Mercer University
Faculty of Business Management

School of Business

Capstone Project

Submitted for the degree of

Master of Business Analytics

Data-Driven Approach to Cost Reduction Through Safety Product Adoption - Preventing Traumatic Injuries

Presented by

Harshan Raghavandar Ramesh 11055132

Quang Nhut Ly 11055575

Siddhanth Polepally 11055510

Submission Date: December 13, 2024

Supervisors: Myriam Quispe Agnoli

Acknowledgments

I would like to express my deepest gratitude to my teammates, Siddhanth, Quang Nhut, for their dedication, collaboration, and unwavering support throughout this project. Their hard work, insightful contributions, and commitment to excellence made this journey both productive and enjoyable. It has been a privilege to work alongside such talented individuals.

I am sincerely thankful to my advisor, Myrian Quispe Agnoli, for their invaluable guidance, constructive feedback, and continuous encouragement. Their expertise and mentorship played a crucial role in shaping this project and ensuring its successful completion.I would also like to extend my heartfelt appreciation to Mercer University for providing a supportive and intellectually stimulating environment. The faculty, staff, and resources offered by the university have been instrumental in my academic and professional growth.

Lastly, I am grateful for the learning experiences this journey has provided. Completing this project has been a rewarding endeavor that I will cherish as a significant milestone in my academic career.

Contents

Abstract

This project analyzes injury trends from the National Electronic Injury Surveillance System (NEISS), focusing on Traumatic Brain Injuries (TBIs) among elderly populations. Given the increased risk of head injuries in older adults due to environmental hazards, the study examines critical patterns linking TBI to demographic, medical, and situational variables. The analysis emphasizes head-related trauma in elderly patients. Statistical models and predictive analyses identified key risk factors such as age, injury mechanisms, and location of the incident. The findings reveal that floors are the leading cause of TBIs in older adults, highlighting the urgent need for fall-prevention strategies, including safer flooring designs and enhanced support systems in living environments. Visualizations and trend analyses underscore the importance of targeted safety interventions. This project demonstrates the potential of data-driven approaches in shaping public health policies aimed at reducing TBIs and improving safety for elderly populations.

INTRODUCTION

1.1. Background

This project explores injury patterns using data from the National Electronic Injury Surveillance System (NEISS). The primary focus is on understanding injury dynamics, particularly Traumatic Brain Injuries (TBIs) among elderly populations. Given the increasing number of older adults prone to falls and related injuries, this research highlights critical factors contributing to TBIs. By analyzing a comprehensive dataset, the project aims to uncover trends in injury occurrences, severity, and demographic characteristics. Understanding these patterns can help inform public health initiatives, safety regulations, and healthcare policies aimed at reducing TBIs and enhancing injury prevention efforts.

1.2. Previous Study

The field of injury research has evolved significantly with technological advancements, enabling deeper exploration of injury patterns and prevention strategies. Below is a concise overview of two key studies relevant to this research:

Study 1: This study examined body part grouping methodologies using NEISS data. It categorized injury codes to facilitate meaningful variable groupings in injury-related research. The study highlighted critical classification approaches that improve data analysis accuracy. (Jennissen, C. A., Koos, M., & Denning, G., 2018, ”Playground Slide-Related Injuries in Preschool Children: Increased Risk of Lower Extremity Injuries When Riding on Laps,” Injury Epidemiology, 5(Suppl 1), 13.)

Study 2: This study explored the definition and medical context of TBIs. It provided a comprehensive overview of how TBIs are identified and managed in healthcare settings, emphasizing the importance of accurate data collection for injury prevention research. (University of Massachusetts Medical School, 2024, ”What is a TBI?” National Center for Cognitive-Behavioral Research.)

These studies underscore the complexity of injury classification and the necessity of robust data analysis frameworks. By incorporating findings from such research, this project aims to enhance understanding of injury mechanisms and contribute to developing data-driven preventive strategies.

1.3. Purpose

The purpose of this project is to explore injury trends using data from the National Electronic Injury Surveillance System (NEISS). We aimed to identify key patterns in injury occurrences, focusing on demographic, medical, and environmental variables. By understanding the drivers of injury incidents, particularly Traumatic Brain Injuries (TBI) among elderly individuals, we hope to provide actionable insights for safety improvements. The analysis emphasizes head-related trauma in elderly patients. Statistical models and predictive analyses identified key risk factors such as age, injury mechanisms, and location of the incident.

1.4. Future Scope & Limitation

Our study reflects current data trends and assumes that injury patterns will remain consistent in the near term. However, the model may not fully capture future anomalies due to changes in healthcare practices, public safety measures, or population demographics.

Additionally, our analysis was limited to the available NEISS dataset, focusing on injury records from a specific time frame. Expanding the dataset or incorporating additional health records could enhance the model's predictive capabilities and generalizability.

1.5. Problem Statement

The problem statement of this study is:

What are the main contributing factors behind Traumatic Brain Injuries (TBIs) among elderly populations based on NEISS injury data?

This research question guides our analysis, supported by injury data trends, and informs the development of predictive models and actionable recommendations.

METHOD

2.1. Dataset

The dataset used in this study is derived from the National Electronic Injury Surveillance System (NEISS), covering injury records from 2019 to 2023. This dataset includes approximately 1.6 million rows and 23 variables, with a consistent number of entries each year, averaging around 350,000 records annually. These records contain key variables such as age, race, body parts affected, diagnosis, and incident location, providing a rich foundation for analysis.

Given the scope of this study, specific measures were taken to clean, group, and refine the data to ensure accuracy and relevance for Traumatic Brain Injury (TBI) analysis among elderly populations. The dataset was presumed to be reliable and comprehensive, but various steps were performed to eliminate incomplete, irrelevant, and redundant data entries.

2.1.1 Variable Grouping

We performed variable grouping using domain-specific knowledge and referenced established research. Specifically, variables such as age, race, body parts, diagnosis, and incident location were grouped using the methodology outlined by *Jennissen, C. A., Koos, M., & Denning, G. (2018)*[1].

Body Group	Body Part	Count	Percentage	Total Percentage
Upper Limb	Finger	130137	7.80%	22.19%
	Hand	67845	4.06%	
	Lower Arm	57490	3.44%	
	Wrist	54667	3.28%	
	Elbow	38258	2.29%	
	Upper Arm	22014	1.32%	
Lower Limb	Ankle	83712	5.02%	20.61%
	Knee	81781	4.90%	
	Foot	63651	3.81%	
	Lower Leg	61313	3.67%	
	Toe	30923	1.85%	
	Upper Leg	22688	1.36%	
Trunk	Lower Trunk	130417	7.81%	18.91%
	Upper Trunk	95511	5.72%	
	Shoulder	63454	3.80%	
	Neck	26241	1.57%	
Head	Head	300361	18.00%	18.00%
Face	Face	146762	8.79%	13.74%
	Mouth	34252	2.05%	
	Ear	26343	1.58%	
	Eyeball	22005	1.32%	
Other	All Parts Body	47112	2.82%	6.54%
	Not Stated/Unk	31660	1.90%	
	Internal	20615	1.24%	
	Public Region	9742	0.58%	
	25-50% of Body	78	0.00%	

Figure 1: Body Parts

DiagGroup	Diagnosis	Code	Count	Percentage	Total Percentage
Internal	Internal Organ Injury	62	202940	12.16%	23.97%
	Contusions, Abrasions	53	184807	11.07%	
	Nerve Damage	61	7381	0.44%	
	Anoxia	65	4899	0.29%	
Foreign	Laceration	59	274902	16.47%	22.83%
	Foreign Body	56	35548	2.13%	
	Poisoning	68	22992	1.38%	
	Ingestion	41	18592	1.11%	
	Puncture	63	12859	0.77%	
	Dermatitis, Conjunctivitis	74	11260	0.67%	
	Aspiration	42	2021	0.12%	
	Submersion	69	1967	0.12%	
	Electric Shock	67	901	0.05%	
Other	Other/Not Stated	71	308952	18.51%	20.26%
	Burns, Thermal	51	13442	0.81%	
	Burns, Scald	48	11945	0.72%	
	Burns, Chemical	49	2271	0.14%	
	Burns, Radiation	73	697	0.04%	
	Burns, Electrical	46	453	0.03%	
	Burns, Not Specified	47	343	0.02%	
Fracture	Fracture	57	278250	16.67%	16.67%
Trauma	Strain, Sprain	64	158534	9.50%	16.27%
	Concussions	52	32136	1.93%	
	Dislocation	55	26461	1.59%	
	Hematoma	58	22088	1.32%	
	Avulsion	72	12728	0.76%	
	Dental Injury	60	8821	0.53%	
	Crushing	54	3786	0.23%	
	Hemorrhage	66	3726	0.22%	
	Amputaion	50	3330	0.20%	

Figure 2: Diagnosis

Race Group	Race	Count	Percentage	Group Percentage
Not Stated	0 - Not Stated in ED Record	494,369	29.6%	**30.6%**
White	1 - White	794,589	47.6%	**49.2%**
Black/African American	2 - Black/African American	285,399	17.1%	**17.7%**
Asian	4 - Asian	31,481	1.9%	**2.0%**
Other	5 - American Indian/Alaska Native	5,932	0.4%	**3.8%**
	6 - Native Hawaiian/Pacific Islander	2,513	0.2%	
	2 - Other	55,154	3.3%	

Figure 3: Race

Disposition Group	Disposition	Count	Percentage	Group Percentage
Released	1 - Treated/Examined and Released	1,428,207	85.6%	**87.6%**
	6 - Left Without Being Seen	34,860	2.1%	
Hospotalized	2 - Treated and Transferred	16,970	1.0%	**12.3%**
	4 - Treated and Admitted/Hospitalized	172,405	10.3%	
	5 - Held for Observation	15,584	0.9%	
Fatality	8 - Fatality, Incl. DOA, Died in ER	1,373	0.1%	**0.1%**
Unknown	9 - Unknown, Not Stated	38	0.0%	**0.0%**

Figure 4: Disposition

Location Group	Incident Location	Count	Percentage	Group Percentage
Not Recorded	0 - Not Recorded	544,992	32.6%	**32.6%**
Home	1 - Home	685,640	41.1%	**41.1%**
Workplace	2 - Farm/Ranch	650	0.0%	**0.0%**
	7 - Industrial	166	0.0%	
Public	4 - Street or Highway	39,503	2.4%	**9.8%**
	5 - Other Public Property	123,742	7.4%	
	6- Mobile/Manufactured Home	316	0.0%	
School/Daycare	8 - School/Daycare	65,136	3.9%	**3.9%**
Place of Recreation/ Sport	9 - Place of Recreation or Sports	209,292	12.5%	**12.5%**

Figure 5: Location

Fire Involvement Group	Fire Involvement	Count	Percentage	Group Percentage
0	0 - No Fire Involved or Fire Involv	1660614	99.5%	**99.5%**
1	1 - Fire Involved and Fire Departm	2538	0.2%	**0.5%**
	2 - Fire Involved and Fire Departm	2050	0.1%	
	3 - Fire Involved and Unknown Fire Department Attendance	4235	0.3%	

Figure 6: Fire Involvement

This grouping allowed us to create meaningful categories, reducing data complexity while retaining essential analytical value.

2.1.2 Data Transformation

In addition to grouping, several data transformations were performed using domain knowledge to enhance the dataset's analytical depth:

- **Body Parts:** Variables 'Body Part 1' and 'Body Part 2' were consolidated into 'Body Part 1' and 'Multi-Body Part.' Multi-Body Part is represented as a binary variable where 1 indicates more than one body part involved and 0 indicates only one body part affected.
- **Diagnosis:** Diagnosis variables were similarly grouped into 'Diagnosis 1' and 'Multi-Diagnosis' using binary encoding.
- **Products:** Variables 'Product 1,' 'Product 2,' and 'Product 3' were combined into 'Product 1' and 'Multi-Product.'
- **Treatment Date Splitting:** The treatment date variable was split into year, month, week, date, and day variables.
- **COVID Indicator:** A binary variable 'COVID' was created, where 1 indicates incidents occurring during the COVID-19 pandemic and 0 indicates pre-pandemic records.

2.2. Initial Selection of Variables

The dataset contained 23 variables after grouping, cleaning, and transforming the data. The variables were selected based on their relevance to understanding TBIs among elderly populations, as well as their representation of demographic, environmental, and medical characteristics.

The selection process followed a three-step iterative approach:

- **Demographic Factors:** Age, race, and gender variables were selected to explore population-specific risks.
- **Injury Characteristics:** Body part, diagnosis, and incident location variables were included to capture the injury context.
- **Incident Context:** Variables related to product involvement and the timing of incidents (e.g., year, month) were retained.

2.2.1 Initial Data Cleaning

Comprehensive data cleaning was performed to remove incomplete and irrelevant records:

- **Missing Data Removal:** Entries with missing data for critical variables like race were deleted.
- **Data Consistency Checks:** Rows with missing values for essential variables like number of injuries or age were also removed.
- **Irrelevant Data Exclusion:** Records involving unlikely or irrelevant injury scenarios, such as product types not related to the study's focus, were excluded.

These cleaning and transformation steps ensured a high-quality dataset suitable for exploratory data analysis (EDA) and predictive modeling aimed at understanding TBIs in elderly populations.

2.3. Data Dictionary

S.No	Variable	Description
1	CPSC_Case_Number	CPSC case number
2	Treatment_Date	Date of treatment
3	Age	Age of patient: 0 - Not recorded 2-120 - Age in years 201-223 - Age 1 to 23 months
4	Sex	Sex of patient: 0 - Not recorded 1 - Male 2 - Female 3 - Non-binary/Other (NA before 2021)
5	Race	Race of patient: 0 - Not stated 1 - White 2 - Black/African American 3 - Other 4 - Asian 5 - American Indian/Alaska Native 6 - Native Hawaiian/Pacific Islander
6	Hispanic	Hispanic, Latino/Latina, or of Spanish origin (from 2019): 0 - Unknown 1 - Yes 2 - No
7	Body_Part	Injured body part: 0 - Internal 30 - Shoulder 31 - Upper trunk 32 - Elbow 33 - Lower arm 34 - Wrist 75 - Head 76 - Face 77 - Eyeball 79 - Lower trunk 85 - All parts
8	Diagnosis	Injury diagnosis: 41 - Ingestion 42 - Aspiration 52 - Concussions 53 - Contusions, abrasions 57 - Fracture 59 - Laceration
9	Disposition	Disposition: 1 - Treated/Examined and released 2 - Treated and transferred 8 - Fatality, incl. DOA, died in ER

Figure 7: Data Dictionary

EDA

3.1. Age and Body Part Classification

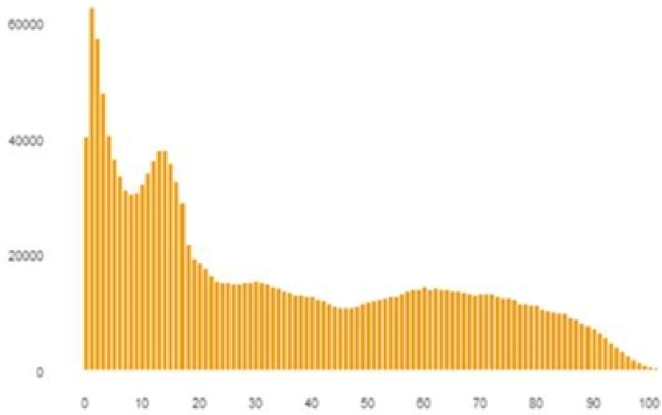

Figure 8: Injuries by Age

Fig 1 - Illustrates the distribution of injuries by age before grouping. The histogram shows high variability, with noticeable peaks in younger age groups and a gradual decline as age increases. This complexity underscored the need for a simplified classification system to improve analysis and interpretability.

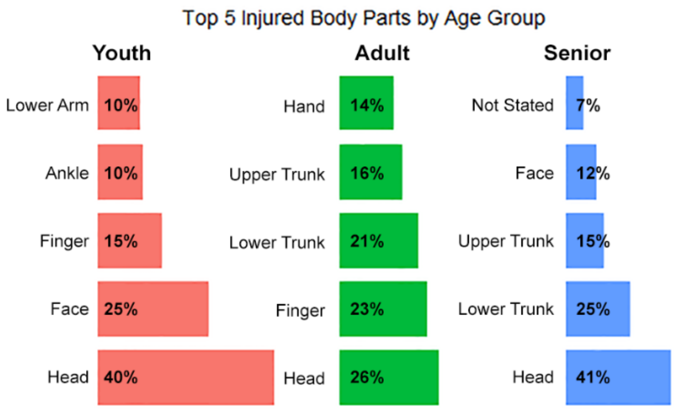

Figure 9: Top 5 Injured Body Parts by Age Group

Figure 2 - categorizes injuries into three age groups—Youth, Adult, and Senior—and highlights the top five injured body parts for each group.

- **Youth:** Head injuries are most common (40%), followed by face and finger injuries.
- **Adult:** Head injuries remain significant (26%), but injuries to fingers and the lower trunk also emerge prominently.
- **Senior:** Head injuries are the most prevalent (41%), with lower trunk and face injuries also notable.

Age Groups	Percentage
Youth (0-17)	40%
Adult (18-64)	40%
Senior (65+)	20%

Figure 10: Age Group Distribution After Grouping

Figure 3 presents the refined age classification system:

- Youth: Ages 0–17 (40% of the dataset)
- Adult: Ages 18–64 (40% of the dataset)
- Senior: Ages 65+ (20% of the dataset)

3.2. Injury Trends by Body Group

BodyGroup	Younger	Adult	Senior
Face	21%	9%	9%
Head	20%	12%	25%
Lower Lim	19%	25%	15%
Others	8%	5%	7%
Trunk	9%	24%	31%
Upper Lim	24%	25%	13%
	100%	100%	100%

Figure 11: Distribution of Injuries by Body Group Across Age Groups

Figure 4 - shows the proportion of injuries distributed across body groups for youth, adults, and seniors.

Among seniors, head injuries accounted for 25% of the total injuries, making it the most significant category. Upper limb injuries were prevalent among youth (24%) and adults (25%), but these were not the focus of this study since upper limb injuries encompass multiple body parts such as the hand, fingers, and arms.

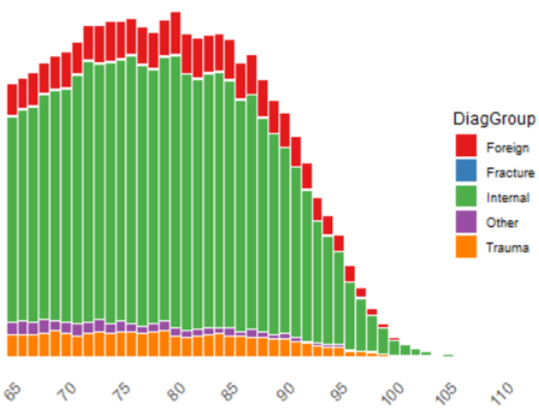

Figure 12: Injury Trends for Seniors by Diagnosis Group

Figure 5 - highlights the breakdown of injury types among seniors, with internal injuries to the head being the most common. This is followed by trauma and foreign body injuries. This trend underscores the vulnerability of seniors to severe head injuries and the need for focused preventive measures in this demographic.

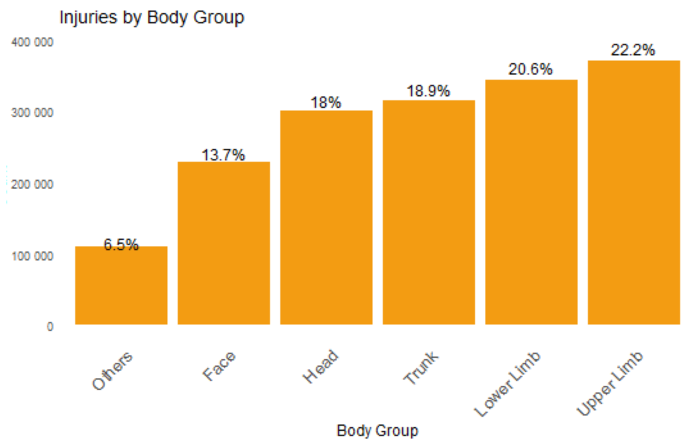

Figure 13: Injuries by Body Group in the Entire Dataset

Figure 6 illustrates the injury distribution across all body groups in the dataset. While upper limb injuries constitute the highest percentage overall (22.2%), head injuries still emerge as a significant category (18%). This aligns with the emphasis on head injuries for seniors, as observed in Figure 4.

3.3. Injury Trends by Gender and Location

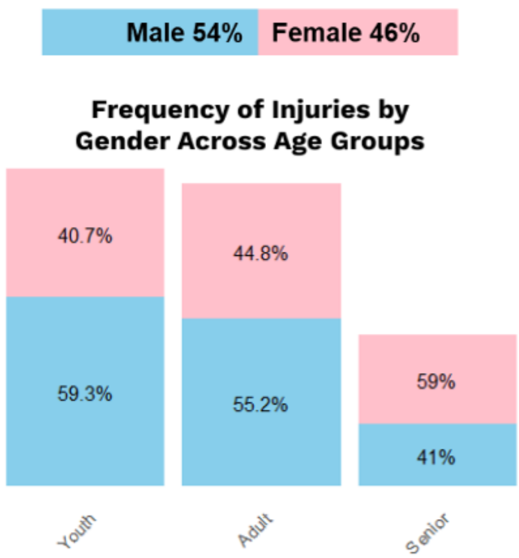

Figure 14: Frequency of Injuries by Gender Across Age Groupst

Figure 7 - provides a comprehensive view of how gender influences injury rates across three age groups—Youth, Adults, and Seniors:

- **Youth and Adults:** Males are more likely to get injured (59.3% in youth and 55.2% in adults) compared to females.
- **Seniors:** Females are more prone to injuries, accounting for 59% of injuries in this age group.

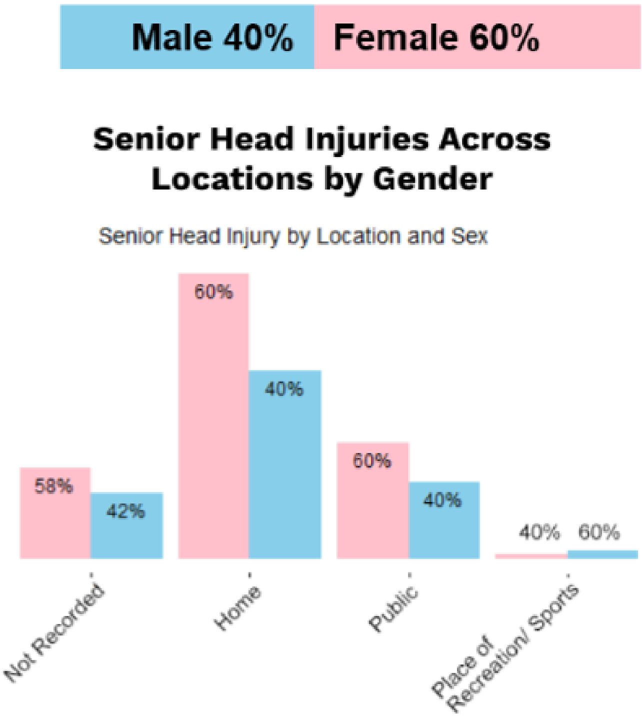

Figure 15: Senior Head Injuries Across Locations by Gender

Figure 8 - Focuses on senior head injuries, revealing location-specific gender trends:

- **Home and Public Places:** Females experience more injuries (60% of injuries in both locations).
- **Recreational/Sports Places:** Males dominate, accounting for 60% of injuries.
-

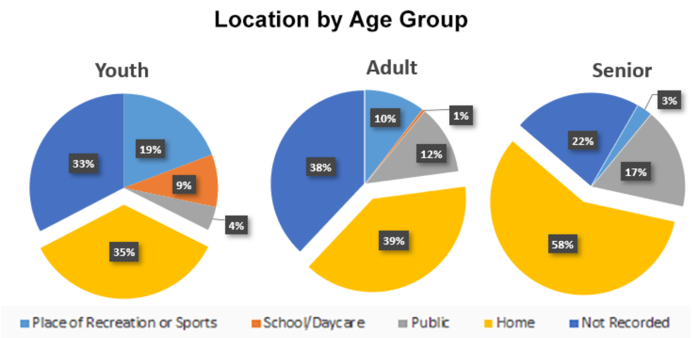

Figure 16: Location by Age Group

Figure 9 - Depicts the proportion of injuries across different locations for youth, adults, and seniors:

- **Youth:** Injuries are evenly distributed between recreation/sports (35%) and home (33%).
- **Adults:** Home remains the most common location for injuries (39%), followed by public areas (38%).
- **Seniors:** Home emerges as the primary site for injuries (58%), with a noticeable reduction in injuries in recreational/sports settings

3.4. Injury Trends by Diagnosis Group

	Foreign	Fracture	Internal	Other	Trauma	
Face	55%	6%	18%	12%	8%	100%
Head	12%	1%	71%	3%	13%	100%
Lower Lim	13%	21%	13%	24%	28%	100%
Others	48%	0%	7%	44%	1%	100%
Trunk	2%	23%	17%	37%	20%	100%
Upper Lim	31%	31%	10%	14%	14%	100%

Figure 17: Injury Distribution by Body Part and Diagnosis Group

Figure 10 - presents the percentage distribution of injuries by body parts across different diagnosis groups:

- **Head:** Internal injuries dominate this category, accounting for 71% of all head injuries. This underscores the critical need for protective measures, as internal head injuries are typically severe and can lead to long-term consequences.
- **Face:** Foreign object injuries are the most common (55%), followed by internal injuries (18%).
- **Lower Limb and Upper Limb:** Trauma and fractures are significant contributors, with trauma accounting for 28% of lower limb injuries and fractures constituting 31% of upper limb injuries.

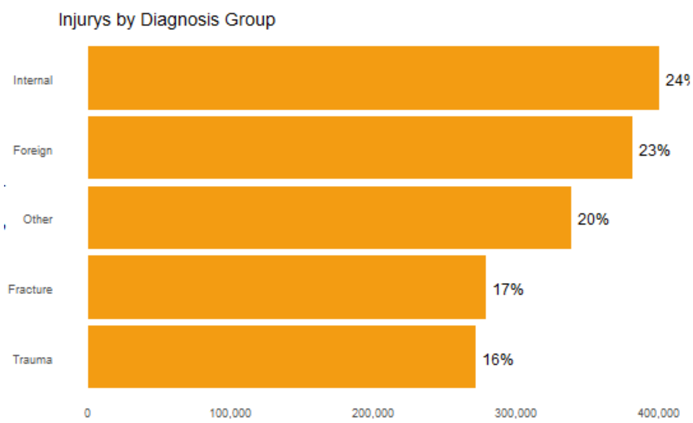

Figure 18: Overall Injuries by Diagnosis Group

Figure 11 - Summarizes the distribution of injuries by diagnosis group across the entire dataset:

- **Internal Injuries:** The most common type of injury (24%), with a significant focus on head injuries.
- **Foreign Object Injuries:** These make up 23% of the total injuries, primarily impacting the face and upper limbs.
- **Trauma and Fractures:** Together, these account for over a third of the injuries (33%), emphasizing the need for interventions targeting physical impacts.

3.5. Injury Trends by Product

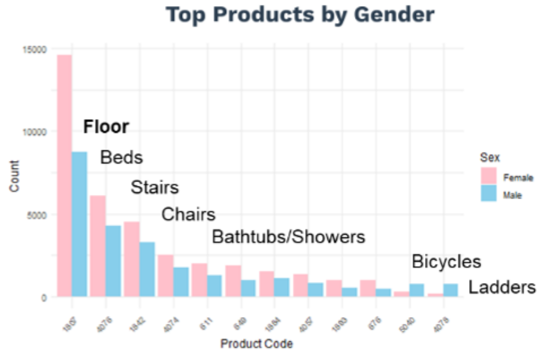

Figure 19: Top Products by Gender (Bar Chart)

Figure 12 - Highlights the products most associated with injuries, segregated by gender:

- **Females:** Injuries are predominantly caused by home-related products such as floors, stairs, and beds. These are consistent with the earlier findings that females are more likely to get injured at home or public places.
- **Males:** Recreational products, including bicycles, basketballs, and soccer equipment, are the leading causes of injuries for males. This aligns with their higher activity levels in sports and outdoor activities.

Top Products by Gender

Product	Count	Female	Male
Floor	142581	59%	41%
Stair	130351	58%	42%
Beds	101536	54%	46%
Bicycles	54004	26%	74%
Basketball	48122	18%	82%
Chairs	41465	54%	46%
Football	38368	6%	94%
Bathtubs or showers	36737	57%	43%
Ceilings and walls	34565	42%	58%
Knives	33519	43%	57%
Exercise without equipment	33458	47%	53%
Tables	32290	47%	53%
Doors	30591	51%	49%
Soccer	28910	28%	72%

Figure 20: Top Products by Gender (Table)

Figure 13 - provides a numerical breakdown of the most injury-causing products by gender:

- **Home Products:** Floors account for the largest share of injuries, with females (59%) significantly more affected than males (41%). Similarly, stairs, beds, and bathtubs show higher injury counts for females.
- **Recreational Products:** Items like bicycles, basketballs, soccer, and football dominate male injuries, with males accounting for 74% to 94% of injuries in these categories.
- The data underscores the need for targeted safety measures: improving home safety for females and enhancing protective equipment or guidelines for males in recreational activities.

3.6. Disposition and Severity of Injuries

Senior Head Injuries
(TBI)

Disposition by Diagnosis Group

DiagGroup	Released	Hospitalized	Fatality
Foreign	8297	1770	14
Fracture	47	276	9
Internal	42970	21676	158
Other	1452	809	3
Trauma	4600	1282	6

DiagGroup	Released	Hospitalized	Fatality
Foreign	82%	18%	0.1%
Fracture	14%	83%	2.7%
Internal	66%	33%	0.2%
Other	64%	36%	0.1%
Trauma	78%	22%	0.1%

Figure 21: Disposition by Diagnosis Group for Senior Head Injuries

Figure 14 - Highlights the outcomes of senior head injuries, categorized by diagnosis group:

- **Internal Injuries:** These account for the highest hospitalization rate (33%) and fatality rate (0.2%). This emphasizes the severity of internal head injuries among seniors.
- **Fractures:** A significant portion (83%) of these cases result in hospitalization, though the fatality rate is lower (2.7%).
- **Other Diagnoses:** While most cases (64%-78%) are released, a smaller fraction leads to hospitalization and fatalities, showcasing the varying severity levels.

Entire Dataset

Disposition by Age Group

AgeGroup	Released	Hospitalized	Fatality
Youth	95.2%	4.8%	0.0%
Adult	90.1%	9.8%	0.1%
Senior	66.9%	32.9%	0.2%

Disposition by Diagnosis Group

	Released	Hospitalized	Fatality
Foreign	94.41%	5.54%	0.04%
Fracture	72.36%	27.61%	0.03%
Internal	87.75%	12.16%	0.09%
Other	85.71%	14.07%	0.21%
Trauma	96.05%	3.94%	0.01%

Figure 22: Disposition by Age Group and Diagnosis Group (Entire Dataset)

Figure 15 - provides a broader view of disposition outcomes across the entire dataset:

1. **By Age Group:**
   - **Youth:** The majority (95.2%) are released, with only 4.8% hospitalized and virtually no fatalities.
   - **Adults:** Hospitalization rates increase to 9.8%, with a slight uptick in fatalities.
   - **Seniors:** This group exhibits the most severe outcomes, with 32.9% hospitalized and a fatality rate of 0.2%.
2. **By Diagnosis Group:**
   - **Internal Injuries:** A high percentage (12.16%) of these cases lead to hospitalization, with a notable fatality rate (0.09%).
   - **Fractures:** Over 27% are hospitalized, though fatalities remain relatively low (0.03%).
   - Foreign and Trauma Injuries: These exhibit higher release rates (94%-96%) and minimal hospitalization or fatality rate.

3.7. Internal Head Injuries as Traumatic Brain Injuries (TBIs)

Findings Across the Dataset

From our analysis of senior injuries and the entire dataset, internal head injuries emerged as the most severe and impactful diagnosis group. Here's how internal head injuries correlate with TBIs, supported by research and findings from the **University of Massachusetts Medical School (2024):**

1. **Definition of TBI:** The University of Massachusetts Medical School defines Traumatic Brain Injury (TBI) as "an injury that occurs when a sudden trauma causes damage to the brain." TBIs often result from impacts or penetrative forces to the head, leading to internal damage such as bleeding, swelling, or fractures. Internal injuries to the head in our dataset align with these characteristics and can be considered TBIs [2].

Evidence from the Analysis

1. **Senior Head Injuries (Figure 14):**
   - Internal injuries account for 71% of senior head injuries.
   - Hospitalization rates for these injuries stand at 33%, the highest among all diagnosis groups.
   - Fatality rates, although low (0.2%), emphasize the critical severity of internal head injuries compared to other categories.
2. **Entire Dataset Disposition Analysis (Figure 15):**
   - Internal injuries exhibit a hospitalization rate of 12.16%, higher than most other diagnosis groups.
   - This consistent pattern across age groups reinforces that internal head injuries lead to severe outcomes, which are hallmarks of TBIs.

3. **EDA Insights (Figures 10 and 11):**
   - Internal injuries consistently emerge as the dominant diagnosis in head-related cases.
   - Seniors experience the highest proportion of internal injuries compared to other demographics, correlating with their vulnerability to TBIs due to falls and other incidents

.

4. **Connection Between Seniors and TBIs:**
   - The dataset identifies falls as a leading cause of head injuries among seniors, a major contributor to TBIs.
   - TBIs often result in cognitive and physical impairments, necessitating improved safety measures, which align with our findings on hospitalization and fatality rates.

RESULTS

4.1. Logistic Regression

We created a new binary variable, "Head", where:

- Head = 1 indicates a patient has a head injury with an internal diagnosis, also known as a traumatic brain injury (TBI).
- Head = 0 otherwise.

4.1.1 Initial Logistic Regression Model

We ran a logistic regression model using a sample of 100,000 observations. The initial results showed a Specificity of 0.007201, highlighting the model's poor ability to correctly identify non-TBI cases. This result suggested an imbalance in the dataset, where the distribution of Head = 1 (TBI cases) and Head = 0 (non-TBI cases) was significantly skewed.

4.1.2 Balancing the Dataset

To address this imbalance, we utilized the ROSE (Random Over-Sampling Examples) package, which generates synthetic data to balance the number of observations where Head = 1 and Head = 0. The balanced dataset allowed for more reliable and robust modeling.

4.1.3 Co-efficient Discussion

To analyze differences across age groups, we ran separate logistic regression models for each group to examine the relationship between predictors and the likelihood of TBI. Key significant variables included:

Variable	**Balanced**	**Younger**	**Adult**	**Senior**
(Intercept)	-0.27 ***	-0.43 ***	-0.28 ***	-0.34 ***
Age				
SexMale (vs Female)	-0.07 ***		-0.28 ***	
RaceBlack (vs Asian)	-0.37 ***		-0.36 ***	-0.39 ***
RaceWhite (vs Asian)	-0.18 ***		-0.22 ***	-0.28 ***
LocationNot Recorded (vs Home)	-0.32 ***	-0.35 ***	-0.28 ***	-0.07 ***
LocationPublic (vs Home)	0.32 ***		0.39 ***	0.41 ***
LocationWorkplace (vs Home)		-1.08 *		
Fire_InvolvementYes	-4.77 ***		-3.76 ***	-3.99 ***
Alcohol1			0.64 ***	0.18 **
Drug1	-0.23 ***	-2.61 ***	0.34 ***	0.10 *
Multi_Injuries1	1.30 ***	1.51 ***	1.30 ***	1.00 ***
Multi_Products1	0.81 ***	1.17 ***	0.41 ***	0.27 ***
Year2022	0.09 ***	0.18 ***	0.12 ***	0.16 ***
Year2023	0.12 ***	0.07 *	0.15 ***	
Weekend1			-0.06 **	-0.14 ***
Covid1			-0.16 ***	0.03

The above is the final consolidated coefficients of three separate models where the dependent variable is ”internal head injury (TBI).”

Likelihood of Internal Head Injuries (TBIs)

sample interpretation: An individual is more likely to have an internal head injury (TBI) if they go to public locations compared to staying at home. The likelihood for such injuries increases by 38.3% for seniors when public locations are involved

1. **All Balanced**
   - Multi_Injuries - 117.1% increased likelihood for injuries.
   - Fire_Involvement - 215.3% decreased likelihood for fire involvement.
2. **Younger**
   - Multi_Injuries - 141.9% increased likelihood for injuries.
   - Multi_Products - 107.8% increased likelihood for multiple products.
3. **Adult**
   - Fire_Involvement - 261.7% decreased likelihood for fire involvement.
   - Drug & Alcohol - 77.7% increased likelihood for alcohol use.
4. **Senior**
   - Fire_Involvement - 229.7% decreased likelihood for fire involvement.
   - Multi_Injuries - 92.5% increased likelihood for injuries.
   - Location - 38.3% increased likelihood for public locations.

4.2. Classification Tree

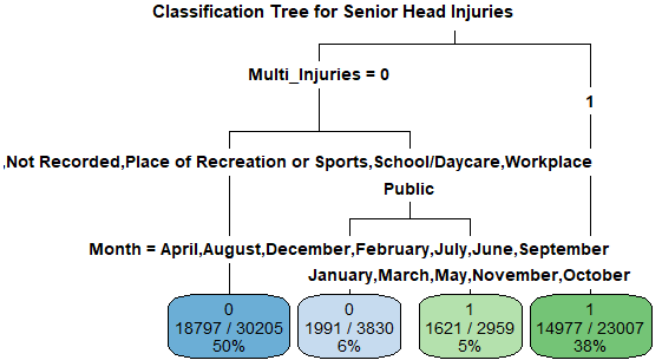

Figure 23: Classification Tree

4.2.1 Key Decision Points on classification tree (Related to TBI)

Critical Variable Identification: The classification tree identifies "Multi_Injuries" as the most critical variable associated with the likelihood of internal head injuries (TBI). This highlights that individuals sustaining multiple injuries are at a significantly higher risk for TBIs.

Environment-Based Differentiation: Within cases involving multiple injuries, the environment (e.g., public spaces or other locations) emerges as a key differentiator. Public locations, in particular, show a stronger correlation with TBIs, underscoring the need for targeted preventive measures in these areas.

Product Involvement: The involvement of "Multi_Products" plays a secondary role in influencing the occurrence of head injuries, including TBIs. Cases involving multiple products are more likely to result in TBIs, indicating that environmental and product-related hazards compound the risk.

Actionable Insights: The tree provides actionable insights by identifying risk clusters, such as individuals in public spaces with multiple injuries. This information can be used to develop tailored safety measures and interventions aimed at reducing the incidence of TBIs, particularly in high-risk environments.

These findings emphasize the interplay between injury multiplicity, environment, and product involvement in predicting TBIs, providing a framework for targeted preventive strategies

4.3. Random Forest

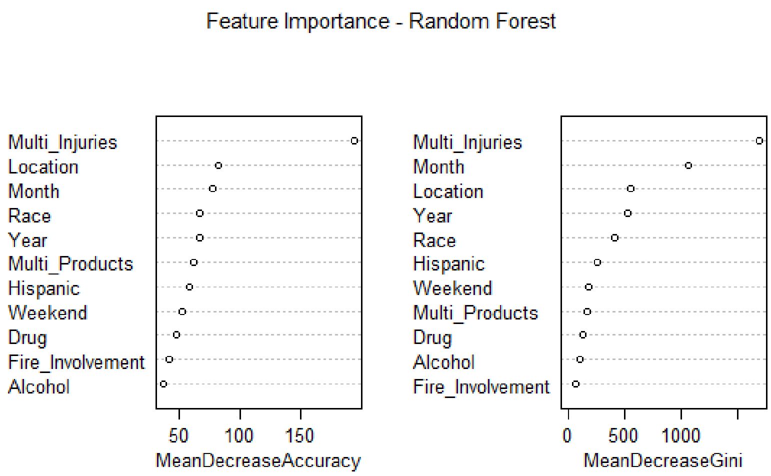

Figure 24: Random Forest

4.3.1 Key Features Influencing Internal Head Injuries on random forest (TBI)

Top Feature: "Multi_Injuries" emerges as the most influential feature in predicting internal head injuries (TBIs) based on both the **Mean Decrease in Accuracy and the Gini Index**. This underscores its critical role in understanding the likelihood of TBIs, particularly in scenarios where multiple injuries occur simultaneously.

Other Influential Factors:

- **Location:** Public and other environmental settings significantly contribute to TBI risks, emphasizing the importance of environmental safety.

- **Month:** Temporal factors, such as seasonal variations, play a role in injury patterns.
- **Race:** While the dataset highlights race as a factor, the focus should remain on ensuring equitable preventive measures and interventions.

Broader Context:

- **"Multi_Products":** The involvement of multiple products moderately influences the likelihood of TBIs, suggesting that certain combinations of products contribute to injury risks.
- **"Weekend":** Injuries occurring on weekends indicate behavioral or environmental changes during these days that elevate risks.
- **"Alcohol":** Alcohol use is a notable factor, linking substance use to increased injury severity and likelihood.

Insights: The random forest analysis validates the critical role of multi-injury scenarios and environmental factors in predicting TBIs. These findings reinforce the need for **multi-dimensional preventive** strategies that address injury multiplicity, environmental risks, and behavioral contributors. By leveraging these insights, interventions can be tailored to high-risk groups and situations to effectively reduce TBI occurrences.

4.4. Comparison Models

We also run Naïve Bayes model to predict TBI before comparison the accuracy of the 4 models by this table:

Model	**Accuracy**	**Sensitivity**	**Specificity**
Logistic Regression	62.4%	69.2%	55.6%
Decision Tree	62.1%	69.1%	55.1%
Random Forest	**64.4%**	**69.8%**	**59.0%**
Naive Bayes	62.3%	69.3%	55.3%

Table 2: Model Evaluation Metrics

When comparing model performance, Random Forest achieved the highest accuracy at 63%, closely followed by Naive Bayes and Logistic Regression at 62%. Naive Bayes excelled in sensitivity, making it better at identifying positive cases, while Random Forest led in specificity, effectively identifying negative cases.

The ROC curve highlights that while the models performed similarly, Random Forest stands out as the most balanced choice for scenarios where accuracy and specificity are equally important.

4.5. ROC Comparison

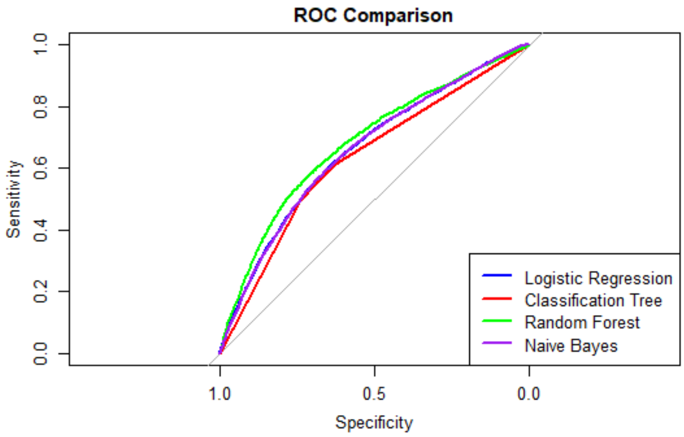

Figure 25: ROC Comparison

- **Visualization:** The ROC curve comparison shows that all models performed similarly in terms of sensitivity vs. specificity trade-offs.
- **Model Distinction:** Logistic Regression and Random Forest slightly edge out others in the higher sensitivity range at equivalent specificity.
- **Area Under Curve (AUC):** The close overlap in curves suggests marginal differences in model discrimination abilities.
- **Insights:** The ROC analysis highlights that the choice of models may depend on specific priorities like sensitivity or specificity rather than substantial performance differences.

4.6. Model Selection

Random Forest:

Highest Accuracy: Random Forest achieves the highest overall accuracy (64.4%), indicating it's the most balanced model in terms of correctly classifying both positive and negative cases.

Strong Performance: It also demonstrates strong performance in both sensitivity (69.8%) and specificity (59.0%). This suggests that Random Forest is effective at identifying both true positives and true negatives.

RECOMMENDATION

5.1. Assumptions and Key Factors for Recommendation

Assumptions for Traumatic Brain Injury (TBI) Calculations

1. **Population Growth Projections:** The elderly population (65 and older) will increase from 17% of the U.S. population in 2022 to 22% by 2040. This increase is expected to influence the incidence of TBIs due to aging-related factors [1].

2. **Incident Rates and Risk Multipliers:** More than one in four older adults (65+) experience falls annually, and those who fall are two to three times more likely to fall again. Additionally, TBIs among older adults are commonly caused by falls, as highlighted in recent research [2, 3].

3. **Cost Calculations:** The annual medical costs of fall-related injuries, including TBIs, were estimated at $50 billion in recent analyses. These cost baselines were adjusted for inflation to project future healthcare burdens [2].

4. **Preventive Measures and Effectiveness:** Interventions such as exercise programs and home modifications were assumed to reduce fall incidence by approximately 30–40% based on controlled trials. These reductions factored into the cost-benefit calculations for preventive strategies [2].

5.2. Preventive Measures for Reducing Traumatic Brain Injury (TBI) in the Elderly

To effectively reduce the incidence of TBI among the elderly, we propose the following preventive measures:

1. **Exercise Programs:** Encourage regular participation in balance and strength training exercises to reduce fall risks.

2. **Smart Devices in Homes:** Utilize technology such as fall detection systems, smart lighting, and non-slip surfaces to create a safer environment.

3. **Awareness Campaigns:** Educate caregivers, healthcare providers, and older adults through targeted campaigns about the risks of TBI and prevention strategies.

5.3. Projected Impact Over Five Years

If these preventive measures are implemented and followed consistently over five years, the potential outcomes can be analyzed and validated using Monte Carlo simulations, which account for uncertainty and variability in the projections. The following outcomes are anticipated:

a. **Savings from TBI Prevention:** Using Monte Carlo simulations, we can model the adoption growth rate (e.g., 2% in Year 1 to 10% in Year 5) as a probabilistic distribution rather than a fixed value. This allows for estimating a range of potential savings from reduced TBI cases, with confidence intervals representing the variability in outcomes.

b. **Future Reduction in TBI Cases:** The simulations project that annual prevention of injuries could range from approximately 200 to 300 cases in Year 1 and grow to 1,200 to 1,400 cases in Year 5, depending on the adoption rate and the effectiveness of the measures. The variability captures uncertainties in behavior change and implementation.

Practical Implications: Monte Carlo simulations can help identify scenarios where adoption rates grow steadily from 2% to 10% over five years. They also provide insights into the likelihood of achieving certain levels of injury reduction, helping decision-makers prioritize resources. For example, the analysis may reveal that a higher investment in awareness campaigns yields better adoption rates and a higher probability of significant cost savings.

5.4. Role of Monte Carlo Simulations

Monte Carlo simulations introduce robustness to the analysis by:

- Accounting for uncertainties in adoption rates, costs, and effectiveness of measures.
- Providing probabilistic outcomes instead of single-point estimates.
- Allowing decision-makers to evaluate different scenarios and assess risks associated with implementation strategies.

By incorporating Monte Carlo simulations, the projected impacts are more reliable, helping stakeholders make informed decisions about resource allocation and program implementation.

References

[1] Administration for Community Living, *2023 Profile of Older Americans*, May 2024.

[2] Centers for Disease Control and Prevention, *A CDC Compendium of Effective Fall Interventions: What Works for Community-Dwelling Older Adults*, 2022.

[3] Centers for Disease Control and Prevention, *Traumatic Brain Injury: Data and Research*, accessed from `https://www.cdc.gov/traumatic-brain-injury/data-research/index.html`.