**Exploration of Factors Associated with Car Accidents in the U.S from 2016-2023**

**Abstract**

Despite the increasing technological advancement in the United States, car accidents continue to grow. Using data from the 2016 to 2023 countrywide traffic accident dataset, this study examines the factors associated with car accidents in the United States and whether these accidents vary by geographical location. The result indicates that factors such as weather and year are related to the occurrence of an accident. The result also shows that most road accidents occur in metropolitan areas. The findings show the need for policymakers and stakeholders to ensure roads are safer for the American people.

**Introduction**

Road travel remains one of the most prevalent modes of transportation in the United States, with most households owning at least one car (Giuliano & Dargay, 2006). However, car accidents pose a significant concern, resulting in thousands of fatalities and injuries yearly. Beyond the personal toll on individuals, car accidents impose substantial economic burdens on both citizens and the government. Individuals face increased insurance premiums, vehicle losses, and time disruptions while the government allocates resources for road repairs and emergency services. Moreover, the impact of accidents extends beyond the immediate individuals involved, affecting other road users directly or indirectly through asset loss or time wastage. According to the U.S. Department of Transportation, over 370,000 people in the U.S. died in road accidents between 2011 and 2020, with young adults being particularly susceptible (Stewart, 2023).

**Literature Review**

Accidents are characterized by unexpected or unintended occurrences that often result in injury, damage, or loss of life (Hollnagel, 2016). Road accidents typically involve collisions between vehicles or incidents where a driver damages property or injures someone. Recognizing that accidents can happen unexpectedly, even to skilled drivers, is crucial. Understanding the factors associated with experiencing a car accident is essential for preventing future incidents.

One significant factor linked to car accidents is driver’s behavior (Ivers et al., 2009; Paleti et al., 2010). Many accidents occur due to distracted driving or driving under the influence of alcohol or drugs. Drivers who are not fully focused pose risks to themselves and other road users (Strayer et al., 2006). Law enforcement agencies strive to combat reckless driving and conduct alcohol tests to deter intoxicated driving. While enforcement efforts have reduced accident rates, they have not eradicated accidents.Apart from driver’s behavior, road infrastructure and environmental factors contribute to accidents (Kaplan et al., 2015; Khan et al., 2020). Poorly designed and poorly maintained roads and inadequate signage can confuse drivers and increase the risk of accidents. Adverse weather conditions, such as rain or fog, can impair visibility and contribute to accidents.

In recent years, car manufacturers have focused on producing safer vehicles, incorporating advanced safety features to comply with regulations. Introducing self-driving cars holds promise for enhanced road safety (Britton et al., 2011). This study explores the factors associated with car accidents in the United States, considering both geographic and environmental factors. Additionally, understanding the geographical distribution of accidents can assist policymakers in allocating resources effectively to address vulnerabilities in specific areas.

**Research Question**

1. What is the trend in the number of car accidents in the United States from 2016-2023?
2. Are there spatial differences in the number of car accidents in the United States?
3. Does the distribution of accident severity level vary by weather in the United States?
4. What is the average polarity and subjectivity score of the description of car accidents in the U.S?
5. What factors are associated with the occurrence of car accidents in the United States?

**Data & Methods**

This study used data from the Countrywide Traffic Accident Dataset which was collected from February 2016 to March 2023 (Moosavi, Samavatian, Parthasarathy, & Ramnath, 2019; Moosavi, Samavatian, Parthasarathy, Teodorescu, et al., 2019). The data is available publicly online from Kaggle and has approximately 7.7 million accident records in 49 states in the United States. The data is exhaustive because it includes information on traffic, location, weather, and time, which this study used for the analysis. There were 46 variables in the original dataset, but this data subset the data to 8 variables for the purpose of this analysis. The result of this study was shown using bar charts, maps, line graphs, boxplots, tables, scatterplots, correlation, and linear regression models.

**Measures**

* **Severity:**Severity shows an accident's impact on traffic and other road users. It was measured as a categorical variable ranging from 1 to 4, with a higher number corresponding to greater impact.
* **Year:**The year of the data collection was derived from the start time of the accident. The data were wrangled to produce the year for each accident, and the years ranged from 2016 to 2023.
* **Street:** This shows the street name where the accident occurred. The data accounts for 336306 streets, including I-95N, I-95S, I-5N, I-10E, and I-10W.
* **State:**The data comprises 49 states in the U.S in which the accident occurred
* **City:** This shows the city in which the accident occurred. Some popular cities in the dataset include Houston, Orlando, Austin, and Los Angeles. There are 13678 cities in the record.
* **Description:** Description is the natural language description of the accident
* **Wind Chill**: Wind chill is measured in Fahrenheit and corresponds to the current wind chill at the time of the accident.
* **Weather Condition:**This indicates the current weather at the time of the accident. Some of the weather recorded include snow, rain, and cloudy.

**Results**

***Descriptive Analyses***

Figure 1 shows the trend in the number of car accidents in the U.S. from 2016 to 2023. In 2016, the number of accidents was 410,821; in 2017, it was 718,093; in 2018, it was 893,426; in 2019, it increased to 954,303 and kept growing till 2022 to 1,762,452 and experienced a sharp decrease to 246,633 in 2023. The sharp decline in 2023 could be due to the incomplete data available for 2023. The average yearly accident is about 966,050.

**Figure 1: Trends in the Number of Car Accidents in the U.S from 2016-2023**

**A graph with a line and a dotted line

Description automatically generated**

According to the data, there are 13,678 cities in the U.S., with the average city having about 565 accidents. The below graph (figure 2) shows the top ten cities with the highest number of accidents. Miami is taking the lead with 186,971 accidents, Houston with 169,609 accidents, and Los Angeles with 156491 accidents. Other cities in the top ten include Charlotte, Dallas, Orlando, Austin, Raleigh, Nashville, and Baton Rouge.

**Figure 2: Number of Accidents by Cities in the United States**

**A graph of a number of road accidents

Description automatically generated**

In Figure 3, the result shows the number of road accidents in the U.S. across the states. The darker shade corresponds to a higher number of accidents. The graph shows that California has the highest number of accidents, followed by states like Texas, Florida, Oregon, Carolina, New York, etc.

**Figure 3: U.S Road Accidents by State from 2016- 2023**

A blue and white map

Description automatically generated

Figure 4 shows the number of road accidents by weather type. The top five road accidents based on weather types are fair, with 2,569,802 accidents, mostly cloudy, with 1,016,195 accidents, cloudy, with 817,082 accidents, and clear and partly cloudy, with 698,972 accidents.

**Figure 4: Road Accidents by Weather Type in the United States**

***A graph of a weather type

Description automatically generated***

The severity of an accident's impact on other road users can be correlated with the weather. Figure 5 shows the average wind chill in Fahrenheit for each accident severity level. The Average wind chill for each severity level is relative, with a severity one having the highest mean wind chill compared to other accident severity groups. A higher wind chill is better regarding weather (Lankford & Fox, 2021). The total average severity is about 59F.

**Figure 5: Distribution of Wind Chill based on Accident Severity**

**A diagram of different colored squares

Description automatically generated**

Table 1 shows the average polarity and subjectivity score of the description of car accidents. The negative polarity score of -0.00485 shows that, on average, the description of accidents comprises a negative tone. The subjectivity score of 0.22 shows that the text in the description of car accidents is predominantly objective and not based mostly on emotions.

**Table 1: Average Polarity and Subjectivity Score of the Description of Car Accidents in the U.S**

|  |  |
| --- | --- |
| **Measure** | **Score** |
| Polarity | -0.00485 |
| Subjectivity | 0.2284 |

***Regression Analyses***

Table 2 and Figure 6 show the linear relationship between the number of accidents and the accident year. The result shows that as year increases, the number of accidents increases. Figure 6 also shows the predicted number of accidents for 2030. The p-value (0.39), however, is not significant.

**Table 2: Linear Regression Results between Number of Accidents and Accident Year**

|  |  |
| --- | --- |
| **Result** | **Coefficient** |
| Slope | 75096.07 |
| Intercept | -150690467.0 |
| R-value | 0.35 |
| P-value | 0.39 |
| Stderr | 81957.57 |
| Intercept stderr | 165513424.41 |

**Figure 6: Regression line Scatter Plot between Number of Accidents and Accident Year**

**A graph with a line going up

Description automatically generated**

**Conclusion & Limitation**

The study shows the distribution of the number of accidents across the United States from February 2016 to March 2023. The study results show that the number of accidents is not evenly distributed, and metropolitan or major cities are more likely to experience an increase in accidents compared to rural areas, which is consistent to the result of Yang et al. (2021). Also, the severity of an accident depends on the weather and differs based on geography. It is imperative to note that as technology advances with greater car features and safety, that has not reduced to the number of accidents in the U.S. Another possible explanation for accidents despite increase in technological advancement is that there could be more road users as the year goes by. Nevertheless, there is a need for more safety precautions as travel by land is one of the most common ways of commuting in the United States.

This study comes with limitations. First, the data did not include demographic factors like age and sex of the person who experienced the accident to see whether the number of accidents would differ by such demographics. Also, the data is only available for 49 states. One better approach is to report the rate of accidents rather than the absolute accident number. The study is limited because the numbers do not account for changing populations and different population sizes, which could have affected the number of accidents in various geographical location. Nonetheless, this study sheds light on the distribution of accidents in the United States.

**REFERENCES**

Britton, J., Gregg, P., Macmillan, L., & Mitchell, S. (2011). The early bird… preventing young people from becoming a neet statistic. *Report commissioned by Praxis*.

Giuliano, G., & Dargay, J. (2006). Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice, 40*(2), 106-124.

Hollnagel, E. (2016). *Barriers and accident prevention*. Routledge.

Ivers, R., Senserrick, T., Boufous, S., Stevenson, M., Chen, H.-Y., Woodward, M., & Norton, R. (2009). Novice drivers' risky driving behavior, risk perception, and crash risk: findings from the DRIVE study. *American journal of public health, 99*(9), 1638-1644.

Kaplan, S., Guvensan, M. A., Yavuz, A. G., & Karalurt, Y. (2015). Driver behavior analysis for safe driving: A survey. *IEEE Transactions on Intelligent Transportation Systems, 16*(6), 3017-3032.

Khan, K., Zaidi, S. B., & Ali, A. (2020). Evaluating the nature of distractive driving factors towards road traffic accident. *Civil Engineering Journal, 6*(8), 1555-1580.

Lankford, H. V., & Fox, L. R. (2021). The wind-chill index. *Wilderness & Environmental Medicine, 32*(3), 392-399.

Moosavi, S., Samavatian, M. H., Parthasarathy, S., & Ramnath, R. (2019). A countrywide traffic accident dataset. *arXiv preprint arXiv:1906.05409*.

Moosavi, S., Samavatian, M. H., Parthasarathy, S., Teodorescu, R., & Ramnath, R. (2019). Accident risk prediction based on heterogeneous sparse data: New dataset and insights. Proceedings of the 27th ACM SIGSPATIAL international conference on advances in geographic information systems,

Paleti, R., Eluru, N., & Bhat, C. R. (2010). Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention, 42*(6), 1839-1854.

Stewart, T. (2023). *Overview of motor vehicle traffic crashes in 2021*.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human factors, 48*(2), 381-391.

Yang, C., Chen, M., & Yuan, Q. (2021). The geography of freight-related accidents in the era of E-commerce: Evidence from the Los Angeles metropolitan area. *Journal of transport geography, 92*, 102989.

**Appendix:**

**Link to data:**

<https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents>