Quantitative Techniques

Report

on

"Risk Analysis in Women Safety: Using Monte Carlo Simulation"

Submitted in partial fulfillment of requirements for the degree of

5-Years Integrated MBA (Tech.) Program in

COMPUTER ENGINEERING & TECHNOLOGY MANAGEMENT

By

Chaitanaya Kusumakar

MBA Tech (CE), 70472100327

Paragi Chauhan

MBA Tech (CE), 70472100429

Under the Guidance of

Dr. Shubhangi Jore

(Assistant Professor, SBM NMIMS Indore)

School of Technology Management & Engineering
SVKM'S-NMIMS (Deemed-to-be-University), Indore
[2024-25]

INTRODUCTION

In recent years, women's safety has become an increasingly important issue, necessitating the development of advanced analytical tools to assess and mitigate risks. Traditional methods of assessing safety often rely on static, historical data, but fail to account for the dynamic nature of risk factors, such as time of day, surrounding environment, and proximity to safety resources. To address this challenge, this report utilizes Monte Carlo Simulation to model and analyse the potential risk factors that affect women's safety in real-world scenarios.

The Monte Carlo Simulation is a powerful statistical technique used to understand the impact of risk and uncertainty in predictive models. By simulating thousands of possible scenarios, it helps in identifying patterns and estimating the likelihood of adverse outcomes based on a variety of input factors. This approach allows us to go beyond deterministic risk assessment by incorporating randomness and variability into the analysis.

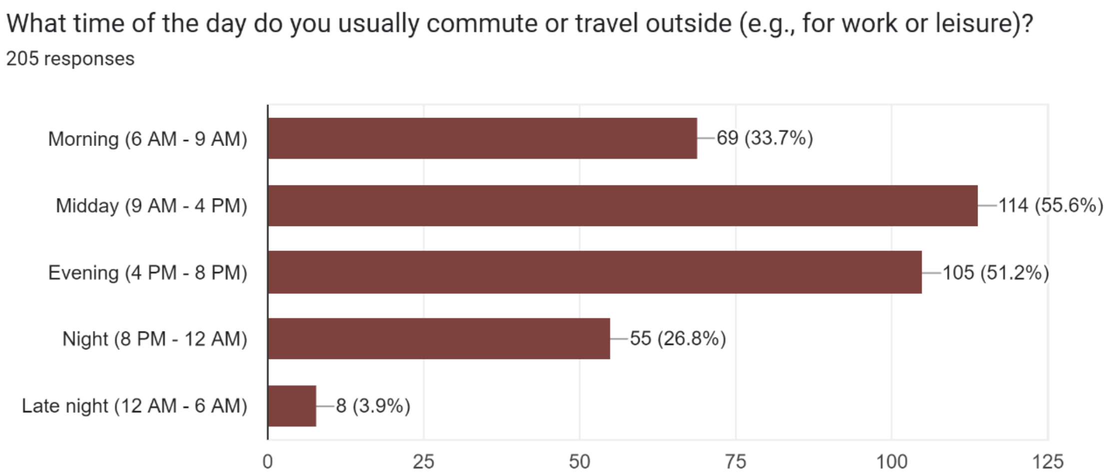
In this report, the simulation focuses on identifying critical factors such as time, the presence of people nearby, and distances from safety resources (home, police stations, and populated areas). By assigning probability distributions to these variables and simulating real-life situations, the model provides insights into when and where women are at the highest risk, offering data-driven recommendations for enhancing safety measures.

This analysis forms the foundation for developing targeted safety solutions and policies that can be implemented in high-risk environments. The outcomes of this study are essential for authorities, urban planners, and public safety organizations to proactively address safety concerns, ensuring that women feel secure in public spaces, particularly during vulnerable hours or in isolated areas.

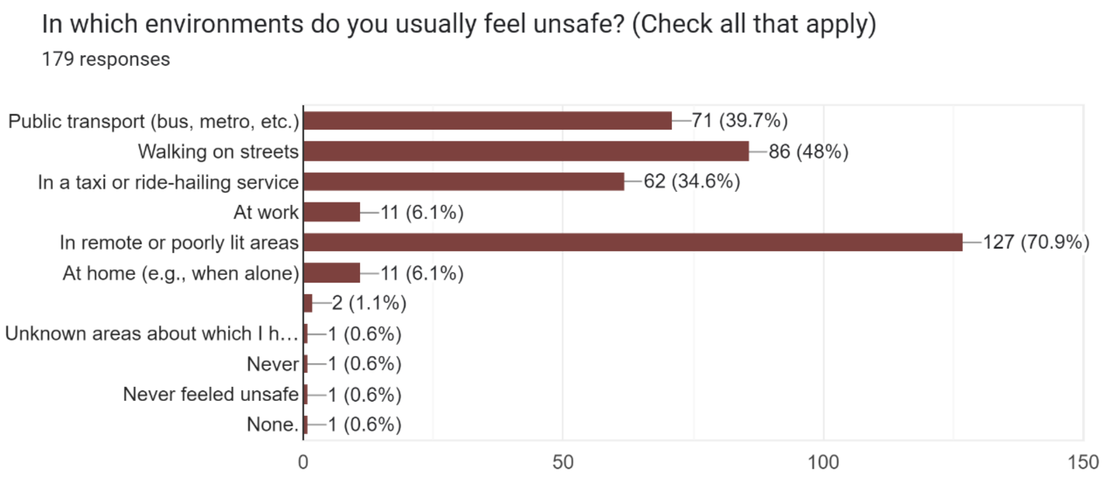
METHODOLOGY

Data for this study was collected through a survey targeting various demographics, focusing on women's experiences and perceptions of safety in public spaces. From the survey, 3 major factors were identified:

- Commute Time: The time of day when respondents typically travel.



- Proximity to People: The number of people present during travel.
- Distances to Safety Resources: This majorly includes the distance from
    - Home
    - Nearest Police Station
    - Nearest Public Areas



As per the data of google maps and the understanding, it was assumed that if a location is far from 5 km, then these factors are independent and directly corelated to the risk. The dataset was cleaned to remove incomplete or irrelevant entries, ensuring accuracy for the simulation. Monte Carlo Simulation was employed to model the variability and uncertainty of different risk factors impacting women's safety. This technique allows for the random generation of thousands of possible real-world scenarios, simulating how various factors - such as time of day, presence of people, and distance to safety resources - can influence overall risk. A total of 10,000 simulations were run, where each iteration drew random values for commute time, proximity to people, and distances from safety resources. These variables were combined to calculate an overall risk score for each scenario. The risk score was determined by weighing the influence of each factor on the perceived danger, with greater risks assigned to scenarios involving late-night commutes, fewer people, and greater distances from safety.

MODEL DESIGNING

The model for this study was designed to simulate real-world scenarios that women might encounter in public spaces, with the objective of calculating a total safety risk score based on several key environmental and situational factors. These factors include time of day, proximity to other people, and distances to safety resources such as home, police stations, and populated areas.

A. Assigning Probability Distributions

For each risk factors identified as mentioned above, a suitable probability distribution was assigned to reflect the variability observed in real-world conditions:

- **Time of Day**: Modelled as a uniform distribution between 0 and 24 hours, allowing for an equal probability of travel at any time. The model accounts for increased risk during late-night hours using a cosine function to capture cyclical risk patterns.
- **Number of People Nearby**: This factor was modelled using a normal distribution with a mean of 5 people and a standard deviation of 2. This reflects a typical public environment where fewer people nearby are associated with higher risk, and larger crowds reduce risk.
- **Distances to Home, Police, and Populated Areas**: These were modelled using normal distributions based on expected average distances. For example, distance to home was assigned a mean of 21 km with a standard deviation of 5 km, while distance to police stations was modelled with a mean of 2 km and a standard deviation of 0.5 km.

B. Developing the Risk Equations

Each variable was associated with a custom risk equation, designed to capture how different levels of that variable contribute to overall safety risk. The following equations were developed:

- ***Risk Based on Time of Day***: A cosine function was used to model the cyclic nature of risk over 24 hours, with higher risk at night and lower risk during the day.

$$risk\_time = (4 \times \cos(\pi/13 \times (time - 3))) + 6$$

This equation captures the variability of risk based on the time of day. The cosine function reflects the cyclic nature of risk, where late-night and early-morning hours typically present higher danger. The constant values scale the risk to a reasonable range, with the highest risk during nighttime (e.g., midnight to 3 a.m.) and the lowest during midday.

- ***Risk Based on Proximity to People:*** An exponential decay function was used, where risk decreases rapidly as the number of people nearby increases.

$$risk\_people = (8 \times e - 0.05 \times people) + 2$$

The risk decreases exponentially as the number of people increases, based on the assumption that more people present reduces the likelihood of unsafe situations. The exponential decay ensures that the risk drops rapidly with even a moderate number of people, but a minimum baseline risk is maintained (scaled by adding 2).

- ***Risk Based on Distance to Home***: Linear functions were used to model how risk increases as the distance to home.

$$risk\_home = (0.32 \times distance\_home) + 2$$

This linear equation accounts for the distance from home, with greater distances implying a higher risk. The coefficient of 0.32 reflects the rate at which risk increases as the distance grows, while the constant 2 ensures that a baseline risk is present even when close to home.

- ***Risk Based on Distance to Police Station***: Exponential functions were used to model how risk increases as the distance nearest Polce Station.

$$risk\_police = (e^{0.4158 \times distance\_police}) + 2$$

The risk of harm increases exponentially as the distance from a police station increases. The exponential function models the rapid escalation of danger when far from immediate law enforcement assistance, while the baseline risk is again maintained by adding 2.

- ***Risk Based on Distance from Large Population***: Exponential functions were used to model how risk increases as the distance to nearest highly populated area.

$$risk\_pop = (e^{0.4158 \times distance\_pop}) + 2$$

Like the risk from distance to the police, this equation models the danger associated with being far from a populated area. The further a woman is from crowds or busy areas, the higher the risk, particularly at night or in isolated locations.

- ***Total Risk Calculation (total_risk)***: The final component of the model is the total risk score, which is a weighted sum of the individual risk factors. Weights were assigned based on the relative importance of each factor:
    - Time of Day (10%)
    - Proximity to People (20%)
    - Distance to Home (20%)
    - Distance to Police (20%)
    - Distance to Populated Areas (30%)

The weights were assigned to emphasize factors that have a greater impact on safety, with isolation (distance to populated areas) carrying the most weight due to its significant role in vulnerability.

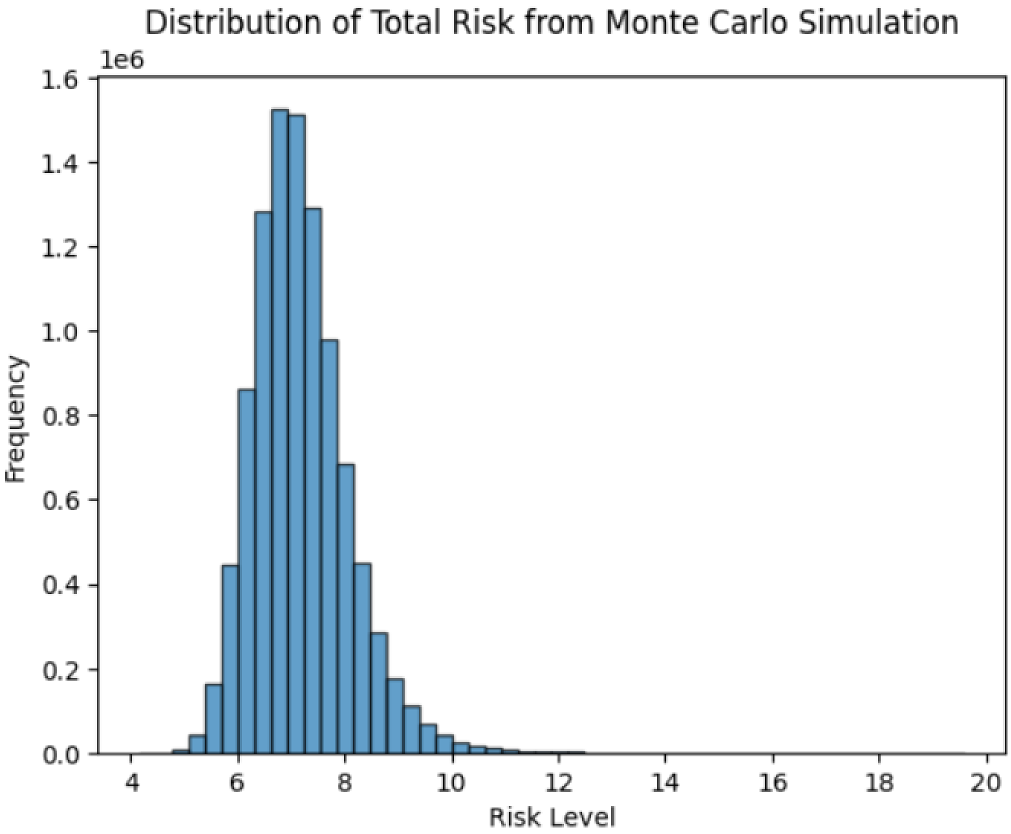
$$total\_risk = (0.1 \times risk\_time) + (0.2 \times risk\_people) + (0.2 \times risk\_home) + (0.2 \times risk\_police) + (0.3 \times risk\_pop)$$

C. Simulation Process

The Monte Carlo Simulation was used to run thousands of simulations, each time randomly drawing values for time, number of people, and distances from their respective distributions. For each scenario, the risk factors were calculated, and a total risk score was produced. This process allowed us to generate a distribution of possible risk outcomes, providing insights into the likelihood of encountering high-risk situations under various conditions.

By integrating these factors into the model, we were able to assess how changes in any variable - such as traveling at night, being far from a populated area, or being alone - affect a woman's safety. The model offers a robust framework for understanding and quantifying risk, which can be applied to real-world safety interventions and policy decisions.

```python
import numpy as np
import math
import matplotlib.pyplot as plt

# Constants for number of simulations
num_simulations = 10000000

# Define random variables based on assumed distributions
def random_time():
    # Time of day between 0 and 24 (assuming higher risk at night)
    return np.random.uniform(0, 24)

def random_people():
    # Number of people nearby (assuming fewer people = higher risk)
    return max(1, np.random.normal(5, 2))  # Mean of 5 people, standard deviation of 2

def random_distance(mean, std_dev):
    # Distances for home, police, populated area in km
    return np.random.normal(mean, std_dev)

def monte_carlo_simulation(num_simulations):
    total_risks = []

    for _ in range(num_simulations):
        time = random_time()
        people = random_people()
        d_home = random_distance(21, 5)  # Mean 21 km, Std dev 5 km
        d_police = random_distance(2, 0.5)  # Mean 2 km, Std dev 0.5 km
        d_pop = random_distance(4, 1)  # Mean 4 km, Std dev 1 km

        risk_time = (4 * math.cos((math.pi / 13) * (time - 3))) + 6
        risk_people = (8 * math.exp(-0.05 * people)) + 2
        risk_home = (0.32 * d_home) + 2
        risk_police = math.exp(0.4158 * d_police) + 2
        risk_pop = math.exp(0.4158 * d_pop) + 2

        total_risk = (0.1 * risk_time) + (0.2 * risk_people) + (0.2 * risk_home) + (0.2 * risk_police) + (0.3 * risk_pop)

        total_risks.append(total_risk)

    return total_risks

# Running the Monte Carlo Simulation
results = monte_carlo_simulation(num_simulations)

plt.hist(results, bins=50, edgecolor='k', alpha=0.7)
plt.title("Distribution of Total Risk from Monte Carlo Simulation")
plt.xlabel("Risk Level")
plt.ylabel("Frequency")
plt.show()

print(f"Mean Risk: {np.mean(results)}")
print(f"Median Risk: {np.median(results)}")
print(f"Standard Deviation: {np.std(results)}")
print(f"Risk > 8 Probability: {np.sum(np.array(results) > 8) / num_simulations}")
```




```
Mean Risk: 7.171190576486968
Median Risk: 7.072774993964423
Standard Deviation: 0.8781284806774327
Risk > 8 Probability: 0.1554471
```

This is the simulation of 10000 random scenarios with different values of risk factors.

	Time	People	Distance_Home	Distance_Police	Distance_Pop	Risk_Time	Risk_People	Risk_Home	Risk_Police	Risk_Pop	Total_Risk	Frequency
0	8.589964	6.469368	23.858840	2.155659	4.258942	6.872607	7.789078	9.634829	4.450576	7.875952	7.424943	0.225558
1	9.244596	6.005964	18.905720	1.753106	1.828507	6.246728	7.924779	8.049830	4.072885	4.138903	5.875843	-1.151420
2	0.133480	5.430852	18.248686	1.607171	2.767143	9.078032	8.097623	7.839580	3.950843	5.160019	6.433418	-0.655798
3	4.918582	6.058621	20.880134	2.946411	2.820951	9.577711	7.909200	8.681643	5.404566	5.231517	6.926308	-0.217673
4	13.657019	3.673622	18.443689	1.974075	2.440220	2.624233	8.657609	7.901980	4.272363	4.758380	5.856328	-1.168767
...	...	...	...	...	...	...	...	...	...	...	...	...
9995	7.380126	5.922699	26.485655	2.243990	3.619605	7.960701	7.949496	10.475410	4.542254	6.504297	7.340791	0.150756
9996	15.111977	2.602882	20.898030	2.560675	4.418508	2.091754	9.023751	8.687370	4.900052	8.279031	7.215119	0.039048
9997	22.237370	4.732887	14.580305	2.525234	2.987007	5.746301	8.314175	6.665698	4.857629	5.462524	6.180888	-0.880269
9998	22.556105	5.453487	22.026816	2.370569	4.845990	6.054232	8.090725	9.048581	4.679641	9.500446	7.819346	0.576139
9999	4.019916	1.177360	17.416735	1.637732	3.282938	9.879115	9.542650	7.573355	3.975791	5.915902	6.981041	-0.169021

10000 rows × 12 columns

RESULTS AND ANALYSIS

The results of the Monte Carlo Simulation provided a detailed distribution of risk levels under various scenarios, highlighting key factors that contribute to women's safety in public spaces. Over 10,000 simulations, the model generated a range of total risk scores based on randomly varying inputs for time of day, number of people nearby, and distances from safety resources.

1. Risk Score Distribution

The histogram of the risk scores shows a bell-shaped distribution, with most scenarios resulting in moderate risk levels, while a smaller percentage of cases fell into the high-risk category:

- **Mean Risk Score**: 7.171
- **Median Risk Score**: 7.072
- **Standard Deviation**: 0.87
- **High-Risk Scenarios (Risk > 8)**: Approximately 15% of all simulations resulted in risk scores greater than 8, indicating significant danger.

2. Findings from the Model

- **Time of Day**: As expected, risk levels were higher during late-night and early-morning hours. Scenarios between 11 p.m. and 4 a.m. consistently produced higher risk scores, confirming the heightened danger associated with these times.
- **Proximity to People**: Scenarios with fewer than three people nearby resulted in notably higher risks. The risk score dropped significantly when five or more people were present, indicating that crowded areas provide a strong protective effect.
- **Distance to Safety Resources**: The further the distance from home, police stations, or populated areas, the higher the risk score. Scenarios involving long distances from all three safety resources consistently led to high-risk outcomes, especially when compounded by other risk factors (e.g., late-night travel).

3. Probability of High-Risk Scenarios

The simulation revealed that the probability of encountering a high-risk situation (risk score > 8) was **15%**, driven largely by combinations of late-night travel, isolation, and distance from populated areas. The risk was most pronounced in scenarios where all three factors were at their worst, such as traveling late at night, being alone, and being far from both police stations and populated areas.

CONCLUSION

The Monte Carlo Simulation provided valuable insights into the dynamics of women's safety in public spaces, confirming that time of day, proximity to other people, and distance from safety resources are critical determinants of risk. The findings indicate that about 15% of situations pose a significant safety risk, primarily in scenarios where multiple risk factors alignsuch as late-night travel in isolated locations.

This study underscores the need for targeted interventions to reduce the likelihood of high-risk scenarios. By improving public safety infrastructure, enhancing surveillance, and promoting community involvement, authorities can significantly reduce the danger women face in public spaces, particularly during vulnerable times of the day.