IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Patterns And Predictors Of Motorcycle Collisions: A Data-Driven Statistical Approach

[1] Akshita Vijayvergiya

Researcher

Vodafone Intelligent Solutions, Pune, India

Abstract: The problem of motorcycle collisions is one of the leading issues of road safety, but the danger is not constant all the time and all day. vehicle characteristics. This paper conducts statistical research into the data on motorcycle crashes to address three useful questions: (1) at what times of the day collisions were concentrated (and are still concentrated) (peak hours), (2) are collisions. more probable on weekends than on weekdays, and (3) patterns of collision varying between vehicle makes and whether make and day type are independent variables. We start with descriptive profiling of collision frequency by car model and time of collision, and a sharp peak in late-afternoon hours. Next, we use Bayes' Rule to compare the conditional likelihood of observing a crash on weekends versus weekdays. Finally, we test the independence of vehicle make and day type using Pearson's Chi-Square test, finding strong evidence of dependence between make and whether a collision occurs on a weekday or weekend. Together, these results suggest that motorcycle collision occurrence varies meaningfully with temporal context and vehicle-make composition, motivating targeted safety messaging and enforcement strategies that align with peak-hour and weekend-driven risk patterns

Index Terms - Motorcycle collisions; peak hours; weekend effect; Bayes' Rule; chi-square test; independence testing

I. BACKGROUND

Motorcycle collisions are shaped by a mix of behavioural, environmental, and contextual factors. Compared with passenger vehicles, motorcycles provide less physical protection and require greater stability control from the rider, which can amplify the impact of situational risks such as traffic density, visibility, and surface conditions. As a result, collision patterns often show strong variation across time-of-day, day-of-week, and weather—dimensions that can be measured directly from collision records and analysed statistically.

One commonly observed pattern in road-safety analytics is temporal clustering, where incidents concentrate during specific windows of the day. For motorcycles, late-afternoon and early-evening periods can be especially important because they combine higher traffic volume with changing light conditions, commuting behaviour, and fatigue effects. Identifying "peak hours" is valuable operationally: if collisions consistently spike at predictable times, prevention messaging, enforcement, and resource deployment can be aligned with those windows rather than spread evenly across the day.

A second practical question is whether collisions are more likely on weekends than on weekdays. Weekends may involve different trip purposes (recreation vs. commuting), different riding routes, and potentially different rider populations. However, "higher weekend collision likelihood" must be interpreted carefully: collision datasets typically record crashes, not exposure (e.g., total miles ridden). A weekend increase in crash counts could reflect increased riding volume, higher-risk riding contexts, changes in traffic composition, or a combination of these. This study therefore treats weekend–weekday differences as an empirical pattern in observed collisions and discusses exposure as an important limitation and direction for future work.

Finally, motorcycle collisions may differ by vehicle make, either because different make are associated with different rider segments, riding styles, or usage patterns (commuting vs. leisure), or simply due to market share (more common makes appear more often in collision records). From a statistical perspective, an important step is testing whether the distribution of makes is similar across weekdays and weekends. If make and day type are dependent, it suggests that the mix of motorcycles involved in crashes changes meaningfully between weekdays and weekends—an insight that can guide targeted interventions and deeper follow-up analysis[1-3].

To address these questions with transparent, reproducible methods, this paper uses two core statistical tools. First, Bayes' Rule is used to frame weekend and weekday comparisons in terms of conditional likelihoods (weekend vs weekday given a crash). Second, a Pearson Chi-Square test for independence is used to test whether vehicle make and day type are statistically independent. Together with descriptive profiling (counts and proportions), these methods provide a clear, interpretable way to quantify patterns in motorcycle collision occurrence and to distinguish "looks different" from "is statistically different."

II. INTRODUCTION

Motorcycle collisions remain a serious road-safety concern because riders have limited physical protection and often face higher injury severity even in crashes that would be considered moderate for passenger vehicles. At the same time, motorcycle riding is highly context-dependent: when people ride, why they ride, and what conditions they ride in can vary substantially across the week and across seasons. That is, there is hardly ever uniform occurrence of collision. Instead, it is apt to congregate around foreseeable windows--as is the case with commuting time, leisure riding time and conditions. That makes transparent or traction. Understanding these patterns is significant since there are a lot of interventions in traffic safety that are operational by nature. Public enforcers, awareness messaging, and rider education are even greater. effective as they concur with the times and circumstances.
in which risk concentrates [5]. This research presents statistical research on motorcycles. collisions with a calculated intent to interpretability and decision relevance. Instead of collisions being viewed as an abstract prediction task, a series of grounded questions are posed. which can be responded to immediately based on the records of collision and communicated in a manner that is understandable to the stakeholders. First, we examine time of day trends to determine high collision periods and to measure the degree of concentration of collisions in late afternoon. periods as compared to other hours [4].
Second, we examine day type through the comparison of weekends and weekdays, posing the question. as a conditional likelihood problem: conditioned on a collision having occurred, what is the probability that that collision will be on a weekend vs. the possibility of it occurring on a weekday?
Third, we discuss the effects of vehicle makes by determining whether. distribution of motorcycles manufactures observed in crashes. varies systematically during the weekend and the weekdays. The reason behind this is that making effects can be a reflector of a blend. of market share, segmentation of the riders and the context of use; even in absence of complete exposure information, changes in composition between day. types may indicate significant behavioral or contextual. differences that should be examined more closely. The paper is methodologically a combination of descriptive profiling. and two standard statistical devices which are clear and. easy to validate. Weekend is expressed using the Bayes Rule. compared to weekday comparisons in a regular probabilistic
form, to avoid any possible misinterpretation of raw counts as probability out of context. To evaluate whether vehicle make and day are both independent, we employ Pearson Chi-Square test of independence. This test provides a formal how to differentiate patterns which only appear different and not identical? patterns statistically unlikely under an independence. assumption. Besides this, the weather profiles are profiled to characterize the environmental situation under which collisions take place. and as a measure of helping in a wider interpretation of conditions linked to greater perceived occurrence. The output of this work has tripled.
First, it provides a definite risk-profile of motorcycle accidents by discovering intervals of high frequency collisions in the day, which can guide specific intervention like rider alerts, enforcement of commuter-period, and enforcement of safety-timing campaigns [6-8].
Second, it presents the weekend–weekday comparison in a probability framework that is intuitive and defensible, producing a direct statement about weekend likelihood in the collision data while remaining honest about what the data can and cannot imply about underlying exposure. Third, it tests and confirms whether vehicle make and day type are statistically associated, which is useful for practical segmentation and for

designing follow-up analyses such as identifying which contributes most to the observed dependence or whether the relationship holds under alternative groupings and preprocessing choices.
The remainder of the paper proceeds as follows. We next describe the dataset, variables, and preprocessing steps used to construct analysis-ready fields for time-of-day, weekend versus weekday classification, vehicle make standardization, and weather categorization. We then present the statistical methodology and assumptions behind the probability comparison and the independence test. Finally, we report results, discuss interpretation and limitations, particularly the absence of direct riding exposure measures and conclude with implications for safety practice and directions for future work.

III.DATA

The data in SWITRS (California Traffic Collision Data) presented in the form of relational SQLite database is analyzed. The records of SWITRS collisions are stored in a set of connected tables, which is convenient as it disaggregates the context of a collision (what, where, when, under what conditions), and the details concerning the participants and victims (who, what was involved). In this research, we will steal out three tables in the database namely collisions, parties and victims and will limit the scope to specifically identify collisions that are explicitly marked as involving a motorcycle. This is operationally restricted by filtering the collision table on motorcycle collision = 1, and extraction of the records of the respective participants and victims sharing the same case set of case id. Case-id field represents a unique collision identifier that is the primary key in all the pipelines which are used to relate collision context with party-level information like vehicle type and make [9,10].

```
There are 270376 records from collisions.
There are 489862 records from parties.
There are 322694 records from victims.
```

Fig.1 SWITRS data

The filtered tables are exported to CSV after extraction and reloaded to analyze. Columns are read in the form of strings in order to minimize the memory overhead and eliminate accidental parsing errors (such as time fields which appear numeric but have missing values or irregular formatting). The types of data are then processed selectively where they are required to be calculated. This requires the steps of cleaning to be both explicit and traceable, which is important in statistical work where an implicit choice to parsing could change distributions and test outcomes down the line [11-12].
In this paper, the research questions focus on several variables. The time analysis is based on collision date and collision time. Collision dates are put into a date representation, which is in the form of a datetime, which allows the calendar-based feature construction. The collision times are transformed into a time-duration format such that events can be grouped in a clean manner into hourly bins which aids the profiling of the peak hours. To compare weekend and weekday patterns, a derived variable day type is constructed to be directly based off the collision date with records marked as a weekday when the date is Monday to Friday and weekends otherwise. Notably, the day type of this dataset is constructed feature that was created to be analyzed; it is the calendar type of the date of the collision, and not the subjective label in the original data.

vehicle_make	case_id
HONDA	45131
HARLEY-DAVIDSON	43593
YAMAHA	32247
SUZUKI	30275
KAWASAKI	23151
MISCELLANEOUS	10077
OTHER - MOTORCYCLE	8614
HARL	7580
BMW	7013
YAMA	6322
HD	5121
DUCATI (MOTORCYCLE)	3829
HOND	3688
TRIUMPH	3312
KAWA	3054
FORD	2556
TOYOTA	2429
SUZU	2406
OTHER - MOPED	2243
CHEVROLET	1897

Fig 2. Vehicle data

To investigate the effects, we need to have a good representation of the manufacture of the motorcycle. This is obtained in the parties table, which contains party-level vehicle descriptors like statewide vehicle type, vehicle make. Since the parties table contains numerous kinds of vehicles, we further limit party records to those containing entries whose statewide vehicle type is motorcycles (in particular, the categories of "motorcycle or scooter" and moped). The field of manufacturers in the real-world collision data is frequently noisy - abbreviations, variations in spellings, and incomplete strings may result in the same manufacturer being represented in various categories. To solve this, the pipeline standardizes the vehicle through a mapping that brings similar variants into a stable set of canonical makers. Once the data has been standardized, the analysis is performed on the N most common motorcycle made in the subset of collisions (set to 20 in the workflow). It is a pragmatic decision to use this top N restriction: it leads to less sparsity, more stable contingency-table computations, and a more interpretable independence testing with no long tail of ultra-rare categories[14].
The collisions table captures the weather indicators as a way of capturing the environmental context. The data set offers two weather variables, weather, 1 and weather, 2 that depict primary and secondary reported weather. Since the field of categorical contexts frequently contains missingness, a uniform default is employed in the workflow (temporarily replacing missing values with a sentinel like -) so the workflow can be plotted and combined without dropping records silently. It is possible to later convert these placeholders to missing values. In some analyses combinations of conditions are also examined in terms of the two weather fields as an ordered pair to avoid imposing an artificial simplification on a single label[13].
Lastly, to facilitate severity-conscious summaries (particularly to profile the condition), several count variables in the collision table are turned into numeric values, such as killed victims, count of related injury/fatality, and other fields. Ignorance of values in these numeric count fields is represented as zero as indicative of no count is recorded in the extracted subset (though reporting practice and missingness may differ across records). It allows regular aggregation and calculating conditional summaries, e.g. average deaths or injuries by various conditions, without type errors or irregular treatment of null values. Upon completion of the data preparation phase, the dataset that is ready to be analyzed is efficiently structured based on collision identifiers (case_id), collision context (collision date, collision time, derived day type, and weather fields), and party-related vehicle attributes (statewide vehicle type, standardized vehicle make). The core objectives of this paper are supported directly by this structure; to determine peak-hour clustering, test weekend likelihood versus weekday likelihood, and test whether manufacturer composition varies systematically between types of days.

III.METHODOLOGY

In this study, an interpretable, statistics-first workflow is applied to answer three practical questions: (1) at what time of the day motorcycle-involved incidents are clustering, (2) are motorcycle-involved incidents more common on weekends than on weekdays, and (3) is the distribution of the motorcycle makes involved in an incident any different on weekdays and weekends. The model focuses on transparency rather than complex models. Rather than parameterizing a prediction model, both descriptive profiling and two classical statistical tests, the Bayes Rule and Pearson Chi-Square test of independence are used in the analysis to measure patterns as well as to decide whether any observed differences could be explained by chance alone.

Using Baye's Rule repeatidly we have that

$$P(\text{crash}|\text{weekend}) = \frac{P(\text{weekend}, \text{crash})}{P(\text{weekend})}$$

$$P(\text{weekend}, \text{crash}) = P(\text{weekend}|\text{crash}) \cdot P(\text{crash})$$

$$\rightarrow P(\text{crash}|\text{weekend}) = \frac{P(\text{weekend}|\text{crash})}{P(\text{weekend})} \cdot P(\text{crash})$$

Similarly, we can calculate,

$$P(\text{crash}|\text{weekday}) = \frac{P(\text{weekday}|\text{crash})}{P(\text{weekday})} \cdot P(\text{crash})$$

Fig.3 Naïve Bayes Rule

Peak-hour identification is performed through descriptive time-of-day profiling. Collision time is converted into a consistent time representation and then aggregated into hourly bins. The analysis computes collision counts per hour and, where needed, converts counts to proportions so the shape of the distribution can be compared across subsets (for example, weekday versus weekend or across different makes). The main output of this step is an hourly frequency distribution that highlights the hours with the highest collision concentrations, along with a ranked list of the most common collision hours. Because the database does not

include exposure measures (such as total motorcycle traffic volume by hour), these peak-hour results are interpreted as peaks in observed collision occurrence rather than peaks in per-mile or per-trip risk. Even with this limitation, identifying high-frequency windows is operationally useful for timing interventions such as enforcement, messaging, and rider advisories[15-17].

$$P(crash|weekend) = 1.0955299286918958.P(crash)$$
$$P(crash|weekday) = 0.9617880285232417.P(crash)$$

Fig.4 Probability Distribution

Weekend likelihood is quantified using a probability framing guided by Bayes' Rule. Let D denote day type, where D is Weekend or Weekday, and let C denote the event that a record is a motorcycle collision within the filtered dataset. Since the dataset is restricted to motorcycle collisions, the analysis focuses on the empirical distribution of collisions across day types by estimating P(D = Weekend | C) and P(D = Weekday | C). These conditional probabilities are computed directly from counts: the number of motorcycle collision records occurring on weekends divided by the total number of motorcycle collision records, and the analogous fraction for weekdays. Bayes' Rule provides a consistent interpretation framework and helps avoid a common pitfall: interpreting raw weekend counts as "higher weekend risk" without clarifying what the probability is conditioned on and without accounting for missing exposure information. The results are therefore presented as an observed likelihood statement given that a motorcycle collision is recorded in the dataset, what fraction occurs on weekends versus weekdays. Where relevant, the same comparison can be repeated within specific measures to assess whether certain makes show stronger weekend skew than others [19].

Make effects are evaluated by testing whether vehicle make and day type are statistically independent. A contingency table is constructed with standardized motorcycles made as rows and day type (weekday vs weekend) as columns. Each cell contains the number of collisions associated with a given make on a given day type. The null hypothesis assumes independence, meaning the weekday–weekend split should be consistent across makes up to random variation. Expected counts under the null are computed using the product of row totals and column totals divided by the grand total. Pearson's Chi-Square statistics are then computed by summing, over all cells, the squared difference between observed and expected counts divided by the expected counts. A small p-value provides evidence against independence and indicates that the make distribution differs between weekdays and weekends beyond what would be expected by chance[18]

Vehicle Make & Weekend crash independent?

Perform Pearson χ-Square Test for independence on vehicle_make & day_type. Our statistic is,

$$\chi_q = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

where we sum over all rows & columns and
$q = (\#\text{columns} - 1) \cdot (\#\text{rows} - 1)$ is our degrees of freedom.

Our null hypothesis is: vehicle_make & day_type are independent.

Fig.5 Vehicle Make Data

To ensure the Chi-Square test is applied responsibly, the workflow includes safeguards for sparsity. Manufacturer names are standardized to reduce fragmentation caused by inconsistent spelling or abbreviations. The analysis focuses on the top N most frequent makes to reduce the long tail of rare categories that can lead to small, expected counts, unstable test behavior, and poor interpretability. The contingency table is checked for very small, expected counts; if necessary, rare makes can be grouped into an "Other" category or the top N threshold can be adjusted to maintain reasonable expected frequencies. In addition to the p-value, an effect-size measure such as Cramér's V can be reported to communicate the strength of association, since statistical significance can occur even for weak relationships when the sample size is large. When deeper interpretation is needed, standardized residuals can be inspected to identify which contributes most to the overall Chi-Square statistics and whether the dependence is driven by a few highly skewed makes or by a broader shift in composition.

Chi-squared test for independence

Test	Df	Chi-square	P-value
Pearson	19	2989.53	0
Log-likelihood	19	2910.44	0

Fig.6 Chi-squared test

Weather is incorporated as contextual profiling rather than causal inference. Collisions are grouped by primary weather category and, when informative, by combinations of primary and secondary weather fields. For each group, collision counts and proportions are computed to describe which conditions are most frequently observed among motorcycle collisions [17,20,21]. Severity-aware summaries can also be produced by aggregating injury and fatality counts within each weather group. These weather results are interpreted carefully: weather conditions may correlate with riding decisions, seasonal patterns, reporting practices, and traffic dynamics, so the goal is to provide context for temporal and day-type patterns rather than to claim direct causal effects .

P(Crash|Condition)

Let's model the $P(\text{crash}|\text{condition}_i)$.

$$P(\text{crash}|\text{condition}_i) = \frac{P(\text{crash} \cap \text{condition}_i)}{P(\text{condition}_i)}$$

$$P(\text{crash} \cap \text{condition}_i) = P(\text{condition}_i|\text{crash}) \cdot P(\text{crash})$$

$$\Rightarrow P(\text{crash}|\text{condition}_i) = \frac{P(\text{condition}_i|\text{crash})}{P(\text{condition}_i)} \cdot P(\text{crash})$$

We can model $P(\text{condition}_i)$ as a marginal distribution using month:

$$P(\text{condition}_i) = \sum_{j=January}^{December} P(\text{condition}_i \cap \text{month}_j) = \sum_{j=January}^{December} P(\text{condit}$$

where as one might expect $P(\text{month}_j) = \frac{1}{12}$.

Altogether we have that,

$$P(\text{crash}|\text{condition}_i) = \frac{P(\text{condition}_i|\text{crash})}{\sum_{j=January}^{December} P(\text{condition}_i|\text{month}_j) \cdot P(\text{month}_j)}$$

Fig.7 Vehicle Crash Probability

Overall, the methodology quantifies observed patterns in collision occurrence and composition using reproducible, interpretable statistical tools. The analysis explicitly acknowledges a key constraint of collision-record data: it captures incidents but does not directly measure exposure (miles ridden, rider population on the road, or trips taken). Therefore, findings are presented as patterns in recorded collisions and as shifts in distribution across time, day type, and make, with exposure-based risk estimation treated as a priority for future work.

	Condition	P(Crash|Condition)/P(Crash)	P(Condition)
0	clear	1.009620	0.881364
3	other	1.000577	0.001406
1	cloudy	0.942404	0.094724
6	wind	0.935701	0.003346
2	fog	0.892661	0.002988
5	snowing	0.882969	0.000237
4	raining	0.845590	0.015934

Fig.8 Conditional probability data

IV. RESULTS AND DISCUSSION

The motorcycle-collision subset shows a highly concentrated distribution of vehicle makes. After standardizing labels to reduce fragmentation from abbreviations and partial strings, the most frequent make in the cleaned output are HARLEY-DAVIDSON (57,053 collisions), HONDA (48,819), YAMAHA (38,625), SUZUKI (34,494), and KAWASAKI (28,044). A mid-tier group follows with categories such as MISCELLANEOUS (10,077) and OTHER – MOTORCYCLE (8,614), and then a long tail that includes BMW (7,013), TRIUMPH (4,133), and DUCATI (MOTORCYCLE) (3,829).

This concentration is convenient to work with as it encourages the focus on the best set of makes to make the inference and explainable contingency tables. It also explains the importance of standardization: prior to cleaning, several makes seem to be divided between multiple near-duplicates (such as HARLEY-DAVIDSON versus abbreviated forms), which may swell the number of categories and misrepresent both descriptive summaries and statistical testing. The time of collision shows a distinct afternoon peak. When the dataset is aggregated to the hour, the most frequent hour of the dataset is 17:00 (27,142 collisions), then 16:00 (24,563), 15:00 (22,911), and 18:00 (20,419). The curve subsequently decays although it is high in the early afternoon and at 14:00 (18,956) and 13:00 (17,241) it also has significant numbers and declines to further in the late evening and early morning. This trend shows that motorcycle accidents are not uniformly spread throughout the day, but instead, there is a high concentration around a window that corresponds to the activity in late-day traffic. As an operational safety outcome, this is a significant finding since it implies that the highest number of collision incidents is detected within a small daily period, which is a natural focus of time-based interventions and awareness communication.

Collision timing exhibits a clear late-afternoon peak. When collision times are aggregated by hour, the most common time in the dataset is 17:00 (27,142 collisions), followed by 16:00 (24,563), 15:00 (22,911), and 18:00 (20,419). The distribution then tapers but remains elevated through the early afternoon, with 14:00 (18,956) and 13:00 (17,241) still contributing substantial counts, before decreasing further into late evening and early morning. This pattern indicates that motorcycle collisions are not evenly distributed across the day; instead, they cluster strongly around a window that aligns with late-day traffic activity. From an operational safety perspective, this is a meaningful result because it suggests that the largest volume of collision events is observed during a narrow daily interval, making it a natural target for time-specific interventions and awareness messaging.

When the hourly distribution is stratified by day type (weekday versus weekend), the late-afternoon concentration persists in both groups. In one day-type subset, 17:00 represents roughly 11.1% of that subset's collisions, with 16:00 contributing about 9.5% and 15:00 about 8.3%. In the other subset, the largest shares still occur in the afternoon, with 15:00–16:00 each contributing around 8.2%–8.8% and 17:00 around 7.7%.

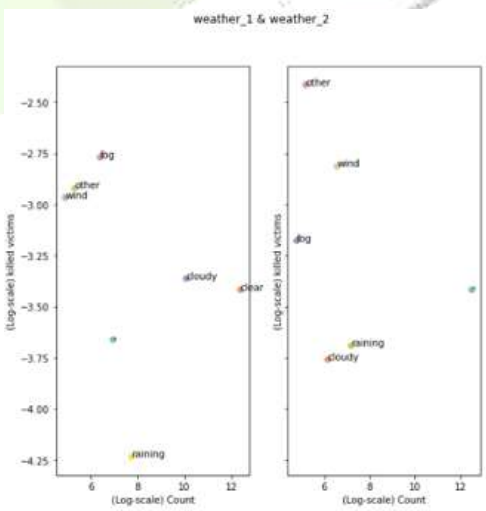

Fig. 8 Collision Time Log

.

Table 4.1 shows that at 5 % level of confidence, the null hypothesis of normality cannot be rejected. KSE-100 index and macroeconomic variables inflation, exchange rate, oil prices and interest rate are normally distributed.

The descriptive statistics from Table 4.1 showed that the values were normally distributed about their mean and variance. This indicated that aggregate stock prices on the KSE and the macroeconomic factors, inflation rate, oil prices, exchange rate, and interest rate are all not too many sensitive to periodic changes and

speculation. To interpret, this study found that an individual investor could not earn higher rate of profit from the KSE. Additionally, individual investors and corporations could not earn higher profits and interest rates from the economy and foreign companies could not earn considerably higher returns in terms of exchange rate. The investor could only earn a normal profit from KSE.

What we learn here is that the timing peak cannot be dominated by any single component of the week, but it seems to be a consistent phenomenon of collision occurrence between the types of day, with late afternoon habitually taking the position of the highest frequency. The likelihood of weekends is assessed based on a Bayes-rule framing to compare the distribution of collision observations across weekend and weekdays. In this framework, it is found that the crashing likelihood is greater on the weekends than on the weekdays, given the correlation between the percentage of collisions on weekends and the percentage of weekend days as the baseline. This finding must be viewed with caution: collision data represents the incidence of an event, not its exposure (the amount of people who ride in a particular setting) There is also the problem of reporting levels on conditions, and the interpretation of make categories as a combination of market share and riding behavior instead of strictly mechanical risk.

the observed collision data and the calendar structure used in the analysis, rather than a definitive per-mile or per-trip risk statement. Even with that caveat, the finding is valuable because it flags weekends as a high-priority context for rider-focused safety communication and follow-up analysis, especially when paired with the observed afternoon peak.

Vehicle make and day type are then tested for statistical independence using Pearson's Chi-Square test on the top 20 motorcycle make. The top 20 make collectively account for 340,991 collisions in the dataset subset used for this test. The Pearson Chi-Square statistic is 2989.53 with 19 degrees of freedom, and the reported p-value is effectively 0 (i.e., extremely small). This provides strong evidence against the null hypothesis of independence, meaning the distribution of makes involved in collisions differs between weekdays and weekends more than would be expected from random variation alone.

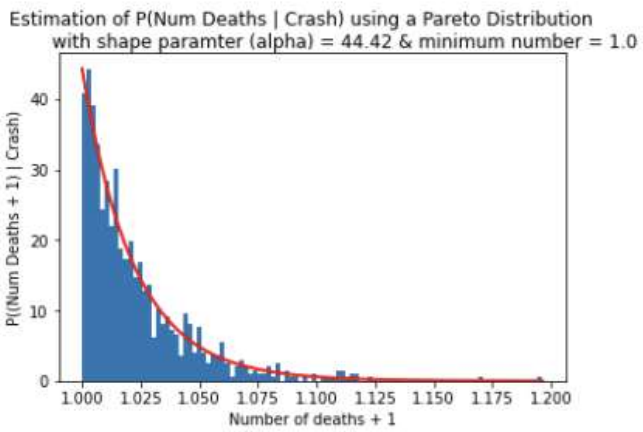

Fig.9 Pareto Distribution

Weather profiling provides additional context on the conditions under which motorcycle collisions are recorded. The most frequent weather combination by a wide margin is clear weather with no secondary condition recorded, with 240,954 collisions. Cloudy conditions (with no secondary condition recorded) account for 22,279 collisions and raining conditions (with no secondary condition recorded) account for 2,297 collisions.

	Condition	P((Crash & Fatality)|Condition)/P(Crash & Fatality)	P(Condition)
3	other	2.118012	0.001406
5	wind	1.678019	0.003346
2	fog	1.610284	0.002988
0	clear	1.004220	0.881364
1	cloudy	0.991830	0.094724
4	raining	0.474565	0.015934

Fig.10 Resulted Crash & Fatality data

Other combinations such as cloudy with rain (1,293) and fog (596) appear far less frequently. This distribution shows that most recorded collisions occur under "normal" or non-extreme reported weather, which is consistent with exposure-driven interpretation: riders are more likely to be on the road in clear conditions, and

clear conditions dominate total collision counts. The weather scatter summaries in the notebook further indicate that less frequent weather categories can show different severity patterns, but those signals require caution because rare categories are more sensitive to small-number effects and missingness in condition reporting.

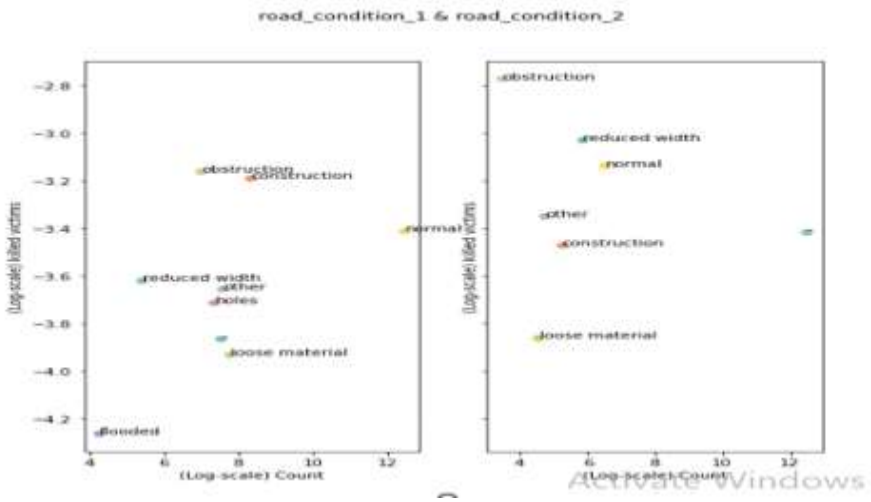

Fig.11Central Empirical Pattern

in late afternoon, weekend likelihood is higher than weekday likelihood under the Bayes-rule comparison used, and vehicle make composition differs significantly between weekday and weekend collisions.

Practically, this implies that the "mix" of motorcycles observed in collisions shifts across day types, which is consistent with the idea that weekday riding and weekend riding may reflect different rider segments, trip purposes, and motorcycle usage patterns. This dependence result justifies deeper follow-up work (for example, residual analysis) to identify which makes are most overrepresented on weekends versus weekdays and to test whether the relationship remains after alternative grouping or additional covariates are considered

REFERENCES

[1] Morinaga, T., Kanda, S. and Nomi, R. 1986. Lipase production of a new thermophilic fungus, *Humicola lanuginosa* var. *catenulata*. J. Ferm. Technol. 64: 451–453.

[2] Limkar, S., Kulkarni, S., Chinchmalatpure, P., Sharma, D., Desai, M., Angadi, S. and Jadhav, P. 2020. Classification and prediction of rice crop diseases using CNN and PNN. In Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Vol. 2. Springer, Singapore, pp. 31–40.

[3] Nagao, T., Shimada, A., Sugihara, A. and Tominaga, Y. 1998. C-terminal peptide of *Fusarium heterosporum* lipase is necessary for its increasing thermostability. J. Biochem. 124: 1124–1129.

[4] Vayadande, K., Bhosle, A.A., Pawar, R.G., Joshi, D.J., Bailke, P.A. and Lohade, O. 2024. Innovative approaches for skin disease identification in machine learning: A comprehensive study. Oral Oncology Reports 10: 100365.

[5] Kale, D.R., Mane, T.S., Buchade, A., Patel, P.B., Wadhwa, L.K. and Pawar, R.G. 2024. Federated learning for privacy-preserving data mining. In Proc. International Conference on Intelligent Systems and Advanced Applications (ICISAA 2024). IEEE, pp. 1–6.

[6] Limkar, S., Hunashimarad, S., Chinchmalatpure, P., Baj, A. and Patil, R. 2020. Potential of robust face recognition from real-time CCTV video stream for biometric attendance using convolutional neural network. In Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Vol. 2. Springer, Singapore, pp. 11–20.

[7] Nakashima, T., Fukuda, H., Kyotani, S. and Morikawa, H. 1988. Culture conditions for intracellular lipase production by *Rhizopus chinensis* and its immobilization within biomass support particles. J. Ferm. Technol. 66: 441–448.

[8] Namboodiri, V.M.H. and Chattopadhyaya, R. 2000. Purification and characterization of a novel thermostable lipase from *Aspergillus niger*. Lipids (Champaign) 35(5): 495–502.

[9] Nishio, T., Takahide, C. and Minoru, K. 1987. Purification and some properties of lipase produced by *Pseudomonas fragi* 22.39 B. Agric. Biol. Chem. 51: 181–187.

[10] Obradors, N. and Montesinos, J.L. 1993. Effects of different fatty acids in lipase production by *Candida rugosa*. Biotechnol. Lett. 15(4): 357–360.

[11] Omar, I.C., Hayashi, M. and Nagai, S. 1987. Purification and some properties of a thermostable lipase from *Humicola lanuginosa* No. 3. Agric. Biol. Chem. 51: 37–45.

[12] Palekar, A.A., Vasudevan, P.T. and Yan, S. 2000. Purification of lipase: A review. Biocatal. Biotransform. 18: 177–200.

[13] Razak, C.N.A., Salleh, A.B., Musani, R., Samad, M.Y. and Basri, M. 1997. Some characteristics of lipases from thermophilic fungi isolated from palm oil mill effluent. J. Mol. Catal. B: Enzymatic 3: 153–159.

[14] Saxena, R.K., Davidson, W.S., Sheoran, A. and Giri, B. 2003. Purification and characterization of an alkaline thermostable lipase from *Aspergillus carneus*. Process Biochem. 39: 239–247.

[15] Shimada, Y., Sugihara, A., Maruyama, K., Nagao, T., Nakayama, S., Nakano, H. and Tominaga, Y. 1996. Production of structured lipid containing docosahexaenoic and caprylic acids using immobilized *Rhizopus delemar* lipase. J. Ferm. Bioeng. 81: 299–303.

[16] Taipa, M.A., Aires-Barros, M.R. and Cabral, J.M.S. 1992. Purification of lipases. J. Biotechnol. 26: 111–142.

[17] Pawar, R.G., Chinchmalatpure, P., Chinchmalatpure, S., Pujeri, R.V., Wategaonkar, D., Jadhav, N., Kapare, S. and Naik, D.N. 2025. Social threat identification on X via sentiment analysis: An SVM and LSTM-based approach. Int. J. Appl. Math. 38(3s): 565–581.

[18] Dixit, B., Pawar, G., Gayakwad, R., Joshi, M., Mahajan, R. and Chinchmalatpure, S.V. 2023. Challenges and a novel approach for image captioning using neural network and searching techniques. Int. J. Intell. Syst. Appl. Eng. 11(3): 712–720.

[19] Toida, J., Kondou, K., Fukuzawa, M., Ohnishi, K. and Sekiguchi, J. 1995. Purification and characterization of a lipase from *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. 59: 1199–1203.

[20] Limkar, S., Baser, S., Jhamnani, Y., Shinde, P., Jithu, R. and Chinchmalatpure, P. 2019. Chatbot: Integration of applications within Slack channel. In Proc. International Conference on Intelligent Computing and Communication. Springer, Singapore, pp. 551–558.

[21] Valero, F., Ayats, F., Lopez-Santin, J. and Poch, M. 1988. Lipase production by *Candida rugosa*: Fermentation behaviour. Biotechnol. Lett. 10: 741–744.