

Predictive Modeling of Road Accident Severity Across Tamil Nadu: A Query Optimization Techniques

Dr. I. Shahina Begam¹, Dr. K. Tajudin², A. Ramya³

¹Head, Assistant Professor, ³Assistant Professor, Department of Computer Science, Adhiyaman Arts and Science College for Women, Uthangarai, Krishnagiri (Dt), Tamil Nadu State, India

²Assistant Professor, PG Department of Computer Science, The New College, Royapettah, Chennai-600014, Tamil Nadu State, India

Abstract— Road traffic accidents represent a critical public safety issue requiring data-driven evaluation and strategic intervention. This study applies a data science framework to analyze district-wise road accident and fatality data across 38 districts of Tamil Nadu for the period 2021–2023. The dataset is structured using relational database architecture, consisting of district-level attributes and annual accident and mortality records. Data preprocessing and structured query operations were performed for extraction, transformation, and aggregation of relevant variables. Exploratory Data Analysis (EDA) techniques were employed to identify spatial and temporal patterns, while statistical measures such as percentage growth rate, proportional contribution, and comparative trend analysis were utilized to examine variations in accident frequency and fatality severity across districts and years. The results indicate notable inter-district disparities, with 2022 emerging as the peak year in terms of fatalities. Several districts exhibited a measurable increase in accident severity compared to 2021, whereas others maintained relatively stable trends. The integration of data analytics techniques enables systematic pattern recognition, risk stratification, and preliminary forecasting potential. By transforming raw accident records into actionable insights, the study demonstrates the effectiveness of data science methodologies in supporting evidence-based decision-making.

Keywords—Query Analysis – Predictive Calculation-Result Summary-Data Science.

I. INTRODUCTION

Every year, road accidents create serious public safety challenges, resulting in heavy human casualties and economic losses. To design effective safety regulations and preventive frameworks, it is necessary to examine accident data through systematic and scientific analysis.

The present study evaluates district-wise accident and fatality statistics in Tamil Nadu for the years 2021, 2022, and 2023, covering all 38 districts of the state. Annual totals of accidents and deaths were organized within a relational database system to ensure structured storage and retrieval.

Various data science techniques—such as data cleaning, aggregation, percentage growth measurement, comparative assessment, and trend evaluation—were applied to transform raw numerical data into interpretable insights.

Through a data-driven methodology, the research classifies districts into higher-risk and lower-risk categories, examines fluctuations in accident severity across years, and tracks variations in fatality rates over time. Descriptive statistical summaries and structured database queries were used to highlight distribution patterns, while comparative evaluations supported the identification of temporal changes.

Across modern research domains, Data Science serves as an essential interdisciplinary field that integrates statistics, mathematics, computer science, machine learning, and domain expertise. It enables researchers to systematically manage and interpret large datasets, ultimately converting unprocessed data into reliable evidence for informed decision-making.

The data science workflow begins with organized data collection and storage, followed by preprocessing steps such as handling missing entries, removing duplicates, correcting inconsistencies, and restructuring variables into suitable analytical formats. Well-prepared data improves the validity and accuracy of research findings.

In addition, predictive analytics and simulation techniques expand the scope of research by enabling future trend estimation and scenario analysis. For example, healthcare studies apply predictive models to anticipate disease spread, while transportation research uses accident forecasting models to identify vulnerable zones and enhance safety planning.

II. RELEATED WORK

In recent years, several research studies have applied data science techniques to analyze large-scale road accident datasets for improving road safety and decision-making.

Similar to these studies, the present research focuses on collecting and analyzing accident data from all 38 districts of Tamil Nadu for a period of three years (2021–2023)[1].

Query processing and analytical techniques are widely used in prior research to extract specific insights from large datasets. Structured Query Language (SQL) and analytical queries help answer important questions such as:

- Which district recorded the highest number of accidents?
- What is the year-wise trend of accident growth?
- Which districts show increasing or decreasing fatality rates?
- What are the major causes contributing to deaths?

Exploratory Data Analysis (EDA), comparative statistics, and visualization methods such as bar charts, trend lines, and heatmaps are applied to interpret the processed results. These approaches help in identifying high-risk districts, accident-prone zones, and dominant contributing factors.

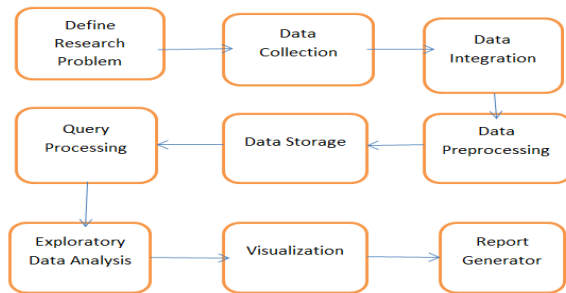


Figure 1. Query Optimization Techniques

III. QUERY OPTIMIZATION AND RESULT

Create a two table trans1 and trans2, the table 1 create a unique field and table 2 references of first table inserted 38 district entries entered in the first table, the second table based on the references of first table entries inserted district no,year,total number of accident and total number of death.

```

SQL> desc trans1;
+-----+-----+-----+
Name                                         Null?   Type
+-----+-----+-----+
NO                                           NOT NULL NUMBER(10)
NAME                                         VARCHAR2(30)

SQL> desc trans2;
+-----+-----+-----+
Name                                         Null?   Type
+-----+-----+-----+
NO                                           NUMBER(10)
YEAR                                         VARCHAR2(5)
TOT_ACC                                     NUMBER(10)
TOT_DEATH                                    NUMBER(10)

SQL> |
  
```

Figure.2 Table Creation

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

The database contains two tables: **TRANS1** and **TRANS2**.

The two tables are connected using the **district number (NO)**, enabling district-wise and year-wise analysis of road accidents and fatalities through SQL queries.

Table I: Yearwise Accident and Death Data

	2021	2022	2023
Total Accident	55682	64105	67213
Total Death	15384	17884	19347

The total number of accidents has shown a consistent upward trend from 2021 to 2023. In 2021, the total accidents were 55,682, which increased to 64,105 in 2022, reflecting a significant rise of 15.13%. In 2023, the figure further increased to 67,213, showing an additional growth of 4.85% compared to 2022. Overall, accidents increased by approximately 20.71% from 2021 to 2023, indicating a steady rise over the three-year period.

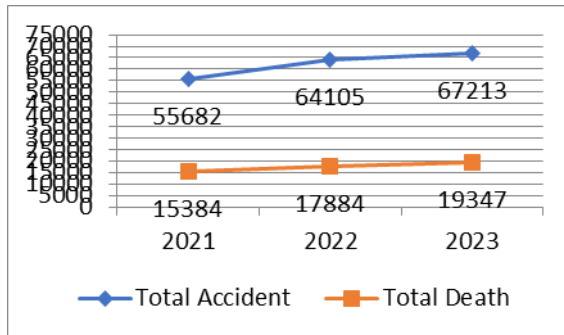


Figure 3. Comparison Chart Yearwise

Similarly, the total number of deaths also recorded a continuous increase during these years. The deaths rose from 15,384 in 2021 to 17,884 in 2022, marking a substantial increase of 16.25%. In 2023, the number further increased to 19,347, representing an 8.18% rise from the previous year. Overall, deaths increased by approximately 25.76% between 2021 and 2023, showing a comparatively higher growth rate than accidents.

Overall, both accidents and deaths show a consistent upward trend, with deaths increasing at a relatively higher percentage than accidents during the period 2021–2023.

```
SQL> select sum(tot_acc),sum(tot_death) from trans2;
SUM(TOT_ACC) SUM(TOT_DEATH)
```

```
-----
55682      15382
```

```
SQL> select *from trans2 where tot_acc=(select
max(tot_acc) from trans2);
```

```
NO YEAR    TOT_ACC TOT_DEATH
```

```
-----
3 2021      5035     999
```

```
SQL>select *from trans2 where tot_acc=(select
max(tot_acc) from trans2 where year=2021);
```

```
NO YEAR    TOT_ACC TOT_DEATH
```

```
-----
3 2021      5035     999
```

```
SQL>select *from trans2 where tot_death=(select
max(tot_death) from trans2 where year=2023);
```

```
NO YEAR    TOT_ACC TOT_DEATH
```

```
-----
4 2023      3657    1044
```

Same way we found 2022 and 2023 data also

Table II. Maximum Accident and Death (from 2021-2023)

District No	Name	Year	Max_acc	Max_death
3	Chennai City	2021	5035	999
4	Coimbatore	2022	3544	1057
4	Coimbatore	2023	3657	1044

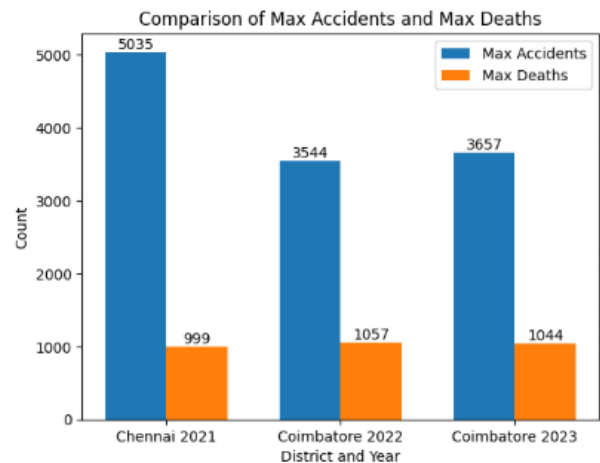


Figure 4. Comparison of Max Accidents and Max Death

The data shows that in 2021, Chennai City recorded the highest number of accidents (5,035) with 999 deaths. In 2022 and 2023, Coimbatore reported the maximum accidents, increasing from 3,544 to 3,657 cases. The number of deaths in Coimbatore slightly decreased from 1,057 in 2022 to 1,044 in 2023.

Overall, Chennai had the highest accident figure in 2021, while Coimbatore recorded higher fatalities in the subsequent years.

```
SQL> select *from trans2 where no=3;
NO YEAR    TOT_ACC TOT_DEATH
-----
3 2021      5035    999
3 2022      3453    507
3 2023      3654    504
```

In District No. 3, deaths decreased from 999 in 2021 to 507 in 2022, showing a sharp decline of 49.25%. In 2023, deaths slightly reduced to 504, marking a marginal decrease of 0.59%. Overall, there was a significant reduction in deaths over the three-year period.

```
SQL> select *from trans2 where no=4;
NO YEAR    TOT_ACC TOT_DEATH
-----
4 2021      2792    841
4 2022      3544   1057
4 2023      3657   1044
```

In District No. 4, deaths increased from 841 in 2021 to 1,057 in 2022, showing a significant rise of **25.68%**. In 2023, deaths slightly decreased to 1,044, reflecting a marginal decline of **1.23%** compared to 2022. Overall, there was an overall increase of about **24.14%** in deaths from 2021 to 2023.

TABLE III Maximum Death Yearwise

District No	District Name	Year	Max_Acc	Max_Death
3	Chennai City	2021	5035	999
3	Chennai City	2022	3453	507
3	Chennai City	2023	3654	504
4	Coimbatore	2021	2792	841
4	Coimbatore	2022	3544	1057
4	Coimbatore	2023	3657	1044

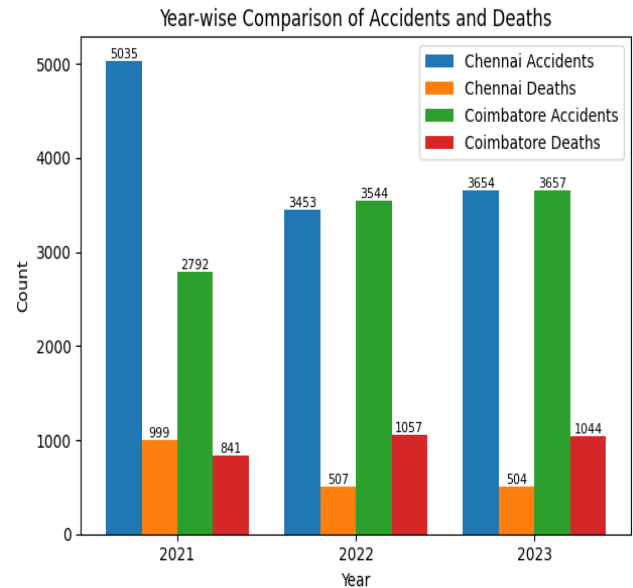


Figure 5. Year-Wise Comparison of Accidents and Details

The data compares accident and death statistics of Chennai and Coimbatore from 2021 to 2023. In 2021, Chennai recorded higher accidents (5,035) and deaths (999) than Coimbatore. However, in 2022 and 2023, Coimbatore reported slightly higher accident cases and significantly higher deaths compared to Chennai. While Chennai showed a reduction in deaths over the years, Coimbatore maintained a higher fatality count despite minor fluctuations.

Overall, the trend indicates that Chennai had higher figures initially, but Coimbatore surpassed Chennai in both accidents and deaths during the later years.

```
SQL>select sum(tot_acc),sum(tot_death) from trans2 where no=4;
SUM(TOT_ACC) SUM(TOT_DEATH)
```

```
-----
9993      2942
SQL>select sum(tot_acc),sum(tot_death) from trans2 where no=3;
SUM(TOT_ACC) SUM(TOT_DEATH)
```

```
-----
12142     2010
SQL> select sum(tot_acc),sum(tot_death) from trans2 union
select sum(tot_acc),sum(tot_death) from trans2 where no=4;
SUM(TOT_ACC) SUM(TOT_DEATH)
```

```
-----
9993      2942
187000    51559
```

District No. 4 recorded a total of 9,993 accidents out of 1,87,000 overall accidents, contributing approximately **5.34%** to the total accidents. In terms of fatalities, the district reported 2,942 deaths out of 51,559 total deaths, accounting for about **5.71%** of the overall deaths.

District No. 4 has a slightly higher share in total deaths (5.71%) compared to its share in total accidents (5.34%).

```
SQL> select sum(tot_acc),sum(tot_death) from trans2
union select sum(tot_acc),sum(tot_death) from trans2
where no=3;
```

```
SUM(TOT_ACC) SUM(TOT_DEATH)
```

```
-----
12142      2010
187000     51559
```

For District No. 3, the total accidents are 12,142 out of 1,87,000 overall accidents, which accounts for approximately **6.49%** of the total accidents. Similarly, the total deaths are 2,010 out of 51,559 overall deaths, representing about **3.90%** of the total deaths.

District No. 3 contributes 6.49% of total accidents and 3.90% of total deaths to the overall figures.

```
SQL> select *from trans2 where tot_death=(select
max(tot_death) from trans2);
```

```
NO YEAR   TOT_ACC TOT_DEATH
```

```
-----
4 2022    3544   1057
```

The data indicates that the maximum number of deaths was recorded in District No. 4 during the year 2022, with 1,057 deaths and 3,544 accidents. Compared to 2021, where the district reported 841 deaths, there was a sharp increase in fatalities in 2022. However, in 2023, the number of deaths slightly decreased to 1,044, though it remained higher than the 2021 level.

Overall, 2022 stands out as the most critical year in terms of deaths, showing a peak when compared to both 2021 and 2023.

```
SQL> select no,sum(tot_acc),sum(tot_death) from trans2
group by no order by no;
```

Table IV: District-wise Total Number of Accidents and Deaths for the Years 2021, 2022, and 2023

(Each cell represents District Number, Total Accidents, and Total Deaths for the respective year.)

1	1691	548	12	3344	1157
2	8060	2345	13	5905	1891
3	12142	2010	14	7432	2411
4	9993	2942	15	1431	299
5	9474	1613	16	2145	493
6	4017	1049	17	5844	1535
7	6045	1974	18	1590	535
8	6403	1873	19	4781	1324
9	3383	886	20	3628	1135
10	2960	840	21	2602	796
11	4077	977	22	9165	2386

23	3348	1124	34	9142	2570
24	3199	772	35	5110	1784
25	6400	1591	36	2787	868
26	659	152	37	7125	1667
27	3258	904	38	4384	1453
28	6493	1788			
29	2241	611			
30	3939	1337			
31	6601	2077			
32	4105	1182			
33	2097	660			

The dataset shows significant variation in accident frequency and fatality counts across 38 districts, with a few districts contributing disproportionately to total accidents and deaths. A few districts reported exceptionally high accident cases (above 9,000–12,000 cases), indicating higher traffic density, urbanization, or transportation activity. These districts also recorded relatively high fatality numbers, suggesting greater accident severity. In contrast, several districts reported lower accident counts (below 2,000–3,000 cases) with comparatively fewer deaths.

From 2021 to 2023, both total accidents and deaths show a continuous increase. Accidents rose from **55,682 in 2021** to **67,213 in 2023**, while deaths increased from **15,382** to **18,293**. This indicates a steady upward trend in road accidents and fatalities over the three-year period.

The death ratio is calculated as:

$$\text{Death Percentage} = (\text{Total Deaths} / \text{Total Accidents}) \times 100$$

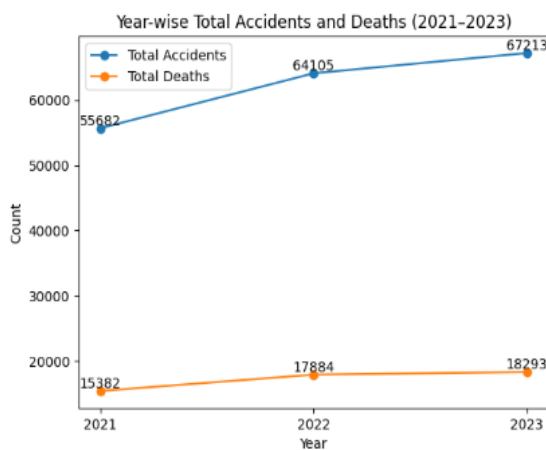


Figure 6. Year-Wise Total Accident Details [2021-2023]

The accident–death ratio remained almost consistent during 2021–2023. In 2021, the death percentage was **27.62%**, which slightly increased to **27.90%** in 2022 and then marginally decreased to **27.22%** in 2023. Overall, the fatality rate stayed around **27–28%**, indicating a relatively stable proportion of deaths compared to total accidents over the three-year period.

IV. RESULT DISCUSSION AND CONCLUSION

The analysis of district-wise accident data in Tamil Nadu for the years 2021–2023 highlights an overall increasing trend in road accidents and fatalities. Using **data science techniques** such as data cleaning, aggregation, percentage analysis, and comparative trend evaluation, meaningful insights were derived from 114 district-level records.

The year 2022 recorded the highest number of deaths across many districts, making it the most critical year in terms of fatality severity. Although 2023 showed slight improvements in certain districts, overall accident levels remained high.

Districts such as Chennai and Coimbatore consistently reported higher accident and death figures, indicating greater accident severity in urban and high-traffic regions. In contrast, Nilgiris reported comparatively lower fatality counts, reflecting regional variations in traffic density and road conditions.

From a quantitative perspective, the accident–death ratio remained almost consistent during 2021–2023. In 2021, the death percentage was **27.62%**, which slightly increased to **27.90%** in 2022 and then marginally decreased to **27.22%** in 2023. The fatality rate stayed around **27–28%**, indicating a relatively stable proportion of deaths compared to total accidents.

Through data science methods such as trend analysis, percentage calculation, and comparative district-level evaluation, the study provides evidence-based insights for policymakers. These findings emphasize the need for predictive analytics, hotspot identification, and data-driven traffic management strategies to reduce accident frequency and fatality rates in the coming years.

REFERENCES

- [1] Tamil Nadu Road Accidents Reports - Dataset - CKAN
- [2] https://www.statista.com/statistics/1084285/india-road-d-accidents-in-tamil-nadu/?srsltid=AfmBOopFkJKoHxlcIl_rQeOVe_IxYP0LnprZnFwt7R2tzfNk7aPrGel
- [3] Shefali Rani, Yogesh Kumar, "A Study and Analysis of Road Accident in Tamilnadu using Data mining Technique," International Journal of Recent Engineering Science (IJRES), vol. 2, no. 4, pp. 6-9, 2015. Crossref, <https://doi.org/10.14445/23497157/IJRES-V2I4P102>.
- [4] https://www.researchgate.net/publication/358666349_Data_Analytic_s_Process_over_Road_Accidents_Data-A_Case_Study_of_Lisbon_City.
- [5] <https://tjier.org/tjier/papers/TIJER2311085.pdf>
- [6] <https://journals.riverpublishers.com/index.php/JRSS/article/view/28829>