



Higher Diploma in Science in Data Analytics

ASSESSMENT

Module Code: **B8IT154**

Module Description: **Platforms for Data Analytics**

Student Name: **Javier Ignacio Astorga Di Pauli**

Student ID: **20023978**

Table Of Contents

Introduction	3
Data Selection and Preparation	3
Dashboard Descriptions	6
Dashboard 1: Demographic Analysis	6
Dashboard 2: Demographic Analysis	9
Final Dashboards	15
Conclusion	16
References	17

Introduction

This report presents an in-depth analysis of the "National Electronic Injury Surveillance System" (NEISS) data for the last quarter of 2022. NEISS, a database maintained by the Consumer Product Safety Commission, offers an invaluable insight into injuries related to consumer products across the United States. By focusing on the data from the final quarter of 2022, this study aims to uncover trends, patterns, and correlations in product-related injuries, offering a window into public safety concerns and product-related risks.

The report is structured to follow a comprehensive approach, starting from data understanding and cleaning to advanced statistical analysis and visual storytelling. Utilizing Business Intelligence tools like Tableau, the study not only addresses common data issues like data type errors and missing values but also leverages these tools for creating insightful visualizations.

Key aspects of this study include the exploration of demographic distributions such as age and gender, the investigation of racial and location-based variations in injury incidents, and the identification of predominant injury causes. This is achieved through a variety of charts and visual representations, including bar charts, treemaps, pie charts, and word clouds. The analysis extends to examining the influence of alcohol and drugs on these injuries and the severity of injuries caused by different products.

Ultimately, this report serves as an exemplar of applying data science techniques in a real-world context, providing a clear and engaging narrative about public safety and product-related risks. The insights derived from this analysis are not only academically relevant to the field of Data Science but also carry significant implications for consumer safety and product design.

Data Selection and Preparation

Choice of Dataset

The "National Electronic Injury Surveillance System" (NEISS) dataset for the year 2022 was selected for this analysis. This annual dataset offers a comprehensive view of injuries attributed to products in the USA, enriched with demographic information and detailed narrative descriptions.

Understanding the Coding Manual

A critical component of this analysis involved referring to the NEISS coding manual, which details the encoding of various database variables like Age, Sex, Race, and Location. This manual served as a guide for interpreting the data correctly.

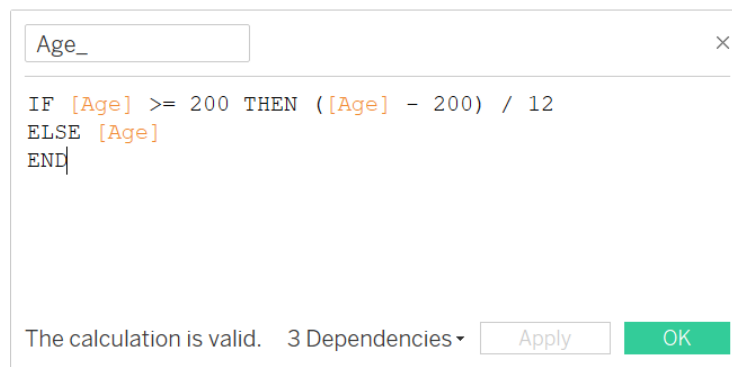
Data Cleaning and Simplification

Several columns were modified or removed to streamline the dataset and focus the analysis:

- The 'Other_Race' column was removed due to a lack of comprehensive data.
- 'Body_part 2', 'Diagnosis_2', and 'Other_Diagnosis_2' columns were eliminated to simplify the dataset and because of incomplete data.
- Columns 'Stratum', 'PSU', and 'Weight' were omitted as they did not contribute relevant information for this analysis.
- The 'Fire_Involvement' column was removed due to its lack of meaningful data.
- 'Product 2' and 'Product 3' columns were eliminated to simplify the dataset without compromising the insights.
- Only data from the last quarter of 2022 (October, November, December) was used, resulting in a dataset of approximately 71,000 rows.

Handling Age Data

The dataset presents a unique challenge in representing ages, especially for children under two years, where age is recorded in months. A calculated field in Tableau was used to transform these values into a more coherent age representation:



This adjustment created a new column with refined age data.

Enhancing Data Visualisation

- 'Bins' were created in the 'Age' column to group age data into decades for easier visualisation.
- Aliases were applied to the 'Sex', 'Race', and 'Location' columns for clarity.
- Quick table calculations were used to show the percentage of the total number of cases.

Top 10 Products Analysis

To identify the top 10 causes of injuries by product, a filter was applied in Tableau:

- Right-click on the dimension on the 'Columns' shelf and select 'Filter'.
- Under the 'Top' tab, select 'Top 10' based on the 'Number of cases'.

Customisation of 'Body Part' Column

A calculated field was used to translate the codes in the 'Body Part' column into descriptive names. The specific code used was:

Body Part Name

```

CASE [Body Part]
  WHEN 75 THEN 'Head'
  WHEN 94 THEN 'Ear'
  WHEN 77 THEN 'Eyeball'
  WHEN 76 THEN 'Face'
  WHEN 88 THEN 'Mouth'
  WHEN 89 THEN 'Neck'
  WHEN 30 THEN 'Shoulder'
  WHEN 31 THEN 'Upper trunk (not shoulders)'
  WHEN 80 THEN 'Upper arm'
  WHEN 32 THEN 'Elbow'
  WHEN 33 THEN 'Lower Arm (not elbow or wrist)'
  WHEN 34 THEN 'Wrist'
  WHEN 82 THEN 'Hand'
  WHEN 92 THEN 'Finger'
  WHEN 79 THEN 'Lower trunk'
  WHEN 38 THEN 'Pubic region'
  WHEN 81 THEN 'Upper leg'
  WHEN 35 THEN 'Knee'
  WHEN 36 THEN 'Lower leg (not knee or ankle)'
  WHEN 37 THEN 'Ankle'
  WHEN 83 THEN 'Foot'
  WHEN 93 THEN 'Toe'
  WHEN 84 THEN '25-50% of Body'
  WHEN 85 THEN 'All Parts of Body'
  WHEN 87 THEN 'Not Stated'
  WHEN 0 THEN 'Internal'
  ELSE 'Other'
END

```

The calculation is valid. 5 Dependencies Apply OK

Narrative Text Analysis For the word cloud in the 'Narrative Text Analysis', two calculated fields were created:

- One to convert all words to lowercase and remove the first six letters (a dataset-specific coding).

Lower Narrative

```

RIGHT(LOWER([Narrative 1]), LEN(LOWER([Narrative 1])) - 6)

```

The calculation is valid. 3 Dependencies Apply OK

- Another to split the text into individual words and place each word in separate rows in the same column.



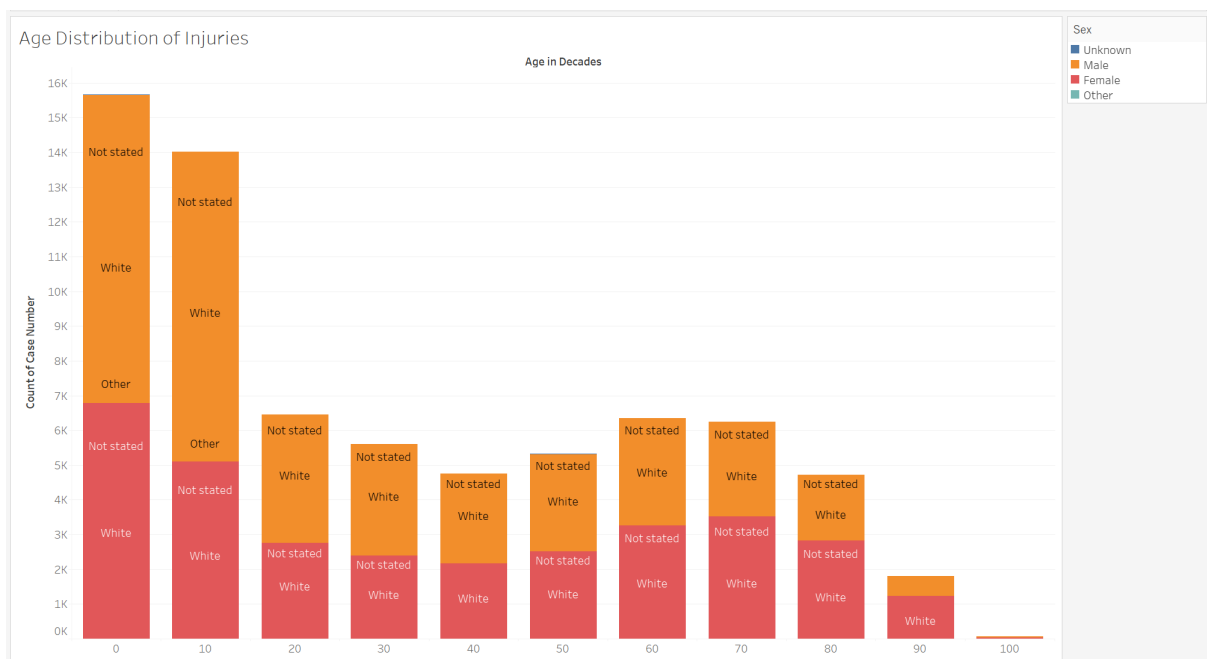
Dashboard Descriptions

Dashboard 1: Demographic Analysis

This dashboard will focus on who is getting injured, considering age, gender, and race.

Age distribution of Injuries

The 'Age Distribution of Injuries' dashboard leverages a bar chart to display the frequency of injuries across different age groups, further segmented by race and gender. This visualisation is designed to quickly identify which demographics are most affected by product-related injuries.



Insights:

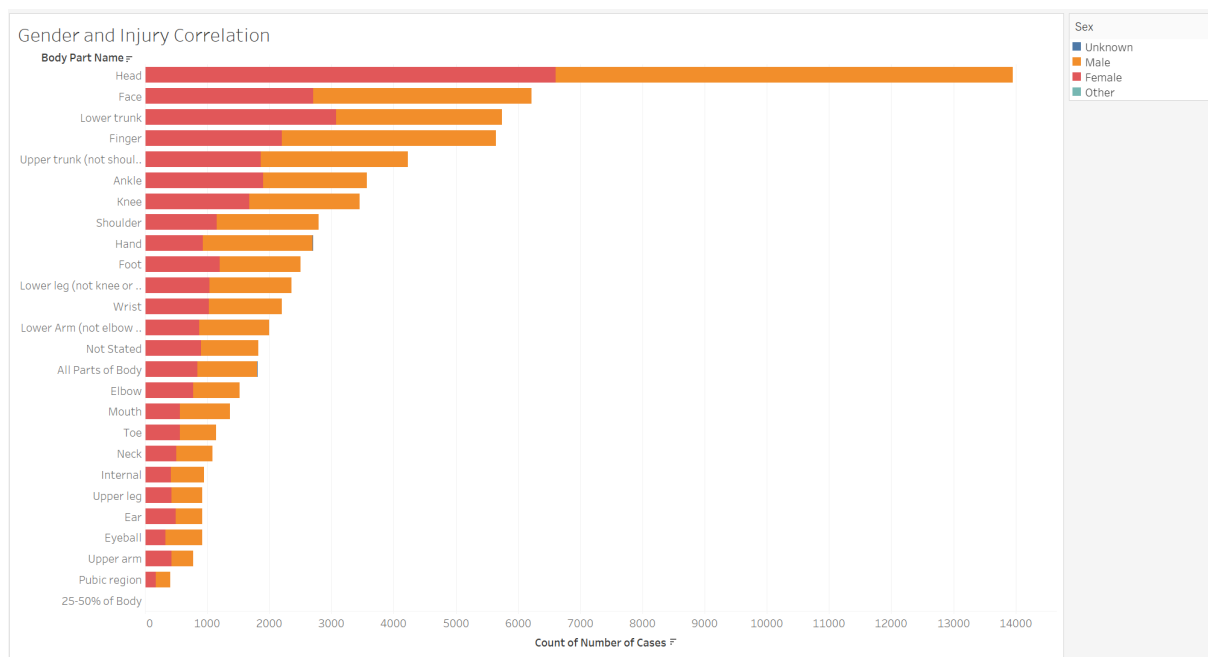
- Demographic Prevalence:** The bar chart indicates that injuries are not uniformly distributed across age groups. Notably, the largest number of injuries is reported in the first decade of life, suggesting that very young individuals may be more prone to

injuries related to products, which could reflect a higher engagement with potentially hazardous products or a developmental stage associated with increased accident risk.

2. **Race and Age Correlation:** The overlay of racial data on the age distribution reveals that within certain age groups, particularly the younger and older demographics, 'White' individuals report more injuries than other races or those where race is not stated. This could signal differences in product use or reporting practices among different racial groups.
3. **Sex Distribution Across Age:** The chart shows that the distribution of injuries between males and females varies with age. In the younger age brackets, there appears to be a higher incidence of injuries among males, which could imply gender-specific interactions with products or risk-taking behaviours. The 'Other' and 'Unknown' categories are minimal, suggesting that gender data is predominantly binary and well-reported.

Gender and Injury Correlation

The 'Gender and Injury Correlation' graph is presented as a horizontal bar chart, which effectively displays the distribution of injuries across different body parts, segregated by gender. Each body part is listed on the Y-axis, while the X-axis represents the count of reported cases. The bars are color-coded to differentiate between male, female, unknown, and other gender identities, allowing for immediate visual segmentation of data.



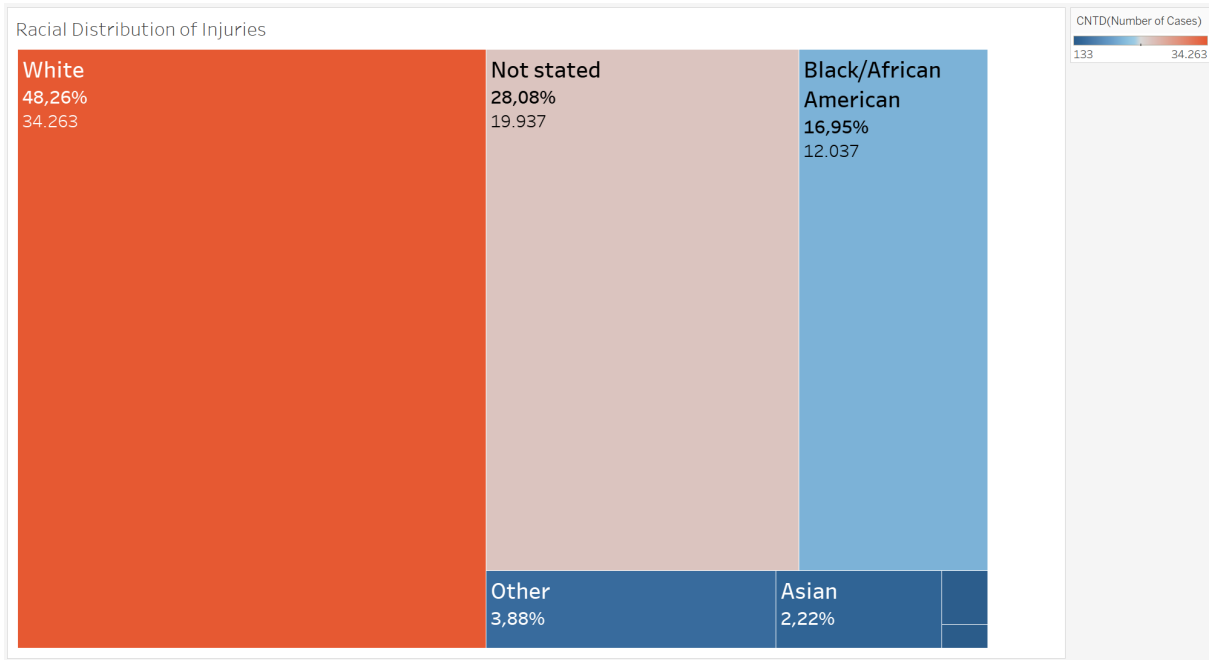
Insights:

1. **Head Injuries Predominate:** The most frequently injured body part across all genders is the head, with a particularly high incidence in males. This suggests the need for increased safety measures in activities or product usage that pose a risk to the head.

- 2. **Gender-Specific Injury Patterns:** Certain body parts, such as the finger, ankle, and knee, show a marked difference in injury cases between genders, potentially indicating different exposure to risk activities or differential product interaction.
- 3. **Lower Extremity Injuries:** Both males and females exhibit a significant number of injuries to the lower trunk, ankle, and foot, which could be indicative of common hazardous environments or products affecting these areas.

Racial Distribution of Injuries

The 'Racial Distribution of Injuries' graph is visualised using a treemap, which provides a clear and immediate hierarchy of the number of injury cases by race. Each segment's size reflects the proportion of injury cases, with annotations displaying both the percentage and the absolute number of cases. The color intensity is uniform, ensuring that size is the primary focus for interpreting the data distribution.



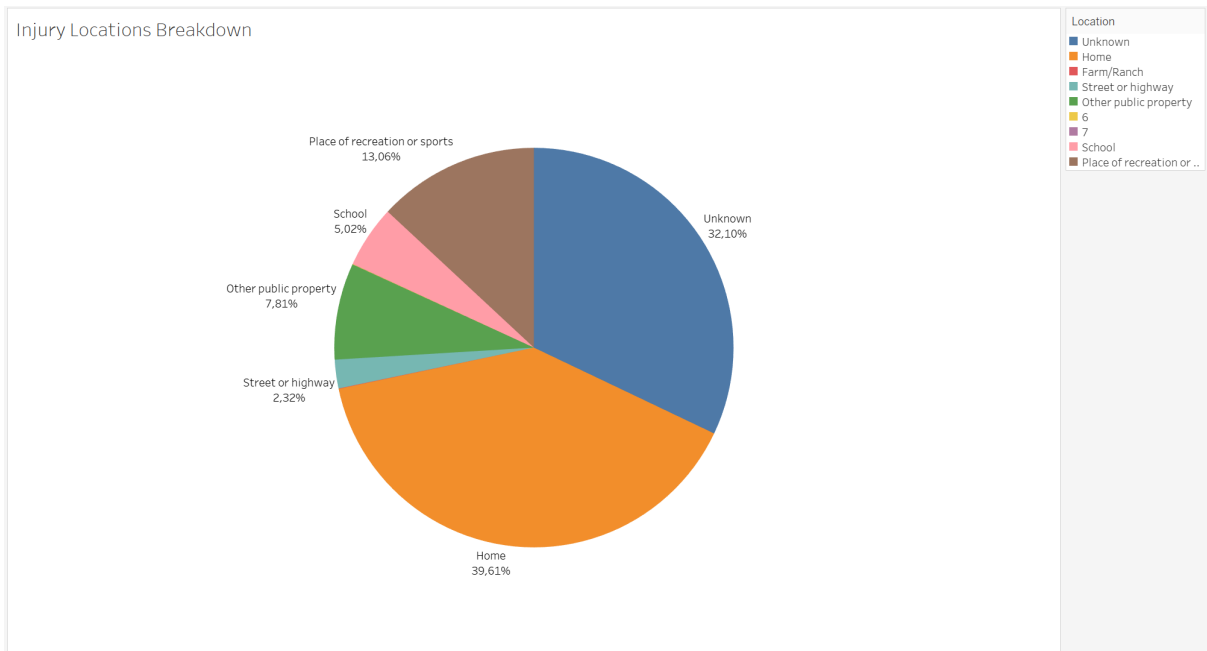
Insights:

- 1. **Prevalence Amongst White Individuals:** White individuals constitute the largest portion of injury cases, accounting for nearly half of the total, which may reflect demographic patterns or could suggest disparities in injury reporting or product usage.
- 2. **Significant Non-disclosed Cases:** A considerable percentage of the data lacks racial information, indicating a potential area for improvement in data collection practices to ensure more comprehensive demographic analysis.
- 3. **Minority Representation:** Black/African American individuals are the next most represented group. However, the proportion of other minorities, such as Asian and Other, is relatively small, which might be due to underreporting or lower rates of injury within these groups.

Injury Locations Breakdown

The 'Injury Locations Breakdown' dashboard is depicted using a pie chart, which intuitively showcases the proportion of injuries occurring in various locations. Each 'slice' represents a different location, with size indicating the percentage of total injuries associated with that

setting. The use of distinct colours enhances the differentiation between locations, making it simple to identify and compare them.



Insights:

- Home as the Primary Location:** The majority of injuries occur at home, comprising nearly 40% of the total cases, highlighting the home as a significant area for safety interventions.
- High Number of Unknown Locations:** A substantial proportion of the data does not specify the injury location, which underscores the need for more thorough documentation and reporting.
- Recreational Spaces and Schools:** Places of recreation or sports, followed by schools, are also notable for injury occurrences, suggesting that these environments should be focal points for safety improvements and education.

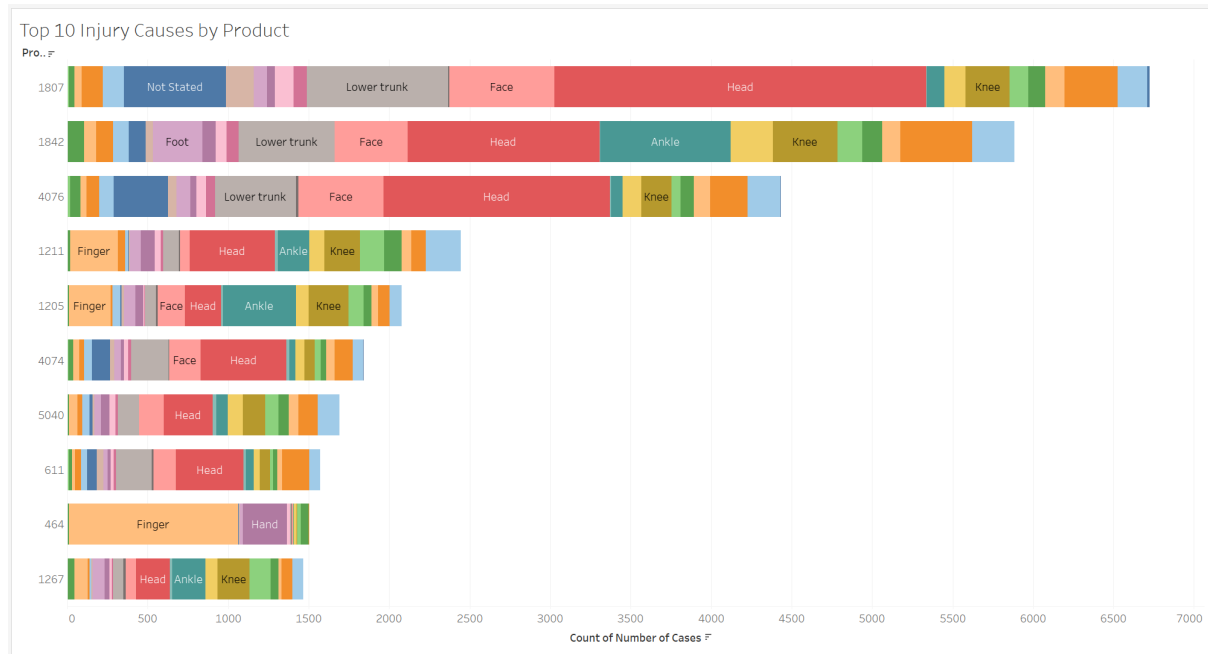
Dashboard 2: Demographic Analysis

This dashboard will delve into the nature of the injuries, their causes, and their severity.

Top 10 Injury Causes By Product

The design of this stacked bar chart employs a diverse color palette to distinguish between the various body parts affected by product-related injuries. While a myriad of colors can sometimes complicate interpretation, in this context, the distinct hues serve a clear purpose: they allow for immediate visual segmentation within each product category, making it possible to discern at a glance the body part distribution of injuries. This choice is deliberate, to facilitate the analysis of complex and varied data, where the multiplicity of colors helps to avoid oversimplification that might obscure important nuances in the data.

To enhance readability, each product is listed vertically along the Y-axis, and the corresponding codes are matched with their product names to aid comprehension. The colors are chosen to contrast well against each other, reducing the likelihood of confusion. Additionally, the use of tooltips can provide detailed information upon hovering over each segment, offering a deeper understanding of the data without cluttering the visual field.



The products corresponding to the codes in the graph are as follows:

- Code 1807: Floors or flooring materials
- Code 1842: Stairs or steps
- Code 4076: Beds or bedframes, other
- Code 1211: Football (activity, apparel, or equipment)
- Code 1205: Basketball (activity, apparel, or equipment)
- Code 4074: Chairs
- Code 5040: Bicycles or accessories
- Code 611: Bathtubs or showers
- Code 464: Knives
- Code 1267: Soccer

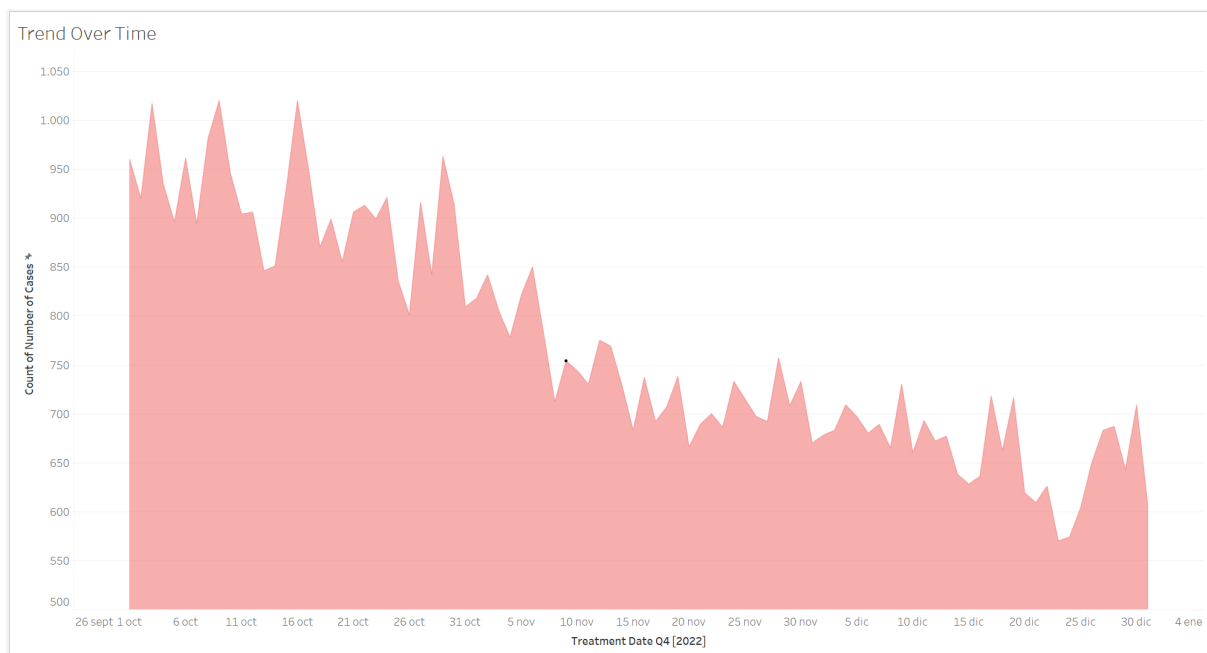
Insights:

1. Products related to home infrastructure and furniture, such as floors (code 1807) and chairs (code 4074), lead to a wide variety of injuries across multiple body parts, suggesting a high incidence of accidents in residential settings.
2. Sporting activities and equipment, including football (code 1211), basketball (code 1205), and bicycles (code 5040), are major contributors to injuries, particularly to the head and lower extremities, indicating the potential need for improved safety measures and equipment in sports.
3. The data reveals that injuries related to bathtubs or showers (code 611) and knives (code 464) are less frequent but are significant enough to appear in the top 10,

highlighting areas where consumer education and product design improvements could enhance safety.

Trend Over Time

This graph displays the trend of injury cases over time, captured from the NEISS dataset for the last quarter of 2022. A line chart with an area fill has been used, providing a clear visual representation of fluctuations over the days within the quarter. The X-axis represents the treatment date, while the Y-axis shows the count of injury cases. The area under the line is shaded, enhancing the visual impact of the data and allowing for a more intuitive understanding of peaks and troughs.



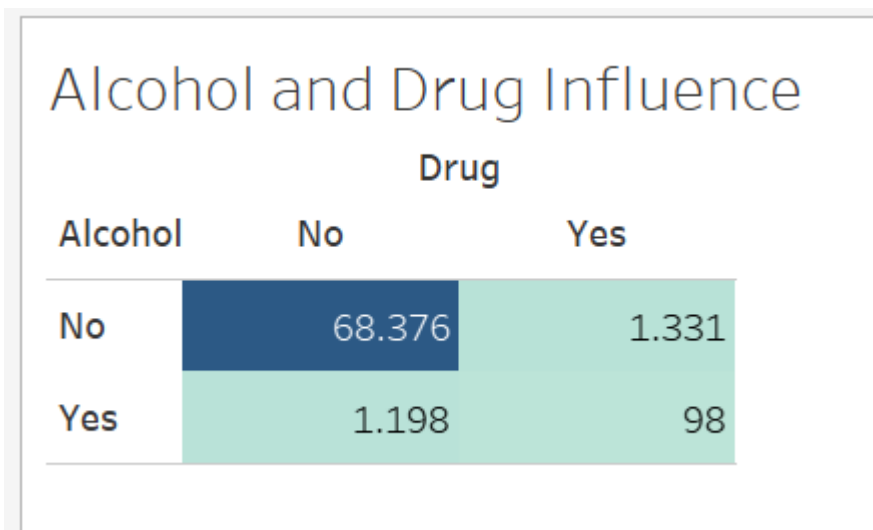
Insights:

1. There are noticeable spikes in the number of cases on certain days, which may correlate with specific events or activities occurring on those dates. Further investigation into these peaks could provide insights into causal factors.
2. The data shows a general decline in the number of cases as the quarter progresses, which could indicate the effectiveness of safety measures or seasonal variations in product use and associated injuries.
3. The periodic peaks suggest a weekly pattern, with higher numbers of injuries on specific days of the week. This could reflect a trend such as increased product usage during weekends or particular days when people engage in activities leading to more injuries.

Alcohol and Drug Influence

The graph is a highlight table that visualizes the correlation between alcohol and drug influence in reported injury cases. It is structured in a matrix format with 'Alcohol' on the Y-axis and 'Drug' on the X-axis, segregating the cases into four categories: neither alcohol

nor drugs, drugs only, alcohol only, and both alcohol and drugs involved. The shading intensity represents the volume of cases, with darker shades indicating higher numbers.



- Insights:
- 1. The majority of injury cases reported no influence of either alcohol or drugs, suggesting that product-related injuries often occur without substance impairment.
 - 2. There are significantly fewer cases involving both alcohol and drugs compared to those involving just one of the substances, indicating that concurrent use is a relatively rare factor in injuries.
 - 3. The number of alcohol-related injuries is slightly higher than those related to drug use, which could reflect broader societal patterns of alcohol consumption versus drug use, or it may suggest that alcohol is a more significant factor in the types of injuries being reported in this dataset.

Injury Severiry by Product

This highlight table showcases the severity of injuries associated with various products. Products are labelled across the top by their code numbers from the NEISS dataset. The vertical axis lists the outcomes or dispositions of the injury cases, such as 'Treated and released', 'Treated and admitted', 'Held for observation', etc. The cell shading intensity corresponds to the number of cases, with darker shading indicating a higher frequency of cases for a given disposition by product code.

Injury Severity by Product										
Disposition	Product 1									
	464	611	1205	1211	1267	1807	1842	4074	4076	5040
Treated and released	1.418	1.237	1.974	2.249	1.383	4.488	4.832	1.451	3.311	1.362
Treated and transferred	5	17	4	9	5	120	64	19	56	9
Treated and admitted	16	239	37	83	34	1.805	699	293	848	222
Held for observation	3	22		4		153	71	31	65	21
Left without being seen	61	55	62	102	46	149	204	44	145	73
Fatality		1	2			10	13	2	6	3

The products corresponding to the codes in the graph are as follows:

- Code 0464: Knives
- Code 0611: Bathtubs or showers
- Code 1205: Basketball
- Code 1211: Football (activity, apparel, or equipment)
- Code 1267: Soccer
- Code 1807: Floors or flooring materials
- Code 1842: Stairs or steps
- Code 4070: Chairs
- Code 4076: Beds or bedframes
- Code 5040: Bicycles or accessories

Insights with Product Codes

1. **Code 1807 (Floors or flooring materials)** and **Code 1842 (Stairs or steps)** show a significant number of cases requiring 'Treated and admitted', indicating these products are often involved in more severe accidents. It is crucial to explore the circumstances under which such injuries occur and to consider interventions for prevention.
2. **Code 4076 (Beds or bedframes)** and **Code 5040 (Bicycles or accessories)** also have a substantial count of injuries but with a higher number of 'Treated and released' cases. This could imply that while these products are frequently involved in accidents, the resulting injuries may not be as severe.
3. Interestingly, sports-related codes, such as **Code 1205 (Basketball)**, **Code 1211 (Football (activity, apparel, or equipment))**, and **Code 1267 (Soccer)**, have a higher incidence of injuries with a range of severity, including hospital admissions and observations. This highlights the inherent risk associated with sports activities.

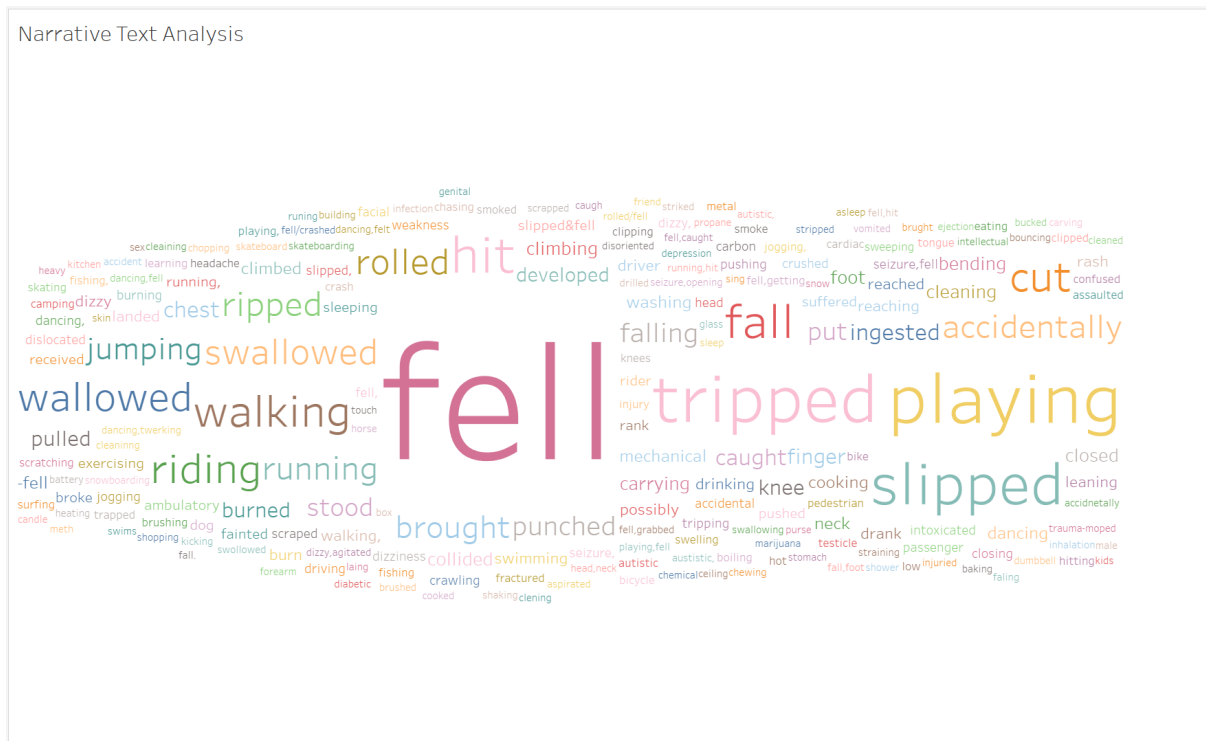
The fatalities column, although generally low across all products, shows that **Code 1205 (Basketball)**, **Code 1807 (Floors or flooring materials)**, and **Code 1842 (Stairs or steps)** have the highest numbers, suggesting a deeper safety analysis for these activities and infrastructure is warranted.

The data suggests a pattern that might be influenced by the nature of the product and the context of its use. For instance, home fixtures like **Code 0611 (Bathtubs or showers)** and **Code 1842 (Stairs or steps)** are associated with both minor and serious injuries, reflecting their everyday use and the risks involved. Meanwhile, the high volume of 'Treated and released' dispositions for products like **Code 0464 (Knives)** and **Code 4070 (Chairs)** suggests frequent minor accidents.

Narrative Text Analysis

The word cloud is an artistic representation of the most frequently occurring words within the narratives of the injury reports. The size of each word within the cloud correlates with its

frequency in the dataset, with larger fonts indicating a higher occurrence. This visualisation effectively highlights the most common themes and activities associated with injuries.

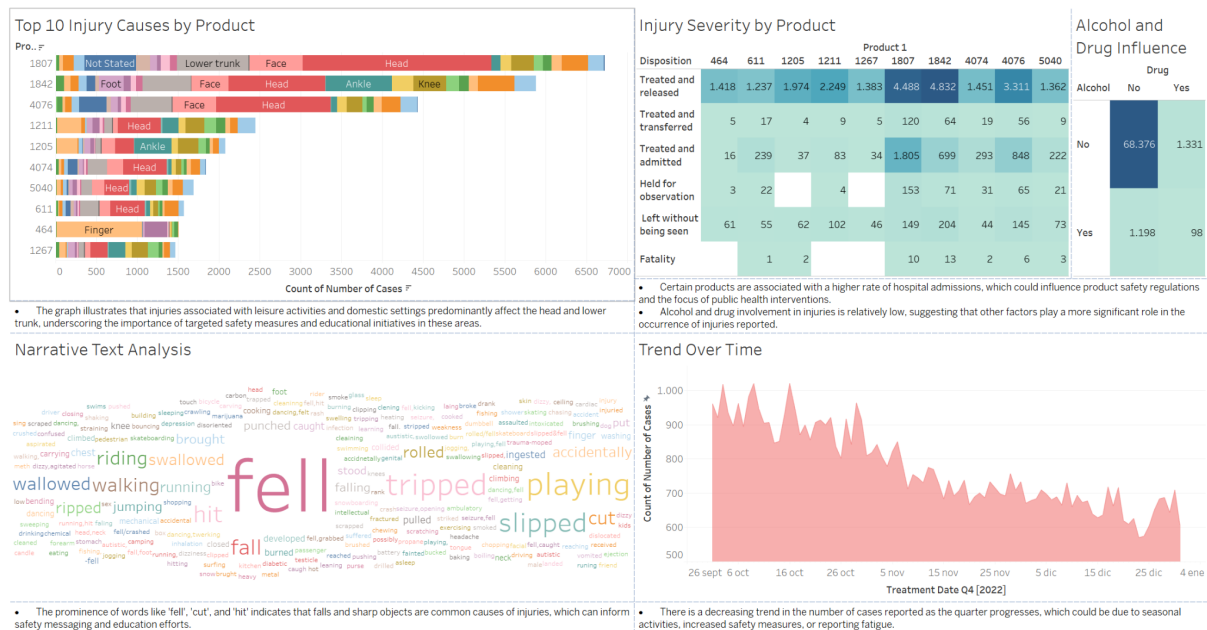
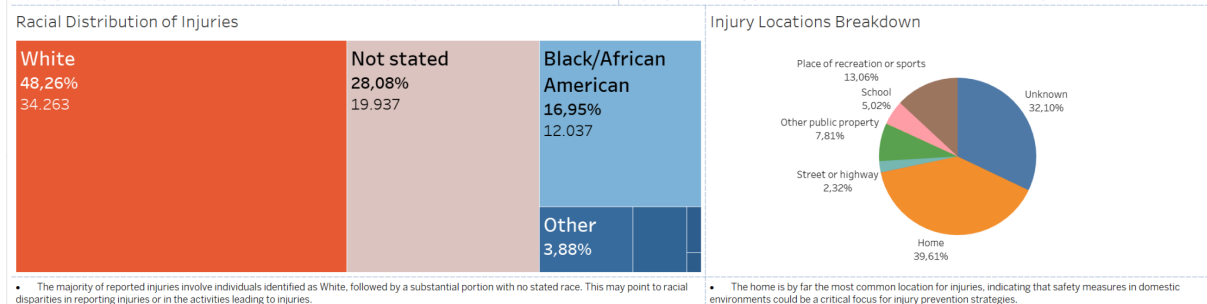
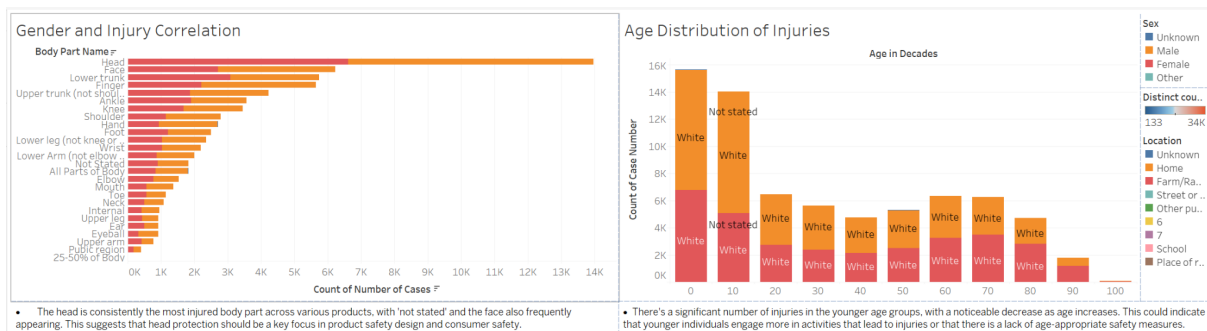


Insights

1. The prominence of words like "fell", "cut", "slipped", and "trip" suggests that falls and slips are among the most common mechanisms of injury, which could point to issues with surface stability or personal mobility.
2. The word "playing" indicates a significant number of injuries occur during recreational activities, which may require a closer look at the safety of play environments and the design of recreational products.
3. Terms like "ingested" and "swallowed" signal incidents of ingestion-related injuries, perhaps emphasising a need for better labeling and packaging of non-edible items to prevent accidental swallowing, particularly among children.

The word cloud provides a quick and intuitive way to assess the narratives and extract key themes, which are critical for understanding the context of injuries and guiding preventive measures.

Final Dashboards



Conclusion

Throughout the analysis of the NEISS 2022 dataset, several crucial insights were uncovered which have significant implications for product safety, public health interventions, and consumer education:

1. **Injury Trends and Demographics:** The majority of injuries occurred in individuals identified as White, with a significant portion not stated, suggesting potential disparities in reporting or activity engagement leading to injuries. The age distribution highlighted a higher incidence of injuries among younger age groups, declining with age.
2. **Common Injury Causes and Locations:** The data showed that falls and slips are common causes of injury, with the head being the most frequently injured body part. The home emerged as the most common location for injuries, indicating that everyday environments hold substantial risk.
3. **Severity and Impact of Product-Related Injuries:** Certain products were associated with a higher rate of severe outcomes, such as hospital admissions, underlining the need for targeted safety measures. For instance, flooring materials and stairs were often linked to more serious injuries, requiring careful consideration in safety standards.
4. **Substance Influence:** The low incidence of alcohol and drug influence in injuries suggests that other factors, such as product design and user behavior, may play a more significant role in the occurrence of injuries.
5. **Narrative Insights:** The word cloud from narrative text analysis underscored the prevalence of terms associated with falls, cuts, and accidental ingestion, which informs safety messaging and education efforts.
6. **Temporal Patterns:** The trend over time indicated a decrease in reported injuries as the quarter progressed, which could be due to seasonal activities, increased safety measures, or reporting fatigue.
7. **Data Analysis and BI Tool Utilization:** The effective use of BI tools like Tableau and Power BI facilitated a deeper understanding of complex datasets, allowing for a comprehensive view of injury trends and the ability to communicate these findings through compelling visualisations.

The insights gathered not only shed light on the patterns of product-related injuries but also emphasize the importance of data analysis in informing safety regulations and consumer safety campaigns. The need for continuous monitoring and analysis of injury data is evident, as is the necessity for a proactive approach to injury prevention and safety education. This report demonstrates the value of data science and Business Intelligence tools in drawing meaningful conclusions that can drive policy and design changes for a safer consumer environment.

References

1. Consumer Product Safety Commission. (2022). *National Electronic Injury Surveillance System (NEISS) Database*. Retrieved from <https://www.cpsc.gov/cgibin/NEISSQuery/home.aspx>.
2. Consumer Product Safety Commission. (2022). *NEISS Coding Manual*. <https://www.cpsc.gov/cgibin/NEISSQuery/Data/Info%20Docs/2022%20NEISS%20Coding%20Manual.pdf>.
3. Tableau Software. (n.d.). *Tableau Desktop*. 2023.2.2.