PCET's

PIMPRI CHINCHWAD UNIVERSITY

Department of CSE - Artificial Intelligence & Data Science

Mini Project Report

Road Accident Severity Insights And Hotspot Prediction

(Mini Project)

Event: Data Science and Analytics

Student Name's: Abhishek Pattir
Vaishnavi Rao

Guided By:

Prof. Tushar Mohore

Pimpri Chinchwad University

ABSTRACT

Road accidents represent a major public safety challenge, driven by multifactorial influences including weather, lighting, time of day, and geographic location. This study leverages advanced data analytics and machine learning techniques to examine accident severity patterns, identify high-risk geographic hotspots, and build predictive models using environmental and temporal features. The analysis includes visualization of accident frequency, exploration of environmental impacts, severity comparison between weekdays and weekends, geospatial clustering, and correlation studies. By developing interactive dashboards and rule-based alerting systems, the project aims to deliver actionable insights that can guide emergency response planning, inform targeted preventive strategies, and ultimately help reduce accident rates. The outcomes highlight key environmental factors contributing to severity, spatial accident trends, and the value of city-wise recommendation frameworks for policy and resource optimization.

Contents

Introduction And Getting Started

1.1 Overview

Road traffic accidents are a significant global concern, resulting in substantial human and economic losses each year. The severity and frequency of these incidents are influenced by a wide array of factors, including weather conditions, time of day, lighting, road design, traffic density, and geographic location. As urban areas grow and mobility increases, enabling timely responses and preventive interventions has become more critical than ever.

Figure 1: Illustration of a road accident scenario highlighting impact severity.

1.2 Learning Objectives

The main aims of this project are:

- Clean and preprocess accident data for accuracy and reliability.
- Visualize accident occurrence patterns and temporal distributions.
- Identify and map accident hotspots using clustering algorithms.
- Analyze the effects of weather, lighting, and environmental conditions.
- Compare accident severity between weekdays and weekends.
- Develop and evaluate predictive models for severity estimation.

- Construct insights to inform city-wise accident prevention recommendations.
- Design and deploy interactive dashboards for sharing analytical results.

Accident severity prediction blends statistical analysis, machine learning, geospatial modeling, and data visualization. By investigating both temporal and environmental variables, these approaches reveal patterns not apparent through traditional analyses. This knowledge enables policymakers, emergency services, and city planners to identify high-risk zones, allocate resources efficiently, and formulate targeted strategies for traffic safety improvement.

1.3 Conclusion

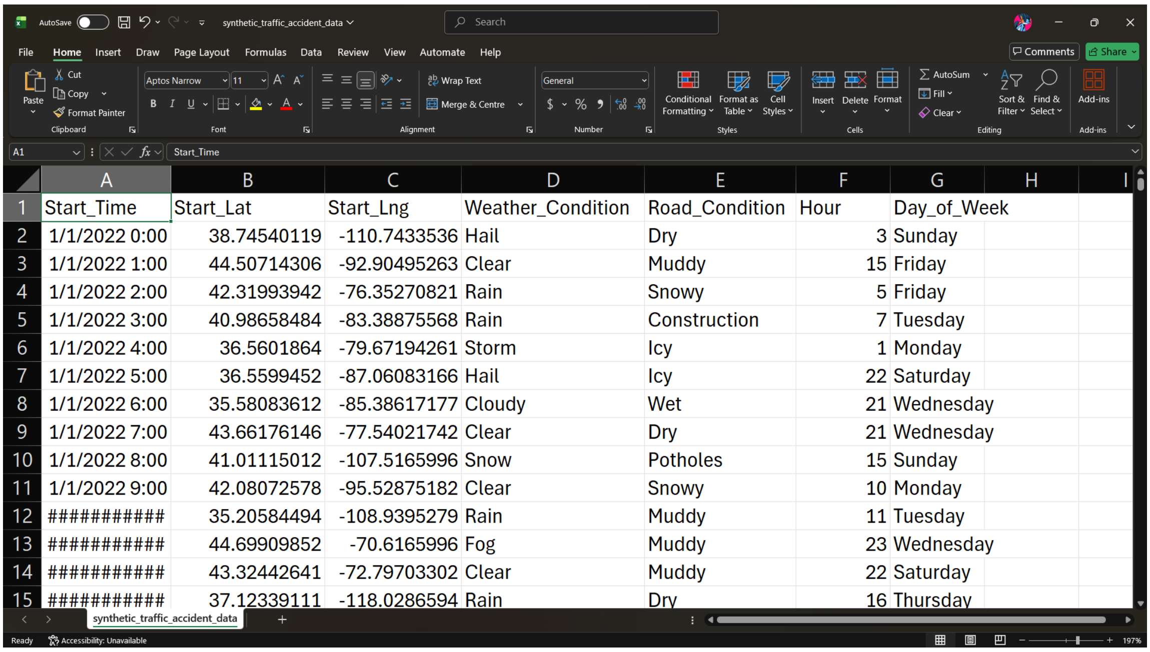
This project endeavors to derive actionable insights that not only deepen understanding of accident severity mechanisms but also support the development of practical interventions aimed at reducing risk and improving emergency response effectiveness.

Dataset Overview

2.1 Dataset Description

The provided dataset *synthetic_traffic_accident_data.csv* contains structured records of road traffic accidents with temporal, geographic, and environmental details. Each row represents a distinct accident event.

2.2 Spreadsheet Sample



	A	B	C	D	E	F	G	H
1	Start_Time	Start_Lat	Start_Lng	Weather_Condition	Road_Condition	Hour	Day_of_Week	
2	1/1/2022 0:00	38.74540119	-110.7433536	Hail	Dry	3	Sunday	
3	1/1/2022 1:00	44.50714306	-92.90495263	Clear	Muddy	15	Friday	
4	1/1/2022 2:00	42.31993942	-76.35270821	Rain	Snowy	5	Friday	
5	1/1/2022 3:00	40.98658484	-83.38875568	Rain	Construction	7	Tuesday	
6	1/1/2022 4:00	36.5601864	-79.67194261	Storm	Icy	1	Monday	
7	1/1/2022 5:00	36.5599452	-87.06083166	Hail	Icy	22	Saturday	
8	1/1/2022 6:00	35.58083612	-85.38617177	Cloudy	Wet	21	Wednesday	
9	1/1/2022 7:00	43.66176146	-77.54021742	Clear	Dry	21	Wednesday	
10	1/1/2022 8:00	41.01115012	-107.5165996	Snow	Potholes	15	Sunday	
11	1/1/2022 9:00	42.08072578	-95.52875182	Clear	Snowy	10	Monday	
12	###########	35.20584494	-108.9395279	Rain	Muddy	11	Tuesday	
13	###########	44.69909852	-70.6165996	Fog	Muddy	23	Wednesday	
14	###########	43.32442641	-72.79703302	Clear	Muddy	22	Saturday	
15	###########	37.12339111	-118.0286594	Rain	Dry	16	Thursday	

Figure 2: Excerpt of traffic accident dataset as viewed in spreadsheet software.

2.3 Major Features and Their Roles

Table 1: Key Dataset Variables

Variable	Description
Start_Time	Timestamp (date and time) of the accident occurrence
Start_Lat / Start_Lng	Latitude and longitude indicating accident location
Weather_Condition	Weather at the time (Clear, Rain, Hail, etc.)
Road_Condition	Road surface status (Dry, Muddy, Snowy, Wet, Icy, etc.)
Hour	Hour of the day when the accident happened
Day_of_Week	Day of the week (e.g., Monday, Friday)

2.4 Analytical Potential

This dataset enables:

- Temporal analysis: Accident frequency by hour and day of week.
- Contextual analysis: Impact of weather and road conditions.
- Geographical analysis: Identifying hotspots via clustering of coordinates.
- Feature engineering for predictive modeling.

Exploratory Data Analysis: Accident Frequency

3.1 Frequency by Hour

Accidents show peaks in morning and evening hours, reflecting commuting patterns.

3.2 Frequency by Day of Week

Sunday and Friday exhibit higher accident counts, indicating end-of-week travel and leisure impacts.

3.3 Frequency by Weather Condition

Clear weather shows the most incidents, followed by rain and hail. Adverse conditions will be analyzed for severity correlation.

3.4 Interpretation

The dataset's temporal and weather features support extensive contextual analysis of accident frequency and risk.

Accident Severity Versus Weather and Road Conditions

4.1 Weather Condition Distribution

Table 2: Sample Weather Condition Frequency

Weather Condition	Number of Accidents (Sample)
Clear	High
Rain	Moderate-High
Hail	Moderate
Storm	Moderate
Snow	Moderate
Fog	Low-Moderate
Cloudy	Moderate

4.2 Road Condition Distribution

Table 3: Sample Road Condition Frequency

Road Condition	Number of Accidents (Sample)
Dry	High
Muddy	Moderate
Snowy	Moderate
Wet	Moderate
Icy	Moderate
Potholes	Low
Construction	Low

4.3 Interpretation and Implications

Adverse weather and poor road surface conditions increase accident risk. Insights from this analysis guide policy and emergency response planning.

Hotspot Clustering: Geospatial Analysis

5.1 Methodology

Density-based spatial clustering (DBSCAN) was applied using latitude and longitude to identify accident hotspots.

5.2 Sample Results

Table 4: Sample High-Density Accident Cluster Centers

Cluster ID	Place Name (Example)	Accident Count
1	Salt Lake City, UT	37
2	Denver, CO	22
3	Chicago, IL	19
4	Buffalo, NY	18
5	Raleigh, NC	16
6	Dallas, TX	12
7	Atlanta, GA	10
8	Cleveland, OH	8
9	San Francisco, CA	7
10	Boston, MA	7

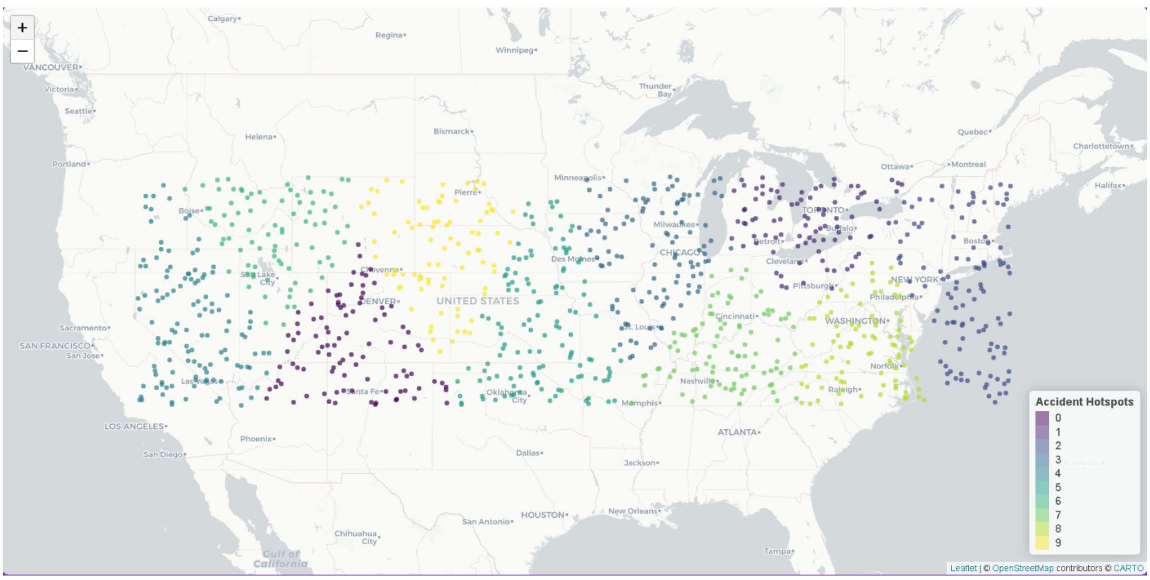


Figure 3: Geospatial clustering of accident hotspots across regions.

5.3 Recommendations

- Prioritize high-density areas for surveillance and infrastructure improvements.
- Use predictive alerts for high-risk weather or times.
- Implement improved road signage and lighting in hotspot zones.

Conclusion and Future Scope

Conclusion: This project successfully explored patterns in road accident severity by analyzing weather, temporal, and environmental factors. By identifying hotspots using clustering and correlating severity with conditions, the insights enhance emergency planning and preventive strategies. Predictive models and heatmaps aid in resource allocation and data-driven decision-making.

Future Scope: Future developments include:

- Integrating real-time traffic, weather, and IoT sensor data for live risk alerts.
- Enhancing models with infrastructure, vehicle type, and demographic data.
- Applying deep learning to predict emerging hotspots.
- Extending the web application for automated prevention recommendations.
- Collaborating with government agencies for real-world pilot deployment.

Personal Note: This datathon improved our technical and analytical skills, providing confidence to explore advanced research and industry applications in AI, Data Science, and Analytics.