

ויזואליזציה של מידע: נסועה בקווי אוטובוס

מגיש: שי גטניר 204617062

הקדמה:

בעבודה זאת נעסוק בתיאור גרפי של קבצי נסועה בקווי אוטובוס במדינת ישראל בין השנים 2022-2024. קובץ נתונים זה, מפורסם ע"י משרד התחבורה ומדעכן אחת לרבעון. הקובץ מתעד את קווי האוטובוס השונים ברמה שבועית במצב שגרה. מכיל תיאור גיאוגרפי של הקו, ממוצע שבועי של נוסעים בפילוחים שונים, עלויות ורווחים ממוצעים.

בעבודה ננסה להשוות בין חברות האוטובוסים השונות וננסה למפות התרחשויות בתחום האוטובוסים ברחבי המדינה.

תיאור נתונים

קבצי הנסועה בקווי אוטובוס זמינים להורדה ב[אתר הממשלתי](#) ומחולקים לשלושה קבצים לפי שנים. המידע הזמין בקבצים אלו הינו מהרבעון האחרון של שנת 2022, הרבעון הראשון והשני של שנת 2023 והרבעון הראשון של שנת 2024. הנתונים נאספים ע"י תיעוד התיקופים באוטובוסים ומכיל את המידע על כלל הקווים בארץ ישראל. מידע זה לא נמדד בזמני חירום ולכן שנת 2023 לא תועדה במלואה.

בשנת 2023 נוספו השדות "מהירות ממוצעת", "משך הנסיעה הממוצע" ו"עלות תפעול ממוצעת" שלא היו קיימים בקובץ של שנת 2022.

מטרת הנתונים הינה לעקוב אחרי שימוש תפקוד ועלויות האוטובוסים במדינת ישראל.

קובץ תיאורי של כלל המשתנים מצורף לעבודה זו. הקובץ מכיל את תיאור המשתנים כפי שמופיע בתיאור של משרד התחבורה. בנוסף מכיל את השדות: ממוצע, סטיית תקן, חציון, ערך מקסימלי ומינימלי.

בנוסף לקובץ הנתונים המקורי. נעשה שימוש גם בקובץ נתונים גיאוגרפיים של הערים בישראל.

השיטה

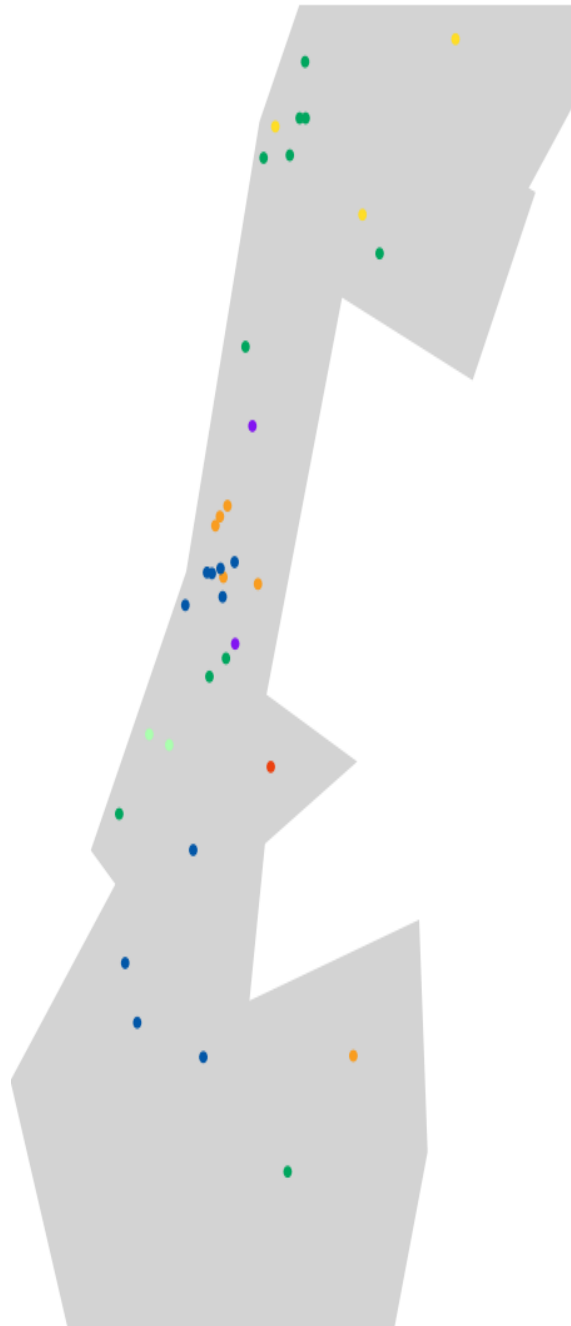
בעבודה זו נעשה שימוש בשפת התכנות Python. נעשה שימוש בספריות pandas, numpy לעיבוד הנתונים. matplotlib, plotly, geopandas כדי לבצע את הויזואליזציות ו-bidi כדי להתמודד עם השפה העברית.

חברות האוטובוסים בישראל ותרבות הנסיעה באוטובוס

המוטיבציה של עבודה זו הינה מצד אחד להשוות בין תפקוד ונתח השוק של חברות האוטובוסים השונות במדינה. ומצד שני, להשוות ולהסתכל על ערים נבחרות בישראל ולנסות להסתכל על ההבדלים בין הנוסעים בהן. את הגרפים ננתח בעזרת המודל של מונזנה.

תוכנית החלוקה

חברת האוטובוסים לה ממוצע הנוסעים הגבוה ביותר בכל עיר



מה?

בגרף, ניתן לראות גאומטריה של מפת ישראל. המיקומים הנבחרים הם ערים אשר הגאומטריה שלהן זמינה בקובץ הערים המצורף. לכל עיר חושבה התכונה: שם חברת האוטובוסים, לה ממוצע הנוסעים הגבוה ביותר בקווים היוצאים ממנה בתוך טבלת הנתונים של הנסועה באוטובוסים.

למה?

הגרף מציג את הטופולוגיה של חברות האוטובוסים על פני המפה. גרף זה נותן לקורא את האפשרות להבין את חלוקת ה"שליטה" של חברות האוטובוסים על פני המפה.

איך?

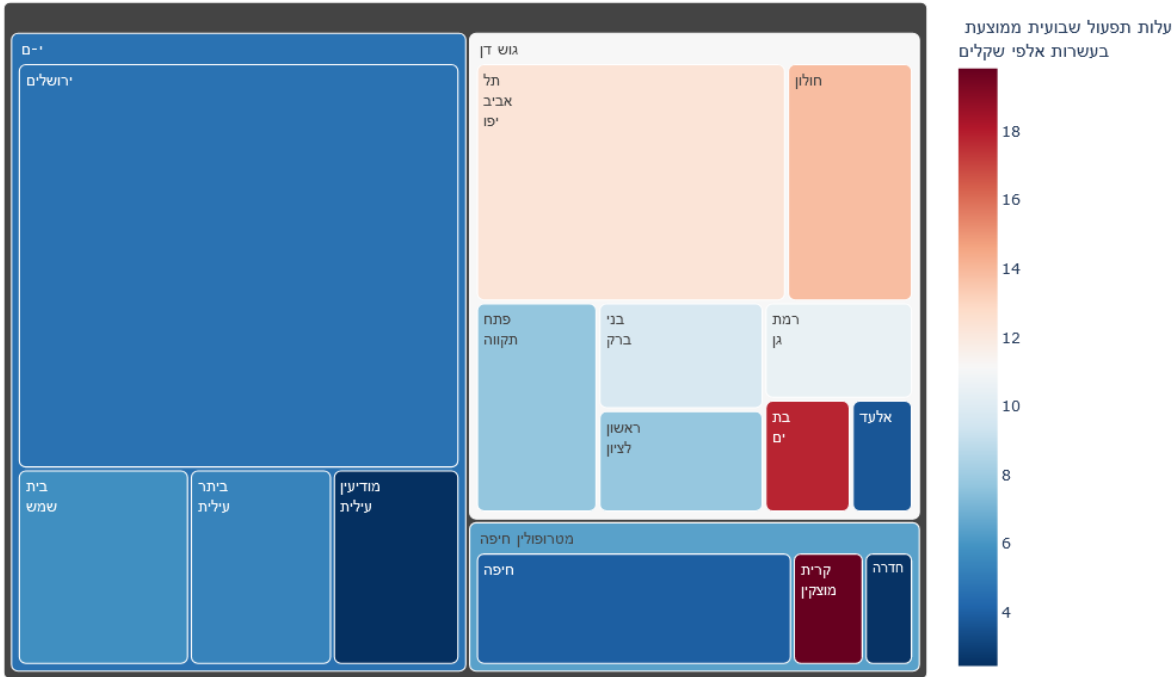
- סינון של תת חברות כמו אגד תעבורה ודן באר שבע והכללה שלהן בתוך חברות האב שלהן. נעשה כדי להקל על העומס בגרף.
- ציור גבולות ארץ ישראל בעזרת

geopandas

- קיבוץ הנתונים לפי עיר וחברה. סכימה של מספר הנוסעים הממוצע.
- איחוד של נתוני הצבעים והקורדינטות על הגרף וזריקת הערים שלא נמצאו בקובץ הערים.
- סינון של 35 הערים להן הממוצע הגבוה ביותר. נעשה כדי להקל על העומס.
- קידוד הצבעים נעשה בהתאם לצבעים המייצגים את החברות השונות. פרט לסופרבוס ואלקטרה אפיקים שצבע האוטובוסים והלוגו שלהן הוא כחול ומתנגש עם הצבע של דן.
- הוספת tooltip. ניתן להצביע עם הסמן על עיר ולראות את שם העיר ואת שם החברה. בנוסף, ניתן לגלול פנימה והחוצה (zoom) בעזרת הגלגלת. שני פיצ'רים אלו הוספו בשל עומס המידע. שני אלמנטים אלו מאפשרים הבחנה טובה בין הנקודות השונות.

עלות התפעול בבת ים וקרית מוצקין: כפול מכל השאר

חמישה עשר הערים עם עלויות התפעול הגבוהות ביותר מחולקות לפי מטרופולין. הגדלים מייצגים את ממוצע הנוסעים השבועי בעיר



מה?

בגרף זה ניתן לראות Tree Map של הערים השונות בחלוקה למטרופולינים השונים. בנוסף, ניתן לראות את התכונות:

- 1) בגודל: ממוצע הנוסעים השבועי בעיר ובמטרופולין.
- 2) בצבע: עלויות ממוצעת לקו אוטובוסים גם עבור העיר וגם עבור המטרופולין.

נתונים אלו זמינים בקובץ הנסועה של שנת 2024.

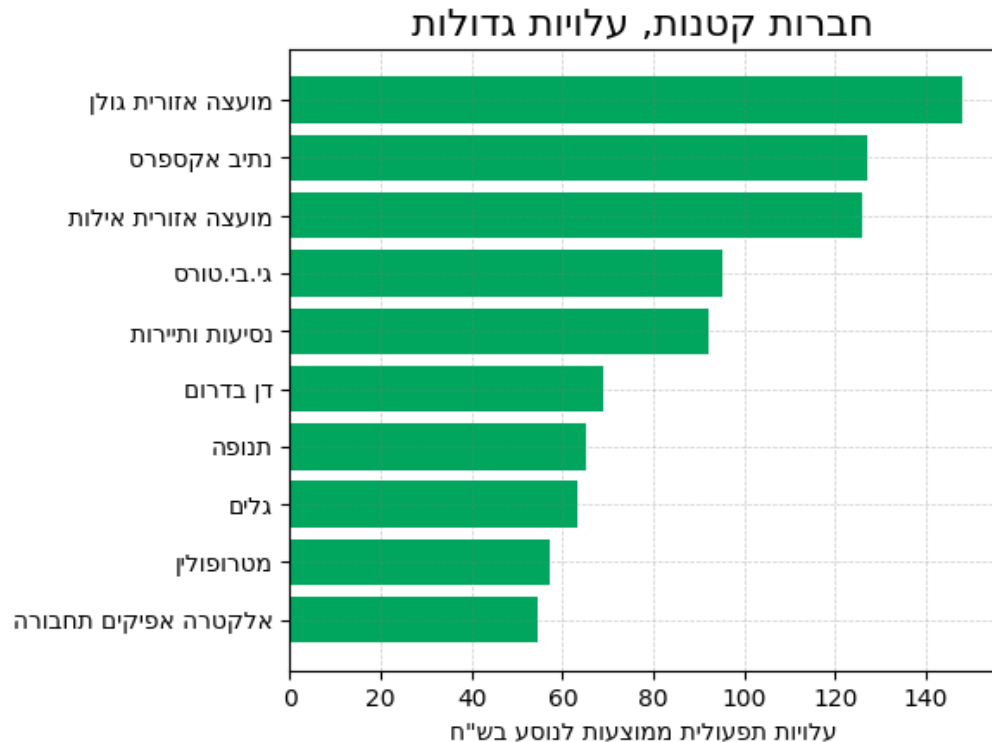
למה?

בגרף זה ניתן לנתונים לדבר. גרף זה מציג את עלויות התפעול בערים השונות. ניתן לזהות את הערים בהן יש עלויות גבוהות ולדעת להצביע עליהן. בנוסף, אפשר להשוות בין הערים. הגרף מצביע על עלות התפעול הגבוהה שבקרית מוצקין ובת ים.

איך?

- סיכון של שנת 2024 בלבד. כדי לקבל עלויות עדכניות.
- חישוב עלות שבועית ממוצעת ע"י הכפלת הערכים ממוצע נוסעים שבועי ו-עלות התפעול הממוצעת פר נוסע (OperatingCostPerPassenger , WeeklyPassengers).
- איחוד המטרופולינים מזרח ירושלים ו-ים. נעשה כדי להקל על העומס.
- סיכון שאר המטרופולינים (שאינם גוש דן, חיפה וירושלים). שאר המטרופולינים מתנהגים פחות או יותר כמו ירושלים ולכן סיכון שלהן לא גורע ממטרת הגרף. בנוסף מקל על העומס.
- קיבוץ לפי השדות מטרופולין ושם העיר וחשוב הממוצע של עלות השבועית הממוצעת.

- התחשבות ב-15 הערים בעלות העלויות הגבוהות ביותר כדי לא ליצר עומס על הגרף.
- חלוקה של הערך עלות שבועית ממוצעת ב-10,000 כדי להקל על עומס המספרים במקרא



מה?

בגרף זה ניתן לראות שוב את הנתונים המגיעים מתוך קובץ הנסועה הסטטי. בציר האנכי ניתן לראות את הקטגוריות: שם חברה האוטובוסים. ואת העלויות התפעוליות הממוצעת של הנוסע.

למה?

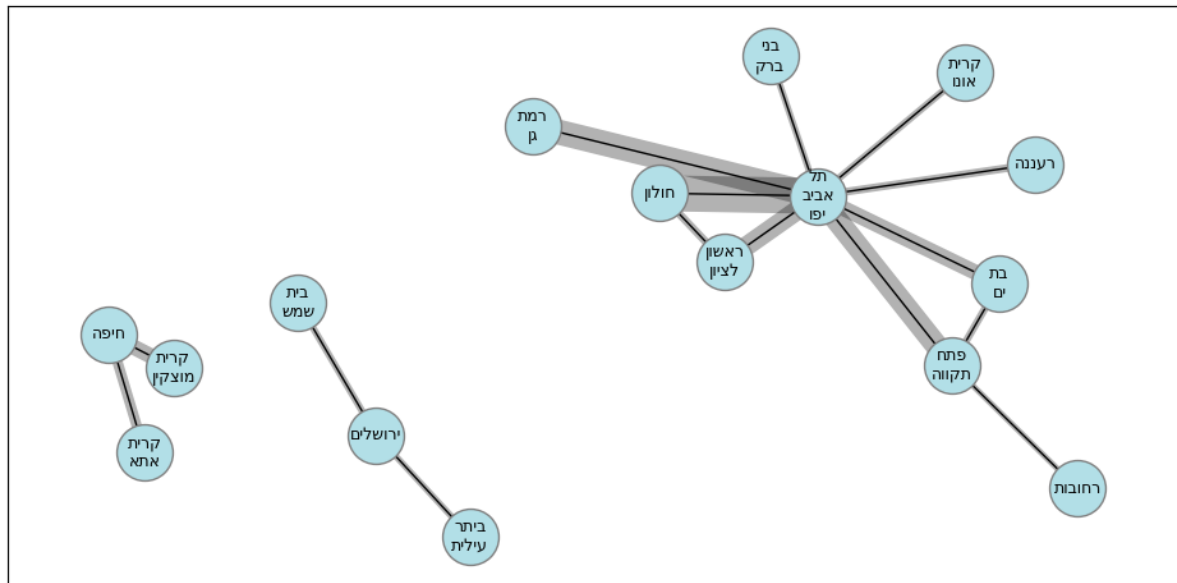
גרף זה מציג את עלויות התפעול בחברות השונות. ניתן לזהות את החברות בהן יש עלויות גבוהות ולדעת להצביע עליהן. בנוסף, אפשר להשוות בין הערים. גם זה הוא גרף משלים לגרפים הראשון והשני המוצגים מעלה. בעזרת גרף זה אפשר להניח שהסיבה לכך שעלויות התפעול בבת ים גדולה כל כך, לא קשורה לחברת האוטובוסים הדומיננטית בה (דן, שלא נמצאת בגרף זה).

איך?

- קיבוץ הנתונים לפי חברה. מיצוע של השדה `OperatingCostPassenger`.
- התחשבות ב-10 החברות בעלות העלויות הגבוהות ביותר כדי לא ליצר עומס על הגרף.
- סיבוב הצירים בגרף נעשה כדי להקל על הקורא בקריאת שמות חברות ההסיעים.
- הסופת `grid` כדי להקל על קריאת הנתונים.

תלויים במטרופולין

מספר הנוסעים הממוצע בין הערים השונות



מה?

גרף זה הינו גרף רשת המציג את שם העיר בצמתים ואת סכימת ממוצעי הנוסעים השבועיים שזורמת בין הערים. גם כאן, הנתונים מוצגים מתוך קובץ הנתונים הסטטי של נסועת האוטובוסים.

למה?

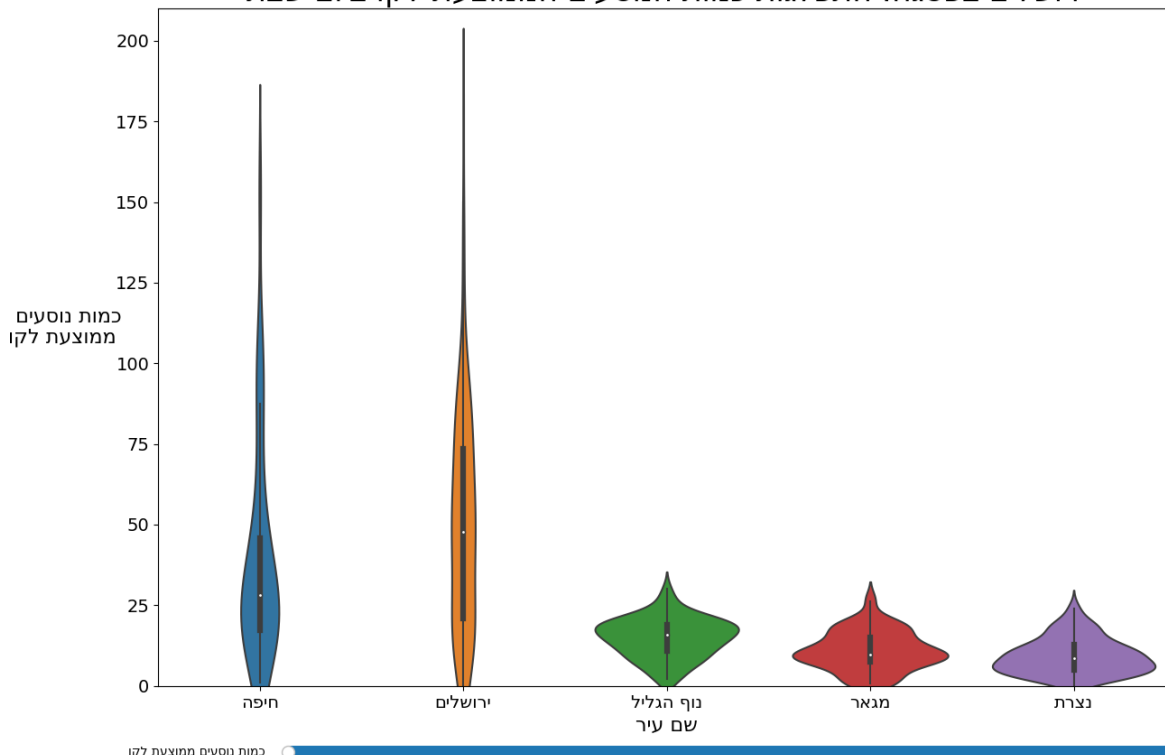
מטרת גרף זה היא לייצר טופולוגיה של הישובים השונים לפי ממוצע הנוסעים שזורם דרכן. בנוסף ניתן למצוא מסלולים בהן אפשר להגיע ממקום אחד לשני.

לאחר התבוננות בגרף זה אפשר לראות שגם לקרית מוצקין וגם לבת ים יש קשרים חזקים עם המטרופולינים שלהם. אולי ניתן הסבר לעלויות התפעול הגבוהות שלהן. כנראה לתחזק קוי אוטובוסים המעבירים את הנוסעים הלך וחזור מערים גדולות עולים יותר, ולא אחוז הנוסעים שיוצא מערים אלו למטרופולינים גבוה.

איך?

- קיבוץ של הנתונים לפי עיר מקור ועיר יעד. ביצוע של ממוצע הנוסעים השבועי.
- סינון המידע כך שלא תהיה עיר שמקושרת לעצמה. עניין לראות אילו ערים הכי מקושרות לערים אחרות ולא לעצמן.
- סינון 30 הקשרים החזקים ביותר. נעשה כדי להקל על העומס בגרף.
- ציור ה-nodes וציור ה-edges לפי האלגוריתם של רינגולד פרוכטרמן. כדי שהאלגוריתם לא ייתכנס יותר מידי הוגבל מקסימום סה"כ האיטרציות של האלגוריתם ל-10.

ירושלים בפסגה: התפלגות כמות הנוסעים הממוצעת לקו ביום שבת



מה?

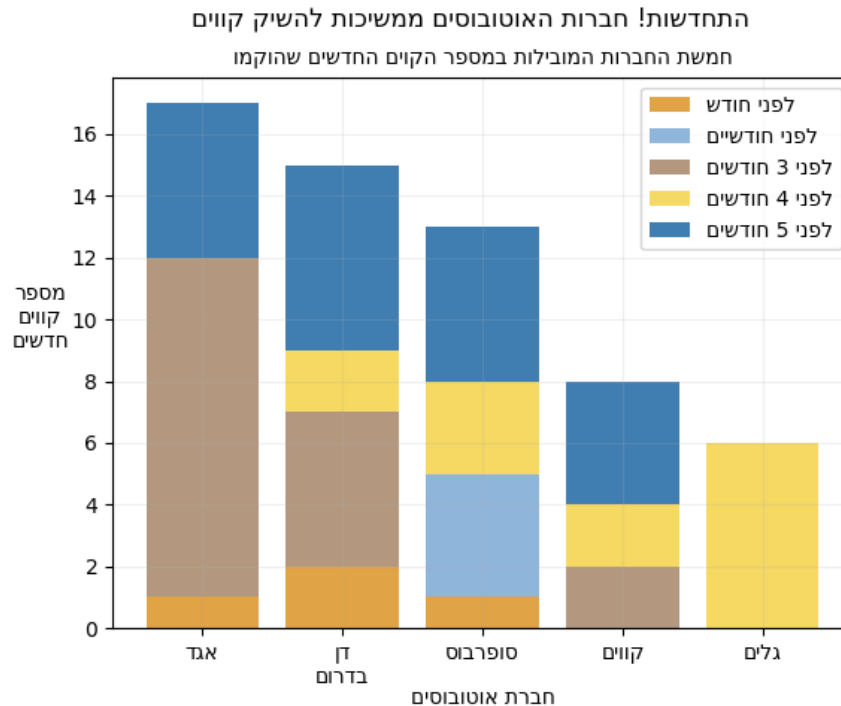
גם גרף זה מציג את המידע הסטטי שנמצא בקובץ הטבלאי של נסועת האוטובוסים. הגרף מציג את הקטגוריה: עיר ואת המשתנה כמות הנוסעים הממוצעת בקו. הנתונים מסודרים כך שלכל תקופה יש ממוצע נוסעים שבועי. ולכן ניתן להתייחס אליהם בסדר כרונולוגי מסויים.

למה?

גרף זה נועד כדי לתאר את הדימיון (או יותר נכון את השוני) בכמות הנוסעים ביום שבת בין ערים אלו. מציג את ירושלים וחיפה כמובילות בתחבורה ציבורית בשבת. הגרף מציג את העיקרון שדווקא ירושלים, שהיא עיר דתית, היא המקום בו יש הכי הרבה משתשים בתחבורה ציבורית בשבת (סליחה על הפוליטיקה).

איך?

- קיבוץ לפי ערים. סכימת כמות הנוסעים הממוצעת ומציאת 5 הערים המובילות.
- סינון הנתונים לפי הערים הנבחרות.
- זריקת שורות עם נתונים חסרים.
- הגבלת ערכי כמות הנוסעים הממוצעת בין 0 ל-140. תחום תחתון הוגבל בשל חוסר עניין במשתנה אם הוא מתחת ל-0. והגבול העליון הוגדר בשביל ראות יותר טובה של החציון.
- זריקת 2 הקטגוריות של שישי משעה 00:00 עד 04:00 ו-שישי משעה 04:00-06:00. בנתונים אלו כמעט אין נתונים השונים מ-0 כך שאין מה להציג.
- הוספת מסנן radio button לפי קטגוריות הזמנים הנתונות בקובץ הנתונים. הוספה זו נתנה שליטה למשתמש. המסקנת הגרף לא משתנה כאשר משנים את הקטגוריה.
- הוספת סנן המאפשר משחק עם הגבולות של ציר ה-y. נעשה בשל הצגת התפלגויות שונות עם טווחים שונים.



מה?

גרף זה הינו גרף עמודות מוערם. מתאר את מספר הקווים החדשים שנוספו עד כ-5 חודשים לפני הדוח, משמע אוקטובר - פברואר.

הנתונים המשוושים ליצירת גרף זה הינם נתונים קטגוריאליים. שם חברת האוטובוסים וקטגורית הותק של הקו. נתונים אלו זמינים בקובץ הנתונים הסטטי הטבלאי של קובץ נסועת האוטובוסים.

למה?

→ Enjoy



את האמת, בשביל הכיף 😊 גרף זה מציג את הנתונים שמופיעים בטבלת הנסועה ואפשר לבצע בעזרתו השוואות בין החברות השונות. בנימה אישית, שמחתי לראות שלמרות (ואולי בגלל) המצב הבטחוני במדינת ישראל, חברות כמו דן בדרום (חברה ששולטת בעיקר באזור באר שבע) וסופרבוס (עם נתח שוק בעיקר בגולן ועמק הירדן) ממשיכות להשיק קווים.

איך?

הכנת הנתונים:

- סינון של שנת 2024 בלבד.
- יצירת 5 עמודות חדשות המציינות האם הרשומה נמצאת באחת מהקטגוריות ע"פ המשתנה OperationSince.
- קיבוץ לפי שם חברת האוטובוסים וספירה של כמות הקווים החדשים.
- סינון 5 החברות הטובות ביותר. כדי להקל על עומס הצבעים. (יש חברות שלא נמצאות בכל הקטגוריות ולכן יש קפיצה של צבעים בן החברות השונות. ממש מורגש אם מוסיפים עוד חברות לגרף).

סיכום

בעבודה זו סקרנו את קובץ הנתונים נסועה בקווי אוטובוס. סקרנו את פיזור החברות על המפה, ראינו את זרימת הנוסעים במטרופולינים, התעסקנו בעלויות ואפילו עשינו מקרה מבחן וניסינו להסביר את הסיבה לקביצת עלויות בעירם בת ים וקרית מוצקין. לאחר מכן, ראינו את התפלגות הנוסעים בשבת בעירים שונות ולבסוף הצגנו 5 החברות המתחשות ביותר בזמן האחרון.

קישורים

[Git Repository of the project](#)

[Gov Data on website](#)