



אוניברסיטת בן גוריון  
קורס 'יזואלייזציה של המידע'

פרויקט יזואלייזציה של תאותות דרכים בעיר ניו יורק

מגישות:

אורן, ת.ז. 315994681  
עדי בכר, ת.ז. 204789473

שם המרצה: פרופ' גלעד רVID

24.02.2026



## תוכן עניינים

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



## תיאור הנתונים והטיפול בהם

מאג'ר הנתונים שבו עשוינו שימוש בעבודה זו, הוא מאג'ר 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 הנитוח שנבחנה היא רמת האירוע (Crash-Level). המאג'ר כולל משתני זמן, משתני מקום, נתוני חומרה כגון מספר פצעים והרוגים, שדות טקסטואליים של גורמים תורמים לתאונה וכן שדות המתארים את סוג כלי הרכב המעורבים.

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

שלב ניקי הנתונים כלל בראש ובראשונה סטנדרטיזציה של שדות טקסטואליים. ערכים בשדות כגון CONTRIBUTING FACTOR VEHICLE NONE, UNKNOWN ו-UNSPECIFIED. מהלך זה נועד להפחית שונות מלאכותית הנובעת מצורות כתיב שונות ולאחרד ערכים זרים מבניה לוגית. נציג כי בשדות החזירים (כגון VEHICLE TYPE CODE 5-1-5) קיימת תופעה של חוסר מבני (Structural Missingness), הנובעת מכך שלא בכל תאונה מעורבים חמייה כלי רכב. לפיכך, ערכים חסרי אינדקסים גבוהים אינם בהכרח בעלי איות נתונם אלא תוצאה טבעית של מבנה הסכמה.

חלק מרכזי בהכנות הנתונים היה הפעחתה ריבוי קטגוריות (High Cardinality) באמצעות ארגזיה מושכלת. בשdots הגורמים התורמים לתאונה בוצע מיפוי לערכי מקאו באמצעות יצירת פונקציה חדשה בשם map\_contributing\_factor . זאת על ידי איחוד ערכים טקסטואליים רבים לקבוצות משמעותיות (כגון DISTRACTION\_PHONE, SPEEDING, FATIGUE\_SLEEP ו-VEHICLE\_TYPE\_CODE 5-1-5). מהלך זה אפשר מעבר מרחב קטגוריאלי רועש ואפוץ ליצוג אינפורטיבי ובר השוואה, אשר שימוש בהמשך ליזואלייזציות של גורמים תורמים לפי חומרת תאונה. באופן דומה, בוצעה ארגזיה של סוג כלי רכב באמצעות יצירת פונקציית VEHICLE\_TYPE\_CODE PRIVATE\_VEHICLE, HEAVY\_VEHICLE, TAXI\_LIVERY או EMERGENCY\_VEHICLE. גם ארגזיה זו צמצמה שונות טקסטואלית גבוהה ואייפשרה ניתוח השוואתי בין קבוצות כלי רכב.

בסוף, בשלב הנדסת המאפיינים בוצע עיבוד למשתני הזמן. השdots CRASH\_DATE ו-CRASH\_TIME אוחדו לייצוג זמן תקני, ומתחום הפקו מאפיינים נגדרים הכלולים את שעת התאונה (hour), יום בשבוע (weekday) וחודש בשנה (month). פירוק זה אפשר מעבר מיצוג טקסטואלי גולמי לייצוג מספרי וקטגוריאלי המאפשר ניתוח דפוסים עונתיים ותבניות זמן.

לבסוף, בוצעו חישובי ארגזיה שונים ברמת התאונה, כגון סכימת מספר הפצועים והרוגים, חישוב מספר כלי רכב מעורבים לכל תאונה, וחישוב מדדים מצטברים לפי קטגוריות שונות (גורמים תורמים, סוג כלי רכב, חומרה ועוד). תהליכי אלו יצרו מערך נתונים עיקבי, מנוקה ומאורגן, אשר אפשר בנית ויזואלייזציות אינפורטטיביות הממחישות דפוסים מרחביים, זמניים וסבירתיים בתאונות הדרכים בעיר ניו יורק.

## דין בעבודות קשורות וויזואלייזציות קיימות לאותם הנתונים

ספרות המחקר בתחום תאונות הדרכים בעיר ניו יורק מראה שימוש מתקדם באנליזה מרחבית, זמנים וסיבות של תאונות, ולעתים משולב גם עם ויזואלייזציות להבנת התנהלות 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 עמוק של תאונות דרכים בעיר ניו יורק בשנים 2013–2019, תוך בחינת גורמים תורמים, דפוסים זמניים והתפלגות חומרה. אחת הויזואלייזציות המרכזיות במחקר (Fig. 5 – Crashes Per Type of Day) מציגה השוואה ביןימי חול לימי שבוע במספר הפציעים והרוגים. בגרף ניתן לראות כי המספר האבסולוטי של נפגעים גבוה יותר ביום חול, הן בעבר פצועים והן בעבר הרוגים. עם זאת, החוקרם מדגישים כי יש לפרש ממצא זה בזיהירות, שכן ימי חול מהווים חמישה ימים בשבוע לעומת שישיים בלבד בסוף השבוע, ולכן ההשוואה אינה מנורמלת לאורך זמן. הויזואלייזציה משתמשת בעמודות מקובצות לפי שנים וחלוקת ברורה בין Weekdays ל-Weekends, ומדגימה כיצד פילוח זמן פשוט וחסיטות יכול לחושף דפוסים חשובים אך גם להטעות אם אינו מלאה בניתוח ויחס (כגון ממוצע יומי).

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

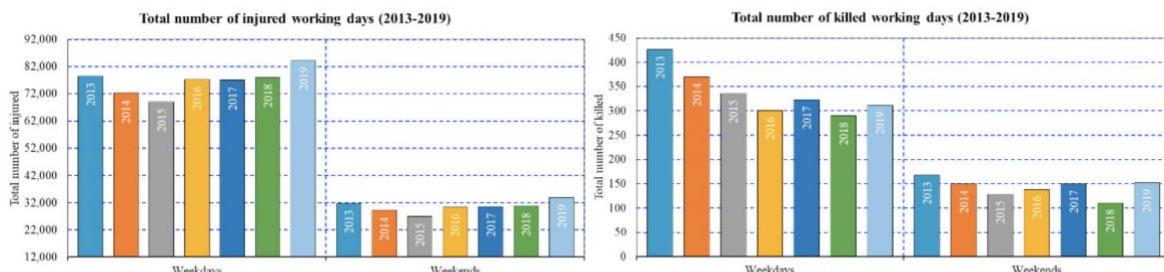


Fig. 5. Total number of injured and killed per type of day (2013-2019).

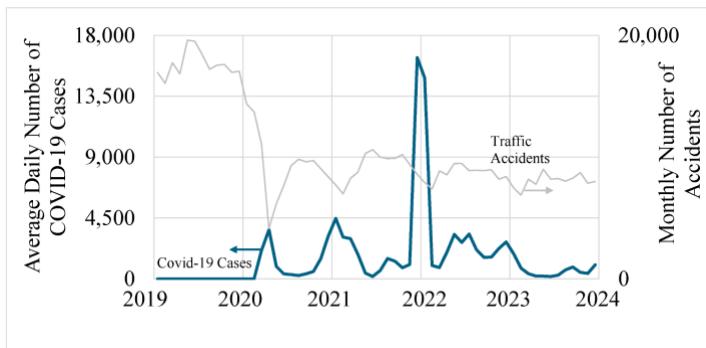
המאמר השני שבחרנו הוא 2013 Patterns and Analysis of Traffic Accidents in New York City between 2013 and 2023, שפורסם ב-2024 Urban Science: 2023 York City between 2013 and 2023

המחקר עוסק בניתוח דפוסי התאונות בעיר ניו יורק לאורך עשור, בשילוב משתני זמן, חומרה וסיבות תאונה. המחקר כולל ויזואלייזציות של מגמות התאונות, ובין היתר מציג גרפים המתארים את הקשר בין מגפת COVID-19 למספר התאונות החודשי (Fig. 9) וכן את ההשפעה של נגיף קורונה מסוימת 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 החודשיות. הויזואלייזציה ממחישה באופן ברור כי ירידת חודה במספר התאונות בתחילת 2020 חלה במקביל לעלייה החדרה במרקעי COVID-19, מה שמרמז על השפעת הסגר והפחיתה תנועת כלי הרכב על פעילות התאונות. עם זאת, לצד התחזקות אחרי הקורונה, ניתן לראות כי לאחר החזרה לתנועה נורמלית מספר התאונות התייצב ואף צמצם

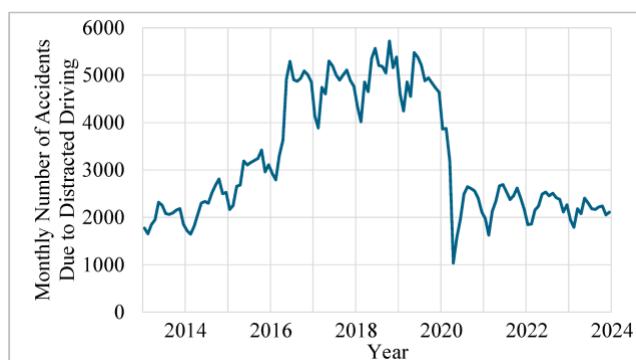
וחזר לرمות קודמות, מה שמדגיש את הקשר המורכב בין גורמים חברתיים (כגון מגפת הקורונה) לדפוסי תאונה בתחוםה.

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



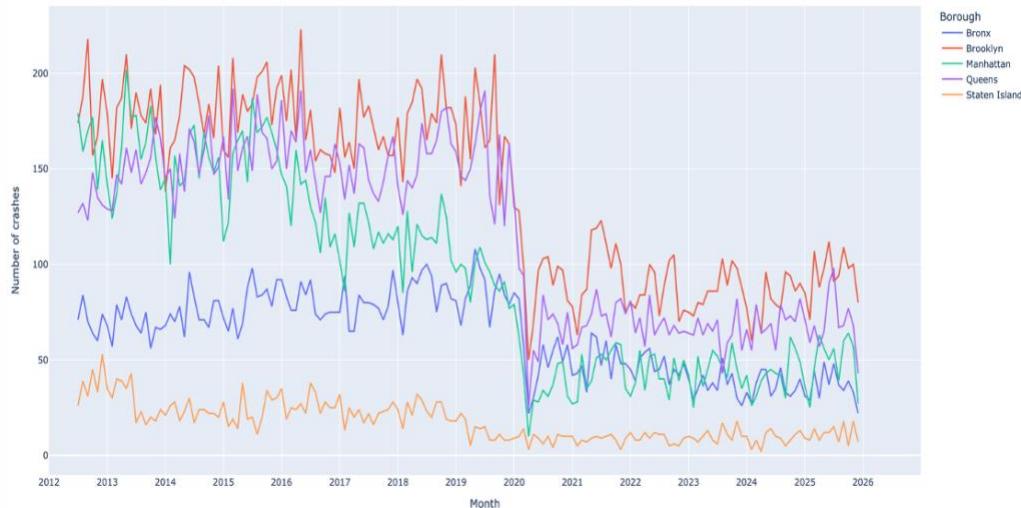
גרף נוסף מכחיק זה הוא Fig. 7, המציג גרפּ קווּי של מספר התאונות החודשי הנובע מנהוגה מוסחת, המתיחס בעיקר לשימוש טלפון נייד. הגרף מדגים מגמת עלייה בשנים 2013–2018, ירידת חדה במהלך 2020 כתוצאה מ מגבלות התנועה, ולאחר מכן יציבות ויזואליזציה זו תומכת במחקר בטיחות תחבורה אחרים שמצאו כי הסחת דעת, ובפרט שימוש טלפונים ניידים, מהוות גורם משמעותי בתאונות דרכים.

ויזואליזציה זו משרות באופן ישיר לפרויקט שלנו, שכן גם בפרויקט זה בוצע מיפוי של גורמים תורמים (Contributing Factors) אשר העלה כי קטגורית הסחת הדעת ובפרט שימוש טלפון נייד, לא רק מהוות אחד הגורמים השכיחים ביותר לתאונות, אלא גם מציגה רמת חומרה גבוהה יחסית בהשוואה לגורמים אחרים. נמצא זה מתוישב עם הספרות המחקרית ועם הויזואליזציה המוצגת במאמר, המחזקת את ההבנה כי שימוש טלפון בזמן נהיגה הוא גורם סיכון מרכזי לתאונות בעלות פוטנציאל פגעה משמעותית. בכך נוצרת עקבות בין הממצאים במחקר החיצוני לבין הדפוסים שהתגלו בניתוח שבוצע במסגרת פרויקט זה.



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

A Rare Bright Spot of COVID-19: Fewer Traffic Crashes on NYC Streets in 2020



### **What - מה הנתונים?**

לצורך גורף זה בוצעה ארגזציה חוזרת של מספר התאונות, תוך חלוקה לפי מחוז (Borough). האובייקטים (Entities) המייצגים בגרף הם יחידות זמן חודשיות, כאשר לכל חודש קיים ערך קבוע של מספר התאונות בכל אחד מחמשת המחווזות: Bronx, Brooklyn, Manhattan, Queens ו-Staten Island.

לכל אובייקט זמן מסוים שלושה מאפיינים עיקריים: משתנה זמן מסוג Temporal (שנה וחודש), משתנה קטגוריאלי מסוג Nominal (מחוז), ומשתנה ככויות מסוג Quantitative (מספר תאונות). בניית הנתונים הוא מבנה של סדרת זמן (Time Series), המאפשר ניתוח של מגמות לאורך זמן והשוואה בין קטגוריות גיאוגרפיות מקבילות.

### **Why - למה מוצגת הויזואלייזציה?**

מטרת הויזואלייזציה היא לאפשר חקירה של דפוסים זמניים והשוואה בין אזורים שונים בעיר ניו יורק. המשמש יכול לזהות מגמות ארוכות טווח במספר התאונות, להשוות בין מחוזות שונים ולבחון האם שינויים מתרכשים באופן אחיד בעיר או באופן מקומי. הויזואלייזציה תומכת במספר משימות אנליטיות מרכזיות: זיהוי מגמות (trend detection), השוואה בין קבוצות (comparison), איתור נקודות חריגות בזמן (anomaly detection), ותקשות מרופד (outliers). התייחסות לתאונות בעיר לאורך השנים.

### **How - כיצד הוויזואלייזציה ממומשת?**

היצוג החזותי נעשה באמצעות גרף קו (Line Chart), אשר כל מחוז מיוצג באמצעות קו צבע שונה. ה-marks המרכזים הם קוים 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 מדדים במקביל: זמן, כמות וקטgorיה.

הפרישה היא לאורך ציר זמן, המתאימה במיוחד להציג סדרות זמן ולהמחשת שינוי רציף. בנוסף, הויזואלייזציה אינטראקטיבית וכוללת אינטראקטיות להציג נתונים מדויקים, וכן אפשרות לSpin-on באמצעות legend, המאפשרת להציג או להסתיר מחוזות שונים. אינטראקטיביות זו מפחיתה עומס חזותי ומאפשרת חקירה מוקדמת של מחוזים מסוימים או השוואה דינמית בין מספר מחוזות.



בחירת גרפּ קווי על פני גרפּ עמודות נבעה מה הצורך להציג רצף זמן ארוך עם מספר רב של נקודות זמן, באופן שמדגיש מגמות ולא רק ערכים נקודתיים. השימוש בצבעים מוחכמים לכל מחודן נועד להקל על זיהוי והשוואה, מבלתי לגען בקריאות הגרפּ.

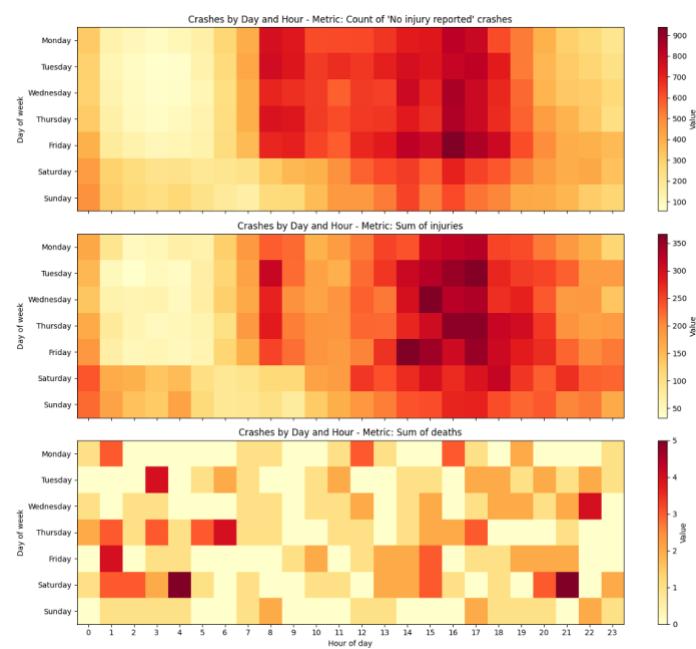
### הסבר הבחירה שנעשָׁה

הבחירה בגרפּ קווי נבעה מהתOLON להציג מגמות לארוך זמן בצורה ברורה, שכן שימוש בגרפּ עמודות היה יוצר עומס חזותי בשל מספר רב של נקודות זמן. בנוסף, השימוש בצבעים שונים לכל מחודן מאפשר הבחנה מיידית בין הקווים ותומך בהשוואה חזותית. הבחירה להציג את הנתונים לפי Borough ולא ברמת העיר בלבד מאפשרת להזזה הבדלים בין אזורים שונים ולהימנע מהציג ממציאות שעלולה להסתדר דפושים מקומיים ממשמעותיים. הבחירה בגרפּ אינטלקטיבי מאפשרת חקירה מעמיקה יותר על ידי המשתמש, כולל אפשרות להתמקדם במחודן מסוים ולהשוותו לאחרים בצורה נוחה.

### תובנות מרכזיות

מהויאלייזציה ניתנת להסיק כי קיימים פערים ברורים בין המחוות, כאשר Brooklyn-Queens מציגים לארוך השנים מספר תאונות גבוהה יחסית, בעוד Staten Island מציג מספר תאונות נמוך משמעותית. בנוסף, ניכרת ירידה חדה בכמות התאונות בשנת 2020 כמעט בכל המחוות, תופעה התואמת את השפעת מגפת הקורונה והטירים אשר הפיחתו באופן משמעותי את תנועת כלי הרכב בעיר. לאחר מכן ניתן לראות חרזה הדרגתית לרמת תאונות דומות לשנים קודמות, מה שמצוין על התוצאות הפעילות העירונית.

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



### What - מה הנתונים?

מתוך המשתנים CRASH TIME ו-CRASH DATE הופקו משתנים קטגוריאליים המייצגים את יום השבוע (weekday) ואת השעה ביום (hour). לפיק, ייחdet הניתוח בגרפּ זה היה תא זמן בדיד המוגדר כשילוב בין יום בשבוע לבין



שעה מסויימת בטוווח 0-23.

לכל שילוב זמן חשובו שלוש מטריקות שונות: מספר התאונות ללא נפגעים, סכום הפגיעהים, וסכום הרוגים. מבנה הנתונים הוא מבנה מטריציוני דו-ממדי, שבו השורות מייצגות את ימי השבוע והעמודות מייצגות את שעות היוםה. כל תא במטריצה משיקן לערך מסווני המיצג את היקף התאונות או את חומרתן באותו חלון זמן. בכך, הייזואלייזציה מאפשרת בחינה סימולטנית של שני ממדים זמינים קטגוריאליים ושל מدد כמותי.

### Why - למה מוצגת הייזואלייזציה?

מטרת הייזואלייזציה זו היא לאזהות דפוסים זמינים (Temporal Patterns) בהתקלגות התאונות הדריכים ובהיקף חומרתן לאורך השבוע והיממה. הייזואלייזציה נוסעת לאפשר חקירה אנאליטית של הבדלים בין ימי חול לסופי שבוע, בין שעות עומס לשעות לילה, ובין חומרת התאונות.

הגרף תומך במספר ממשימות אנאליטיות מרכזיות: זההו תבניות חוזרות, השוואה בין קטגוריות זמן, איתור אירועי זמן חריגים, והדגשת חלונות זמן המאפשרים ברמת סיכון גבוהה יותר. מעבר לכך, הייזואלייזציה מאפשרת תקשורת מסר ברור בדבר חשיבותו מימד הזמן בהසבר דינמית התאונות בעיר ניו יורק.

### How - כיצד הייזואלייזציה מוממשת?

הייזואלייזציה מוממשת באמצעות Heatmap, "ציג מטריציוני שבו כל תא מייצג שטח, והערך הcoresי מוקוד באירועים צבע. הציר האופקי מייצג את השעה ביום, והציר האנכי מייצג את יום השבוע. עצמת הצביע משמשת כערוך קידוד כמותי, כך שכך שהצביע כהה יותר, הערך גובה יותר.

נכחר סולם צבעים בגווני אדום, כאשר אדום כהה מייצג ערכים גבוהים יותר של תאונות, פצעים או הרוגים. הבחירה בגווני אדום אינה מקרית אלא נושא ש商量ות טמנטי, אדום הוא צבע המזוהה תרבותית וחוזותית עם סכנה, אזהרה ופגעה בחיי אדם. לפיכך, השימוש בגוון אדום כהה להדגשת ריכוזי תאונות גבוהים מחזק את הקשר האינטואיטיבי בין עצמת הצביע לבין חומרת האירוע, ותורם להבנה מהירה של מוקדי סיכון. בחירה זו גם תומכת בהיררכיה תיפיסתית ברורה, אזורים כהים יותר מושכים את תשומת הלב באופן מיידי ומשמעותי "אזורים חמימים" (Hot Spots), בעוד אזורים בהירים מושדרים רמת סיכון נמוכה יותר. בכך מושגת הדגשה הייזואלית של מוקדים קריטיים מביי להעשים בפרטם גרפים נוספים.

בנוסף, ההפרדה שלוש מפות נפרדות, תאונות ללא נפגעים, פצעים ורוגים, מאפשרת השוואה אנאליטית בין נפח התאונות לבון חומרתן, מביי לערבב סקאלות שונות באותו יציג חזותי.

### הסביר הבחירה שנעשות

הבחירה ב-Heatmap נבעה מה הצורך לנתח בו-זמןית שני ממדים קטגוריאליים דיסקרטיים (יום ושעה), תוך שמירה על הבחירה חזותית וധויו התבניות. לאחר שמבנה הנתונים הוא מטריציוני מטבחו, Heatmap מהווה התאמת טבעית למבנה זה.

בנוסף, בפרויקט זה בוצעה הנדסת פיצ'רים זמינים מתוך תאריך ושעה, מתוך הנחה כי מימד הזמן מהוות משתנה מרכזי בהסביר חומרת תאונה. לפיכך, הייזואלייזציה אינה רק תיאורית, אלא גם תומכת בהחלות מידול עתידיות, בכך שהיא מדגימה אמפירית את חשיבות משתני הזמן.

### תובנות מרכזיות

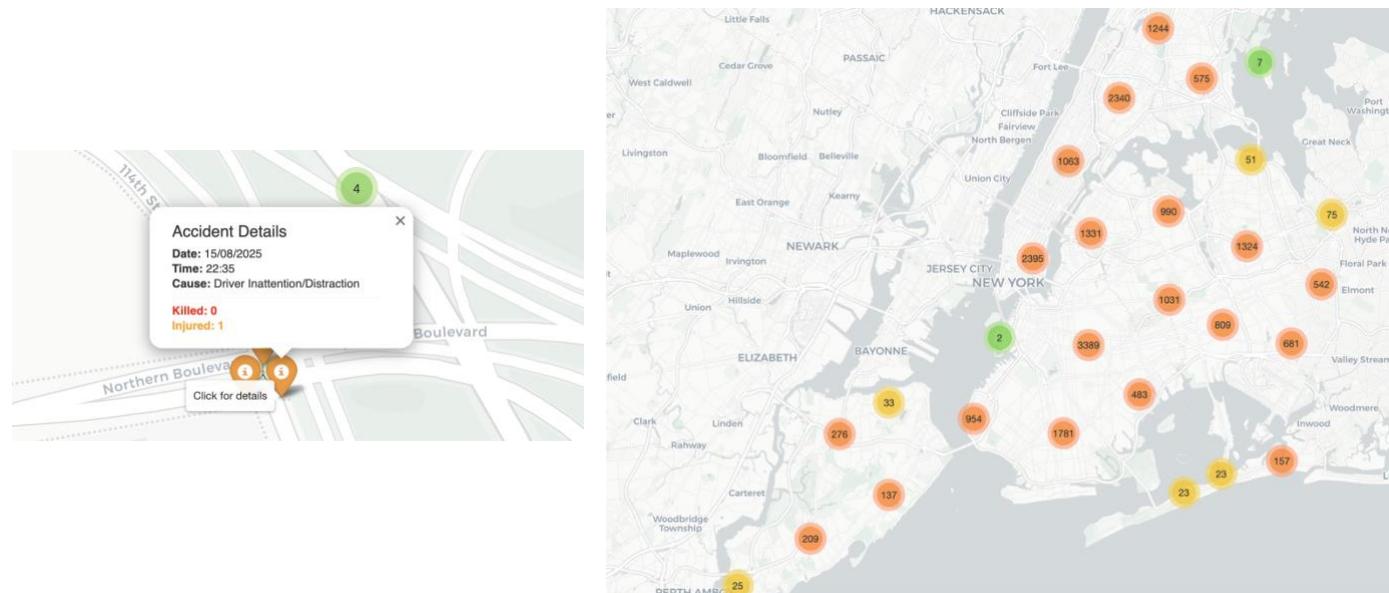
ניתוח הייזואלייזציה מוביל למספר מסקנות מרכזיות. ראשית, קיימת מחזריות ברורה לאורך השבוע והיממה, כאשר שעות אחר הצהרים והערב בימי חול מאופינות בנפח תאונות גבוהה במיוחד. בנווסף, ניתן לאלהות כי גם שעות הבוקר המוקדמות בימי חול מציגות עלייה במספר התאונות. דפוס זה מתישב עם שעות השיא התחבורתיות בעיר ניו יורק, המתאפיינות ביציאה לעבודה בשעות הבוקר ובחזור מכנה בשעות אחר הצהרים. עומסי תנועה, צפיפות גובהה בכבישים וריבוי אינטראקציות בין כלי רכב ומשתמשי דרך נוספים בשעות אלו מגדילים את ההסתברות להתרחשויות



תאונות.

שנית, דפוס ההרוגים אינו חופף לצלולין לדפוס נפח התאונות, דבר המרמז כי תנאים מסוימים, בעיקר בשעותليل ובסופי שבוע, עשויים להיות קשורים לרמת חומרה גבוהה יותר גם אם נפח התאונות נמוך יחסית. ממצוא זה מצביע על כך ש衲ח תאונות גבוה אינן בהכרח מתורגם לחומרה גבוהה, וכי משתנים סביבתיים והתנהגותיים נוספים עשויים להשפיע על תוצאה האירוע.

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



#### **What - מה הנתוניים?**

הויזואלייזציה מתמקדת בתת קבוצה של משתנה המטריה, הכוללת "תאונות חמורות" בלבד, תאונות שהן גרשמו פצועים או הרוגים (Injured/Killed < 0). הנתונים המוצגים הם ברמת האירוע הבודד וכוללים חמישה ממדיים עיקריים: מיקום גיאוגרפי מדויק (Latitude & Longitude), חומרת הפגיעה (מספר פצועים והרוגים), זמן התרחשות (תאריך ושעה), והגורם התורם המרכזי לתאונה (Contributing Factor).

#### **Why - למה מוצגת הויזואלייזציה?**

מטרת הויזואלייזציה היא לאפשר חקירה מרחבית אינטראקטיבית של מוקדי סיכון בעיר. בעוד שגרפים סטטיסטיים מאפשרים להבין מגמות כלליות, מפה אינטראקטיבית מאפשרת למשתמש לבצע משימות של זיהוי ריכוזים (Clustering) וחיפוש נקודתי (drill-down). המשימה האנאליטית היא לאתר "מוקדים חמימים" (Hotspots) של תאונות קטלניות ומחוראות, ולהבין האם קיימת גיאוגרפיה המקשרת בין מיקום התאונה לבין גורמי הסיכון (כגון צמתים عمומיים או עורקי תחבורה ראשיים), ובכך לתמוך בתחום החלטות בתחום תשויות הבטיחות בדרכים.

#### **How - כיצד הויזואלייזציה מוממשת?**

הויזואלייזציה מומשה באמצעות ספריית Leaflet ומושתמשת בקידודים הבאים:

- **מקום (Spatial Position):** הנקודות מוקדדות לפי מיקומן הגיאוגרפי המדויק על גבי מפת בסיס (Base Map).



- צבירת נתונים (Marker Clustering): כדי להתמודד עם בעית ה-Overplotting (עומס חזותי של נקודות חופפות), עשינו שימוש באלגוריתם צבירה המאגד נקודות קרובות לאשכולות ממוספרים. גודל האשכול וצבעו משתנים דינמיים בהתאם לרמת ה-Zoom, מה שמאפשר מעבר רציף ממצגת מאקרו עירונית לתצוגת מיקרו רחוב.
- צבע וסמלים (Color & Glyph Coding): נעשתה הבחנה קטגוריאלית בין דרגות חומרה:
  - כתום (איקון Info): מסמל תאונות עם פצעים בלבד.
  - אדום (איקון Exclamation): מסמל תאונות קטלניות הוללות הרוגים, ובכך מייצר היררכיה חזותית המדגישה את האירועים הקritisטים ביותר.
- אינטראקטיביות: שימוש ב-Tooltips ו-Popups להציג הנחיה ללחיצה, ו-Icons המציגים רקע מפורט על פרטיה התאונה, הגורם לה והנפגעים, מה שמאפשר הנגשת מידע רב מבלי להעמיס על הממשק הראשון.

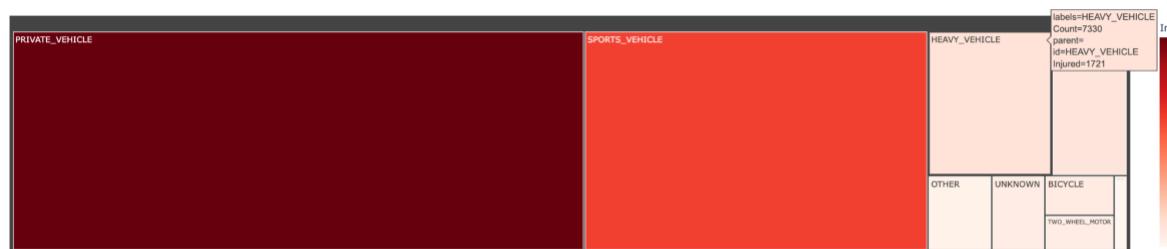
### nymok לבחרת הייזואלייזציה

בחירה במפה אינטראקטיבית מבוססת אשכולות (MarkerCluster) נבעה מה הצורך לגשר בין כמות הנתונים לבין יכולת התפיסה החזותית של המשתמש. מפה סטטיסטית הייתה נכשלת בהציג אלפי נקודות למרחב האובייקטיבי של ניו יורק. המפה מאפשרת גמישות, המשתמש יכול לקבל תמונה שלמה של רובע העיר, ובמקביל לבחון אירועים בודדים ברמת הרחוב. השימוש באירועים אדום מול כתום נבחר כדי לייצר "קריאה לפועלה" חזותית כלפי המוקדים הקטלניים, בהתאם לעקרונות של קשב טרום קשי (Pre-attentive processing).

### מסקנות מההייזואלייזציה

ניתוח המפה מעלה כי תאונות חמורות אין מפוזרות באופן אחיד, קיימת נטייה ברורה לרכיבים באזורי המעבר בין רובעי העיר (גשרים ומנהרות) ובאזורים מרכזיים במנהטן ובברוקלין. מה-Popups עולה כי במקרים רבים של תאונות קטלניות (באדום), הגורם התורם לשנתה דעת או למהירות מופרזת, מה שמעיד על כך שהగורם האנושי מהווים משתנה קריטי בחומרת התאונה במקומות אלו. האינטראקטיביות חשופת כי גם באזורי שנראים בטוחים ממבט עליון קיימים אזורי סיכון נקודותים הדורשים התיחסות תשתיתית ספציפית.

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



### What - מה הנתונים?

הייזואלייזציה הריבעת מנהצת את המסתנה הקטגוריאלי המתאר את סוג כלי הרכב המעורבים בתאונות, תוך יישום תהליכי של הנדסה מאפיינים (Feature Engineering) שנועד להתגבר על חוסר האחדות הקיים בנתונים. במסגרת עיבוד הנתונים, הפעלנו פונקציית מיפוי שביצעה סטנדרטיזציה של הערכים הטקסטואליים, וקידזה שורות תוויות גולמיות לתשע קטגוריות על. תהליך זה אישד לדוגמה רכבים פרטיים, ג'יפים ואננים תחת הקטגוריה PRIVATE\_VEHICLE, ריצ' רכבי מטען, אוטובוסים ותחבורה כבדה תחת HEAVY\_VEHICLE, וסיגר רכבי הצלה EMERGENCY\_VEHICLE וטווילרים (TAXI\_LIVERY) בהתאם, לצד קטגוריות ספציפיות לאופניים, אופניים, ומוניות לרכבות TAXI\_LIVERY ו-EMERGENCY\_VEHICLE.



רכבי ספורט וטיפול בערכיים חסרים (UNKNOWN). היצוג הסופי בייזואלייזציה מציג את השכיחות הרכמיות של אירועי תאותות עבור כל אחת מקבוצות אלו, ובכך מאפשר מעבר ממארג נתונים רועש למידע סטטיסטי בעל משמעות אנליטית.

### Why - למה מוצגת הייזואלייזציה?

המטרה המרכזית היא לבצע ניתוח השוואתי (Comparative Analysis) המאפשר להבין את פלח השוק של כל סוג הרכב במפתח הכלכליים העירוניים. המשימה האנליטית היא דיהוי הפרופורציות של הקטגוריות השונות, ככלומר, להמחיש בצורה הייזואלית מידית אילו סוג רכב מהווים את מרבית נפח האירועים לעומת קטגוריות אחרות. השימוש במיפוי הקטגוריות הכרחי למשימה זו, שכן הוא מאפשר להשוות בין קבוצות בעלות משמעויות בטיחותית ותכונית, ולזהות האם קיימת דומיננטיות חריגה של סוג רכב מסוים (כגון רכבים כבדים או מוניות) ביחס לתפיסה הרווחת של נפח התנועה.

### How - כיצד הייזואלייזציה ממומשת?

הייזואלייזציה מומשה באמצעות תרשימים עץ (Treemap), המשמש בקידודים החזותיים הבאים:

- קידוד שטח (Area Encoding): גודל המלבן של כל קטgorיה מקובד לפי מספר התאותות המשויך אליה. ככל שהשטח גדול יותר, כך שכיחות המעורבות של אותו סוגרכב גבוהה יותר.
- חלוקה מרחבית (Spatial Partitioning): המלבנים מסודרים בתוך פריים מרכזי אחד, מה שמאפשר השוואת מהירות של יחסיו "חלק מתוך חלק" (Part-to-Whole).
- צבעוניות קטגוריאלית: כל מלבן (אשכול) מקבל גוון שונה המאפשר הפרדה הייזואלית, ככל שהגוון כהה יותר הדבר מעיד על כמות פצעים גבוהה יותר עבור כל סוגרכב.
- תוויות (Labelling): שמות הקטגוריות המאוחדות והערכים הרכמיים מוצגים בתוך המלבנים, מה שמנגיש את המידע ללא צורך במעבר חוזר לציריהם או למקרה חיצוני.

### nymok לבחרת הייזואלייזציה

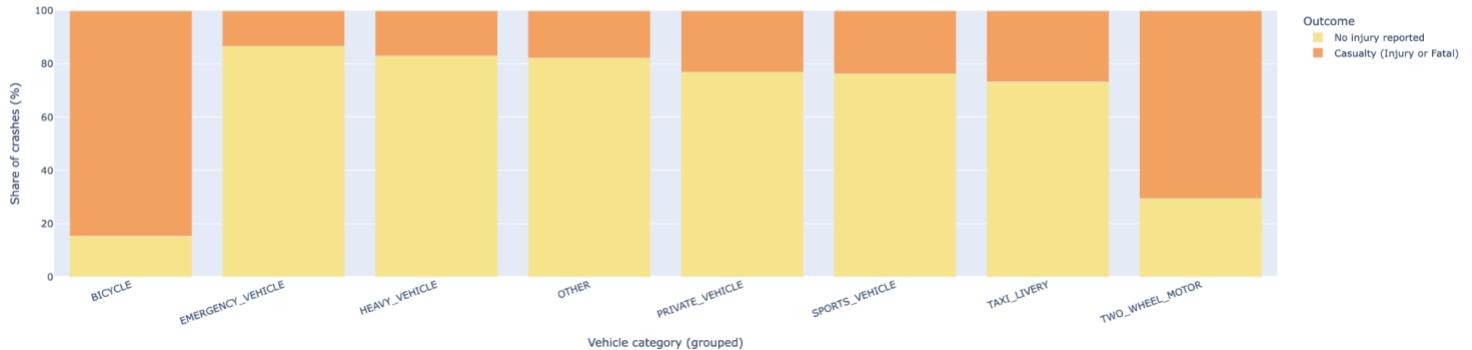
הבחירה בתרשימים עץ (Treemap) על פני גרפים קלאסיים אחרים נבעה מהיכולת שלו לנצל 100% מהשטח הייזואלי להציג פרופורציות מורכבות. בעוד שגרף עמודות מתקדם בהשווה בין ערכים בודדים, תרשימים העץ מאפשר למתבונן לחשוף באופן אינטואיטיבי את ה"משקל" של כל קטgorיה בתוך המערכת הכוללת. המבנה המרובע והדחוס של ה-Treemap מתאים במיוחד להציג הקטגוריות שנוצרו בתהליך המיפוי, שכן הוא מדגיש את הפער הניכר בין הקטגוריות הדומיננטיות (כמו רכבים פרטיים) לבין סוג הרכב המשניים, בצורה עיליה ואסתטית.

### מסקנות מההייזואלייזציה

ניתוח תרשימים העץ מעלה כי רכבים פרטיים (PRIVATE\_VEHICLE) תופסים את נתח השטח הנגדל ביותר, מה שמעיד על מעורבותם הגבוהה ביותר בתאותות מבחינה מספרית. עם זאת, חשוב לסייע לזה ולהציג כי הנתונים בייזואלייזציה זו אינם מנורמלים. הדומיננטיות החזותית של הרכבים פרטיים מושפעת באופן ישיר ובלתי נמנע מהכמויות האבסולוטיות של רכבים אלו בכביש העיר (ນפח התנועה). לכן, שטח המלבנים בgraf משקף את שכיחות האירועים בשטח אך אינו מעיד בהכרח על "ModelProperty" יחסית של סוגרכב אחד על פני אחר. תובנה זו מהווה את הבסיס לצורכי בייזואלייזיה הבאיה, שتنטרל את השפעת הרכמיות ותבחן את הנתונים בצורה מנורמלת.

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

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



### What - מה הנתונים?

הយיאלייזציה החמישית מציגה את נתוני התאונות לאחר תהליך של נרמול (Normalization). הנתונים מוחלקים לממד קטגוריאלי (סוג הרכב) ולמדד של רמת חומרה, המפצל לשתי קבוצות עיקריות: תאונות הכלולות 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 סוגרכב במנוקב מהשכיחות האבסולוטית שלו בכביש.

### Why - למה מוצגת היזואלייזציה?

המטרה המרכזית היא לבצע ניתוח קומפוזיציה יחסית, כדי להשוות את חומרת התאונות בין סוגי רכבי שונים. ייזואלייזציה זו קריטית להסתרת הטוית (Bias) שנבעו מהפרשיות בין הקטגוריות, היא מאפשרת לענות על השאלה: "האם רכבים פרטיים באמת מסוכנים יותר, או שהם פשוט נפוצים יותר?". זהוCLI חיוני לקביעת סדרי עדיפויות בטיחות בדרכים שאינם מבוססים רק על נפח אירועים אלא על רמת סיכון מהותית.

### How - כיצד היזואלייזציה ממומשת?

היזואלייזציה מומשה באמצעות גרף عمמודות מוגרם מנורמל (Stacked Bar Chart 100%), המשמש בקידודים הבאים:

- קידוד אורך ומיקום (Length & Position): כל העמודות מיישרות 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 מוחכניים המייצגים את היחס היישר בין תאונות עם נפגעים (פצועים/הרוגים) לבין תאונות ללא נפגעים. חלוקה זו מאפשרת השוואת ייזואלייזציה 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 Σ מנורמל: בנגוד לגרפים הקודמים, ציר ה-Σ אינו מציג "מספר תאונות" אלא ערך יחסית (Percentage), מה שמייצר בסיס השוואה הוגן בין רכבים נפוצים לרכיבים נדירים.

### nymok לבחירה היזואלייזציה

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

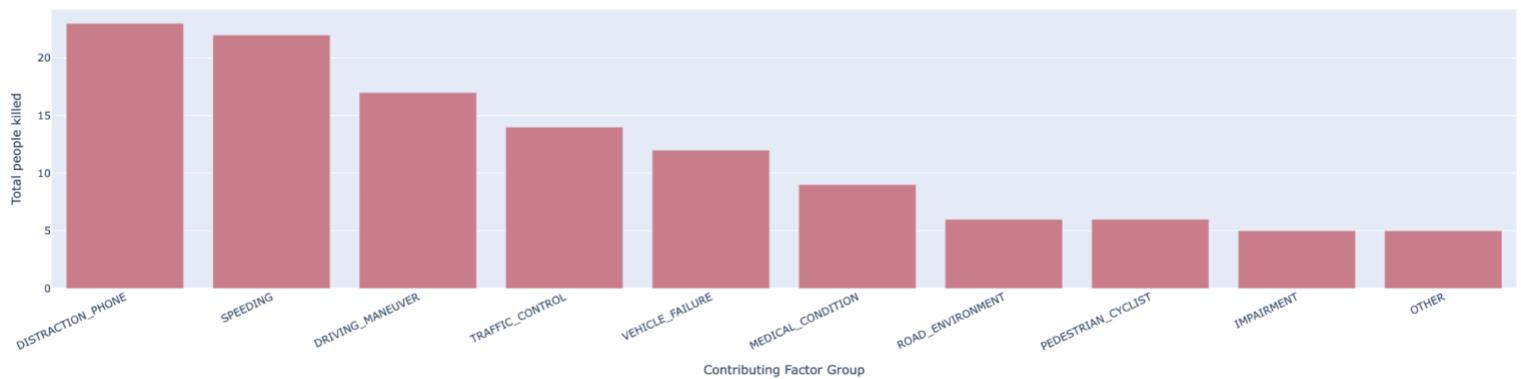
אפשר לעין האנושית להבחין בклות בהבדלים קטנים באחוז הנפגעים. נימוק זה עומד בסיס הចורך לתקן את העייפות הייזואלי שיזכר נפח התנוועה הגדל של הרכבים הפורטיים, ולהבליט במקומם וככבים שמעורבווים בתאונות היא אול' נמכה כמותית, אך חמורה יותר מבחןת תוצאותית.

### מסקנות מהייזואלייזציה

הנרטול מוביל למסקנות הפוכות ומשילימות לאלו של הגרפים הקודמים, ניתן לראות כי קטגוריות כמו TWO\_WHEEL\_MOTOR (כגון אופנוע) ו-BICYCLE מציגות אחוז גבוה משמעותית של תאונות עם נפגעים מתחום סך האירועים שלו, למروת נוכחותן המוצמצמת בגרפים האבסולוטיים. לעומת זאת, רכבים פרטיים, למروת היותם הרוב המספרי, מציגים אחוז גבוה של תאונות ללא נפגעים (Only Property Damage). המסקנה היא שחוරת התאונה מושפעת באופן דרמטי מסוג הרכב ומהגנה הפיזית שהוא מספק לשמשמי הדרך, וכי מאמציו הבטיחות צריכים להתמקד בקטגוריות שבהן "יחס הנפגעים" הוא הגבוה ביותר, ללא קשר לשכיחות הכללית בכביש.

### ויזואלייזיה 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?

הייזואלייזציה השימוש מתמקדת בתוצאה החמורה ביותר של תאונות הדרכים, אובדן חי אדם. הנתונים מציגים את מספר הנפגעים בנפש (הרוגים) מפולחים לפי הגורמים התורמים המרכזיים (Contributing Factors). לצורך כך, ביצענו הצלבה בין משתנה Severity (סיכון למקרי מוות בלבד), לבין קטגוריות המאקרו של הגורמים התורמים שעבורו ארגזיה (כגון DISTRACTION\_PHONE, SPEEDING, ALCOHOL ו-עוד). בינווד לגרפים קודמים שככלו פצעים או נזק לרכוש, הייזואלייזיה זו מבודדת אך ורק את מקרי המוות כדי לבחון אילו התנהלותות נהיגה הן ה"קטלניות" ביותר.

### Why - למה מוצגת הייזואלייזיה?

המטרה המרכזית היא לבצע ניתוח קטלניות (Fatality Analysis) ברור. המשימה האנאליטית היא לדרג את גורמי הסיכון לפי מחיר הנפגעות שהם גבוהים בפועל. מטרת העל של הייזואלייזציה היא תקשורתית ו"חינוך", להמחיש לנוגע כי פעולות שנראות לעיתים "קלות דעת", כמו שליחת הודעה בזמן נהיגה, אין מסתימיות רק בתאונה או בפיצעה, אלא מובילות באופן ישיר למorta. הבחירה להתמקד בהרוגים בלבד נועדה ליצור זעוז ויזואלי והבנה של הסיכון המוחלט.

### How - כיצד הייזואלייזציה מוממשת?

הייזואלייזציה מושהה באמצעות גרף עמודות (Bar Chart) הממוקד במידת הקטלניות:



- קידוד אורך (Length Encoding): אורך העמודה מייצג את מספר האנשים שנרגו כתוצאה מכל גורם. זהו הקידוד הושיר והברור ביותר להשוואת כמיות של אובדן חי אדם.
- מין יורד (Ranking): העמודות מסודרות מהגורם שהוביל להבי הרוגים וכלפי מטה, מה שמדגיש באופן מיידי את הסיבות להרג בכביש.
- צבעוניות סמנטיבית: החלנו להשתמש בצבע אדום המזוהה אינטואיטיבית עם סכנת חיים ודם, כדי לחזק את המסר של "Don't Kill".

### nymok לבחרת הייזואלייזציה

בחירה בגרף עמודות המציג הרוגים בלבד נבעה מה הצורך בפשטות ובעוצמה חזותית. בעוד שגרפים מורכבים (כמו גרפים מעורמים או מנורמלים) טובים לניתוח סטטיסטי רחב, הייזואלייזציה זו המטריה היא להעביר מסר חד משמעי. הצגת המספר האבסולוטי של ההרוגים מנטירה את ניסיון ל"יריכור" הנתונים ומראה את המציאות כפי שהיא – אילו עבירות נהיגה גורמות למוות בפועל ברחובות ניו יורק.

### מסקנות מההייזואלייזציה

הנימוח עולה תמונה מדאייה, גורמים הקשורים להסתחת דעת וניגזה בחומר תשומת לב (ובכל זאת שימוש בטלפון) נמצאים בראש רשימת הגורמים למוות בדרכים. המסקנה המרכזית היא שקיים קורלציה ישירה ומובהקת בין שימוש בטלפון לבין תאונות קטלניות. הייזואלייזציה מוכיחה כי האזהרה "Don't text and drive" אינה סימנה בלבד, אלא מבוססת על נתונים המראים כי הסתחת דעת היא אחד הגורמים המובילים בכביש להרג, יותר מגורמים טכניםים או תשתתיים אחרים.



## ביבליוגרפיה

.1. מאמרי מחקר מודzos:

- 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](https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data)