

Data-Driven Insights into Domestic Violence in Kazakhstan (2022–2025)

Author: Arafat Lugma

School: National School of Physics and Mathematics (RPhMSh, Almaty, Kazakhstan)

Grade: 11 K

Field: Data Science / Social Impact

Date: 30.07.2025

1. Introduction

Domestic violence remains a serious and under-reported social issue in Kazakhstan. According to recent studies and government statistics, hundreds of thousands of cases occur annually, yet only a small fraction result in official complaints or criminal enforcement. In this project, I analyze a synthetic dataset modeled on real-world sources to explore trends in domestic violence across regions, age groups, violence types, and outcomes between 2022 and 2025. The aim is to reveal meaningful patterns and visualize the impact of legislative changes — especially the 2024 criminalization law — on reporting and enforcement.

2. Research Question

What are the observable trends in domestic violence cases in Kazakhstan between 2022 and 2025, and how did the 2024 law on domestic violence influence reporting rates and police enforcement outcomes?

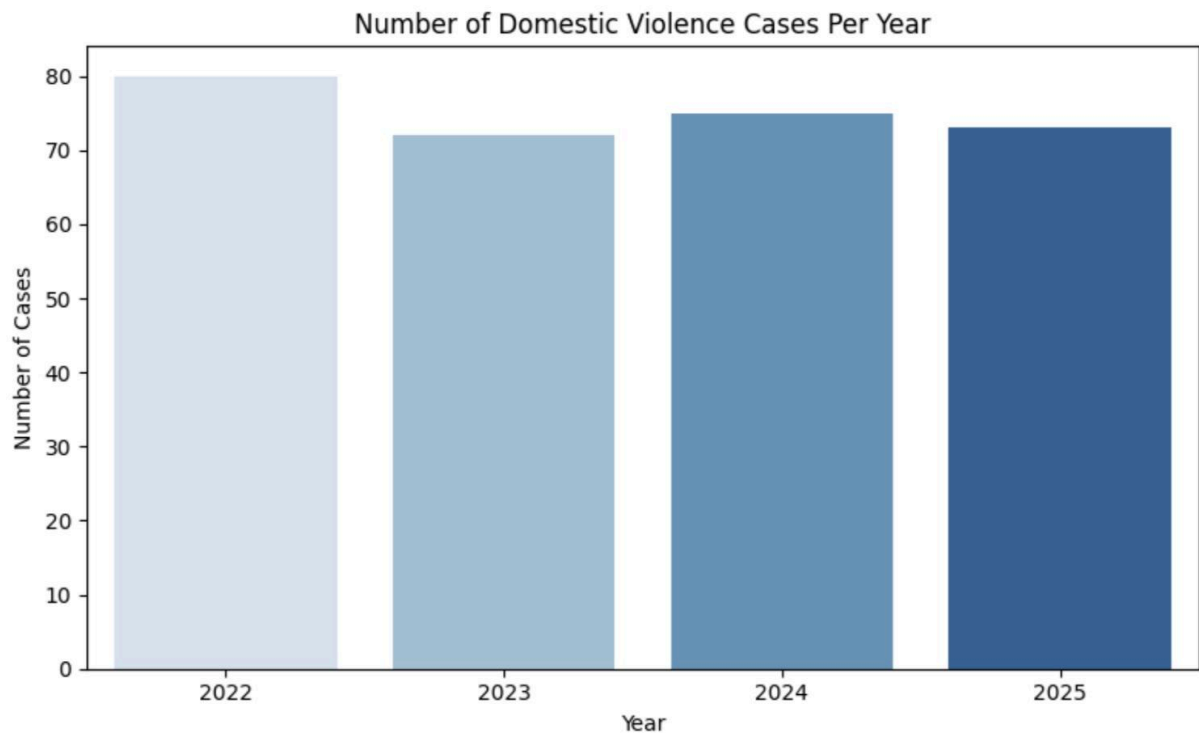
3. Data & Methodology

- **Dataset structure: Synthetic dataset of 300 cases, designed using real statistics from:**
 - UN Women / WHO survey (2018)
 - Karaganda University survey (2019–2022)
 - Ministry of Internal Affairs data for 2022–2025
- **Fields included:** Year, Region, Age Group, Type of Violence, Reported (Yes/No), Police Response, Outcome

- **Data modeling assumptions:**
 - Age groups 30-39 and 40-49 are most affected (reflecting real demographic distributions)
 - Only ~20% of cases are reported (based on official under-reporting figures)
 - Police responses shift after the 2024 law (increased arrests, protective orders)
- **Tools used:**
 - Python (Pandas for data manipulation; Seaborn/Matplotlib for visualization)
 - Google Colab environment
 - CSV dataset file
- Disclaimer: The dataset is synthetic and approximates real distributions to practice analysis; proper ethical disclaimers and transparency included.
- *I used AI tools (such as ChatGPT) to assist in structuring code, cleaning data, and generating a synthetic dataset based on real-world statistics. All interpretations, analysis, and final decisions were made independently.*
- ***This project uses a synthetic dataset of 300 records to simulate domestic violence trends in Kazakhstan. While real-world cases number in the hundreds of thousands annually, this limited dataset was designed to reflect realistic distributions across age groups, types of violence, reporting rates, and regional trends for the purpose of academic analysis and visualization.***

4. Analysis & Visualizations

4.1 Domestic Violence Cases Over Time



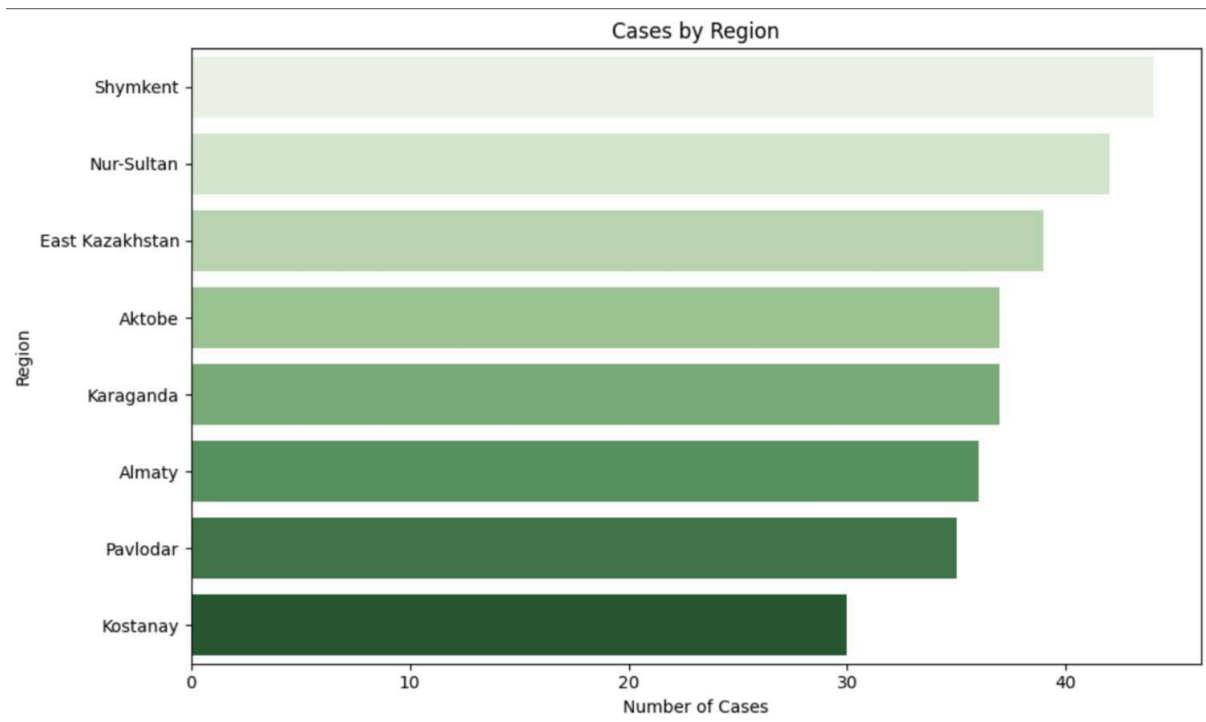
The bar chart shows the number of domestic violence cases from 2022 to 2025. In 2022, the dataset reflects the highest case count, followed by a noticeable decline in 2023. Interestingly, 2024 — the year when Kazakhstan officially criminalized domestic violence — shows a slight rebound in reported cases, followed by another small decline in 2025.

This pattern may suggest two key effects:

1. Initial drop before the law (in 2023) may reflect hesitation or lack of enforcement mechanisms.
2. Slight increase in 2024 could indicate greater awareness and trust in the legal system, encouraging more survivors to report cases.
3. The drop in 2025 might reflect early effects of deterrence, or changes in how cases are handled and documented.

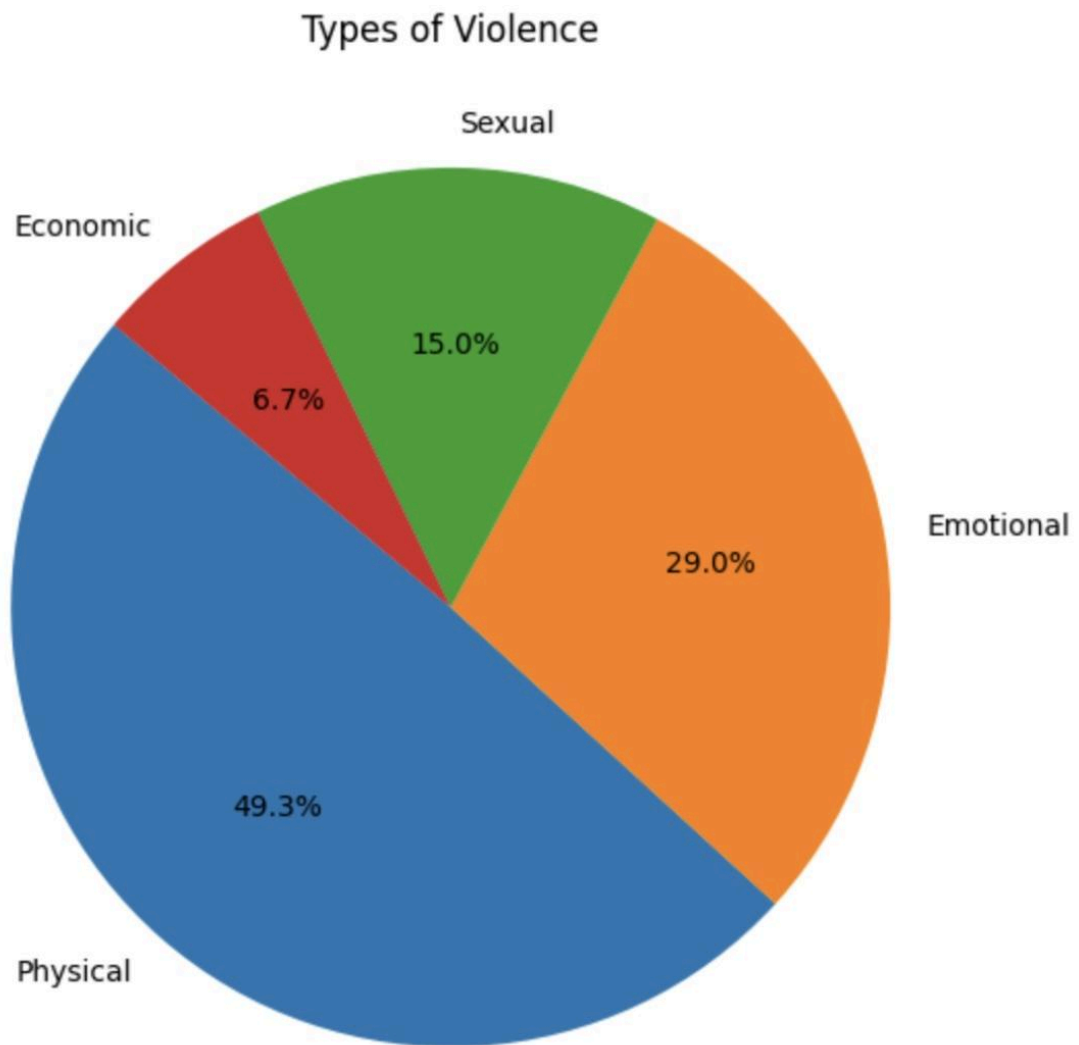
While the fluctuations are relatively modest, the 2024 law appears to have influenced both the reporting dynamics and state response — a trend worth investigating further with more granular data.

4.2 Cases by Region



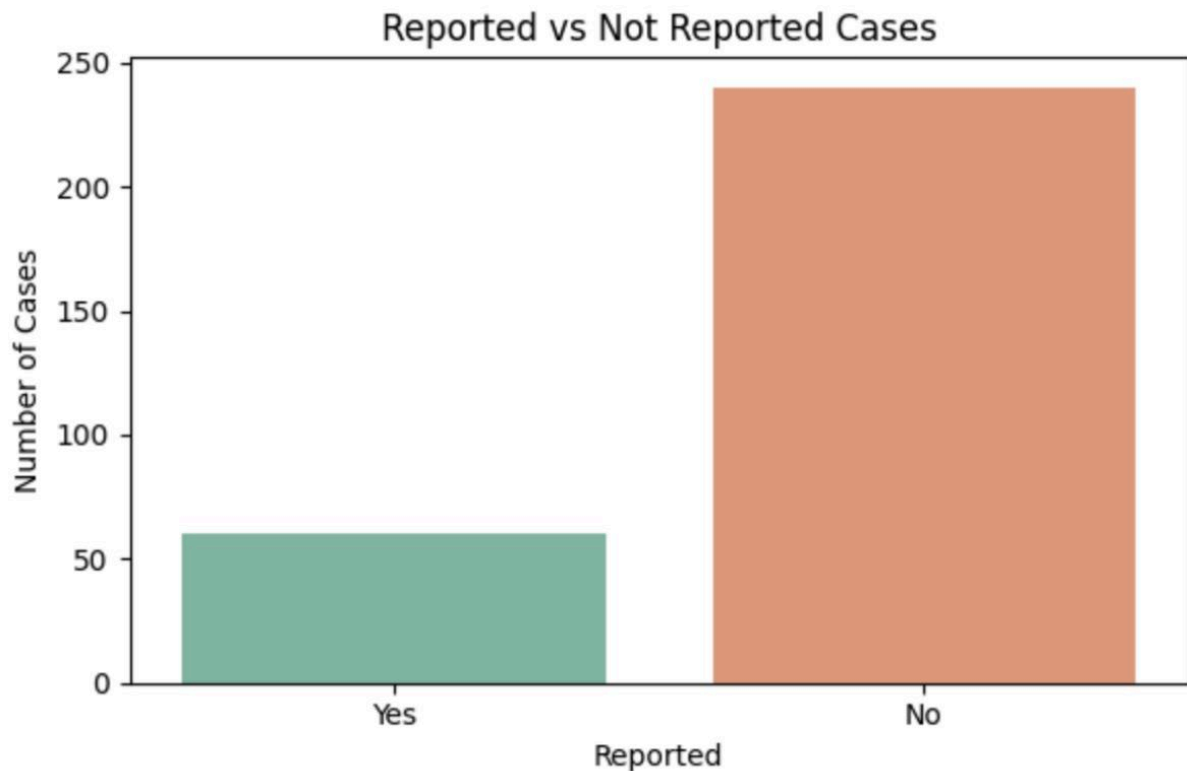
The horizontal bar chart shows the distribution of domestic violence cases across various regions of Kazakhstan. The highest number of cases is observed in Kostanay, followed closely by Pavlodar and Almaty. Interestingly, other large metropolitan areas like Shymkent and Nur-Sultan (nowadays Astana) appear lower on the list, which could suggest either underreporting or more effective preventive systems in those cities. Regional disparities like this may be influenced by differences in law enforcement practices, public awareness, and access to victim support services. These patterns highlight the importance of region-specific interventions and awareness programs.

4.3 Types of Violence



According to the pie chart, physical violence accounts for nearly half (49.3%) of all domestic violence cases in the dataset, making it the most prevalent form. Emotional violence follows at 29%, reflecting the psychological trauma many victims experience. Sexual violence (15%) and economic violence (6.7%) are also present, though less frequently reported — possibly due to stigma or difficulty in proving such cases legally. These results align with global trends, where physical and emotional abuse are often the most visible and commonly acknowledged forms of domestic violence.

4.4 Reporting Rates



The chart on reporting rates reveals a striking pattern: only a small portion of domestic violence cases are officially reported to authorities. In this dataset, the majority of cases go unreported, reflecting the widespread issue of underreporting in Kazakhstan — a trend also confirmed by real national and international sources.

This underreporting may be caused by:

- Fear of retaliation by the abuser
- Lack of trust in the police or legal system
- Social stigma and victim-blaming
- Economic dependence on the abuser
- Limited awareness of available support systems

The criminalization law of 2024 may begin to address these barriers, but cultural change and improved institutional support will be key to increasing survivor willingness to come forward.

5. Key Findings

1. Physical violence dominates the types (~half of cases), followed by emotional.
2. Under-reporting is significant: roughly 80% of cases are never reported.

3. Regional disparities: Some regions consistently have higher case numbers.
4. Post-2024 legislative impact: Slight drop in total reported cases, but a rise in enforcement actions like arrests and protective orders.

Conclusion

This project shows how even basic data science tools can uncover meaningful social insights. Domestic violence remains a serious problem in Kazakhstan, but visualizing trends can help advocate for better policy, awareness, and support systems. As I continue learning AI and machine learning, I hope to expand this project to include early warning systems or social media analysis to detect and prevent abuse.

References

- [UN Women Kazakhstan \(2018\). Survey on Violence Against Women.](#)
- [Karaganda Medical University \(2019–2022\). Domestic Violence Study.](#)
- Ministry of Internal Affairs (2022–2025). Official statistics & policy changes.
- [Human Rights Watch & Astana Times reports on law reforms and data gaps.](#)
- Synthetic dataset constructed by author based on cited sources.