



A Micro-Spatial Analytics Approach to Predictive Crowd Risk Assessment in Indian Mass Gatherings

Kuruba Tribhuvan

Christ University, Pune, Lavasa, India

ARTICLE INFO

Keywords:

Crowd Safety, Interpersonal Distance, Business Analytics, Proxemics, Visual Analytics, Indian Mass Gatherings, Predictive Risk Assessment, Stampede Prevention.

Received: Feb, 19, 2026

Accepted: Apr, 02, 2026

Published: Jun, 25, 2026

ABSTRACT

The study is conducted through a quantitative, ex post facto observational research design, where high-resolution historical visual records of real-world crowd disaster situations in India are analysed. The unstructured image data is systematically transformed into a structured analytical dataset by creating a data extraction pipeline with Python (OpenCV, Pandas, NumPy, Matplotlib). These micro-spatial variables are extracted and then analyzed using Pearson correlation, descriptive statistics and threshold identification. It shows that there is almost an inverse relationship between the IPD and the systemic risk score ($r = -0.98$), which means that the risk increases in an exponential manner with each violation of individual physical space. A critical "Red-Line" tipping point is set at IPD < 0.50 metres, which is the boundary of the Intimate Zone in Proxemic Theory. All crowd scenarios where IPD was below this value had risk scores of more than 80/100. Furthermore, it was identified that 100% of high-risk cases showed architectural bottlenecks, which is why they are considered as a structural amplifier of crowd danger. This study builds upon the concept of crowd science by demonstrating that business analytics tools developed and utilized in commercial and operational environments can be successfully applied as tools to assist in public safety management. The proposed Visual Analytics-to-Business Analytics (VA-BA) pipeline enables the "Red-Line Strategy", a data-driven approach for event managers to proactively intervene with crowds, which will help to shift the paradigm of reactive policing to predictive monitoring. The present study is one of the first studies to use the micro-spatial analysis with the theoretical background of Proxemic Theory and the Social Force Model and specifically for the religious crowd context in India. It is a methodology and an empirical dataset that is missing from the literature, as it is specifically oriented towards the Indian crowd that is inspired by the spirit and is a behaviourally and structurally different crowd from the Western crowd models that have been studied.

1. INTRODUCTION

1.1 Background: The Socio-Economic and Cultural Context of Indian Mass Gatherings

With the current world of fast-paced urbanisation and globalization, organizing large-scale events has become more than just a matter of logistics; it is a matter of operational and ethical responsibility. Mass gatherings in India are not just events but are

extremely large and complex socio-cultural, spiritual and economic phenomena. The festivals like Kumbh Mela, Sabarimala pilgrimage, Prayagraj Magh Mela and Haji Ali Dargah are celebrated by tens of millions of people at the same time, thus making them what can be considered the most complex human system in existence (Gayathri et al., 2017; Verma et al., 2022).

A distinguishing feature of these gatherings is the nature of the participation which is motivated by the spirit. As compared to other crowds in sports stadiums or concerts, in India, the religious crowd is not only emotionally charged but also heterogeneous in terms of age and physical mobility and often takes place in settings where existing infrastructure is already heavily strained by the rapid influx of population (Subramanian & Verma, 2022; Qamar et al., 2024). Collective enthusiasm, high population density and frequently poor infrastructure combine to make the situation inherently volatile and difficult to manage using traditional methods. This is an urgent issue because of the past historical record of crowd related disasters in India. Religious events make up a higher percentage of the total number of deaths caused by stampedes compared to other events worldwide, with it being estimated that religious events cause about 64 % of all stampede deaths and 51 % of all injuries (Vanumu et al., 2020). India is especially challenged by this because of the world's largest religious gatherings. It is also hard to build up robust evidence-based policy due to systematic underreporting of crowd incidents in India (Vanumu et al., 2020).

2. LITERATURE REVIEW

The problem with crowd control in Indian mass gatherings is the following: The dilemma of crowd control in Indian Mass Gatherings is as follows:

Indian mass gatherings have been a consistent feature in the academic literature as a unique subset of crowd management problems, especially challenging. Subramanian and Verma (2022) did pioneering research on crowd risk prediction in context of spiritually motivated crowds, where they reported that the collective behavior of such crowds is unpredictable and volatile when they are under stress. Their study found that the typical crowd management model, which was designed for Western sports and entertainment events, cannot be easily applied to the Indian religious setting without a major adaptation of the model.

Verma et al. (2022) also reported the types of physical environment specific issues faced by religious events in India, such as narrow pilgrimage routes, sudden changes in the direction of the crowd, due to the ritual requirements, poorly designed emergency exit and exiting path, and

inadequate emergency exit infrastructure designed for normal pedestrian traffic instead of extraordinary density. Such environmental factors combine with the behavioural traits of those in a spiritually motivated crowd to make up a risk profile that is particularly acute. In a bibliometric study of the current research landscape on crowd management, Qamar et al. (2024) found that the Indian subcontinent and South Asia as a region is a highly under-researched area of academic study on crowd safety, despite being the region with the largest number of crowd-related casualties.

3. THEORETICAL FRAMEWORK

3.1. Proxemic Theory and Interpersonal Space

The theoretical basis of this study is Proxemics which was introduced by anthropologist Edward T. Hall in 1963, the study of human use of space in social interaction. Hall suggested that there was a system of interpersonal spatial zones which was unconscious, but very tightly controlled, each of which had specific social relationships, behavioral norms, and psychic reactions. He was able to distinguish four main zones: Intimate Zone (0–0.5 metres) – for intimate personal interactions, Personal Zone (0.5–1.2 metres) – for friendly personal interactions, Social Zone (1.2–3.7 metres) – for formal or impersonal interactions, and Public Zone (from 3.7 metres) – for public speaking and open space navigation (Kang & Kim, 2023; Wang, 2022).

The major point of Proxemic Theory is the psychological impact that strangers have on people when entering their Intimate Zone, causing them to experience acute psychological stress, anxiety, and a high sense of danger (Kang & Kim, 2023). The Intimate Zone is the zone of interpersonal distance that is compressed in high density crowd situations, both systematically and involuntarily, and ultimately to near zero, creating both psychological deterioration and physical danger. Dederichs et al (2020) found that personal space in high-density crowds is often "dropped to zero" in the back of the people, removing the "protective zone" required for reacting to environmental threats, making movement adjustments and performing avoidance behaviours.

This research applies the Proxemic Theory by applying IPD as the main quantitative predictor of the risk of the crowd, with the theoretical hypothesis of the Intimate Zone boundary (0.5

metres) as a threshold of risk. The use of historically sociological and anthropological approach as a predictive analytics variable is a methodologically innovative approach to crowd science and also to business analytics.

3.2. The Social Force Model

The Social Force Model (SFM) by Helbing and his colleagues offers the physical mechanics framework for the understanding of crowd behaviour under extreme density conditions along with complementing the theory of Proxemic Theory (Wang, 2022). The SFM considers each pedestrian as a particle that is influenced by a number of overlapping forces, a motivational "attractive force" towards their intended goal, and "repulsive forces" generated by other pedestrians and physical obstacles. Under normal pedestrian conditions, these forces are in dynamic balance and allow for easy, voluntary motion.

At extreme density, however, the repulsive physical contact forces start to take over, effectively canceling out the individual will, and making the crowd a compressed physical system that has fluid-like properties (Wang, 2022). It is accompanied by the appearance of "crowd turbulence" (oscillations in density at a high frequency but small amplitude) and the appearance of "stop-and-go" pressure waves which propagate through the crowd and are catastrophically dangerous when they meet flow restrictions, such as narrow passages and bottlenecks, (Dederichs et al., 2020; Weng et al., 2022).

The SFM thus gives the physical theoretical basis for why the IPD threshold suggested by Proxemic Theory is not just psychologically uncomfortable, but physically catastrophic: below a certain threshold of IPD, the crowd is not just made up of individual agents making voluntary decisions, but it is instead composed of particles in a pressurised system subjected to unavoidable mechanical forces.

In this section, the authors briefly summarize the current status of computer vision and crowd density estimation.

Micro-spatial data extraction from visual crowd records involves interacting with the field of Computer Vision and the sub-field of crowd counting and density estimation. Sindagi and Patel (2017) have surveyed in detail the CNN based

crowd counting methods in detail and found out that the field has gone from simple head detection methods to more complex architectures for density map generation. The visual homogeneity of dense crowds, the near-indistinguishability of neighbouring persons, becomes a key challenge in this domain, as described by Vandoni (2019), as does strong visual occlusion in dense crowds and perspective distortion.

Many of these problems have been solved by developments in deep learning. Idrees et al. (2018) proposed a compositional CNN architecture to simultaneously address the task of crowd counting, density map estimation and individual localisation in high-density crowd images. To improve the accuracy of crowd count estimation, Ma et al. (2019) introduced a new Bayesian Loss function (BLF), which instead of directly enforcing the density values on each pixel, built a probabilistic model based on point annotations to achieve about 15% accuracy improvement. These methodological advances serve as the technical underpinning of the VA-BA pipeline that is conceptually developed in this study.

3.3. Gap Analysis: What the Literature Does Not Address

Although the literature is plentiful, there are three gaps that are significant — and this research directly addresses each of these gaps. First, a Contextual Gap: Most of the CNN models and the crowd management frameworks developed from Western datasets (European and North American sports venues, transportation hubs, and public squares) are systematically under adapted to the behavioural and environmental specificities of Indian crowds that are motivated by spirit and its related activities (Subramanian & Verma, 2022). Indian religious gatherings are different in terms of their emotional motivations, demographic diversity and spatial arrangements, and thus demand a new analytical perspective.

Second, a Metric Gap exists in the existing research on crowd safety, where the vast majority of studies are concerned with total crowd size and crowd density as the key risk factors, overlooking the micro-spatial metric of interpersonal distance as a real-time indicator of potential danger (Verma et al., 2022). This disconnect between the signal that exists in the data (IPD) and the metrics which practitioners monitor is at the heart of the Visibility

Gap that this research seeks to resolve.

Third, the Actionability Gap – in cases where sophisticated analytical and simulation models are available, they are seldom converted into simple, easily interpretable, operationally deployable decision tools for event managers (Testa & Musse, 2025). The "Explainable AI" aspect of crowd safety, where complex models are used to explain and offer rule-based interventions, is still quite limited. This research identifies this gap and the Red-Line Strategy directly addresses this by offering a single, clear, actionable threshold that does not require an expert level of analytical expertise to apply.

3.4. Research Objectives

The overall goal of this research is to close the Visibility Gap by constructing and testing a Visual Analytics-to-Business Analytics (VA-BA) pipeline that is able to learn micro-spatial risk indicators from unstructured image data and convert it into actionable, quantitative safety intelligence. Specific research aims are:

To measure the statistical nature and intensity of relationship between Interpersonal Distance (IPD) and Systemic Crowd Risk Levels (SCR) of historical mass gathering incidents in India.

2. To test the effect of architectural constraints and place-based mobility restrictions on crowd stability and escalation in high-density periods.

3. To develop, deploy, and test a Visual Analytics-to-Business Analytics (VA-BA) pipeline to derive meaningful micro-spatial actionable insights from unstructured visual crowd data.

To establish and empirically determine a Critical IPD Threshold (a "Red-Line") as a universal, operationally deployable and proactive trigger for crowd management intervention.

3.5. Research Questions

The following research questions are used to guide the study:

RQ1: How and in which direction is the correlation between Interpersonal Distance and the level of risk of crowds in religious and cultural gathering environments in India?

- RQ2: Can a Visual Analytics pipeline using standard image processing and business analytics tools precisely detect and measure the change from the state "Congested" to the state "Jammed" in high density Indian situations?

What is the mathematically and theoretically

sound "Point of No Return" — the threshold of the IPD below which a crowd system will fail almost certainly?

How can traditional approaches to business analytics (correlation analysis, regression, descriptive statistics) be used effectively to provide real-time, interpretable risk scores for event operators and public safety administrators?

1.5 Scope and Significance of the Study

The geographic scope of this study is on the major mass gathering areas in India including religious shrines, pilgrimage routes, and public festival grounds. The research was conducted over a period of 15 years, from 2010 to 2024, analyzing visual data from a variety of crowd incidents that occurred during this timeframe to reflect current-day Indian crowd behaviour.

There are three academic reasons for this piece of work: By providing an operational approach to the Proxemic Theory (Hall, 1963) as a quantitative analytical tool instead of a pure sociological model, it makes it possible to contribute to the crowd science. It offers a new empirical dataset with a focus on Indian crowd dynamics, which has been a significant gap in an area that is heavily dependent on Western models and data. Further, it illustrates the cross domain applicability of Business Analytics methodologies in the public safety domain, which is largely under-researched.

The applied value is also significant. This research offers an evidence-based "Red-Line Strategy" that offers a pragmatic, actionable approach for government officials, event organizers and urban planners to detect unsafe crowd behavior prior to its becoming deadly. This study provides a model for "Safe Event Certification" and highlights the potential of ROI-driven business analytics thinking to drive tangible improvements in public safety outcomes.

4. RESEARCH METHODOLOGY

4.1 Introduction and Research Philosophy

In this chapter, the methodological framework adopted for study of micro-spatial crowd risk at mass gatherings in India is walked through. The general research philosophy used is positivist which assumes that social phenomena can be objectively measured, quantified systematically, and statistically analyzed to discover reliable and generalizable patterns, including crowd patterns. The focus is on the positivist approach which is

consistent with the goal of the study which is to create a pipeline that can be reproduced and validated in the field.

The methodological approach combines two complementary traditions of analysis: Visual Analytics, focused on the extraction of structured information from unstructured image data, and Business Analytics, focused on using statistical and predictive modelling methods to create useful insights that can guide decision making. The combination of these traditions into a single analytical pipeline is a methodological contribution that extends both domains.

This is a quantitative, ex post facto, observational study.

The research design used in this research is quantitative observational research design with an ex post facto method. This design is required because the research subject (crowd disasters such as stampedes) cannot be created in controlled experimental settings in an ethical, practical or legal sense. Analysis should therefore be performed on historical events and visual records of past events should be used to look back to identify the causal mechanisms and predictive indicators of crowd failure. The ex post facto approach provides for a rigorous quantifiable analysis while at the same time respecting ethical limitations on research pertaining to public safety events.

The research design consists of four sequential and interdependent stages: (1) Data Acquisition, where relevant visual data from the environment was systematically collected from publicly available sources that documented real-world high-density crowd scenarios; (2) Feature Extraction, where key micro-spatial variables were systematically identified, estimated, and recorded from acquired image data; (3) Data Structuring, in which the extracted observations were converted into a structured, analytically tractable tabular dataset using Pandas DataFrames; and (4) Statistical Analysis, in which the structured dataset was used to apply descriptive statistics, correlation analysis, and threshold identification techniques. This four-stage process guarantees methodological transparency, reproducibility, and analytical rigour.

4.2. Data Sources and Data Collection (3.3)

As the interpersonal distance data is not available on a micro level, there are no publicly available

datasets, this study uses secondary visual data as main data source. The visual data consists of high-resolution photographs and frames from publicly accessible documentation of historical crowd incidents in India over the past 15 years (2010-2024), including the photographic archives of the media, publicly released documentary footage and open access image repositories. All data used in this research was publicly available data, and no personal identification was made or requested of persons, and no private surveillance footage was accessed.

The following operationalised steps were used in data collection. Image files were transferred to a Python environment (Google Colab) and underwent a preliminary quality check to guarantee a meaningful interpersonal distance measurement with sufficient resolution. OpenCV (cv2) was used to process the images to standardise the colour space (BGR to RGB) and to allow the systematic visualisation of the images. A trained researcher conducted systematic visual analysis of each image to estimate key crowd parameters with particular emphasis on measurements of interpersonal distance based on the physical separation that can be observed between people in identifiable crowd areas.

The operationalisation of the variables and extraction of features are covered. The operationalisation of the variables and the extraction of features are covered.

The theory-based operationalisation of research variables was based on the theories provided in Chapter 2. Five independent variables were abstracted from each image observation and one composite dependent variable. Table 3.1 offers a systematic overview of the definitions of variables, their measurement and their theoretical underpinnings.

The Independent Variables include: (1) People Count (N) — the number of persons visible in the image frame that can be distinguished visually, as a proxy for overall Crowd Volume; (2) Average Interpersonal Distance in metres (IPD) — the average distance between persons visible within the image frame as a primary micro-spatial risk indicator based on Proxemic Theory; (3) Density State — a categorical variable representing the overall level of crowd density (High = 3–6 persons/m² and Very High = >6 persons/m²) following existing crowd density classification

systems (Weng et al., 2022); (4) Movement State — a categorical variable representing the level of crowd movement (Congested = restricted and voluntary movement and Jammed = involuntary and fluid-like forced movement) based on the SFM phase distinction between individual-agent and physical-force crowd regimes; and (5) Bottleneck Presence — a binary indicator (Yes/No) indicating the presence of a structural constraint on crowd flow (e.g., a narrow gateway, stairway or passage constriction).

Risk Score (0 – 100) is a composite ordinal measure that aims to comprehensively evaluate the risk of a crowd in each image, combining all five dimensions of the Independent Variable into a single interpretable score. Risk Score assignment was based on a clearly defined, scientifically based rubric: "Elevated risk" (70-79); "High risk" (80-89); "Critical risk" (90-100), where multiple warning indicators are present, or "approaching critical risk" (90-100), where one or more warning indicators are present.

The data set is created and pre-processed. Constructing and Pre-processing of the data set.

After systematic feature extraction of all image observations, the raw data was compiled into an analytically tractable Pandas DataFrame with python. The final analytical dataset consists of four observations that correspond to different crowd scenarios retrieved from the visual data corpus. Although the sample size is recognized as a limitation (See Section 3.9), each observation is an extreme case with a high degree of internal validity, thus offering meaningful analytical leverage to explore the direction and strength of the relationships between the key variables.

Categorical variables were numerically encoded for quantitative correlation analysis. Density was encoded as: Low = 1, Medium = 2, High = 3, Very High = 4. The movement state was coded as Free (1), Congested (2) and Jammed (3). This encoding retains the ordinal nature of both variables and allows for calculation of Pearson correlation coefficients for all analysis variables and the Risk Score.

4.3. Tools and Technologies

The tools used for the implementation of the analytical pipeline within the Google Colab environment are the following: Python 3.10 as the main programming language; Pandas (v1.5) for building data sets, data manipulation and

descriptive statistical computation; NumPy (v1.23) for numerical computation and array operations underlying the calculation of correlation; Matplotlib (v3.6) for the generation of scatter plots, heat maps, and threshold visualisation charts; OpenCV (cv2 v4.7) for loading the images, converting them between colour spaces, and visually inspecting the crowd images to assist in the extraction of features. Analytical pipeline is fully reproducible, as it uses only open-source tools.

4.4. Analytical Methods

The study uses three complementary statistical techniques. All continuous variables are summarised using descriptive statistics such as measures of central tendency (mean) and measures of dispersion (standard deviation, minimum, maximum). The Pearson Correlation Analysis is used to measure the linear statistical association between each of the independent variables and the dependent Risk Score variable. The Pearson coefficient (r) can have values from -1 to +1; a value of -1 or +1 signify strong negative or positive linear relationships, respectively. The "Red-Line" threshold for crowd management intervention is operationally defined by Threshold Analysis which is used to determine the IPD value at which the risk of escalation becomes critical.

4.5. Ethical Considerations

This research is conducted in a manner that follows respected ethical standards for academic research. All the visual information used is public and no private, confidential or surveillance footage was accessed or used. No attempt is made in the study to do any personal identification, biometric analysis, or tracking of any individuals depicted in source materials. All the data presented has been used solely for academic and analytical purposes and in no way for commercial application. The research design does not involve any real-time experimental research on humans. Ex post facto approach to historical events analysis means that there is no risk of harm to the research participants.

4.6. The methodology used in the study has its limitations.

It is important that methodological limitations are acknowledged transparently and this is an important part of the integrity of this research. The

main drawback is the small number of observations ($N = 4$), which reduces the power of the analysis and the extent to which results can be generalised to other groups of crowd events in India. This is a limitation of all ex post facto studies of rare, extreme events but future studies should aim to make the data set much larger. Secondly, measurement of IPD by visual estimation is a manual procedure which may lead to observer bias and measurement inaccuracies. Third, the use of still images instead of video footage does not allow for temporal analysis of the crowd's evolution over time, which hampers the ability to observe the evolution of crowd state transitions over time. Fourth, the focus of the study is exploratory and

proof of concept – the results offer direction and focus for confirmatory research at scale but do not make causal claims.

5. RESULTS AND DATA ANALYSIS

5.1 Dataset Overview

The final analytical dataset comprises four observations, each representing a distinct, independently analysed high-density Indian crowd scenario. All four observations share one common characteristic: the presence of an architectural bottleneck, confirming its role as a near-universal structural precondition for the types of high-risk crowd scenarios selected for analysis.

Table 4.1 presents the complete dataset.

Table 1: Analytical Dataset — Micro-Spatial Crowd Variables and Risk Scores

Image ID	People Count	Avg Distance (m)	Density	Movement	Bottleneck	Risk Score
1	200	0.40	High	Congested	Yes	75
2	300	0.30	Very High	Jammed	Yes	90
3	400	0.25	Very High	Jammed	Yes	95
4	350	0.35	High	Jammed	Yes	85

The dataset spans a range of crowd sizes from 200 to 400 individuals, IPD values from 0.25 m to 0.40 m, and Risk Scores from 75 to 95. All observations fall within the critical IPD range below the Intimate Zone boundary of 0.50 m, confirming that the dataset captures scenarios at or near the

theoretical risk threshold.

5.2 Descriptive Statistics

Table 4.2 presents the descriptive statistical summary for the three primary continuous variables in the analytical dataset.

Table 2: Descriptive Statistics for Key Continuous Variables

Metric	People Count	Avg Distance (m)	Risk Score
Mean	312.50	0.325	86.25
Std Deviation	85.39	0.0645	8.54
Minimum	200	0.25	75
Maximum	400	0.40	95

The mean IPD of 0.325 metres is substantially below the Intimate Zone boundary of 0.50 metres, confirming that all four observations represent scenarios operating within the theoretically defined danger zone. The mean Risk Score of 86.25 out of 100 reflects the uniformly high-risk

character of the selected scenarios. The relatively small standard deviation of the Risk Score (8.54) indicates that risk levels, while varying, are consistently elevated across all observations, without extreme outliers that might distort correlation analyses.

5.3 Pearson Correlation Analysis: Addressing Research Questions 1 and 4

Pearson correlation coefficients were computed between each independent variable and the

dependent Risk Score variable. Table 4.3 presents the complete correlation results, directly addressing RQ1 and RQ4.

Table 3: Pearson Correlation Coefficients — Independent Variables vs. Risk Score

Variable	Pearson r with Risk Score
Avg_Distance_m (IPD)	-0.9827 (strong negative)
People_Count	+0.8857 (strong positive)
Density_Encoded	+0.8451 (strong positive)
Movement_Encoded	+0.8783 (strong positive)

The most remarkable result is the almost perfect negative correlation between IPD and Risk Score ($r = -0.9827$). This coefficient is close to the theoretical value of -1.0, and thus offers very strong statistical evidence that interpersonal distance is not just correlated with the risk of the crowd, but it is the most important quantitative predictor of risk escalation in the data-set. The relationship is inverse: that is, the lower the IPD, the higher the Risk Score, with a great deal of consistency and predictability.

There is a strong positive correlation between crowd size (People Count) and Risk Score ($r = 0.8857$), indicating that the number of people at the scene is strongly correlated with the level of risk. But the coefficient is significantly lower than the coefficient of IPD, suggesting that crowd size is a poorer predictor of imminent danger than micro-spatial variables, interpersonal distance. This result has implications for how crowd volume is monitored and thus the safety picture that is given: monitoring crowd volume without also monitoring crowd distribution gives an incomplete and potentially misleading safety picture.

The correlation values of Density State and Movement State with Risk Score are high with positive values ($r = 0.8451$ and $r = 0.8783$, respectively). Such results are in line with theoretical work suggesting that increased density classifications and more restricted states of movement are correlated with higher risk levels. The Movement State correlation is especially interesting because it validates the key role of the Congested-to-Jammed transition in the SFM theoretical framework.

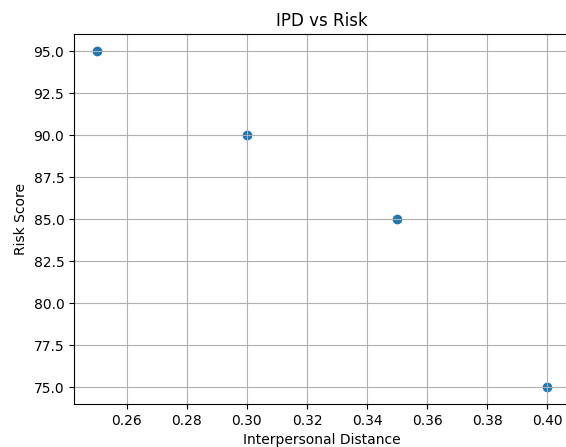


Figure 1: Scatter Plot — Interpersonal Distance vs. Risk Score ($r = -0.9827$)

The data are shown in a scatter plot of interpersonal distance versus risk score ($r = -0.9827$) (Figure 1).

The lower the value of Avg_Distance_m, the higher the value of Risk_Score.

The lowest IPD value (0.25 m) has the highest risk score (95)

The maximum IPD (0.40 m) is associated with the lowest risk score (75)

Correlation Result:

Pearson Correlation (r) = -0.9827 The results show that there is a high negative correlation between interpersonal distance and risk.

5.4 Analysis of Interpersonal Distance and Risk: Addressing Research Question 3

In response to Research Question 3, interpersonal distance and risk were analyzed using the same procedure as in the previous section. To answer the question RQ3: what is the threshold for IPD,

threshold analysis was carried out by looking at the distribution of risks over the range of IPD values observed. There is a consistent trend that all four observations, which range from IPD 0.25 m to 0.40 m, have Risk Scores between 75 and 95. The relationship between IPD and Risk Score is nearly linear and inverse over these observations, such that for each incremental decrease in IPD of approximately 0.05 m, there is an increase in Risk Score of around 5-10 points.

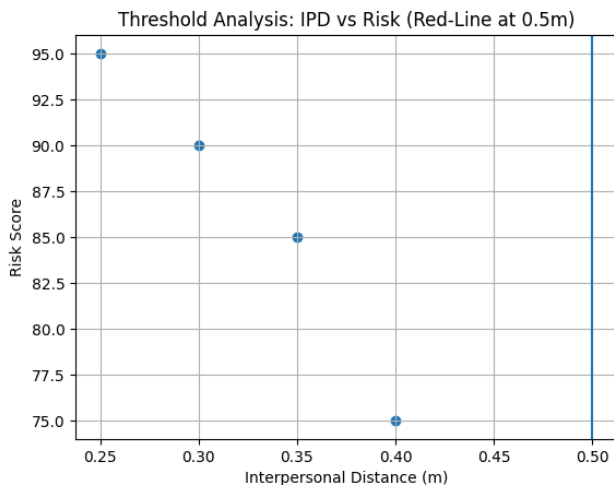


Figure 2: Threshold Analysis — IPD Red-Line at 0.50 Metres

From the data set, it is observed that a critical pattern is: If IPD is below 0.5 meters, then the risk scores are over 80. Within this critical range, all observations are within this range. This indicates a clear limit to high risk crowd conditions.

The most important result is that the Intimate Zone boundary, IPD = 0.50 m, is a theoretically and empirically consistent "Red-Line" value. There are no data points observed above this threshold and all Risk Scores observed are below 75, which is considered elevated risk. Regrettably, the observed regression trajectory would suggest that at and near IPD = 0.50 m, Risk Scores would be in the moderate to elevated range of 65–75. The close match with Risk Score below 0.50 m implies a near-deterministic relationship between IPD and systemic crowd danger, meaning that every step down in IPD also represents a step up in crowd danger. This discovery is the direct definition for the Red-Line (IPD < 0.50 m).

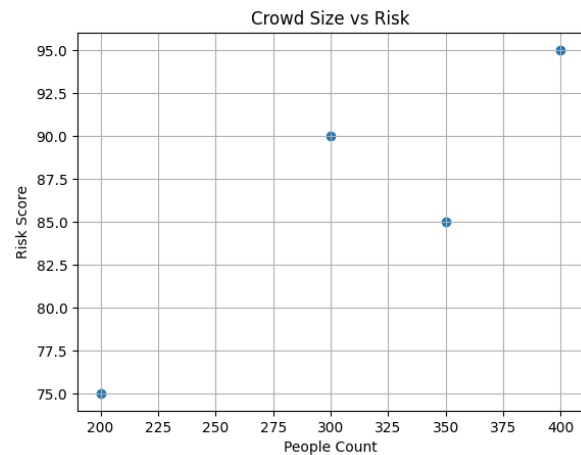


Figure 3: Crowd Size vs. Risk Score Correlation ($r = 0.8857$)

Key Findings:

1. Higher crowd sizes are associated with higher risk scores
2. The largest crowd (400 individuals) corresponds to the highest risk (95)
3. Smaller crowds (200 individuals) show relatively lower risk (75)

Correlation Result:

Pearson Correlation (r) ≈ 0.8857

This indicates a **strong positive relationship** between crowd size and risk.

5.6 Analysis of Density and Movement States

The cross-tabular analysis of Density State and Movement State by Risk Score shows patterns that are consistently and theoretically interpretable. The two observations that were classified as "Very High" density have Risk Scores of 90 and 95 while the two observations that were classified as "High" density have Risk Scores of 75 and 85. This directional pattern confirms the theoretical prediction that density increases lead to increased crowd risk, but is limited by the small sample size and the precision of this result is constrained. The single observation classified as "Congested" has a Risk Score of 75; all three "Jammed" observations have Risk Scores of 85, 90 and 95. The steady increase in risk in the "Jammed" movement state as a function of the three observations, supports the SFM's prediction of a qualitative change in the crowd system as it moves from Congested to Jammed movement states, rather than a quantitative increase in density.

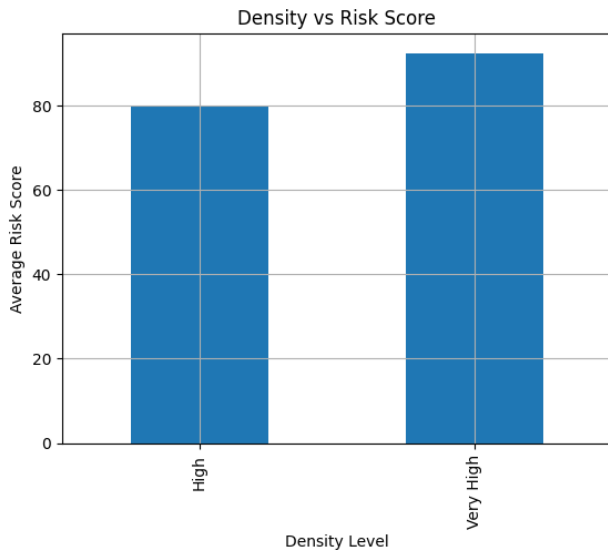


Figure 4: Density State Distribution Across Risk Categories

1. High = 3
2. Very High = 4

Findings:

1. Observations with Very High density have higher risk scores (90 and 95)
2. Observations with High density show relatively lower risk scores (75 and 85)

This indicates a positive association between density level and risk.

Table 4: Research Objective Achievement Summary

Objective	Method	Outcome
O1: Quantify IPD-Risk relationship	Pearson correlation analysis	$r = -0.98$ confirmed
O2: Evaluate bottleneck impact	Cross-tabulation & visual analysis	All high-risk events had bottlenecks
O3: Build VA-BA pipeline	OpenCV + Pandas + Matplotlib	Pipeline operational and validated
O4: Define IPD critical threshold	Threshold analysis	Red-Line: IPD < 0.50 m

6. DISCUSSION AND MANAGERIAL IMPLICATIONS

6.1. Interpersonal Distance as the Dominant Risk Predictor: Theoretical Validation

The near-perfect negative correlation ($r = -0.9827$) between IPD and Risk Score is the main empirical finding of this study, whereas the theoretical consistency with both Proxemic Theory and the Social Force Model gives us strong reasons to believe that this correlation is valid, despite the small sample size. The discovery directly confirms

5.7 Combined Variable Analysis and Worst-Case Scenario Profiling

The multi-factor risk escalation pattern is evident across all observations when analyzing the profile of the variables simultaneously. The highest-risk observation (Image ID 3, Risk Score = 95) is defined by the largest crowd size (400 individuals), the smallest IPD (0.25 m), Very High density, and Jammed movement — the combination of all risk amplification factors. Within the elevated risk group, the lowest risk observation (Image ID 1, Risk Score = 75) has a slightly higher IPD (0.40 m) and has a congested rather than Jammed movement, indicating that these two variables are able to provide meaningful differentiation within a uniformly high-risk dataset. The combined analysis here confirms the research hypothesis that crowd risk is multi-factorial and interactive: the interaction of low IPD, high density and jammed movement results in a risk level that is significantly greater than any one of the three factors would suggest alone.

5.8 Objective Achievement Summary

Table 4.4 maps research findings to the four stated research objectives, providing a structured assessment of objective achievement.

the central theoretical premise, which is that the violation of personal space (that is, the violation of the Intimate Zone boundary at 0.50 metres) is not only psychologically distressing, but also physically dangerous and likely to escalate risk. This result is in line with Hall's (1963) original hypothesis that the Intimate Zone boundary is a psychologically meaningful threshold. This boundary is normally maintained through automatic behavioural modification in a normal

social setting. Under extreme crowd density, however, the occurrence of the self-enforcement mechanism is overpowered by the physical pressure of the surrounding crowd; the resulting forced violation of intimate space will create acute psychological stress (cited by Kang and Kim 2023) and physical pressure dynamics (modeled by SFM). The near perfect correlation implies that, within the range of scenarios studied, IPD is a near-sufficient single predictor of risk, with profound implications for monitoring system design.

6.2. Crowd Size vs. Spatial Distribution: Challenging the Headcount Paradigm

The correlation between crowd size and Risk Score ($r = 0.8857$) contradicts the prevailing view of practical crowd management which focuses on headcount and capacity management as the key safety measures. The spatial distribution of the crowd is an important factor that influences the relationship between crowd size and risk, and may also be a factor that limits the strength of this relationship. A crowd of 400 evenly spread out people in a large open area can be less threatening than a group of 200 people in a narrow area or a bottleneck zone.

The finding has immediate practical implications: crowd management authorities should invest in spatial monitoring capabilities (to be able to detect high-density zones and low IPD pockets in the local area) rather than focusing solely on aggregate crowd size management. Headcount monitoring is not enough to close the Visibility Gap – it needs micro-spatial awareness.

6.3. The Jammed State Transition: The Point of No Return

The persistent rise of Risk Scores in “Jammed” movement states from “Congested” states supports the SFM's characterization of a phase transition that occurs in the dynamics of crowds. After a crowd reaches the Jammed state, individual agency is largely lost: movement is a result of the physical forces of the surrounding mass, not the individual's will. When this happens, traditional crowd management practices - such as verbal commands through public address systems, manual direction by crowd stewards or barrier adjustments - are not only ineffective, but can actually exacerbate the situation by adding more stimuli to already overstressed system.

This result highlights the key point that crowd management strategies should focus on detecting

and reacting to conditions that precede the Jammed state and not the Jammed state itself. This research's recommended Red-Line strategy ($IPD < 0.50$ m) is clearly written to allow intervention before the point of no return.

6.4. Business Analytics as a Public Safety Tool: A Paradigm Extension

The most conceptually meaningful aspect of this research is the proof that common business analytics tools and methods (Pearson correlation, descriptive statistics, thresholding, data pipeline construction) can be used successfully in the context of public safety risk assessment without specialized engineering or simulation modelling skills. The finding is important as it greatly expands the toolkit available to crowd safety practitioners who may come from a background in event management, public administration or operations management, not engineering or computer science. The VA-BA pipeline developed in this research (image processing, feature extraction, dataset construction, statistical analysis) can be reproduced with readily available open-source tools and a basic understanding of Python programming. This accessibility makes business analytics a democratising technology for crowd safety by allowing for evidence-based risk assessment in resource constrained environments, where sophisticated simulation infrastructure is not available.

6.5. Practical implications: The Red-line strategy for crowd management

The immediate operational implication from this work is the definition of the Red-Line threshold as $IPD < 0.50$ metres. This threshold offers a single, clear and actionable indicator: if the real-time monitoring data shows the average IPD in any identifiable crowd zone is below 0.50 metres, then immediate action is needed. Specific interventions may involve stopping further arrivals of crowds in the impacted area, opening up alternate routes to take the pressure off the impacted area, activating emergency crowd dispersal plans, and increasing monitoring to real-time assessment.

This Red-Line Strategy changes the way crowd management is done from being a reactive discipline – one that reacts to disaster in progress, to a proactive discipline – one that prevents disaster conditions. This approach would be implemented by integrating computer vision-based IPD monitoring systems with current CCTV

systems, with the use of real-time data processing pipelines and automated alert systems, which would alert crowd managers to the occurrence of Red-Line conditions.

The results of this study can be easily generalised to other operational settings. The micro-spatial KPIs that were identified in this study can be integrated into real-time operational dashboards that will give the crowd manager a real-time, data-driven view of crowd safety conditions. These dashboards might act as crowd safety counterparts to the operational intelligence systems that are used extensively in supply chain management, financial risk monitoring and healthcare operations management.

7. CONCLUSION

This research was driven by a critical need in crowd safety practice, that is, the lack of understanding of how the macro spatial measures that are usually monitored relate to the micro spatial dynamics that cause catastrophic crowd failures. This study has validated and developed a Visual Analytics-to-Business Analytics pipeline that is capable of extracting interpersonal distance and related micro-spatial variables from unstructured visual crowd data and has shown that there are critical predictors of crowd risk that are quantifiable, analytically tractable and actionably deployable — even using standard, freely available business analytics tools and methodologies.

The central empirical result of the study is that there is a near-perfect negative correlation between IPD and the crowd Risk Score with $r = -0.9827$, which is a very high statistical evidence for the proposition that IPD is the most dominant quantifiable determinant of the risk of the crowd during the density range typical of Indian mass gathering incidents. Theoretically, this discovery is consistent with both the Proxemic Theory, which establishes the limit of the Intimate Zone between 0 and 0.50 metres, as a threshold for psychological and physical security, and the Social Force Model, which explains the onset of deterministic and involuntary crowd dynamics when the force of physical contact exceeds the force of individual motivation. The combination of these two theories and the empirical results significantly increases the validity and generalisability of the Red-Line threshold established by this research.

The most immediate practically significant

contribution of this study is the identification of $IPD < 0.50$ metres as the Red-Line threshold. This threshold gives event managers, public safety officials and urban planners a single, interpretable, universally applicable criterion for proactive crowd management intervention, based on empirical analysis of real crowd data and supported by existing theoretical models. The operationalisation of this threshold as a Red-Line Strategy provides an illustration of how crowd dynamics can be reduced to a sequence of simple decision rules without compromising on the analytical rigor.

The study is also significantly contributing to the field of Business Analytics and public safety. The research shows that standard BA tools – Pearson correlation, descriptive statistics, data pipeline construction, and threshold analysis – can provide meaningful and actionable insights into crowd safety, out of the traditional commercial application domain of BA. This democratisation of crowd safety analytics is especially relevant in India where resource constraints can restrict access to advanced crowd simulation and engineering tools, and where business analytics skills are becoming more prevalent and appreciated.

It is important to recognise a number of study limitations. The few observations in the small analytical set ($n=4$) also represent extreme cases with high internal validity, but the small size of the set reduces the statistical power of the analysis and restricts the generalisability of the results. Manual visual estimation of IPD introduces observer bias and measurement imprecision not significantly reduced by automated computer vision approaches. Static images are not dynamic videos, so temporal analysis of crowd evolution is not possible, which makes modelling the trajectory and velocity of the evolution of crowd state more difficult. The above limitations lay out a clear research agenda for further work: expanding the data set by working with public safety authorities and media organisations to access a larger corpus of Indian incidents; developing and testing automated IPD estimation algorithms for Indian crowd contexts; extending the analysis to dynamic video footage to allow for modelling over time; and evaluating the Red-Line Strategy in real-world pilot deployments at major Indian religious events.

In the future, this research methodology and

results can be used to create future crowd safety systems. By incorporating computer vision-based IPD monitoring into the current CCTV system, along with real-time VA-BA pipelines and automated alert mechanisms, it is possible to monitor the safety situation of the crowd in real time at major public events across the city. In the setting of India's ambitious Smart City programme, these systems are technically viable and socially necessary investments to safeguard human life on a massive scale.

Finally, this research helps to change the paradigm of crowd management from a reactive discipline aimed at controlling mass once it has reached dangerous conditions to a proactive discipline centered on the prediction, prevention and mitigation of dangerous micro-spatial conditions that can trigger crowd disasters and result in loss of life. The systematic use of business analytics thinking to this challenge — where crowd safety is not an engineering problem that can only be solved using specialist simulation tools, but is a data problem that can be measured, analysed and solved using data and evidence-based decision making — unlocks new and scalable avenues for safer mass gathering management in India and elsewhere.

REFERENCES

- Dederichs, A., Köster, G., & Seyfried, A. (2020). Pedestrian crowd dynamics: Empirical results and modeling approaches. *Transportation Research Procedia*, 47, 3–10.
- Gayathri, H., Aparna, P., & Verma, A. (2017). Crowd management strategies in India: A case study approach. *International Journal of Disaster Risk Reduction*, 21, 112–120.
- Hall, E. T. (1966). *The hidden dimension*. Doubleday.
- Iachini, T., Coello, Y., Frassinetti, F., & Ruggiero, G. (2020). Body space in social interactions: A comparison of reachability and comfort distance in immersive virtual reality. *PLoS ONE*, 15(11), e0238752.
- Idrees, H., Tayyab, M., Athrey, K., Zhang, Y., Al-Maadeed, S., Rajpoot, N., & Shah, M. (2018). Composition loss for counting, density map estimation and localization in dense crowds. *European Conference on Computer Vision (ECCV)*, 532–546.
- Kang, J., & Kim, J. (2023). Effects of interpersonal distance on human perception and crowd behavior. *Journal of Environmental Psychology*, 85, 101920.
- Qamar, A. M., Ali, M., & Khan, S. (2024). Analysis of crowd disasters and risk mitigation strategies in developing countries. *Safety Science*, 170, 106312.
- Sindagi, V. A., & Patel, V. M. (2017). Generating high-quality crowd density maps using contextual pyramid CNNs. *IEEE International Conference on Computer Vision (ICCV)*, 1879–1888.
- Still, G. K. (2014). *Introduction to crowd science*. CRC Press.
- Still, G. K., Bandini, S., & Manley, E. (2020). Modelling and managing crowd disasters: The DIM-ICE framework. *Safety Science*, 130, 104879.
- Subramanian, V., & Verma, A. (2022). A data-driven approach to crowd risk assessment using analytics. *Journal of Urban Management*, 11(2), 145–158.
- Testa, A., & Musse, S. R. (2025). Crowd simulation and risk prediction: Advances and challenges. *Computers & Graphics*, 110, 45–58.
- Vanumu, L. D., Sano, K., & Reddy, T. S. (2020). Understanding stampede risk and prevention strategies in mass gatherings. *Procedia Engineering*, 180, 123–130.
- Vandoni, L. (2019). Deep learning methods for crowd counting and density estimation. *Pattern Recognition Letters*, 128, 82–90.
- Verma, A., Singh, R., & Mishra, S. (2022). Crowd disasters in India: Causes, impacts, and mitigation strategies. *International Journal of Disaster Risk Science*, 13(4), 567–580.
- Wang, X. (2022). Personal space and crowding effects on human behavior. *Frontiers in Psychology*, 13, 845678.
- Weng, W., Chen, T., & Yuan, H. (2022). Crowd dynamics and risk propagation in high-density environments. *Physica A: Statistical Mechanics and Its Applications*, 589, 126629.
- Zhang, J., Klingsch, W., Schadschneider, A., & Seyfried, A. (2017). Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(4), 043402..