Exploratory Data Analysis for US Car Accidents

Komal Gogi
gogi0001@rangers.uwp.edu
University of Wisconsin
Parkside, Wisconsin, USA

Navjot Kaur
kaur0049@rangers.uwp.edu
University of Wisconsin
Parkside, Wisconsin, USA

ABSTRACT

There has been an increase in the number of people purchasing vehicles, increasing accidents involving significant losses. Therefore, it is crucial to locate the "root cause" of an accident so that similar accidents can be prevented in the future. Understanding how accidents occur can provide insight into their causes. It may be possible to prevent accidents by identifying patterns in accidents. An Exploratory Data Analysis (EDA) study is described in this paper, which can help take corrective and preventive actions. The data set is collected for the United States of America from 2016 to 2021. Numerous features are included in the dataset, such as the condition of the weather, the state, the city, the severity, the start and end time, the wind direction, the wind speed, and the precipitation. Python libraries such as seaborn, NumPy, matplotlib, and pandas have been employed in the study. An analysis of the data set was performed using descriptive statistics, the findings were compared, correlation coefficients and linear regression coefficients were calculated, and finally, the results were visualized using charts. As a result of the analysis discussed in the paper, we better understand the factors and causes contributing to accidents. The identification and interpretation of patterns and trends can help us draw valuable conclusions from derived findings.

KEYWORDS

EDA

ACM Reference Format:
Komal Gogi and Navjot Kaur. 2022. Exploratory Data Analysis for US Car Accidents. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

A large number of accidents occur every day and are caused by various factors that affect either directly or indirectly the safety of drivers, passengers, and pedestrians who travel on the roads.[9] The World Health Organization estimates that 1.35 million people per year die in car accidents. An average of 3,700 people die on the roads every day. There are between 20 and 50 million people who suffer non-lethal wounds, which affect many more. The main cause of death for children and teenagers between the ages of 5 and 29 is presently related to traffic injuries. Vulnerable road users, including walkers, cyclists, and motorcyclists, are impacted by more than 50% of all road incidents. There is no denying that this is a widespread issue with significant social and economic repercussions. In other words, the average cost of car accidents to a country's GDP is 3%. [18]

In the United States, according to the Association for Safe International Road Travel, 46,000 people are killed in car crashes each year, and 2.35 million more are injured or immobilized by car accidents. 1600 casualties were caused by children under the age of 15, while 8000 individuals between the ages of 16 and 20 died in road accidents. [17]The medical costs associated with car accidents may amount to billions of dollars each year for an American citizen. Around $871 billion was estimated to be the total social harm resulting from traffic accidents in 2010, and the figures have been increasing. The Traffic Safety Administration reports that 2018 saw the highest number of deaths among vulnerable road users since 1990. However, the absolute number of traffic incidents has increased by almost 19%, even though the ratio of lethal or property-damaging accidents did not change considerably. The frequency of road accidents in the United States is concerning when compared to a population increase of 5.1%.[18]

The accident data used in this case study was gathered nationally from 49 different US states. There will be 1.5 million traffic accidents between 2016 and 2020. [Check the number of accidents in your state].

The main objective of this paper is to perform a descriptive statistical analysis of data, investigate the state that has the greatest number of accidents occurring along with the weather conditions at the time of accident, and create a visual display: gather, summarize, and evaluate the data, describe the overall accident condition in the United States, and identify the variables affecting the frequency and severity of accidents; Finally, the accident's severity is predicted and analyzed.

2 LITERATURE REVIEW

The rate of crashes and the percentage of accidents have also increased due to the rapid annual growth in vehicle numbers across different countries worldwide.

T.Lu et al. [11] established the foundation of six factors which includes the driver, the vehicle, the environment, and the road a logistic traffic accident hotspot prediction model was developed, with the location of the car in the road transects, the road safety grade, the condition of the road surface, the condition of the driver's vision, the condition of the vehicle, and the dependent variable being the traffic accident hotspot. T.Jian. [6] has given a statistical analysis on major accident on highway and waterways showing a downward trend in waterways industry in China, stating 77% of accidents are through road transport, the frequency of accidents is occurred in month of August with 25% of accident during the period

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

8:00 – 12:00, the authors also stats that causes of accidents are due to inadequate safety skills, poor technical conditions of vehicle and vessels and illegal behaviors. To address the safety issues raised by severe traffic accidents following countermeasures are proposed: Intrinsic safety level of vehicles, safety education and training, implementation of safety responsibilities. M.Feng et al. [5] considered three sections; first grouped accidents incidents on an interactive Google map to highlight some hotspots, then narratively visualized accident attributes to uncover potentially related factors, and finally investigated several cutting-edge machine learning, deep learning, and time series forecasting models to predict the number of road traffic accidents (RTA) in the future, the article outlines to prove novel big data analytics platform for UK traffic accident analysis. S.A.Evtyukov. et al [4] When a two-wheeled vehicle collides with a car, the kinematics of the accident's movement and the elements that affect how severely hurt a person will be, analysis of how vehicle speed affects the severity of injuries for a certain point of contact and analysis of how vehicle speed affects the severity of injuries for a certain point of contact has been conducted along with regression equation graphs are constructed depending on the collision time and vehicle speed.

M. Mokoatle. et al [13] clustering and association rule mining is used to study the patterns and trends in road traffic accidents, the dataset for accident used in this article is from South Africa. Performed PAM (Partitioning Around Medoids) clustering prior to identifying correlations within the data. Following that, the Ariori algorithm was applied on the entire unsegmented dataset (EUD) and each cluster was subjected to association rule mining. I.A.Novikov. et al [15] uses geographic information systems as a component of intelligent transportation systems (ITS), this article suggests a solution for measuring the level of road safety on specific sections of the road network (RN). The impact of "road," one of the components of the Driver-Car-Road-Environment (DCRE) system, on traffic safety was specifically examined in detail, and solutions for recognizing typical portions in geographic information systems (GIS) were provided. Considering where hazardous locations are located, elements that affect road safety, accident likelihood, and their intensity were considered. N.Puller. et al [16] employed a map-based method to identify more information on accident types which has been recognized where the street data fusion is frequently made available by federal or local authorities. Authors conclude by concentrating on incidents at intersections, many collisions involving Vulnerable Road Users (VRUs) and cars can be avoided.

S. Kaplan et al.[8] The methodologies for detecting the two primary elements contributing to road accidents, distraction and tiredness, are provided in this paper. This research looked on Advanced Driver Assistance Systems (ADAS) and Driver Inattention Monitoring Systems. This study provided a solution for reducing the delays mentioned above by connecting a smartphone to the automobile. The National Highway Road Safety Administration has released a preliminary estimate of traffic deaths for 2021. According to NHTSA predictions, 42,915 individuals were killed in car accidents in 2017, a 10.5 percent rise over the expected 38,824 fatalities in 2020. [12]. Adib, M. A. et al.[1] developed the BT-Heart motive device so that it may be used with a mobile application to prevent car accidents. The motorist may rapidly monitor their heart rate with this smartphone app. This device also alerts the passenger to act quickly to aid the driver.

Ziwen Niu et al.[14] augmented Apriori technique for association analysis to mine the key contributing components based on different accident levels. It uses K-means clustering to categorize the degree of accident depending on the number of deaths. Correlation findings may more clearly show the relationship between accident factors and accident severity, which is ideal for traffic accident profile analysis. Qiuming Jiang et al.[7] used a distributed execution approach of a data mining algorithm based on the idea of IoT is described considering the interaction between data mining and IoT. The suggested methodology enhances data processing performance while minimizing network traffic between endpoint devices and the cloud. Data about traffic accidents is sorted and examined using data mining techniques. Das, S. et al.[3] employed the 'Eclat' algorithm for association rules to find the links between numerous components that led to WWD (wrong way driving) crashes. This investigation's findings confirmed the hypothesis that head-on collisions are more commonly involved in WWD fatalities. Furthermore, male drivers and off-peak hours are commonly involved in fatal WWD crashes. Driver impairment was highlighted as a critical element in one of the top twenty regulations. Daher, J.R. et al.[2] noticed that certain temporal and environmental variables produce the most catastrophic collisions, which are emphasized using the Frequency Pattern Growth algorithm to foster knowledge and establish association rules. A data-driven rule states that "the bulk of accidents occur between 12:00 pm and 6:00 pm." As a result, it is suggested that existing navigation software be changed to notify drivers of the increase in risk factor. I-Ying Wu et al.[19] focuses on anticipating the duration of traffic gridlock in the aftermath of a vehicle collision. To address this issue, we propose a novel model called Dual-Attention Multi-Scale Graph Convolutional Networks (DAMGNet). The proposed model considers and mixes heterogeneous data, such as accident data, urban dynamics, and other highway network properties. Because of the dual attention method developed, the DAMGNet model can efficiently learn the association between features. Wu Lingtao et al.[10] calculaed using the Gaussian function, which is greater in the center, is comparable to the traffic accident index distribution line near junctions. The line visually depicts that traffic accidents are more prevalent near intersection centers and progressively decrease as one moves away from the center, with the distribution of accidents appearing to remain constant once the distance exceeds 400 meters.

The significance of this research study is based on several data elements, exploring the state of road accidents in the United States presently and makes suggestions for minimizing accident severity and frequency. The data is investigated in different outlook, initial analysis led to exploring the geographic distribution of accidents in Unites States. Second, examine the recent volatility in total accidents and the concentrated accident period perspective. The weather dimension is then utilized to evaluate the varying impacts of different meteorological factors on severity. Finally, examine the POI (point of interest) factors that can be easily changed to raise or lower the accident rate from the POI dimension and offer suitable modification ideas.

The panda module of Python will be used to assess and understand whether current road accidents in the United States are

increasing or decreasing year over year, as well as when the peak period for car crashes is most likely to occur. Accidents that happen simultaneously are the most severe kind. Most accidents occur in both good and harsh weather. Based on the results of these data analysis, we can identify the reasons why accidents happen on the road and reduce the accident rate.

3 DESIGN

The pre-processing of data involves data preparation where unwanted features were removed from the dataset and only the important features are taken into consideration for our analysis. The dataset is trained to obtain descriptive statistical analysis and visualization analysis as shown in Figure 1

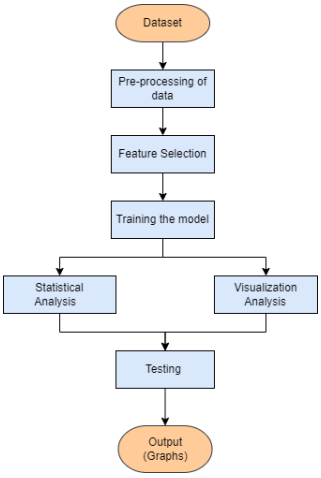

Figure 1: Design and Architecture of the system

4 IMPLEMENTATION

The descriptive statistical analysis and the data visualization are computed using Python language in Google-colab and also Power BI. Below listed are open source libraries used for exploratory data analysis:

(1) **Pandas:** Pandas is a software library for data analysis and manipulation created in the Python programming language. It contains data structures and techniques designed for interacting with time series and numerical tables.
(2) **Matplotlib:** Python's Matplotlib toolkit is a comprehensive tool for building static, animated, and interactive visualizations. Matplotlib is also used to create 2D graphs and plots.
(3) **Seaborn:** Seaborn is a matplotlib-based data visualization library. It offers a high-level interface for creating visually attractive and informative statistical visuals, is effective for visualizing random distribution.

5 EXPERIMENTAL RESULTS

A. **Which State has maximum number of car accidents**
Inferring from the State attribute, we may say that California has the highest rate of car accidents—around 800,000—and Washington has the lowest rate—around 50,000.

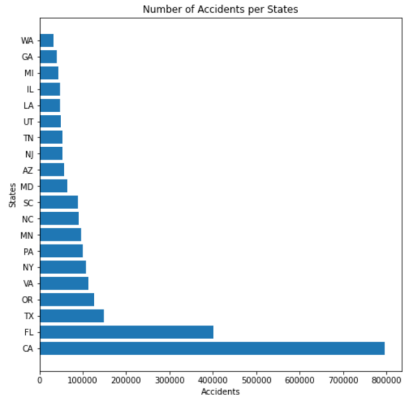

Figure 2: Number of Accidents per States

B. **Severity of the accidents**
The severity feature contains four labels: mid, high, extremely high, and low. Using the pie chart, we can anticipate that 89% of the severity is mid-range, 5.5% is high-range, and 0.9% is low-range.

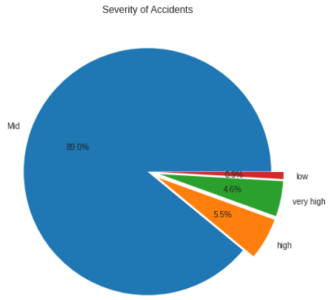

Figure 3: Severity rate of Accidents

The purpose of stacked bar charts is to make comparative category totals while also highlighting major shifts at the component level that are most likely to affect overall category total movements. Figure 4 shows the comparison of two variables by computing the bivariate visualization for the severity per state. We observe that California has the highest severity rate, with frequency ranging from 50,000 to 750,000.

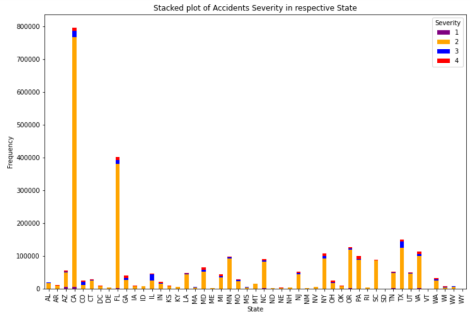

Figure 4: Stacked plot of Accidents Severity per State

Table 1: Comparison Table

Sr	Ref. & Year	Dataset	Techniques used	Results Acquired
1	Prashant Krishnan. et al [9]		Machine Learning Model, Decision Tree,K-means clustering	Approximately 61% of alcohol-related accidents take place on the weekends
2	Reem Elfatih Salman. Et al [18]	Chicago Data Portal (Jan 1st 2018 – May 1st 2020)	Apriori-Algorithm, Association Rules, Market Basket Analysis	Market Basket Analysis findings, severe accident patterns have shown associations with a disregard for traffic rules
3	Sri Siddhartha Reddy. et al [17]	Sobhan Moosavi website	Machine Learning, Big Data Solutions	Using ML predictive crash maps that analyze historical and recent data to identify high risk areas
4	Tao Lu. et al [11]	170 set of Beijing's Traffic accident data (2004-2007)	Logistic Regression Analysis	The occurrence of accident, and the prediction accuracy is approximate 86.67%
5	Mingchen Feng. et al [5]	Dept. of Transport, UK (2005-2017)	Deep Learning, Big Data Analytics, Time series forecasting	Using Prophet, LSTM trends and patterns are discovered
6	Ziwen Niu et al [14]	British government public dataset	Apriori Algorithm, K-Means clustering	Enhanced mining efficiency and depicts relationship between features and severity for analysis
7	Daher, J.R. et al [2]	Genesee Finger Lakes Region of New York	Frequency Pattern growth algorithm, Association Rules	Most of the accidents takes place between noon to 6 pm
8	Das, S. et al [3]	Louisiana(2010-2014)	Eclat association	Driver Impairment, most acciednts involves male drivers and during off-peak hours.
9	Qiuming Jiang et al [7]		Data Mining(Distributed Execution)	boosts data analysis while reducing network traffic between endpoint devices and the cloud.
10	I-Ying Wu et al [19]		Machine Learning	DAMGNet is proposed for better correlation among features.

C. **Heatmap**

A heat map is a two-dimensional representation of the correlation (measure of dependence) between the various variables, which are represented by different colors. The degree of association is indicated by the changing color intensity. A measure of the linear relationship between two variables is correlation. Each square displays the correlation between the elements on each axis, the correlation exists in the range of -1 to +1. There is no linear trend between the two variables if the values are closer to zero. The closer the correlation is to 1, the more positively associated they are; that is, as one rises, the other does as well, and the stronger this relationship is, the closer to 1 the correlation is, similar results can be obtained with a correlation that is closer to -1, but instead of both variables rising, one will fall as the other does.

Figure 5 shows the correlation among all the different attributes. Start_Lat and End_Lat along with Start_Lng and End_Lng presents the highest positive correlaion of 1, followed by Crossing with Traffic signal presenting the positive corelation of 0.42; and Temperature with Start_Lat as well as Temperature with End_Lat persents the most negative corelation of -0.48.

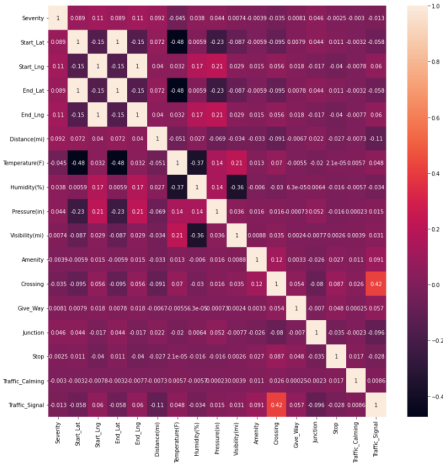

Figure 5: Heatmap

D. **City**

There are a total of 11,682 cities in the dataset; from Figure 6, we can see that Ontario has the fewest accidents and Miami has the most accidents, with a count of 106,966. The top 30 cities are shown by visualization graphs. According to the histogram distribution graph 7, over 1100 cities have reported just one accident over the course of the period. This can be extremely positive news, or it might be the consequence of missing data. As a result, we divided the cities into two categories: those with high accident rates and those with low accident rates, where high accident rates were defined as greater than or equal to 1000 and low accidents were defined as less than 1000. The proportion of cities with more than 1000 accidents is 4.25% of all cities.

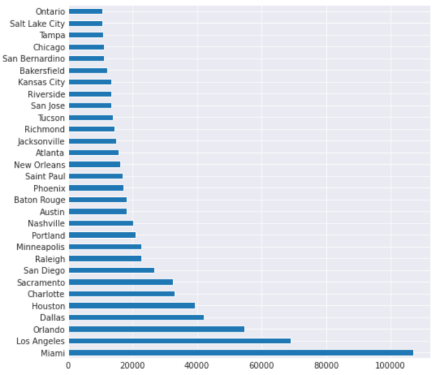

Figure 6: Top 30 Cities

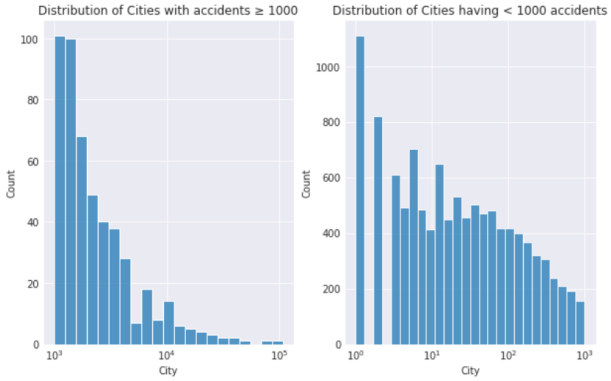

Figure 7: Distribution of cities with High and Low Accidents

In terms of high accident rates, only 100 cities out of 11682 recorded more than 1000 accidents between February 2016 and December 2020. In over 1100 cities, there has only been one accident in the past four years due to low accident rates. The majority of them had seen between 10 and 100 accidents over that time. Cities see a declining proportion of accidents that is exponential.

E. **StartTime**

The StartTime feature describes the precise time the accident happened, letting us predict the proportion of accidents that will occur within 24 hours. Figure 8 shows that there is a high peak for accidents between 3 p.m. and 6 p.m.

Figure 9 shows StartTime comparison for Monday (weekday) and Sunday (weekend). The distribution for Monday follows the normal 24-hour pattern—two peaks coincident with the work rush, whereas the distribution for Sunday shows a peak between 10 a.m. and 4 p.m. It's possible that Sunday leisure activity is the cause of this.

Figure 10 shows the monthly distribution of accidents is depicted in Figure 10. "Monthly." We observe that the second half of the year appears to have a consistent increasing curve; this may be because of the summer break and the start of the new school year in August or September.

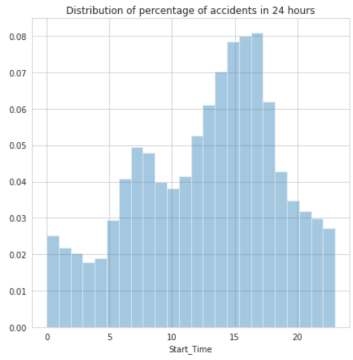

Figure 8: Distribution of percentage of accidents in 24 hours

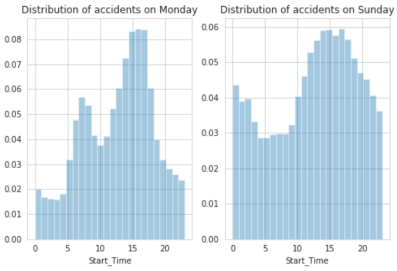

Figure 9: Distribution of accidents on Monday and Sunday

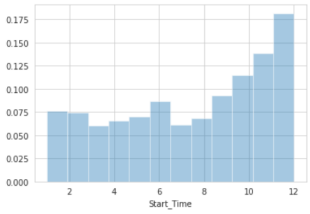

Figure 10: Monthly Distribution of Accidents

F. **Temperature**

Figure 11 depicts the temperature at the time of the accident. This is measured in Fahrenheit. It shows the ideal temperature when most of the accidents are observed, that is in between 50 - 60 degrees Fahrenheit.

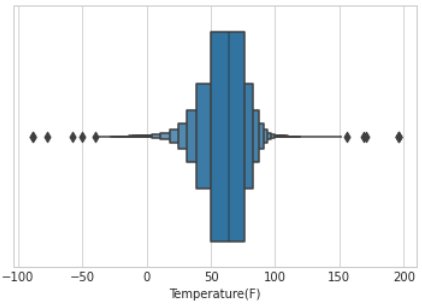

Figure 11: Temperature observed at time of accident

G. **Weather Condition**
A word cloud presents the collection of words. The most frequent the particular word is used, the bolder and bigger it is going to appear in the word cloud formation. Similarly, least the word has appeared the smaller it is going to be in word cloud.

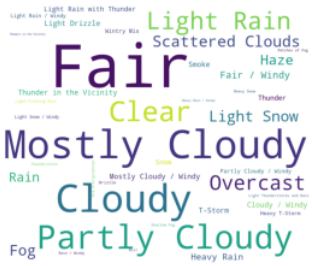

Figure 12: Word Cloud for Weather Condition

Figure 12 shows that the weather condition when the accidents has occured. The weather is "Fair" most of the times when the accident took place, followed by the "mostly cloudy" and "cloudy" conditions.

H. **Wind Direction**
Figure 13 shows that the word cloud for Wind Direction. The most frequent word that appears is "Calm", which means the speed of wind was calm and thus was not in particular direction. Furthermore, we can see that West direction of windspeed was least encountered in table.

Figure 13: Word Cloud for Wind Direction

I. **Side**
Figure 14 sides It can be observed that the majority of the incidents occurred on the right side of the road as opposed to the left side, with the right side having 2353309 accidents and a total of 492032 on the left.

J. **Accidents Reported yearly**
Figure 15 presents the increase in number of accidents with respect to year. It can be seen that the reported number of accidents in 2021 were more than 1.4 million which is way higher than the reported accidents from 2020.

6 CONCLUSION

The study uses exploratory data analysis to investigate various inferences in order to produce results and forecast statistical descriptive measures for the frequency of car accidents. In regards

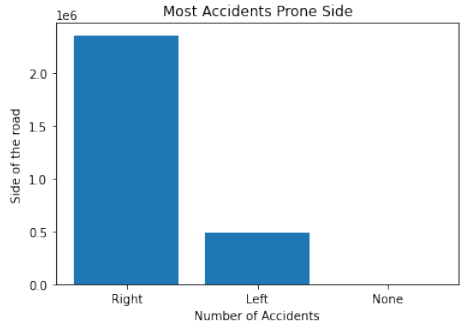

Figure 14: Number of accidents with respect to different sides

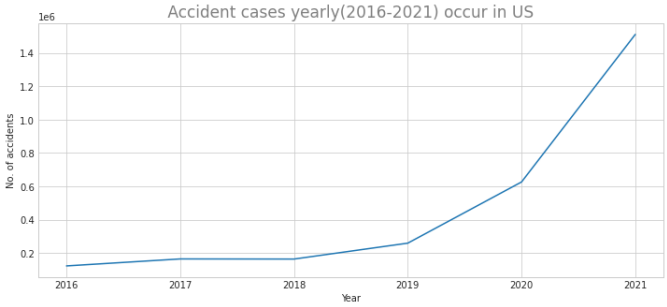

Figure 15: Number of accidents with respect to years

with state California is expected to have the highest severity rate of accidents (with severity = 2, which is classified as mid range) according to plotted stacked bar graph. Out of 11682 cities the highest number of accidents is anticipated in Miami from 2016 to 2021, the histogram distribution shows the high and low accidents. The majority of the accidents have happened during the workday between 3 and 6 p.m., and most of them have occurred during the second half of the year.The data confirms the idea that accidents occur more frequently during the morning and evening rush hours on weekdays. Weekends have the peak throughout the afternoon. The top 5 states by accidents include prominent ones like Los Angeles, Houston, and Florida. Between February 2016 and December 2020, less than 5% (4.35%) of cities had more than 1000 accidents.In most circumstances, the weather seems to be "Fair" or "Mostly Cloudy." Most of the accidents occur when the temperature is in between 50-60 degrees Fahrenheit.

REFERENCES

[1] Mohd Azrul Hisham Mohd Adib, Muhammad Irfan Abdul Jalal, and Nur Hazreen Mohd Hasni. 2020. Improvement of the BT-Heartomotive Device for Avert Car Accident Using MYBradyTachyHeart Mobile Application. In *Proceedings of the 2019 6th International Conference on Bioinformatics Research and Applications* (Seoul, Republic of Korea) *(ICBRA '19)*. Association for Computing Machinery, New York, NY, USA, 105–110. https://doi.org/10.1145/3383783.3383805

[2] Jean Raymond Daher, Satish Chilkaka, Abdunnaser Younes, and Khaled Shaban. 1970. Association rule mining on five years of motor vehicle crashes. https://qspace.qu.edu.qa/handle/10576/20505

[3] Subasish Das, Anandi Dutta, Mohammad Jalayer, Apoorba Bibeka, and Lingtao Wu. 2018. Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology* 7, 2 (2018), 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001

[4] S. A. Evtyukov, I. S. Brylev, M. M. Blinder, V. R. Rogov, and D. A. Ptitsyn. 2021. Modeling of a Road Traffic Accident Using Multivariate Analysis of Injuries in a

Two-Wheeled Vehicle Collision with a Car. In *2021 Systems of Signals Generating and Processing in the Field of on Board Communications*. 1–4. https://doi.org/10.1109/IEEECONF51389.2021.9415998

[5] Mingchen Feng, Jiangbin Zheng, Jinchang Ren, and Yanqin Liu. 2020. Towards Big Data Analytics and Mining for UK Traffic Accident Analysis, Visualization amp; Prediction. In *Proceedings of the 2020 12th International Conference on Machine Learning and Computing* (Shenzhen, China) *(ICMLC 2020)*. Association for Computing Machinery, New York, NY, USA, 225–229. https://doi.org/10.1145/3383972.3384034

[6] Tian Jian. 2021. Statistical Analysis and Countermeasures Study on Major Accidents of Highway and Waterway in China. In *2021 6th International Conference on Transportation Information and Safety (ICTIS)*. 694–697. https://doi.org/10.1109/ICTIS54573.2021.9798527

[7] Qiuming Jiang and Song Li. 2022. Traffic Accident Data Analysis Based on Internet of Things and Cloud Computing Technology. In *Proceedings of the 3rd Asia-Pacific Conference on Image Processing, Electronics and Computers* (Dalian, China) *(IPEC '22)*. Association for Computing Machinery, New York, NY, USA, 1045–1048. https://doi.org/10.1145/3544109.3544995

[8] Sinan Kaplan, Mehmet Amac Guvensan, Ali Gokhan Yavuz, and Yasin Karalurt. 2015. Driver Behavior Analysis for Safe Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 16, 6 (2015), 3017–3032. https://doi.org/10.1109/TITS.2015.2462084

[9] Prashant V. Krishnan, Vivek Chandra Sheel, M.V.S. Viswanadh, Chetan Shetty, and S. Seema. 2018. Data Analysis of Road Traffic Accidents to Minimize the rate of Accidents. In *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*. 247–253. https://doi.org/10.1109/CSITSS.2018.8768736

[10] Wu Lingtao, Yang Manjuan, Tang Chengcheng, Peng Daoyue, Shen Guohua, Zhang Tiejun, and Guo Zhanyang. 2010. Research on Influence Extention of Two-Lane Highway Intersections Based on Traffic Accident Database. In *Proceedings of the 2010 International Conference on Optoelectronics and Image Processing - Volume 02 (ICOIP '10)*. IEEE Computer Society, USA, 244–246. https://doi.org/10.1109/ICOIP.2010.51

[11] Tao Lu, Zhu Dunyao, Yan Lixin, and Zhang Pan. 2015. The traffic accident hotspot prediction: Based on the logistic regression method. In *2015 International Conference on Transportation Information and Safety (ICTIS)*. 107–110. https://doi.org/10.1109/ICTIS.2015.7232194

[12] NHTSA Media. 2022. NHTSA early estimates show overall increase in roadway deaths in first half of 2022. https://www.nhtsa.gov/press-releases/early-estimates-traffic-fatalities-first-half-2022

[13] Mpho Mokoatle and Vukosi Marivate. 2018. Collision Course: Challenges with Road Traffic Accident Data in South Africa. In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. 1–6. https://doi.org/10.1109/ICABCD.2018.8465419

[14] Ziwen Niu, Yanli Wang, and Shibao Sun. 2022. Correlation Analysis of Traffic Accident Factors Based on Mean Clustering. In *Proceedings of the 7th International Conference on Cyber Security and Information Engineering* (Brisbane, QLD, Australia) *(ICCSIE '22)*. Association for Computing Machinery, New York, NY, USA, 569–575. https://doi.org/10.1145/3558819.3565146

[15] I. A. Novikov, A. E. Borovskoy, Y. V. Gorbun, A. V. Terentyev, and M. G. Pletney. 2021. Geographic Information Systems to Improve Road Safety. In *2021 Systems of Signals Generating and Processing in the Field of on Board Communications*. 1–6. https://doi.org/10.1109/IEEECONF51389.2021.9416096

[16] Niklas Puller, Hendrik-Jörn Günther, Gwendal Lucas, André Leschke, and Vittorio Rocco. 2021. Towards Increasing VRU Safety: A Map-Based and Data-Driven Analysis of Accident Black Spots. In *2021 IEEE Vehicular Networking Conference (VNC)*. 60–67. https://doi.org/10.1109/VNC52810.2021.9644619

[17] Sri Siddhartha Reddy, Yen Ling Chao, Lakshmi Praneetha Kotikalapudi, and Ebrima Ceesay. 2022. Accident analysis and severity prediction of road accidents in United States using machine learning algorithms. In *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. 1–7. https://doi.org/10.1109/IEMTRONICS55184.2022.9795852

[18] Reem Elfatih Salman and Ayman Alzaatreh. 2022. Market Basket Analysis of Chicago Road Accidents. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*. 1–7. https://doi.org/10.1109/ASET53988.2022.9734867

[19] I-Ying Wu, Fandel Lin, and Hsun-Ping Hsieh. 2021. Dual-Attention Multi-Scale Graph Convolutional Networks for Highway Accident Delay Time Prediction. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems* (Beijing, China) *(SIGSPATIAL '21)*. Association for Computing Machinery, New York, NY, USA, 554–563. https://doi.org/10.1145/3474717.3484259