Contribution Documentation – Data Engineering Project

Team Members

- **Mahmoud Ayman Khalil** – 13006074

- **Ammar Maged** – 13007170

- **Abd Elrahman Nazmy** – 13006333

- **Omar Bassem** – 13004601

- **Adham Bahaa** – 13004340

---

1. Dataset Processing Contributions

Ammar Maged

- Performed initial data exploration for both datasets.

- Handled missing values.

- Executed and validated the merging of the two datasets.

Mahmoud Ayman Khalil

- Detected and analyzed outliers across both datasets.

- Documented the dataset merging workflow and data exploration findings.

Abd Elrahman Nazmy

- Removed duplicate records.

- Identified and handled new missing values created after dataset joins.

Adham Bahaa

- Standardized data formats.

- Cleaned inconsistent or redundant columns to ensure schema consistency.

Omar Bassem

- Ensured proper data type formatting and corrected mismatches.

- Contributed to standardizing remaining inconsistencies in data structure.

---

2. Website Development Contributions

Ammar Maged

- Responsible for deployment setup and configuration.

Mahmoud Ayman Khalil

- Developed the initial implementation of the website.

Abd Elrahman Nazmy

- Worked on the front-end design and user interface.

Adham Bahaa

- Reviewed all research questions and ensured they were correctly represented on the website.
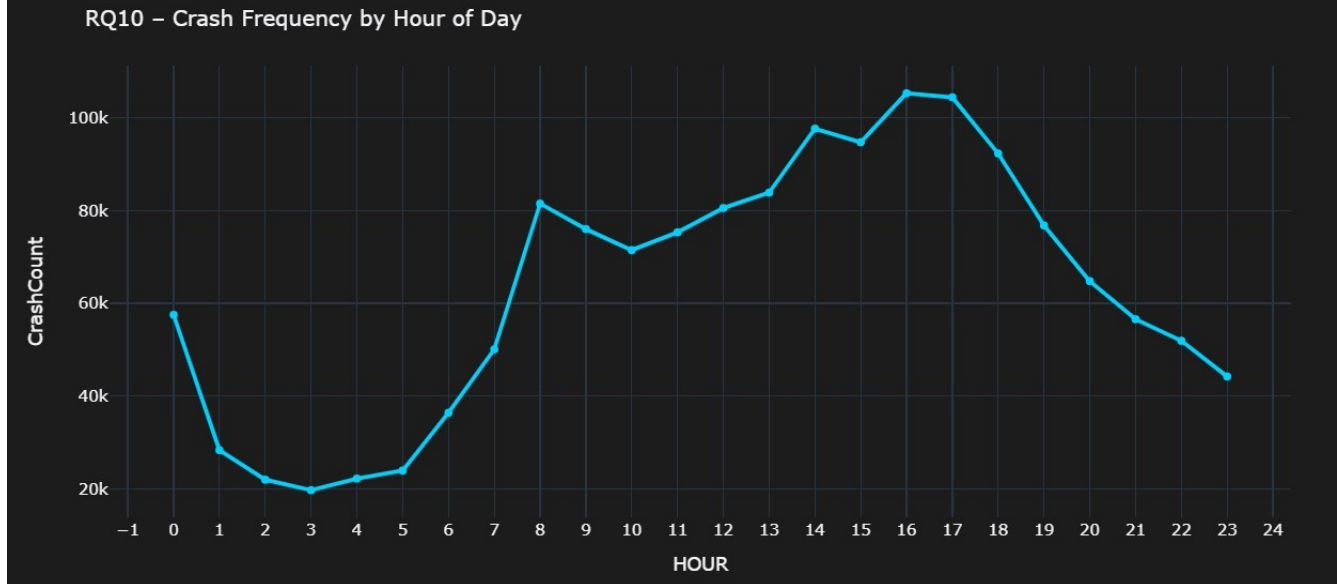
  Omar Bassem

- Selected and designed the appropriate visualizations used in the website.

Research questions

RQ3: Which bodily injuries most frequently result in death?

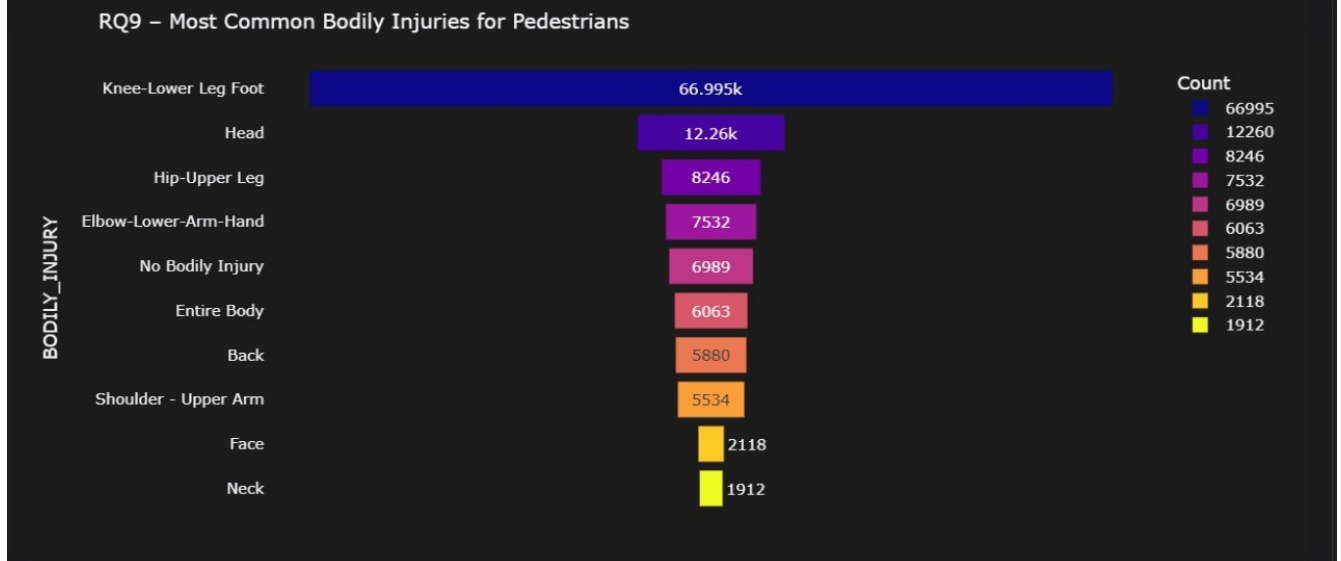RQ3 – Heatmap: Bodily Injuries Resulting in Death

The heatmap suggests Bodily Injury is the most frequent category resulting in death, with all other categories combined being lower. The data label seems obscured/misaligned, making a precise injury category difficult to determine.

RQ10 – Crash Frequency by Hour of Day

Crash frequency is highest between 3 PM and 6 PM (hours 15-18), aligning with typical afternoon rush hour traffic.

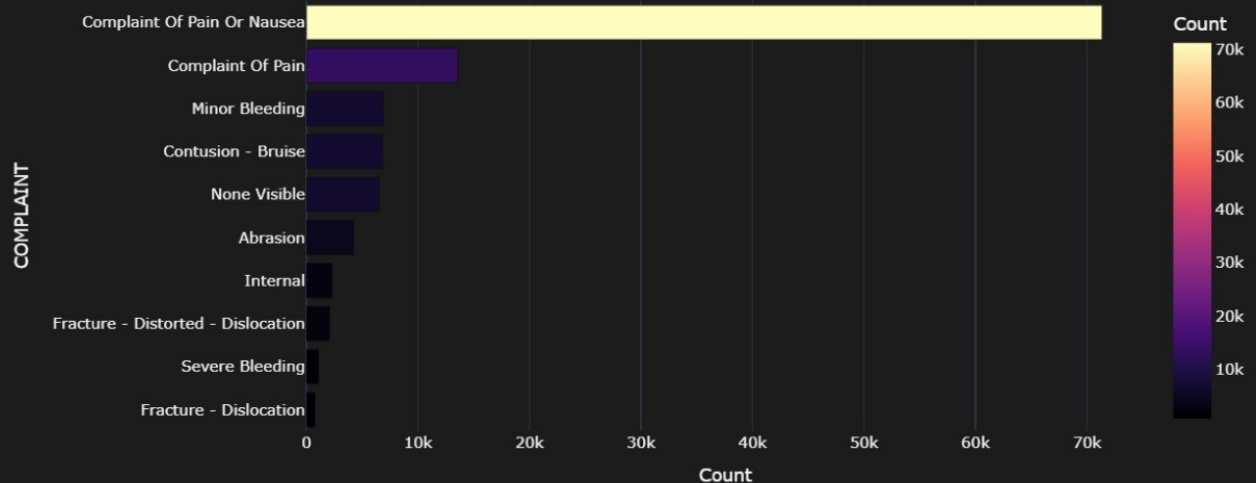**RQ9 – Most Common Bodily Injuries for Pedestrians**

BODILY_INJURY	Count
Knee-Lower Leg Foot	66.995k
Head	12.26k
Hip-Upper Leg	8246
Elbow-Lower-Arm-Hand	7532
No Bodily Injury	6989
Entire Body	6063
Back	5880
Shoulder - Upper Arm	5534
Face	2118
Neck	1912

Count
- 66995
- 12260
- 8246
- 7532
- 6989
- 6063
- 5880
- 5534
- 2118
- 1912

The Knee-Lower Leg Foot is the most common site for bodily injuries sustained by pedestrians, far exceeding other injury locations.
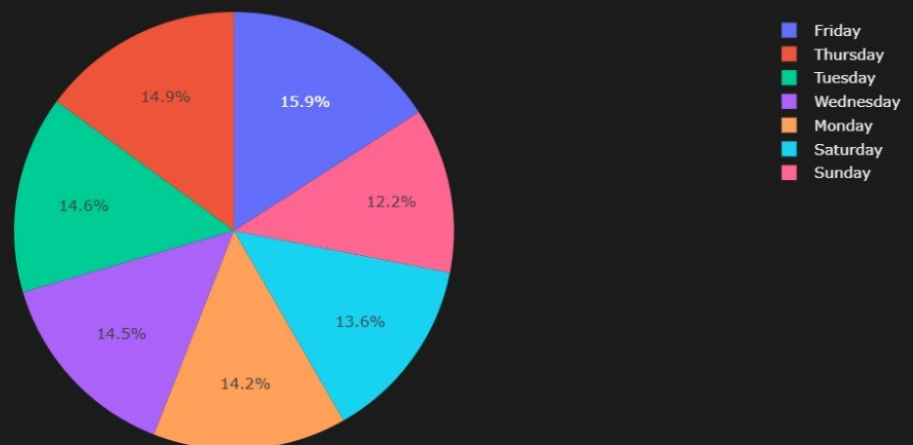
RQ6 – Top Complaints for Injured Pedestrians

Complaint Of Pain Or Nausea is overwhelmingly the most frequent initial complaint among injured pedestrians.
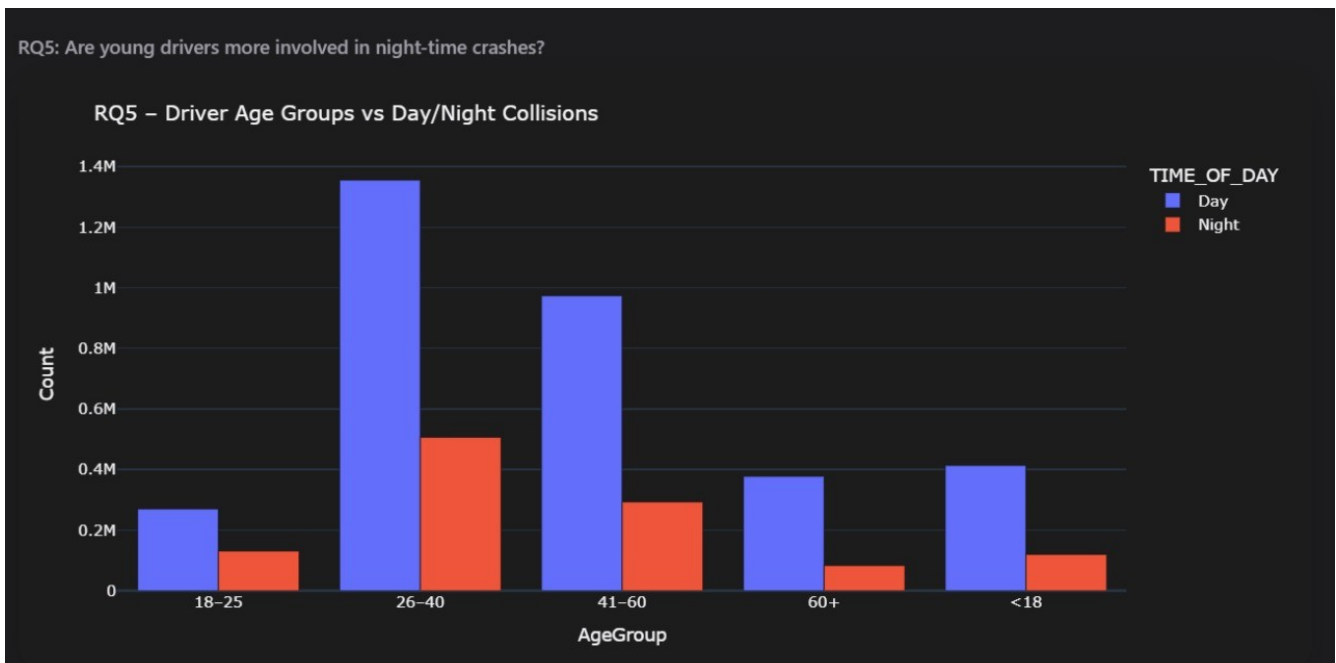
RQ2 – Crash Distribution by Day of Week

Friday sees the highest percentage of crashes at 15.9%, suggesting crash frequency peaks towards the end of the work week
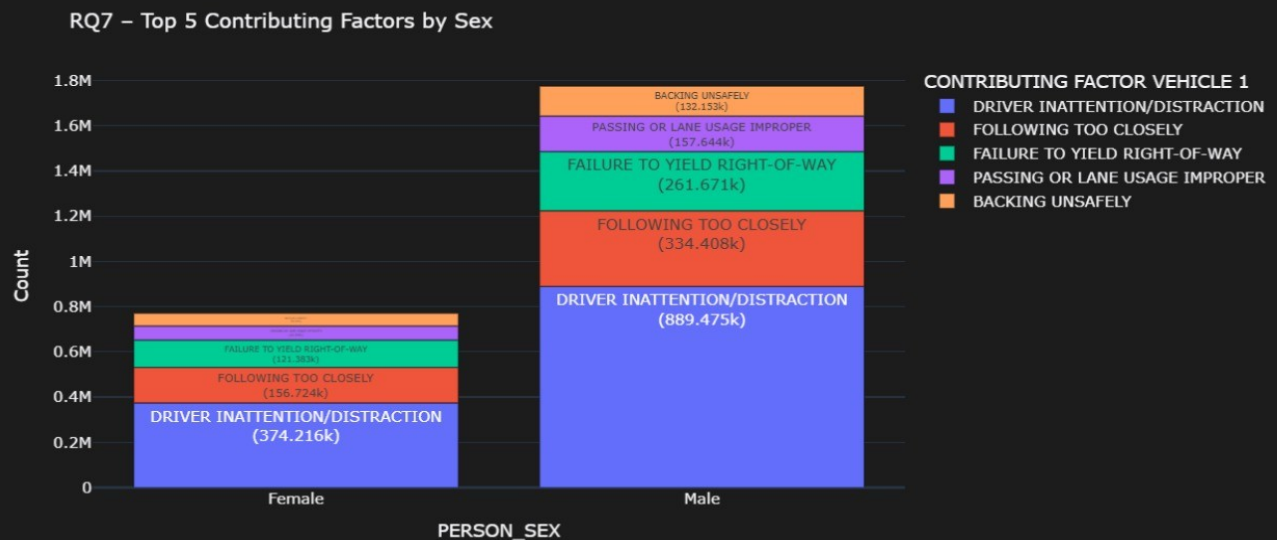
RQ4 – Heatmap: Vehicle Positions Resulting in Death

Status	Big Vehicle Passenger	Driver	Front Passenger	Irregular Position	Middle Front Passenger	Middle Rear Passenger	Outside Vehicle	Rear Passenger
Killed	6	1375	97	3	4	17	61	90

sum of Count

POSITION_IN_VEHICLE

The Driver position accounts for the vast majority of fatalities in vehicle crashes (1375 deaths).

RQ5 – Driver Age Groups vs Day/Night Collisions

TIME_OF_DAY
■ Day
■ Night

Count vs AgeGroup (18–25, 26–40, 41–60, 60+, <18)

Drivers in the **26-40 age group** are involved in the highest count of both day and night collisions compared to other age brackets.
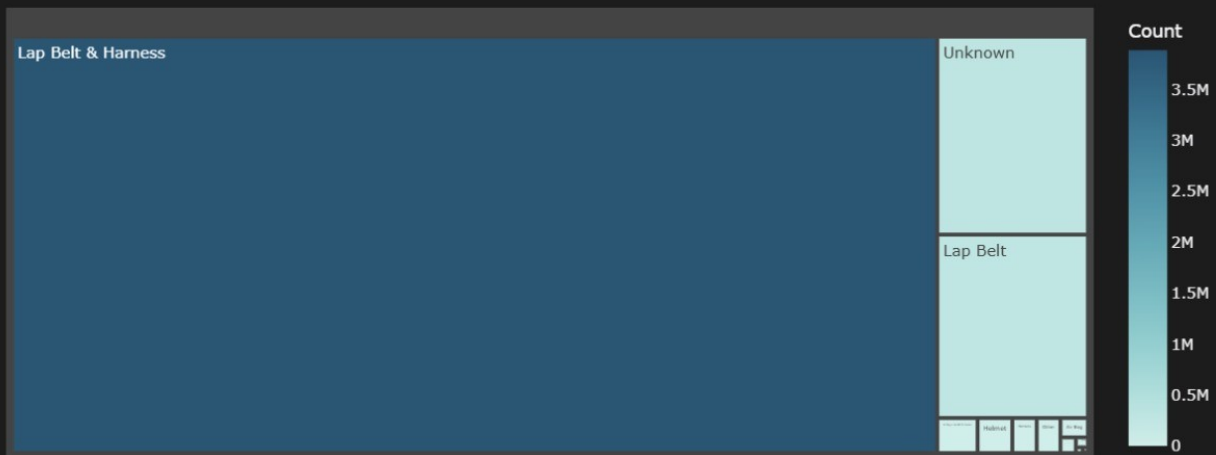
RQ7: Top 5 Contributing Factors for Female and Male

RQ7 – Top 5 Contributing Factors by Sex

CONTRIBUTING FACTOR VEHICLE 1
- DRIVER INATTENTION/DISTRACTION
- FOLLOWING TOO CLOSELY
- FAILURE TO YIELD RIGHT-OF-WAY
- PASSING OR LANE USAGE IMPROPER
- BACKING UNSAFELY

Driver Inattention/Distraction is the top contributing factor for both male and female drivers, but it contributes to a higher count of male-involved crashes.

RQ8: What safety equipment is most frequently used by drivers?

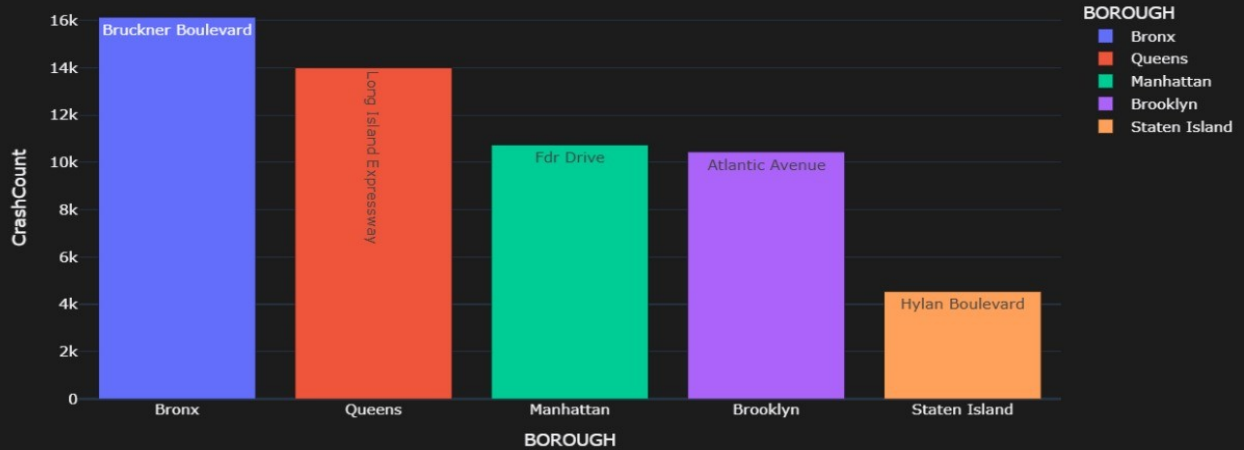RQ8 – Safety Equipment Used by Drivers

Lap Belt & Harness is by far the most frequently used safety equipment by drivers, with usage significantly higher than other options like the Lap Belt alone.

RQ1 – Street with Most Crashes per Borough (All Time)

The street with the highest crash count is Bruckner Boulevard in the Bronx, followed closely by the Long Island Expressway in Queens.