



אוניברסיטת בן גוריון
קורס 'ויזואליזציה של המידע'

פרויקט ויזואליזציה של תאונות דרכים בעיר ניו יורק

מגישות:
אור חן
עדי בכר

שם המרצה: פרופ' גלעד רביד

24.02.2026

תוכן עניינים

3	תיאור הנתונים והטיפול בהם
4	דיון בעבודות קשורות וביזואליזציות קיימות לאותם הנתונים
6	2020 A Rare bright of COVID-19: Fewer Traffic Crashes of NYC streets in - 1 ויזואליזציה
7	The Clock of Danger: Saturday Nights Turn Deadly on NYC Roads - 2 ויזואליזציה
9	Road Roulette: Mapping the Big Apple's Danger Zones - 3 ויזואליזציה
10.....	A Forest of Steel: Mapping the Giants of NYC Collisions - 4 ויזואליזציה
12..	Small Vehicle, Big Consequences: Two Wheel Crashes Turn Dangerous Fast - 5 ויזואליזציה
13.....	If you don't want to Kill, Don't text and drive - 6 ויזואליזציה
15.....	ביבליוגרפיה

תיאור הנתונים והטיפול בהם

מאגר הנתונים שבו עשינו שימוש בעבודה זו, הוא מאגר Motor Vehicle Collisions - Crashes של עיריית ניו יורק, הכולל תיעוד מפורט של תאונות דרכים בעיר לאורך מספר שנים. כל שורה במאגר מייצגת תאונה אחת, ולכן רמת הניתוח שנבחרה היא רמת האירוע (Crash-Level). המאגר כולל משתני זמן, משתני מיקום, נתוני חומרה כגון מספר פצועים והרוגים, שדות טקסטואליים של גורמים תורמים לתאונה וכן שדות המתארים את סוגי כלי הרכב המעורבים.

בהתאם למטרת העבודה, הוגדר משתנה חומרה (severity) על בסיס חומרת הנפגעים בתאונה, בו נעשה שימוש הן בהבחנה קטגוריאלית (ללא נפגעים, פצועים, הרוגים) לצורך ויזואליזציות תיאוריות, והן בהבחנה בינארית לצורך הבנת דפוסים כלליים בין תאונות עם נפגעים לבין תאונות ללא נפגעים. יצירת משתנה חומרה התבססה על השדות המספריים של מספר הפצועים ומספר ההרוגים בכל תאונה.

שלב ניקוי הנתונים כלל בראש ובראשונה סטנדרטיזציה של שדות טקסטואליים. ערכים בשדות כגון 1 CONTRIBUTING FACTOR VEHICLE הומרו לאותיות גדולות, רווחים מיותרים הוסרו, וטופלו ערכים כלליים כגון "UNKNOWN", "NONE", "UNSPECIFIED". מהלך זה נועד להפחית שונות מלאכותית הנובעת מצורות כתיב שונות ולאחד ערכים זהים מבחינה לוגית. נציין כי בשדות החוזרים (כגון 5-1 VEHICLE TYPE CODE ו-5-1 CONTRIBUTING FACTOR VEHICLE) קיימת תופעה של חוסר מבני (Structural Missingness), הנובעת מכך שלא בכל תאונה מעורבים חמישה כלי רכב. לפיכך, ערכים חסרים באינדקסים גבוהים אינם בהכרח בעיית איכות נתונים אלא תוצאה טבעית של מבנה הסכימה.

חלק מרכזי בהכנת הנתונים היה הפחתת ריבוי קטגוריות (High Cardinality) באמצעות אגרגציה מושכלת. בשדות הגורמים התורמים לתאונה בוצע מיפוי לערכי מאקרו באמצעות יצירת פונקציה חדשה בשם `map_contributing_factor`, זאת על ידי איחוד ערכים טקסטואליים רבים לקבוצות משמעותיות (כגון `DISTRACTION_PHONE`, `SPEEDING`, `FATIGUE_SLEEP`). מהלך זה איפשר מעבר ממרחב קטגוריאל רועש ומפוצל לייצוג אינפורמטיבי ובר השוואה, אשר שימש בהמשך לויזואליזציות של גורמים תורמים לפי חומרת תאונה. באופן דומה, בוצעה אגרגציה של סוגי כלי רכב באמצעות יצירת פונקציית `map_vehicle_type`. שדה `VEHICLE TYPE CODE` 1 עבר מיפוי לקטגוריות רחבות (כגון `PRIVATE_VEHICLE`, `HEAVY_VEHICLE`, `TAXI_LIVERY`, `EMERGENCY_VEHICLE`). גם אגרגציה זו צמצמה שונות טקסטואלית גבוהה ואיפשרה ניתוח השוואתי בין קבוצות כלי רכב.

בנוסף, בשלב הנדסת המאפיינים בוצע עיבוד למשתני הזמן. השדות `CRASH DATE` ו-`CRASH TIME` אוחדו לייצוג זמן תקני, ומתוכם הופקו מאפיינים נגזרים הכוללים את שעת התאונה (hour), יום בשבוע (weekday) וחודש בשנה (month). פירוק זה אפשר מעבר מייצוג טקסטואלי גולמי לייצוג מספרי וקטגוריאל המאפשר ניתוח דפוסים עונתיים ותבניות זמן.

לבסוף, בוצעו חישובי אגרגציה שונים ברמת התאונה, כגון סכימת מספר הפצועים וההרוגים, חישוב מספר כלי רכב מעורבים לכל תאונה, וחישוב מדדים מצטברים לפי קטגוריות שונות (גורמים תורמים, סוגי כלי רכב, חומרה ועוד). תהליכים אלו יצרו מערך נתונים עקבי, מנוקה ומאורגן, אשר אפשר בניית ויזואליזציות אינפורמטיביות הממחישות דפוסים מרחביים, זמניים וסיבתיים בתאונות הדרכים בעיר ניו יורק.

דין בעבודות קשורות וביזואליזציות קיימות לאותם הנתונים

ספרות המחקר בתחום תאונות הדרכים בעיר ניו יורק מראה שימוש מתקדם באנליזה מרחבית, זמנים וסיבתיות של תאונות, ולעיתים משולב גם עם ויזואליזציות להבנת התנהגות התאונות לאורך השנים.

המחקר הראשון בו נתמקד הוא של Shaaban & Ibrahim (2021), Analysis and Identification of Contributing Factors of Traffic Crashes in New York City
<https://www.sciencedirect.com/science/article/pii/S2352146521005809>

המחקר עוסק בניתוח סטטיסטי מעמיק של תאונות דרכים בעיר ניו יורק בשנים 2013–2019, תוך בחינת גורמים תורמים, דפוסים זמניים והתפלגויות חומרה. אחת הוויזואליזציות המרכזיות במחקר (Fig. 5 – Crashes Per Type of Day) מציגה השוואה בין ימי חול לסופי שבוע במספר הפצועים וההרוגים. בגרף ניתן לראות כי המספר האבסולוטי של נפגעים גבוה יותר בימי חול, הן עבור פצועים והן עבור הרוגים. עם זאת, החוקרים מדגישים כי יש לפרש ממצא זה בזהירות, שכן ימי חול מהווים חמישה ימים בשבוע לעומת יומיים בלבד בסוף השבוע, ולכן ההשוואה אינה מנומקת לאורך זמן. הוויזואליזציה משתמשת בעמודות מקובצות לפי שנים ובחלוקה ברורה בין Weekdays ל-Weekends, ומדגימה כיצד פילוח זמן פשוט יחסית יכול לחשוף דפוסים חשובים אך גם להטעות אם אינו מלווה בניתוח יחסי (כגון ממוצע יומי).

ויזואליזציה זו רלוונטית במיוחד לעבודה הנוכחית, שכן גם בפרויקט זה בוצעה הנדסת פיצ'רים זמניים (יום בשבוע, שעה, חודש), מתוך הבנה כי מימד הזמן מהווה משתנה מרכזי בהסבר חומרת תאונה. בדומה למחקר, גם כאן נעשה שימוש בגרפים השוואתיים להמחשת דפוסים זמניים.

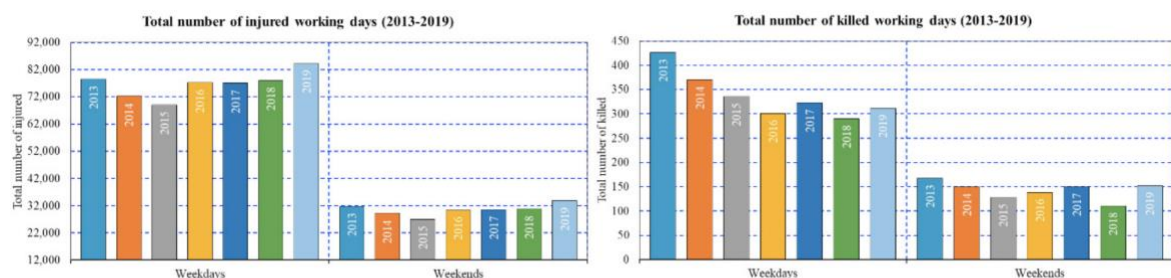


Fig. 5. Total number of injured and killed per type of day (2013-2019).

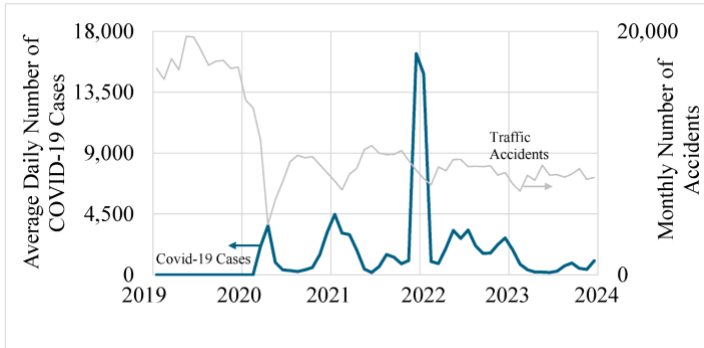
המאמר השני שבחרנו הוא Patterns and Analysis of Traffic Accidents in New York City between 2013 and 2023, שפורסם ב-Urban Science בשנת 2024: Patterns and Analysis of Traffic Accidents in New York City between 2013 and 2023

המחקר עוסק בניתוח דפוסי התאונות בעיר ניו יורק לאורך עשור, בשילוב משתני זמן, חומרה וסיבות תאונה. המחקר כולל ויזואליזציות של מגמות התאונות, ובין היתר מציג גרפים המתארים את הקשר בין מגיפת COVID-19 למספר התאונות החודשי (Fig. 9) וכן את ההשפעה של נהיגה מוסחת על כמות התאונות החודשית על פני השנים (Fig. 7).

בגרף המוצג ב-Fig. 9, המחברים מציגים קו כפול המשווה בין מגמת מקרי קורונה יומיים לבין מגמת כמות התאונות החודשית. הוויזואליזציה ממחישה באופן ברור כי ירידה חדה במספר התאונות בתחילת 2020 חלה במקביל לעלייה החדה במקרי COVID-19, מה שמרמז על השפעת הסגר והפחתת תנועת כלי הרכב על פעילות התאונות. עם זאת, לצד ההתחקות אחרי הקורונה, ניתן לראות כי לאחר החזרה לתנועה נורמלית מספר התאונות התייצב ואף כמעט

וחדר לרמות קודמות, מה שמדגיש את הקשר המורכב בין גורמים חברתיים (כגון מגפת הקורונה) לדפוסי תאונה בתחבורה.

בדומה לפרויקט שלנו, גם כאן נעשה שימוש בגרפים מבוססי זמן על מנת לזהות מגמות חריגות ושינויים משמעותיים לאורך השנים, ובפרט כדי להדגיש כיצד אירועים חיצוניים כגון מגפת הקורונה עשויים לשנות דפוסי תנועה ובכך להשפיע גם על כמות התאונות וגם על רמת החומרה שלהן.

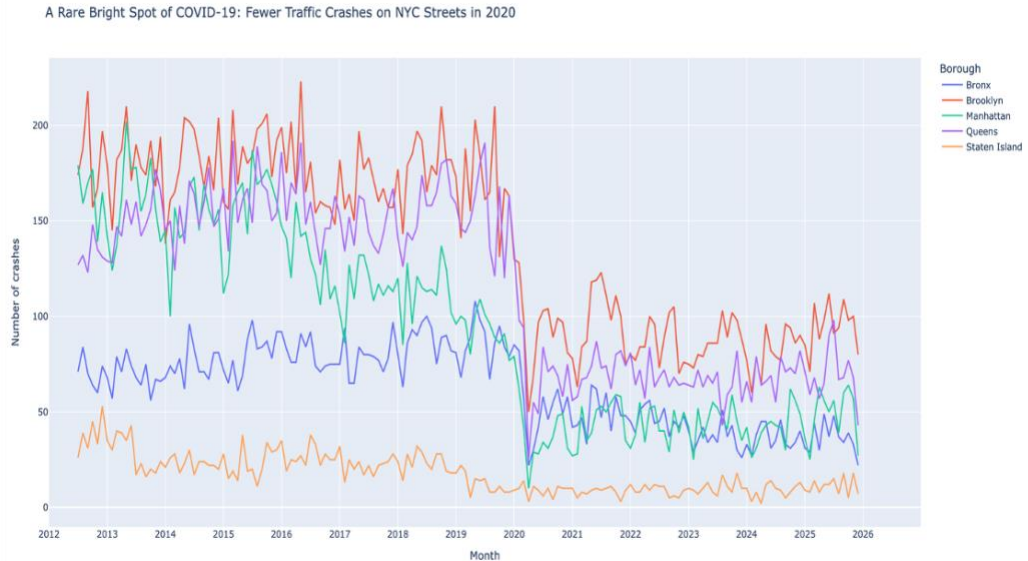


גרף נוסף ממחקר זה הוא Fig. 7, המציג גרף קווי של מספר התאונות החודשי הנובע מנהיגה מוסחת, המתייחס בעיקר לשימוש בטלפון נייד. הגרף מדגים מגמת עלייה בשנים 2013–2018, ירידה חדה במהלך 2020 כתוצאה ממגבלות התנועה, ולאחר מכן יציבות יחסית ברמות התאונות הקשורות להסחת דעת. ויזואליזציה זו תומכת במחקרי בטיחות תחבורה אחרים שמצאו כי הסחת דעת, ובעיקר שימוש בטלפונים ניידים, מהווה גורם משמעותי בתאונות דרכים.

ויזואליזציה זו מקשרת באופן ישיר לפרויקט שלנו, שכן גם בפרויקט זה בוצע מיפוי של גורמים תורמים (Contributing Factors) אשר העלה כי קטגוריית הסחת הדעת ובפרט שימוש בטלפון נייד, לא רק מהווה אחד הגורמים השכיחים ביותר לתאונות, אלא גם מציגה רמת חומרה גבוהה יחסית בהשוואה לגורמים אחרים. ממצא זה מתיישב עם הספרות המחקרית ועם הוויזואליזציה המוצגת במאמר, המחזקת את ההבנה כי שימוש בטלפון בזמן נהיגה הוא גורם סיכון מרכזי לתאונות בעלות פוטנציאל פגיעה משמעותי. בכך נוצרת עקביות בין הממצאים במחקר החיצוני לבין הדפוסים שהתגלו בניתוח שבוצע במסגרת פרויקט זה.



ויוזאליזציה 1 - 2020 A Rare bright of COVID-19: Fewer Traffic Crashes of NYC streets in



What - מה הנתונים?

לצורך גרף זה בוצעה אגרגציה חודשית של מספר התאונות, תוך חלוקה לפי מחוז (Borough). האובייקטים (Entities) המיוצגים בגרף הם יחידות זמן חודשיות, כאשר לכל חודש קיים ערך כמותי של מספר התאונות בכל אחד מחמשת המחוזות: Staten Island ו-Bronx, Brooklyn, Manhattan, Queens. לכל אובייקט זמן משויכים שלושה מאפיינים עיקריים: משתנה זמן מסוג Temporal (שנה וחודש), משתנה קטגוריאל מסוג Nominal (מחוז), ומשתנה כמותי מסוג Quantitative (מספר תאונות). מבנה הנתונים הוא מבנה של סדרת זמן (Time Series), המאפשר ניתוח של מגמות לאורך זמן והשוואה בין קטגוריות גיאוגרפיות מקבילות.

Why - למה מוצגת הויוזאליזציה?

מטרת הויוזאליזציה היא לאפשר חקירה של דפוסים זמניים והשוואה בין אזורים שונים בעיר ניו יורק. המשתמש יכול לזהות מגמות ארוכות טווח במספר התאונות, להשוות בין מחוזות שונים ולבחון האם שינויים מתרחשים באופן אחיד בעיר או באופן מקומי. הויוזאליזציה תומכת במספר משימות אנליטיות מרכזיות: זיהוי מגמות (trend detection), השוואה בין קבוצות (comparison), איתור נקודות חריגות בזמן (anomaly detection), ותקשורת מסר כולל לגבי התנהגות התאונות בעיר לאורך השנים.

How - כיצד הויוזאליזציה ממומשת?

הייצוג החזותי נעשה באמצעות גרף קווי (Line Chart), כאשר כל מחוז מיוצג באמצעות קו בצבע שונה. ה-marks המרכזיים הם קווים המחברים נקודות זמן רציפות, וה-channels העיקריים הם מיקום על ציר ה-X (זמן), מיקום על ציר ה-Y (מספר תאונות) וצבע (הבחנה בין מחוזות). בחירה זו מאפשרת קידוד ברור של שלושה ממדים במקביל: זמן, כמות וקטגוריה.

הפריסה היא לאורך ציר זמן, המתאימה במיוחד להצגת סדרות זמן ולהמחשת שינוי רציף. בנוסף, הויוזאליזציה אינטרקטיבית וכוללת אינטראקציות להצגת ערכים מדויקים, וכן אפשרות לסנן באמצעות legend, המאפשרת להציג או להסתיר מחוזות שונים. אינטראקטיביות זו מפחיתה עומס חזותי ומאפשרת חקירה ממוקדת של מחוז מסוים או השוואה דינמית בין מספר מחוזות.

בחירת גרף קווי על פני גרף עמודות נבעה מהצורך להציג רצף זמן ארוך עם מספר רב של נקודות זמן, באופן שמדגיש מגמות ולא רק ערכים נקודתיים. השימוש בצבעים מובחנים לכל מחוז נועד להקל על זיהוי והשוואה, מבלי לפגוע בקריאות הגרף.

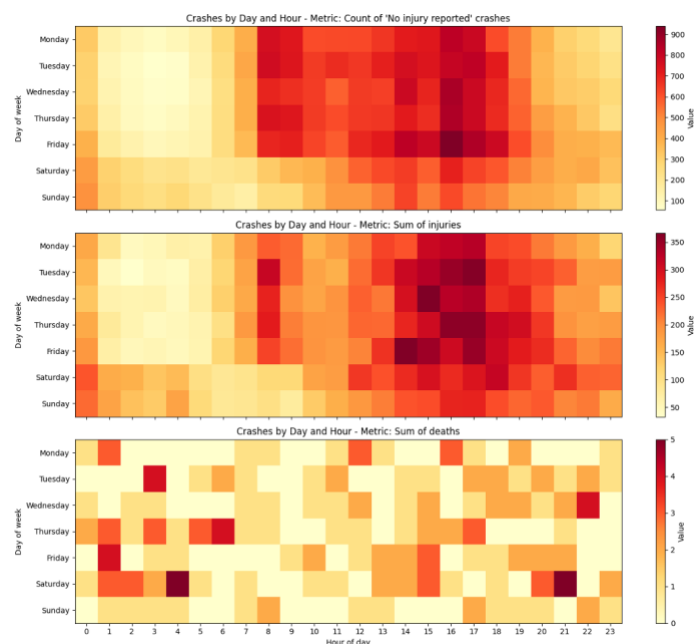
הסבר הבחירות שנעשו

הבחירה בגרף קווי נבעה מהרצון להציג מגמות לאורך זמן בצורה ברורה, שכן שימוש בגרף עמודות היה יוצר עומס חזותי בשל מספר רב של נקודות זמן. בנוסף, השימוש בצבעים שונים לכל מחוז מאפשר הבחנה מיידית בין הקווים ותומך בהשוואה חזותית. הבחירה להציג את הנתונים לפי Borough ולא ברמת העיר בלבד מאפשרת לזהות הבדלים בין אזורים שונים ולהימנע מהצגה ממוצעת שעלולה להסתיר דפוסים מקומיים משמעותיים. הבחירה בגרף אינטרקטיבי מאפשרת חקירה מעמיקה יותר על ידי המשתמש, כולל אפשרות להתמקד במחוז מסוים ולהשוותו לאחרים בצורה נוחה.

תובנות מרכזיות

מהויזואליזציה ניתן להסיק כי קיימים פערים ברורים בין המחוזות, כאשר Brooklyn ו-Queens מציגים לאורך השנים מספר תאונות גבוה יחסית, בעוד Staten Island מציג מספר תאונות נמוך משמעותית. בנוסף, ניכרת ירידה חדה בכמות התאונות בשנת 2020 כמעט בכל המחוזות, תופעה התואמת את השפעת מגפת הקורונה והסגרים אשר הפחיתו באופן משמעותי את תנועת כלי הרכב בעיר. לאחר מכן ניתן לראות חזרה הדרגתית לרמות תאונות דומות לשנים קודמות, מה שמצביע על התאוששות הפעילות העירונית.

ויזואליזציה 2 - The Clock of Danger: Saturday Nights Turn Deadly on NYC Roads



What - מה הנתונים?

מתוך המשתנים CRASH DATE ו-CRASH TIME הופקו משתנים קטגוריאליים המייצגים את יום השבוע (weekday) ואת השעה ביממה (hour). לפיכך, יחידת הניתוח בגרף זה היא תא זמן בדיד המוגדר כשילוב בין יום בשבוע לבין

שעה מסוימת בטווח 0-23.

לכל שילוב זמן חושבו שלוש מטריקות שונות: מספר התאונות ללא נפגעים, סכום הפצעים, וסכום ההרוגים. מבנה הנתונים הוא מבנה מטריצוני דו-ממדי, שבו השורות מייצגות את ימי השבוע והעמודות מייצגות את שעות היממה. כל תא במטריצה משויך לערך כמותי המייצג את היקף התאונות או את חומרתן באותו חלון זמן. בכך, היוזואליזציה מאפשרת בחינה סימולטנית של שני ממדים זמניים קטגוריאליים ושל מדד כמותי.

Why - למה מוצגת היוזואליזציה?

מטרת ויוזואליזציה זו היא לזהות דפוסים זמניים (Temporal Patterns) בהתפלגות תאונות הדרכים ובהיקף חומרתן לאורך השבוע והיממה. היוזואליזציה נועדה לאפשר חקירה אנליטית של הבדלים בין ימי חול לסופי שבוע, בין שעות עומס לשעות לילה, ובין חומרת התאונות.

הגרף תומך במספר משימות אנליטיות מרכזיות: זיהוי תבניות חוזרות, השוואה בין קטגוריות זמן, איתור אזורי זמן חריגים, והדגשת חלונות זמן המאופיינים ברמת סיכון גבוהה יותר. מעבר לכך, היוזואליזציה מאפשרת תקשורת מסר ברור בדבר חשיבות מימד הזמן בהסבר דינמיקת התאונות בעיר ניו יורק.

How - כיצד היוזואליזציה ממומשת?

היוזואליזציה ממומשת באמצעות Heatmap, ייצוג מטריצוני שבו כל תא מיוצג כשטח, והערך הכמותי מקודד באמצעות צבע. הציר האופקי מייצג את השעה ביממה, והציר האנכי מייצג את יום השבוע. עוצמת הצבע משמשת כערוץ קידוד כמותי, כך שככל שהצבע כהה יותר, הערך גבוה יותר. נבחר סולם צבעים בגווי אדום, כאשר אדום כהה מייצג ערכים גבוהים יותר של תאונות, פצעים או הרוגים. הבחירה בגווי אדום אינה מקרית אלא נושאת משמעות סמנטית, אדום הוא צבע המזוהה תרבותית וחזותית עם סכנה, אזהרה ופגיעה בחיי אדם. לפיכך, השימוש בגוון אדום כהה להדגשת ריכוזי תאונות גבוהים מחזק את הקשר האינטואיטיבי בין עוצמת הצבע לבין חומרת האירוע, ותורם להבנה מהירה של מוקדי סיכון. בחירה זו גם תומכת בהיררכיה תפיסתית ברורה, אזורים כהים יותר מושכים את תשומת הלב באופן מיידי ומסמנים "אזורים חמים" (Hot Spots), בעוד אזורים בהירים משדרים רמת סיכון נמוכה יותר. בכך מושגת הדגשה ויוזואלית של מוקדים קריטיים מבלי להעמיס בפרטים גרפיים נוספים. בנוסף, ההפרדה לשלוש מפות נפרדות, תאונות ללא נפגעים, פצעים והרוגים, מאפשרת השוואה אנליטית בין נפח התאונות לבין חומרתן, מבלי לערבב סקאלות שונות באותו ייצוג חזותי.

הסבר הבחירות שנעשו

הבחירה ב-Heatmap נבעה מהצורך לנתח בו-זמנית שני ממדים קטגוריאליים דיסקרטיים (יום ושעה), תוך שמירה על בהירות חזותית וזיהוי תבניות. מאחר שמבנה הנתונים הוא מטריצוני מטבעו, Heatmap מהווה התאמה טבעית למבנה זה.

בנוסף, בפרויקט זה בוצעה הנדסת פיצ'רים זמניים מתוך תאריך ושעה, מתוך הנחה כי מימד הזמן מהווה משתנה מרכזי בהסבר חומרת תאונה. לפיכך, היוזואליזציה אינה רק תיאורית, אלא גם תומכת בהחלטות מידול עתידיות, בכך שהיא מדגימה אמפירית את חשיבות משתני הזמן.

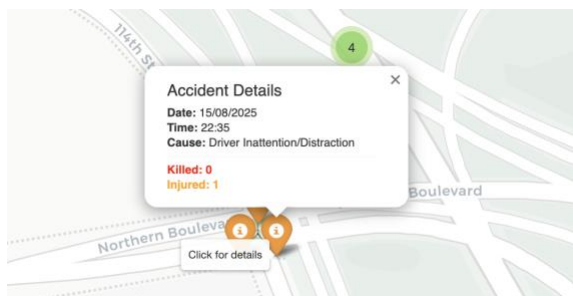
תובנות מרכזיות

ניתוח היוזואליזציה מוביל למספר מסקנות מרכזיות. ראשית, קיימת מחזוריות ברורה לאורך השבוע והיממה, כאשר שעות אחר הצהריים והערב בימי חול מאופיינות בנפח תאונות גבוה במיוחד. בנוסף, ניתן לזהות כי גם שעות הבוקר המוקדמות בימי חול מציגות עלייה במספר התאונות. דפוס זה מתיישב עם שעות השיא התחבורתיות בעיר ניו יורק, המתאפיינות ביציאה לעבודה בשעות הבוקר ובחזרה ממנה בשעות אחר הצהריים. עומסי תנועה, צפיפות גבוהה בכבישים וריבוי אינטראקציות בין כלי רכב ומשתמשי דרך נוספים בשעות אלו מגדילים את ההסתברות להתרחשות

תאונות.

שנית, דפוס ההרוגים אינו חופף לחלוטין לדפוס נפח התאונות, דבר המרמז כי תנאים מסוימים, בעיקר בשעות לילה ובסופי שבוע, עשויים להיות קשורים לרמת חומרה גבוהה יותר גם אם נפח התאונות נמוך יחסית. ממצא זה מצביע על כך שנפח תאונות גבוה אינו בהכרח מתורגם לחומרה גבוהה, וכי משתנים סביבתיים והתנהגותיים נוספים עשויים להשפיע על תוצאת האירוע.

ניזואליזציה 3 - Road Roulette: Mapping the Big Apple's Danger Zones



What - מה הנתונים?

הויזואליזציה מתמקדת בתת קבוצה של משתנה המטרה, הכוללת "תאונות חמורות" בלבד, תאונות שבהן נרשמו פצועים או הרוגים ($0 < \text{Injured/Killed}$). הנתונים המוצגים הם ברמת האירוע הבודד וכוללים חמישה ממדים עיקריים: מיקום גיאוגרפי מדויק (Latitude & Longitude), חומרת הפגיעה (מספר פצועים והרוגים), זמן התרחשות (תאריך ושעה), והגורם התורם המרכזי לתאונה (Contributing Factor).

Why - למה מוצגת הויזואליזציה?

מטרת הויזואליזציה היא לאפשר חקירה מרחבית אינטראקטיבית של מוקדי סיכון בעיר. בעוד שגרפים סטטיים מאפשרים להבין מגמות כלליות, מפה אינטראקטיבית מאפשרת למשתמש לבצע משימות של זיהוי ריכוזים (Clustering) וחיפוש נקודתי (Drill-down). המשימה האנליטית היא לאתר "מוקדים חמים" (Hotspots) של תאונות קטלניות וחמורות, ולהבין האם קיימת חוקיות גיאוגרפית המקשרת בין מיקום התאונה לבין גורמי הסיכון (כגון צמתים עמוסים או עורקי תחבורה ראשיים), ובכך לתמוך בתהליכי קבלת החלטות בתחום תשתיות הבטיחות בדרכים.

How - כיצד הויזואליזציה ממומשת?

הויזואליזציה מומשה באמצעות ספריית Leaflet ומשתמשת בקידודים הבאים:

- מיקום (Spatial Position): הנקודות מקודדות לפי מיקומן הגיאוגרפי המדויק על גבי מפת בסיס (Base Map).

- צבירת נתונים (Marker Clustering): כדי להתמודד עם בעיית ה-Overplotting (עומס חזותי של נקודות חופפות), עשינו שימוש באלגוריתם צבירה המאגד נקודות קרובות לאשכולות ממוספרים. גודל האשכול וצבעו משתנים דינמית בהתאם לרמת ה-Zoom, מה שמאפשר מעבר רציף מתצוגת מאקרו עירונית לתצוגת מיקרו רחוב.
- צבע וסמלים (Color & Glyph Coding): נעשתה הבחנה קטגוראלית בין דרגות חומרה:
 - כתום (אייקון Info): מסמל תאונות עם פצועים בלבד.
 - אדום (אייקון Exclamation): מסמל תאונות קטלניות הכוללות הרוגים, ובכך מייצר היררכיה חזותית המדגישה את האירועים הקריטיים ביותר.
- אינטראקטיביות: שימוש ב-Tooltips להצגת הנחיה ללחיצה, ו-Popups המציגים טקסט מפורט על פרטי התאונה, הגורם לה והנפגעים, מה שמאפשר הנגשת מידע רב מבלי להעמיס על הממשק הראשוני.

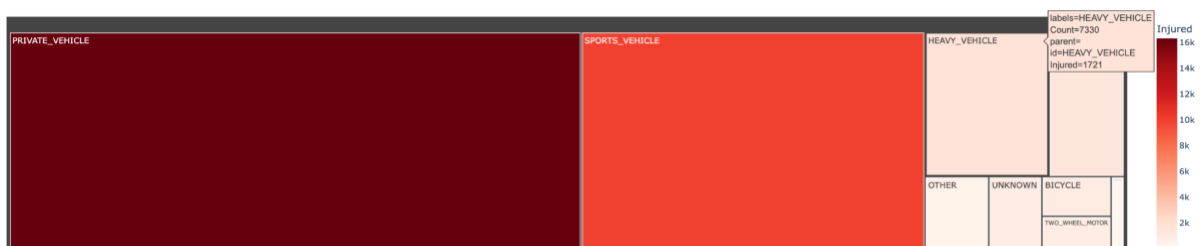
נימוק לבחירת היוזאליזציה

הבחירה במפה אינטראקטיבית מבוססת אשכולות (MarkerCluster) נבעה מהצורך לגשר בין כמות הנתונים לבין יכולת התפיסה החזותית של המשתמש. מפה סטטית הייתה נכשלת בהצגת אלפי נקודות במרחב האורבני הצפוף של ניו יורק. המפה מאפשרת גמישות, המשתמש יכול לקבל תמונה שלמה של רובעי העיר, ובמקביל לבחון אירועים בודדים ברמת הרחוב. השימוש בצבעים אדום מול כתום נבחר כדי לייצר "קריאה לפעולה" חזותית כלפי המוקדים הקטלניים, בהתאם לעקרונות של קשב טרום קשבי (Pre-attentive processing).

מסקנות מהיוזאליזציה

ניתוח המפה מעלה כי תאונות חמורות אינן מפוזרות באופן אחיד, קיימת נטייה ברורה לריכוזים באזורי המעבר בין רובעי העיר (גשרים ומנהרות) ובצמתים מרכזיים במנהטן ובברוקלין. מה-Popups עולה כי במקרים רבים של תאונות קטלניות (באדום), הגורם התורם קשור להסחת דעת או למהירות מופרזת, מה שמעיד על כך שהגורם האנושי מהווה משתנה קריטי בחומרת התאונה במוקדים אלו. האינטראקטיביות חושפת כי גם באזורים שנראים בטוחים במבט על, קיימים אזורים סיכון נקודתיים הדורשים התייחסות תשתיתית ספציפית.

ויזואליזציה 4 - A Forest of Steel: Mapping the Giants of NYC Collisions



What - מה הנתונים?

היוזאליזציה הרביעית מנתחת את המשתנה הקטגוריאלי המתאר את סוגי כלי הרכב המעורבים בתאונות, תוך יישום תהליך של הנדסת מאפיינים (Feature Engineering) שנועד להתגבר על חוסר האחידות הקיים בנתונים. במסגרת עיבוד הנתונים, הפעלנו פונקציית מיפוי שביצעה סטנדרטיזציה של הערכים הטקסטואליים, וקיבצה עשרות תוויות גולמיות לתשע קטגוריות על. תהליך זה איחד לדוגמה רכבים פרטיים, ג'יפים וואנים תחת הקטגוריה PRIVATE_VEHICLE, ריכוז רכבי מטען, אוטובוסים ותחבורה כבדה תחת HEAVY_VEHICLE, וסיווג רכבי הצלה ומוניות לקבוצות EMERGENCY_VEHICLE ו-TAXI_LIVERY בהתאמה, לצד קטגוריות ספציפיות לאופנועים, אופניים,

רכבי ספורט וטיפול בערכים חסרים (UNKNOWN). הייצוג הסופי בויזואליזציה מציג את השכיחות הכמותית של אירועי תאונות עבור כל אחת מקבוצות אלו, ובכך מאפשר מעבר ממאגר נתונים רועש למידע סטטיסטי בעל משמעות אנליטית.

Why - למה מוצגת הויזואליזציה?

המטרה המרכזית היא לבצע ניתוח השוואתי (Comparative Analysis) המאפשר להבין את פלח השוק של כל סוג רכב במפת הסיכונים העירונית. המשימה האנליטית היא זיהוי הפרופורציות של הקטגוריות השונות, כלומר, להמחיש בצורה ויזואלית מיידית אילו סוגי רכב מהווים את מרבית נפח האירועים לעומת קטגוריות משניות. השימוש במיפוי הקטגוריות הכרחי למשימה זו, שכן הוא מאפשר להשוות בין קבוצות בעלות משמעות בטיחותית ותכנונית, ולזהות האם קיימת דומיננטיות חריגה של סוג רכב מסוים (כגון רכבים כבדים או מוניות) ביחס לתפיסה הרווחת של נפח התנועה.

How - כיצד הויזואליזציה ממומשת?

הויזואליזציה מומשה באמצעות תרשים עץ (Treemap), המשתמש בקידודים החזותיים הבאים:

- קידוד שטח (Area Encoding): גודל המלבן של כל קטגוריה מקודד לפי מספר התאונות המשוך אליה. ככל שהשטח גדול יותר, כך שכיחות המעורבות של אותו סוג רכב גבוהה יותר.
- חלוקה מרחבית (Spatial Partitioning): המלבנים מסודרים בתוך פריים מרכזי אחד, מה שמאפשר השוואה מהירה של יחסי "חלק מתוך שלם" (Part-to-Whole).
- צבעוניות קטגוריאלי: כל מלבן (אשכול) מקבל גוון שונה המאפשר הפרדה ויזואלית, ככל שהגוון כהה יותר הדבר מעיד על כמות פצועים גבוהה יותר עבור כל סוג רכב.
- תוויות (Labelling): שמות הקטגוריות המאוחדות והערכים הכמותיים מוצגים בתוך המלבנים, מה שמנגיש את המידע ללא צורך במעבר חוזר לצירים או למקרא חיצוני.

נימוק לבחירת הויזואליזציה

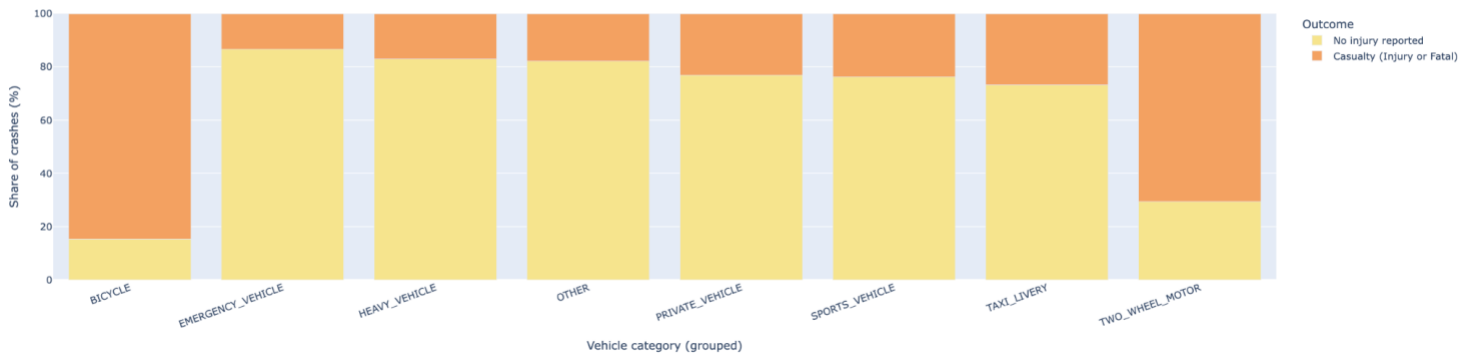
הבחירה בתרשים עץ (Treemap) על פני גרפים קלאסיים אחרים נבעה מהיכולת שלו לנצל 100% מהשטח הויזואלי להצגת פרופורציות מורכבות. בעוד שגרף עמודות מתמקד בהשוואה בין ערכים בודדים, תרשים העץ מאפשר למתבונן לתפוס באופן אינטואיטיבי את ה"משקל" של כל קטגוריה בתוך המערכת הכוללת. המבנה המרובע והדחוס של ה-Treemap מתאים במיוחד להצגת הקטגוריות שנוצרו בתהליך המיפוי, שכן הוא מדגיש את הפער הניכר בין הקטגוריות הדומיננטיות (כמו רכבים פרטיים) לבין סוגי הרכב המשניים, בצורה יעילה ואסתטית.

מסקנות מהויזואליזציה

ניתוח תרשים העץ מעלה כי רכבים פרטיים (PRIVATE_VEHICLE) תופסים את נתח השטח הגדול ביותר, מה שמעיד על מעורבותם הגבוהה ביותר בתאונות מבחינה מספרית. עם זאת, חשוב לסייג ממצא זה ולהדגיש כי הנתונים בויזואליזציה זו אינם מנורמלים. הדומיננטיות החזותית של הרכבים הפרטיים מושפעת באופן ישיר ובלתי נמנע מהכמות האבסולוטית של רכבים אלו בכבישי העיר (נפח התנועה). לכן, שטח המלבנים בגרף משקף את שכיחות האירועים בשטח אך אינו מעיד בהכרח על "מסוכנות" יחסית של סוג רכב אחד על פני אחר. תובנה זו מהווה את הבסיס לצורך בויזואליזציה הבאה, שתנטרל את השפעת הכמות ותבחן את הנתונים בצורה מנורמלת.

ויוזאליזציה 5 - Small Vehicle, Big Consequences: Two Wheel Crashes Turn Dangerous Fast

"Small Vehicle, Big Consequences: Two-Wheel Crashes Turn Dangerous Fast"



What - מה הנתונים?

הויוזאליזציה החמישית מציגה את נתוני התאונות לאחר תהליך של נרמול (Normalization). הנתונים מחולקים לממד קטגורי (סוג הרכב) ולממד של רמת חומרה, המפוצל לשתי קבוצות עיקריות: תאונות הכוללות נפגעים (הרוגים או פצועים) ותאונות ללא נפגעים (Property Damage Only). בעוד שהגרף הקודם הציג ספירה אבסולוטית, בגרף זה הנתונים חושבו כמדד יחסי המנטרל את השפעת גודל האוכלוסייה של כל סוג רכב. נרמול זה מאפשר לבחון את ה"פרופיל הבטיחותי" של כל סוג רכב במנותק מהשכיחות האבסולוטית שלו בכביש.

Why - למה מוצגת הויוזאליזציה?

המטרה המרכזית היא לבצע ניתוח קומפוזיציה יחסי, כדי להשוות את חומרת התאונות בין סוגי רכב שונים. ויוזאליזציה זו קריטית להסרת הטיות (Bias) שנבעו מהפרשי הכמויות בין הקטגוריות, היא מאפשרת לענות על השאלה: "האם רכבים פרטיים באמת מסוכנים יותר, או שהם פשוט נפוצים יותר?". זהו כלי חיוני לקביעת סדרי עדיפויות בבטיחות בדרכים שאינם מבוססים רק על נפח אירועים אלא על רמת סיכון מהותית.

How - כיצד הויוזאליזציה ממומשת?

הויוזאליזציה מומשה באמצעות גרף עמודות מוערם מנורמל (Stacked Bar Chart 100%), המשתמש בקידודים הבאים:

- קידוד אורך ומיקום (Length & Position): כל העמודות מיושרות לאותו גובה (100%), כאשר אורך המקטע בתוך העמודה מייצג את אחוז התאונות באותה רמת חומרה מתוך סך התאונות של אותו סוג רכב.
- חלוקה פנימית (Internal Division): כל עמודה מחולקת לשני חלקים מובחנים המייצגים את היחס הישיר בין תאונות עם נפגעים (פצועים/הרוגים) לבין תאונות ללא נפגעים. חלוקה זו מאפשרת השוואה ויוזאלית של "יחס הקטלניות" בין הקטגוריות השונות.
- קידוד צבע (Color Encoding): נעשה שימוש בניגודיות צבעונית, שימוש בצבע כתום עבור תאונות עם נפגעים, ובצבע צהוב עבור תאונות ללא נפגעים. בחירה זו ממקדת את תשומת הלב של המשתמש בסטייה מהנורמה ובקטגוריות בעלות סיכון גבוה.
- ציר Y מנורמל: בניגוד לגרפים הקודמים, ציר ה-Y אינו מציג "מספר תאונות" אלא ערך יחסי (Percentage), מה שמייצר בסיס השוואה הוגן בין רכבים נפוצים לרכבים נדירים.

נימוק לבחירת הויוזאליזציה

הבחירה בגרף עמודות מוערם מנורמל ל-100% היא הדרך האפקטיבית ביותר להציג השוואה בין פרופורציות. בניגוד לתרשים עץ (Treemap) שמדגיש גודל אבסולוטי, המבנה הנוכחי "מיישר" את כל הקטגוריות לקו בסיס אחד, ובכך

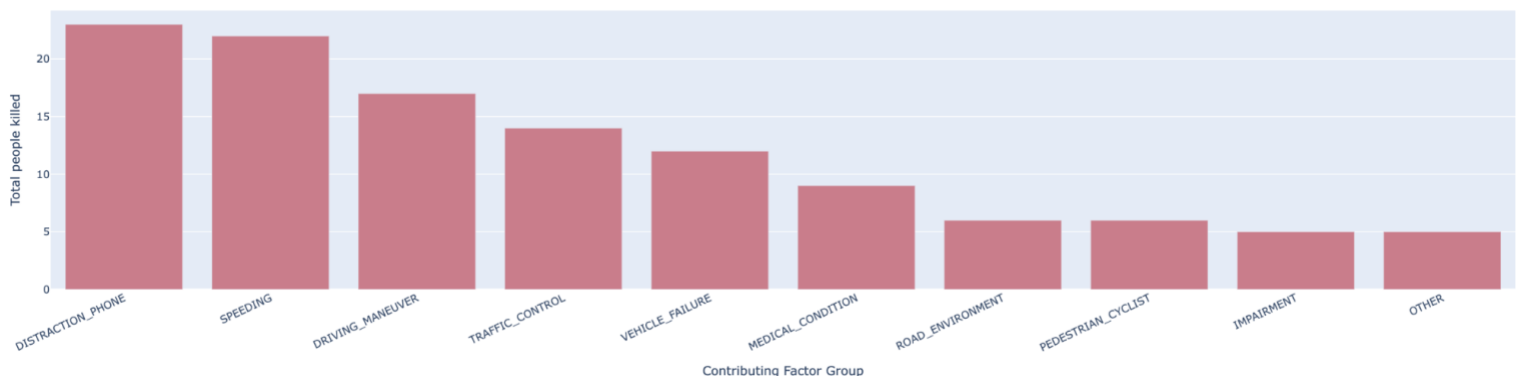
מאפשר לעין האנושית להבחין בקלות בהבדלים קטנים באחוז הנפגעים. נימוק זה עומד בבסיס הצורך לתקן את העיוות הוויזואלי שיוצר נפח התנועה הגדול של הרכבים הפרטיים, ולהבליט במקומם רכבים שמעורבותם בתאונות היא אולי נמוכה כמותית, אך חמורה יותר מבחינה תוצאתית.

מסקנות מהוויזואליזציה

הנרמול מוביל למסקנות הפוכות ומשלימות לאלו של הגרפים הקודמים, ניתן לראות כי קטגוריות כמו TWO_WHEEL_MOTOR (כגון אופנוע) ו-BICYCLE מציגות אחוז גבוה משמעותית של תאונות עם נפגעים מתוך סך האירועים שלהן, למרות נוכחותן המצומצמת בגרפים האבסולוטיים. לעומת זאת, רכבים פרטיים, למרות היותם הרוב המספרי, מציגים אחוז גבוה של תאונות ללא נפגעים (Property Damage Only). המסקנה היא שחומרת התאונה מושפעת באופן דרמטי מסוג הרכב ומההגנה הפיזית שהוא מספק למשתמשי הדרך, וכי מאמצי הבטיחות צריכים להתמקד בקטגוריות שבהן "יחס הנפגעים" הוא הגבוה ביותר, ללא קשר לשכיחות הכללית בכביש.

וויזואליזציה 6 - If you don't want to Kill, Don't text and drive

If You Don't Want to Kill, Don't Text and Drive



What - מה הנתונים?

הוויזואליזציה השישית מתמקדת בתוצאה החמורה ביותר של תאונות הדרכים, אובדן חיי אדם. הנתונים מציגים את מספר הנפגעים בנפש (הרוגים) מפולחים לפי הגורמים התורמים המרכזיים (Contributing Factors). לצורך כך, ביצענו הצלבה בין משתנה ה-Severity (סיכון למקרי מוות בלבד), לבין קטגוריות המאקרו של הגורמים התורמים שעברו אגרגציה (כגון DISTRACTION_PHONE, SPEEDING, ALCOHOL ועוד). בניגוד לגרפים קודמים שכללו פצועים או נזק לרכוש, וויזואליזציה זו מבודדת אך ורק את מקרי המוות כדי לבחון אילו התנהגויות נהיגה הן "קטלניות" ביותר.

Why - למה מוצגת הוויזואליזציה?

המטרה המרכזית היא לבצע ניתוח קטלניות (Fatality Analysis) ברור. המשימה האנליטית היא לדרג את גורמי הסיכון לפי מחיר הנפשות שהם גובים בפועל. מטרת העל של הוויזואליזציה היא תקשורתית ו"חינוכית", להמחיש לנהג כי פעולות שנראות לעיתים "קלות דעת", כמו שליחת הודעה בזמן נהיגה, אינן מסתיימות רק בתאונה או בפציעה, אלא מובילות באופן ישיר למוות. הבחירה להתמקד בהרוגים בלבד נועדה לייצר זעזוע וויזואלי והבנה של הסיכון המוחלט.

How - כיצד הוויזואליזציה ממומשת?

הוויזואליזציה מומשה באמצעות גרף עמודות (Bar Chart) הממוקד במדד הקטלניות:



- קידוד אורך (Length Encoding): אורך העמודה מייצג את מספר האנשים שנהרגו כתוצאה מכל גורם. זהו הקידוד הישיר והברור ביותר להשוואת כמויות של אובדן חיי אדם.
- מיון יורד (Ranking): העמודות מסודרות מהגורם שהוביל להכי הרבה הרוגים וכלפי מטה, מה שמדגיש באופן מיידי את הסיבות להרג בכביש.
- צבעוניות סמנטית: החלטנו להשתמש בצבע אדום המזוהה אינטואיטיבית עם סכנת חיים ודם, כדי לחזק את המסר של "Don't Kill".

נימוק לבחירת הויזואליזציה

הבחירה בגרף עמודות המציג הרוגים בלבד נבעה מהצורך בפשטות ובעוצמה חזותית. בעוד שגרפים מורכבים (כמו גרפים מוערמים או מנורמלים) טובים לניתוח סטטיסטי רחב, בויזואליזציה זו המטרה היא להעביר מסר חד משמעי. הצגת המספר האבסולוטי של ההרוגים מנטרלת כל ניסיון ל"ריכוך" הנתונים ומראה את המציאות כפי שהיא – אילו עבירות נהיגה גורמות למוות בפועל ברחובות ניו יורק.

מסקנות מהויזואליזציה

מהניתוח עולה תמונה מדאיגה, גורמים הקשורים להסחת דעת ונהיגה בחוסר תשומת לב (ובכלל זה שימוש בטלפון) נמצאים בראש רשימת הגורמים למוות בדרכים. המסקנה המרכזית היא שקיימת קורלציה ישירה ומובהקת בין שימוש בטלפון לבין תאונות קטלניות. הויזואליזציה מוכיחה כי האזהרה "Don't text and drive" אינה סיסמה בלבד, אלא מבוססת על נתונים המראים כי הסחת דעת היא אחד הגורמים המובילים בכביש להרג, יותר מגורמים טכניים או תשתיתיים אחרים.

ביבליוגרפיה

1. מאמרי המחקר מהדוח:

- Shaaban, K., & Ibrahim, H. (2021). Analysis and identification of contributing factors of traffic crashes in New York City. *Transportation Research Interdisciplinary Perspectives*, 12, 100466. <https://doi.org/10.1016/j.trip.2021.100466>
- : Zheng, L., et al. (2024). Patterns and analysis of traffic accidents in New York City between 2013 and 2023. *Urban Science*, 8(1), 14. <https://doi.org/10.3390/urbansci8010014>

2. מקור הנתונים (NYC Open Data):

- New York City Police Department (NYPD). (2024). *Motor Vehicle Collisions - Crashes*. NYC Open Data. https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data