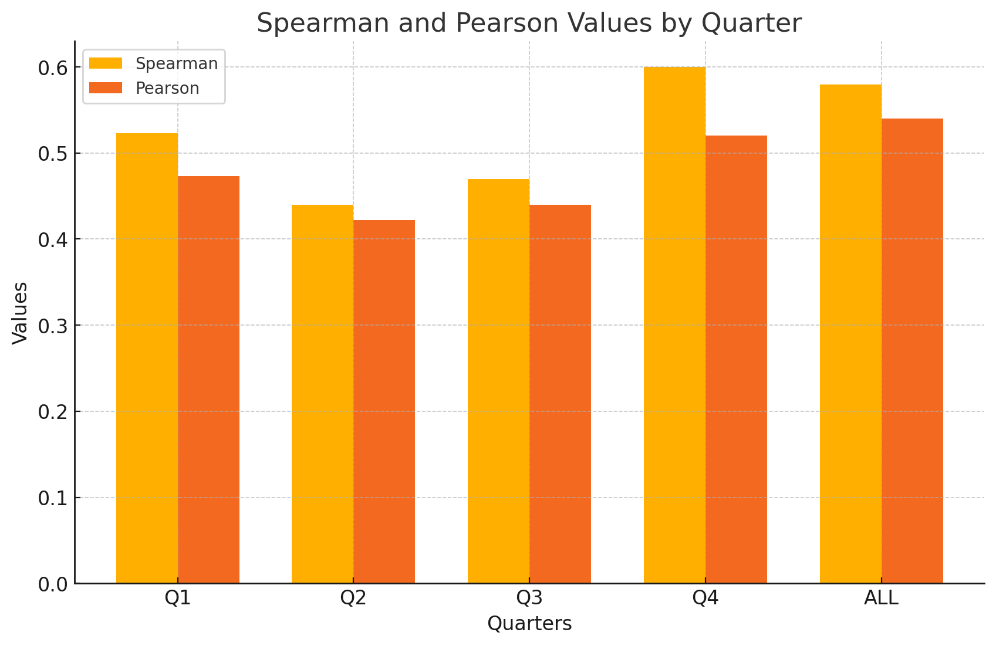


תרשים 1: מפת החום של תחנות האופניים בעיר שיקגו

ניתן לראות שהחום נמצא באזור מרכז העיר מכיוון ששם נמצאות הרבה תחנות, מה שעזר לנו לחשוב ששאלת המחקר הראשונה נכונה, שאותה נבדוק עכשיו.

## שאלת המחקר הראשונה – האם תחנות הקרובות למקומות מרכזיים בעיר הן תחנות מרכזיות ברשת

למענה על שאלה זו, אנו בודקים את המתאם בין מדד הקרבה ל-Betweenness של התחנות (הצמתים). נעשה זאת לכל אחת מהרשתות בנפרד וכך נוכל גם לראות אם יש הבדל בין רבעוני השנה. בדקנו מתאם מסוג פירסון (קשר לינארי) ומתאם ספירמן (קשר מונוטוני, לא בהכרח לינארי) וקיבלנו את התוצאות הבאות:



תרשים 2: תוצאות הקורלציה על כל הרשתות

בכל הרשתות קיבלנו בין 42% ל-60% (מתאם חיובי בינוני), עם p-value אפסי (לדוגמה 6\*10^-18) ולכן קבענו שההשערה שלנו נכונה, ואכן יש התאמה בין מדד הקרבה שבנינו לערך ה-Betweenness של הצמתים.

**כבונוס:** מכיוון שההשערה נכונה נוכל לחזות את חשיבות הצומת ברשת (ה-Betweenness של הצומת) בעזרת מדד הקרבה, מה שיכול לעזור לחברה להבין היכן שווה לה להוסיף תחנות חדשות לפני שבונים את התחנה ומגלים שזו הייתה טעות. השתמשנו במודל Random Forest על מנת לעשות זאת, כאשר האימון נעשה על הרבעון הרביעי של השנה והמבחן בעזרת הרשת השלמה.

A graph with red and blue dots

Description automatically generated

תרשים 3: תוצאות מודל ה-Random Forest שלנו

ניתן לראות שערכי ה-Predictions קרובים לערכים האמיתיים, מה שמראה שהמודל טוב, וניתן לאשש זאת בעזרת המדדים: MAE שנתן 0.0006, MSE שנתן 0 ו-R2 שנתן 0.7641 שמראים שהמודל טוב.

## שאלת המחקר השניה – האם ואיך המצב הסוציואקונומי של השכונה בה ממוקמת התחנה משפיע על נסיעות המשתמשים

מצאנו מחקר דומה לשלנו, עם רשת אופניים בספרד שבו החוקרים מצאו כי הסיכוי ששתי שכונות יהיו מחוברות ברשת עולה ככל שרמת העושר בהן גדולה יותר. ובנוסף, למסלול המחבר שתי שכונות עשירות יש סיכוי גבוהה פי שניים להיות קיים ברשת ממסלול המחבר שתי שכונות עניות.[[5]](#footnote-5) מכאן קיבלנו השראה לכתיבת שאלת מחקר דומה שאותה נבדוק עכשיו. בשאלה זו ובשאלה הבאה השתמשנו בקהילות כחלק מבדיקת ההשערה ולכן הדבר הראשון שעשינו הוא למצוא חלוקה טובה של הרשת לקהילות. ניסינו את כל הדרכים שנלמדנו בכיתה לחלוקה לקהילות כמו Louvain, Girvan-Newman, ועוד (ראה קוד להסבר המלא), ובדקנו כל חלוקה כזו בעזרת מדד המודולריות. החלוקה הטובה הביותר נוצרה בעזרת האלגוריתם Louvain עם מודולריות של 40% ולכן איתה המשכנו לשאלת המחקר הזו וגם לבאה. לבדיקת שאלת המחקר, קודם כל חילקנו את התחנות לקטגוריות לפי ההכנסה החציונית בשכונתן (בעזרת העמודות SourceMedianIncome ו-TargetMedianIncome) כאשר תחנות עם פחות מ-50,450$ של הכנסה חציונית שנתית יחשבו לעניות, תחנות עם יותר מ-121,080$ יחשבו לעשירות, וכל התחנות שהכנסתן בין שני סכומים אלו יחשבו למעמד הביניים. המספרים נבחרו לפי נתונים שמצאנו באתר של עיריית שיקגו. לאחר מכן, חילקנו את הקהילות לקטגוריות הכנסה לפי אחוז התחנות המקסימלי של אותה קטגוריה. לדוגמה, אם 40% מהתחנות בקהילה נחשבות לעניות, 30% לעשירות, ו-30% לביניים, הקהילה תיחשב כקהילה עניה. בעזרת החלוקה הזו בנינו Contingency Table הסופרת את מספר הנסיעות שיצאו והגיעו מכל קטגוריה של תחנה.

A blue squares with numbers and a number on it

Description automatically generated with medium confidence

תרשים 4: טבלת השכיחויות של רשת הרבעון השני

השתמשנו בטבלה זו על מנת לבצע מבחן חי-בריבוע הבודק אם יש קשר בין תחנות היציאה לתחנות היעד. בכל הרשתות (רשתות ארבעת הרבעונים והרשת המכילה את הנסיעות) קיבלנו שאכן יש קשר מובהק עם p-value שקרוב לאפס, ולכן נראה שהשערתינו נכונה. לצורך בדיקה נוספת ולהיות בטוחים שאין לנו הטיה בגלל נסיעות שיוצאות ונכנסות לאותה קהילה, בדקנו את אותו הדבר שוב לכל הרשתות אך הפעם רק לנסיעות בין קהילות (להלן, Inter-community trips). ושוב, לפי מבחן חי-בריבוע קיבלנו קשר מובהק בין תחנות המקור לתחנות היעד. ניתן גם לראות שברוב טבלאות השכיחויות שקיבלנו יש הרבה יותר נסיעות בין קהילות מאותו סוג, או שלפחות אחד מתחנות הנסיעה נחשב לעשיר. ולכן, קבענו שכמו במחקר שקראנו, גם כאן להכנסה באזורי התחנה יש השפעה על נסיעות הנוסעים. נראה שיש יותר נסיעות למקומות שנחשבים לעשירים, או מקומות מאותו סוג של הכנסה.

## שאלת המחקר השלישית – האם בקהילות עם יחס גדול יותר של מנויים יש יותר נסיעות והאם המנויים "מנצלים" את זוגות התחנות באופן טוב יותר

למענה על שאלה זו, השתמשנו בחלוקת Louvain כדי לחלק את הרשת לקהילות (משום זו החלוקה הטובה ביותר שמצאנו), חישבנו את אחוז נסיעות המנויים בכל קהילה, וניסינו להשוות את אחוז זה לצפיפות (Density) של כל קהילה, ומספר הנסיעות בקהילה. עשינו ויזואליזציה של המידע לכל רשת בעזרת Bar Charts (ראו במחברת הפרויקט), ואז חישבנו ובדקנו מתאם בין אחוז המנויים בקהילה למידע הנ"ל.



תרשים 5: תוצאות הקורלציה שקיבלנו ברשתות השונות

ניתן לראות שאין מתאם מובהק עם מדד הצפיפות של הקהילות משום שבחלק מהרשתות יש מתאם חיובי ובחלק שלילי. לעומת זאת, בכל הרשתות קיבלנו מתאם חיובי בינוני/חלש (13%-59%) בין מספר הנסיעות ליחס המנויים בקהילות. התוצאות שקיבלנו לא מספיקות בכדי לדחות את השערת האפס, ומכאן ניתן לקבוע שהשערתנו לא נכונה, המנויים לא "מנצלים" את זוגות התחנות באופן טוב יותר כי הצפיפות בקהילות המכילות יותר מנויים לא גדולה יותר, ובנוסף אין מספיק מידע שהמנויים נוסעים יותר (למרות שכן יש כמה רשתות שתומכות בזה, יכול להיות שעם נסיעות נוספות או רשת אופניים שונה כן היינו דוחים את השערת האפס).

# מסקנות

מסיכום שאלות המחקר, מצאנו מסקנות חדשות ומעניינות על רשת האופניים בהם תכף נשתמש ליצירת רשת נוירונים התמליץ היכן למקם תחנות חדשות. ראשית, גילינו שתחנות הקרובות למקומות מרכזיים בעיר הן תחנות מרכזיות ברשת, והחברה יכולה להשתמש במידע זה בשביל לחזות אם תחנות חדשות שתוסיף יהיו תחנות מרכזיות ברשת, מה שיכול לעזור לה לעשות החלטות נכונות יותר. שנית, ראינו שכמו במחקר שקראנו, גם אצלנו המצב הסוציואקונומי באזור שבו התחנות ממוקמות משפיע על נסיעות המשתמשים. גילינו שיש נטייה לשכונות מאותו מצב סוציואקונומי להיות עם יותר נסיעות ביניהן משכונות עם הכנסה שונה, וראינו שקיימים יותר מסלולים בין תחנות בשכונות ממצב אקונומי שונה כשלפחות אחד מתחנות המסלול ממוקמת בשכונה עשירה.

# בונוס – מודל Deep Learning הממליץ על אזורים לתחנות חדשות

בנוסף למודל ה-Random Forest שהוספנו כבונוס בשאלת המחקר הראשונה, בנינו מודל Deep Learning המשתמש ברשת נוירונים בשביל להמליץ על מיקומים לתחנות חדשות, ובנינו לו גם אפליקציה ידידותית למשתמש התסמן את ההמלצות על מפת העיר שיקגו. את המודל וכל קבציו צירפנו בזיפ הנקרא “Model.zip”. משום שדחינו את השערת האפס בשאלה הראשונה והשניה, נוכל להשתמש במסקנות שקיבלנו משאלות אלו, ולכן האפליקציה מבקשת מהמשתמש להכניס ערך מדד קרבה (ProximityIndex) רצוי, והכנסה חציונית שנתית בשכונת התחנה. האפליקציה תכניס את הנתונים האלו למודל והוא ימליץ על אזור למיקום התחנה החדשה. המודל שקיבלנו נתן דיוק של 80%. ראו את בניית המודל במחברת “DLModel.ipynb”, המכילה גם תיעוד של הקוד עם הסברים מפורטים יותר.

A map with a blue circle

Description automatically generated

תרשים 6: המלצת המודל עם ProximityIndex של 90% והכנסה חציונית שנתית של 120,000$

A map with a blue circle

Description automatically generated

תרשים 7: המלצת המודל עם ProximityIndex של 25% והכנסה חציונית שנתית של 40,000$

A map of a city

Description automatically generated

תרשים 8: המלצת המודל עם ProximityIndex של 60% והכנסה חציונית שנתית של 60,000$

להרצת האפליקציה הוציאו את קבצי המודל מהזיפ והריצו את הקובץ app.py. כמובן שצריך ש-Python יהיה מותקן במחשב שלכם עם ספריות הקוד. להתקנת הספריות יש להכניס את הפקודה הבאה בטרמינל (כי זו לא מחברת ובאפליקציה אין אפשרות להכניס את זה כחלק מהקוד):

pip install Flask folium pandas numpy tensorflow scikit-learn pickle5

לאחר ההרצה של הקובץ יפתח שרת לוקלי שמריץ את האפליקציה. פתחו את כתובת השרת שכתובה בחלון ה-CMD שנפתח.

כתובת זו היא בדר"כ 127.0.0.1:5000 אבל יכול להיות שפורט זה תפוס אצלכם אז חשוב להסתכל בחלון השרת.

A screenshot of a computer screen

Description automatically generated

תרשים 9: כתובת השרת הכתובה בחלון שנפתח בהרצת האפליקציה

# תרומת חברי הצוות

נתן יצחק פור: חיפוש נתונים, עיבוד מקדים, שאלת המחקר הראשונה, שאלת המחקר השניה, שאלת המחקר השלישית, כתיבת הדו"ח, כתיבת הקוד, מודל ה-Random Forest, מודל ה-Deep Learning, הכנת המצגת

כאלד כלף: שאלת המחקר השניה, מודל ה-Deep Learning

לירז ליזי כהן: חיפוש נתונים

**תודה רבה לכם!**

1. Berlusconi, Giulia. (2017). Social Network Analysis and Crime Prevention. 10.1007/978-3-319-27793-6\_10. [↑](#footnote-ref-1)
2. Suárez-Vega, Rafael, Yolanda Santana-Jiménez, Juan M. Hernández, and José Juan Santana-Figueroa. 2023. "Assessment of the Bike-Sharing Socioeconomic Equity in the Use of Routes" ISPRS International Journal of Geo-Information 12, no. 8: 320. https://doi.org/10.3390/ijgi12080320 [↑](#footnote-ref-2)
3. Yao, Yi, Yifang Zhang, Lixin Tian, Nianxing Zhou, Zhilin Li, and Minggang Wang. 2019. "Analysis of Network Structure of Urban Bike-Sharing System: A Case Study Based on Real-Time Data of a Public Bicycle System" Sustainability 11, no. 19: 5425. https://doi.org/10.3390/su11195425 [↑](#footnote-ref-3)
4. Dobrzyńska, Ewa & Dobrzyński, Maciej. (2016). Structure and dynamics of a public bike-sharing system. Case study of the public transport system in Białystok. Ekonomia i Zarzadzanie. 8. 10.1515/emj-2016-0033. [↑](#footnote-ref-4)
5. Suárez-Vega, Rafael, Yolanda Santana-Jiménez, Juan M. Hernández, and José Juan Santana-Figueroa. 2023. "Assessment of the Bike-Sharing Socioeconomic Equity in the Use of Routes" ISPRS International Journal of Geo-Information 12, no. 8: 320. https://doi.org/10.3390/ijgi12080320 [↑](#footnote-ref-5)