Analysis of Vehicle Alert Data

A Comprehensive Report on Vehicle Alert Patterns and Trends

Intel Unnati Grand Challenge Summer 2023

St. Vincent Pallotti College of Engineering and Technology, Nagpur

September 2023

Authors

Om Sanjay Rode

Prajwal Vijay Atkale

Prathamesh Lakhe

Introduction

In this report, we present a comprehensive analysis of vehicle alert data collected over a period of three months, from June 1, 2022, to August 31, 2022. The data was obtained from five vehicles equipped with Intel's ADAS CAS (Collision Avoidance System), which meticulously logged various types of alerts that occurred during the data collection period. The primary objective of this project was to unearth hidden patterns within the dataset and present them in an easily understandable manner, ensuring that readers, including those without extensive technical knowledge, can grasp the insights effectively.

Key Findings

1. Alert Frequency
   We found that the most frequent alert type in the dataset is "cas_hmw" (headway monitoring and warning). We used **EDA** (Exploratory Data Analysis) to extract valuable insights from specific data points like total count, uniqueness and frequency. EDA helped us narrow down our focus and further helped in performing different data wrangling operations.
2. Time-Based Trends
   Alerts vary by day and time, with higher occurrences during certain hours and days of the week. These trends can reveal how various factors, such as the occurrence of different types of alerts, vehicle speed, or alert frequency, change or fluctuate as time progresses. Identifying time-based trends helped us extracting key metrics of data.
3. Alert Severity
   Although we do not have explicit severity information, warnings such as "cas_fcw" (collision warning) are generally considered more severe. An alert is issued when the system detects that the vehicle is rapidly approaching an object or another vehicle in its lane. This indicates an impending collision hazard, which is an important warning to the driver.

Detailed Analysis

1. *Summary of Data*
   - Latitude ('Lat'):

The data has a count of **21,325**, indicating that there are no missing values for this column. The mean latitude is approximately **12.9005**, suggesting that the data is centred around this value. The standard deviation (std) of **0.1472** is relatively small, indicating that the latitude values are closely packed around the mean. The minimum and maximum latitude values are **12.3387** and **13.1828**, respectively, suggesting that the data spans a geographical range.

	Lat	Long	Speed
count	21325	21325	21325
mean	12.900458	80.11853	38.40385
std	0.147163198	0.107352	16.84761
min	12.338685	79.77412	0
25%	12.850645	80.06598	27
50%	12.942436	80.12999	41
75%	13.006814	80.20568	54
max	13.182797	80.31608	65

- Longitude ('Long'):

Similar to latitude, there are no missing values for the longitude column. The mean longitude is approximately **80.1185**, indicating a central tendency around this value. The standard deviation (std) of **0.1074** suggests that longitude values are relatively closely clustered. The minimum and maximum longitude values are **79.7741** and **80.3161**, respectively, indicating a geographical span.

- Vehicle Speed ('Speed'):

The data for vehicle speed also has no missing values. The mean speed is approximately **38.404**, indicating a central speed value. The standard deviation (std) of **16.848** suggests a moderate level of variability in vehicle speeds. The minimum speed value is 0, which could indicate that some data points represent stationary or very slow-moving vehicles. The maximum speed value is **65**, indicating the highest recorded vehicle speed.

2. *Descriptive Analysis of specific columns and insights*

- Date-Time

For the simplicity of analysis, we have parsed two individual columns of date and time into single column, Parsing date and time column enabled us using pythons default date time format *(%d-%m-%Y %H: %M: %S).* Which further reduces complexity of existing DA (Data Analysis) operations and lowers down the redundancy of the dataset. There are **20,674** unique datetime values in the column. This means that there are some instances where multiple records have the same timestamp. The most frequently occurring datetime value in the dataset is "**15-08-2022 17:48**," which appears **5 times**. This suggests that this specific timestamp had the highest frequency of recorded events. The earliest timestamp in the dataset is "**01-06-2022 05:36**," indicating the start date and time of data collection. The latest timestamp in the dataset is "**31-08-2022 18:19**," indicating the end date and time of data collection.

	Datetime
count	21325
unique	20674
top	15-08-2022 17:48
freq	5
first	01-06-2022 05:36
last	31-08-2022 18:19

- Alert

In the 'Alert' column, we have a total of **21,325** records, indicating that there are no missing values in this category. Among these records, we find four distinct alert types. Notably, "**cas_hmw**" alerts are the most frequent, occurring a remarkable 12,328 times, making them the most common alert type in our dataset. This insight highlights that "**cas_hmw**" alerts dominate the recorded events compared to the other alert types.

	Alert
count	21325
unique	4
top	cas_hmw
freq	12328

- Vehicle

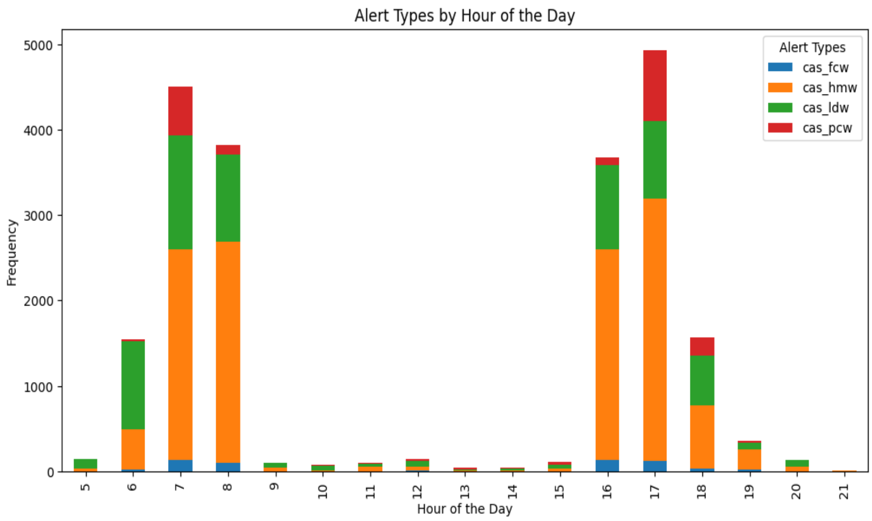
In the 'Vehicle' category, we have a total of **21,325** records, indicating that all data points are complete in this column. This dataset includes five unique vehicle identifiers. *Among these vehicles, '805' is the most frequently observed*, with a total of **6,875** instances. This suggests that vehicle '805' is the most active or frequently monitored vehicle in the dataset, potentially deserving special attention in further analysis or investigations.

	Vehicle
count	21325
unique	5
top	805
freq	6875

Detailed Alert Type Analysis

- Time based Trends

In the process of alert-based analysis, we initially segmented the data by the 'hour' component derived from the previously parsed 'Date-Time' column. This segmentation provided us with the necessary data points for further exploration. Subsequently, we introduced a pivotal variable that forms the core of our stacked chart visualization. Upon plotting the graph depicting the frequency of alerts against the hours of the day, distinct patterns emerged. Our observations reveal that



alerts are notably more frequent during specific periods, primarily occurring between **6 to 9 AM** in the morning and from **4 to 7 PM** in the evening. *Among these alerts, the orange bars, symbolizing 'cas_hmw' incidents, stand out as the most prevalent throughout the day*. To gain a deeper understanding of why 'cas_hmw' incidents are so prominent, it's imperative to comprehend the nature of this alert. 'cas_hmw' signifies situations where the vehicle ahead fails to maintain a safe distance, which raises concerns about road safety. *This pattern in 'cas_hmw' alerts is indicative of the typical traffic conditions in Chennai, a bustling metropolis notorious for its traffic congestion. In the subsequent sections of this report, we will delve into the specific areas within the city that frequently experience 'cas_hmw' alerts and traffic congestion.* To visualize these trends effectively, we will employ heatmap representations of alerts.

- Analysis of Alert Hotspots

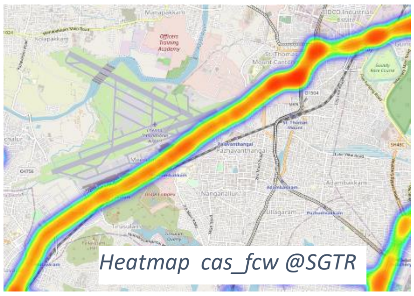
In our thorough data analysis, we've unveiled intriguing trends that shed light on specific locations in Chennai where severe alerts occur more frequently than in others. To satiate our curiosity and enhance the data's visual impact, we opted for visual representations such as heatmaps (utilizing Python's Folium library), real-time traffic data from Google Maps, and pertinent articles that corroborate our findings.

Heatmap cas_fcw @SGTR

Our investigation uncovered that certain areas in Chennai exhibit a higher frequency of severe alerts compared to others. *Notably, the "Southern Grand Trunk Road" emerged as a hotspot for alerts, particularly those categorized as "cas_fcw," "cas_hmw," and "cas_ldw."*

To provide a clearer picture of our findings:

1. Heatmaps: We've used heatmaps to visually represent the density of alerts along the "Southern Grand Trunk Road."



Heatmap cas_fcw @SGTR

Typical Traffic @ SGTR

Live tarffic @SGTR

2. Real-time Traffic Data: We've cross-referenced our data with real-time traffic conditions from Google Maps, revealing a correlation between alert hotspots and areas prone to traffic congestion.

Six killed in accident on Grand Southern Trunk Road

May 06, 2014 01:44 pm | Updated 01:44 pm IST - KANCHEEPURAM:

STAFF REPORTER

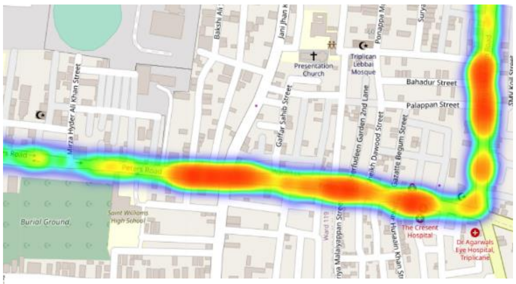
COMMENTS SHARE READ LATER

Six people travelling in a car, five of whom were of the same family, were killed in a road accident on Grand Southern Trunk Road at Madurantakam on Monday. The victims include an 18-month-old boy

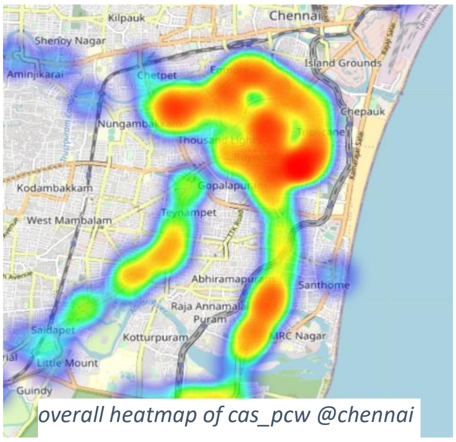
Insights and Implications:

Traffic Vulnerability: These findings underscore the vulnerability of the "Southern Grand Trunk Road" to severe alerts, signifying potential road safety challenges in this area. In conclusion, our comprehensive analysis not only identifies areas of heightened alert activity but also illustrates the significance of visual aids, such as heatmaps and real-time traffic data, in enhancing data interpretation. This empirical evidence emphasizes the need for targeted safety measures and heightened awareness on the "Southern Grand Trunk Road" to address traffic-related issues effectively. Articles Involving accidents @Southern Grand Trunk road (bit.ly/3qYulpn)

- Analysis of Alert Relevance to Population Density



Heatmap of cas_pcw along St. peters road and Triple cane road



overall heatmap of cas_pcw @chennai

In our analysis, we've uncovered a noteworthy pattern: Most "cas_pcw" alerts are found in areas around Chennai, particularly in busy spots such as *Peters Road and Triplecane Highway Road*. Close proximity to *Marina Beach, Arulmigu Sri Parthasarathy Perumal Temple Tiruvallikeni, and MA Chidambaram Stadium, Chepauk,* characterizes these regions. Urban centers with high population density require stringent safety protocols, an insight that underscores this fact. These bustling urban centers serve as vital arteries of Chennai, teeming with the vibrancy of city life. The data suggests that this alert type is especially relevant in areas where people congregate in significant numbers.

Speed and Alert Analysis

	Lat	Long	Vehicle	Speed	Alert_cas_fcw	Alert_cas_hmw	Alert_cas_ldw	Alert_cas_pcw
Lat	1	0.958751	0.231491	-0.37855	0.038213145	0.326797188	-0.461388346	0.152053695
Long	0.958751407	1	0.227512	-0.41938	0.039043167	0.315130577	-0.476654668	0.195618388
Vehicle	0.231490903	0.227512	1	-0.06997	0.028046542	0.070499461	-0.144476264	0.092733645
Speed	-0.378553566	-0.41938	-0.06997	1	-0.020403064	-0.214506756	0.488312294	-0.395999044
Alert_cas_fcw	0.038213145	0.039043	0.028047	-0.0204	1	-0.197456502	-0.11084282	-0.053906135
Alert_cas_hmw	0.326797188	0.315131	0.070499	-0.21451	-0.197456502	1	-0.769185401	-0.374077563
Alert_cas_ldw	-0.461388346	-0.47665	-0.14448	0.488312	-0.11084282	-0.769185401	1	-0.209989601
Alert_cas_pcw	0.152053695	0.195618	0.092734	-0.396	-0.053906135	-0.374077563	-0.209989601	1

The correlation matrix grants access to information regarding dataset relationships. Addressing this case requires calculating the correlation matrix for Lat(Latitude), Long(Longitude), Vehicle, Speed, and Alert types. Lat and Long Correlation: Strongly related through a positive correlation of nearly 0.96, 'Lat' and 'Long' columns. Longitude increases alongside increasing latitude, signifying a spatial link. Co ordinates of a particular location are identified by both 'Lat' and 'Long'.Speed and Alert Types Correlation: The Speed column's weak correlation exists only with one-hot encoded alert types (Alert_cas_fcw, Alert_cas_hmw, Alert_cas_ldw, Alert_cas_pcw). Closeness to zero implies no significant linear connection between speed and type of alert Alert Types Correlation: Among themselves, correlations exist within the one-hot encoded alert types. For example: With approximately -0.77 correlation, 'Alert_cas_hmw' and 'Alert_cas_ldw' are negatively linked. When one type of alert happens more, the other happens less frequently, as indicated by this.Other alert types have weak correlations with 'Alert_cas_fcw', implying little influence.

Notably, "cas_hmw" alerts were the most frequent, occurring 12,328 times, often associated with speeds averaging 35.32 km/h. Conversely, "cas_ldw" alerts, signaling lane departure warnings, were recorded 6,431 times, with speeds around 50.92 km/h. "cas_fcw" alerts for forward collision warnings appeared 590 times at an average speed of 36.37 km/h. In contrast, "cas_pcw" alerts, denoting pedestrian collision warnings, occurred 1,976 times, with a lower average speed of 17.53 km/h. These insights illuminate distinct driving scenarios and underscore the significance of tailored safety systems.

Alert	Speed	Count
cas_fcw	36.36610169	590
cas_hmw	35.31659637	12328
cas_ldw	50.92349557	6431
cas_pcw	17.52732794	1976

Recommendations

Based on our comprehensive analysis, we propose the following recommendations: Prioritize road maintenance in alert-prone areas like the Grand Southern Trunk Road to improve road conditions and reduce alerts. Promote public awareness campaigns on responsible driving to mitigate alerts related to unsafe distances (cas_hmw) and disobedient road behaviour.

Conclusion

Through our in-depth analysis of the dataset, we have unearthed insights that challenge conventional perceptions of road safety. While speed is often correlated with road accidents, our findings reveal a different narrative. We observe that the majority of alerts do not occur at higher speeds. Instead, factors such as road conditions and traffic management emerge as critical determinants of road safety. The urban landscape, population density, and the behaviour of citizens all play significant roles in shaping the occurrence of these alerts. In light of these revelations, it is imperative to reconsider our approach to road safety, moving beyond a singular focus on speed. Our study underscores the multifaceted nature of road safety challenges and the need for holistic, context-aware solutions.