Open Voices:

An Open Data Project about Socio-Cultural Dynamics of Violence Against Women in Italy

Project link: https://github.com/Asemica-me/OADE_OpenVoices.

1. Introduction

This project outlines the integration and reuse of datasets from Istituto Nazionale di Statistica (ISTAT) to analyze violence against women in Italy. Its goal is to create a fair, legally valid, and ethically sound scenario that provides actionable insights into societal attitudes, demographic patterns, and reported cases. The study complies with the EU Open Data Guidelines and addresses privacy, sustainability, and technical challenges.

To approach this complex issue, we have established a few fundamental parameters:

- **Year span**: 2018, providing a focused and detailed analysis based on data from a single year to ensure consistency and clarity in interpreting the results.
- **Gender**: The study includes data on both women and men, as well as aggregated totals, to provide a comprehensive understanding of the phenomenon and examine differences and commonalities across genders.
- **Victims**: The analysis specifically targets victims who reach out to the national helpline 1522 and categorizes the data by geographic regions across Italy.
- **Factors of interest**: The study examines public opinions on gender roles, the acceptability of intimate partner violence, causes of violence, and victim demographics.

To communicate these findings effectively, a variety of visualizations have been employed to present the data in an accessible and engaging manner, helping to highlight key trends and patterns.

2. Scenario

The perception and occurrence of violence against women in Italy are shaped by demographic factors, including geographical location, which influence societal stereotypes and the acceptability of violence. A data-driven approach that integrates population statistics, societal attitudes, and reports of violence can pinpoint intervention opportunities to reduce violence and challenge harmful stereotypes.

This project aims to raise public awareness about violence against women, promote gender equality, and enhance support services for victims. The documentation is based on statistical data from ISTAT, the Ministry of Justice, and the Ministry of Security, with a particular focus on public opinions regarding gender roles and the utilization of the anti-violence hotline 1522.

This research explores the relationship between regional stereotypes, public perceptions, and the prevalence of sexual and domestic violence against women in Italy. It investigates how stereotypes about women and sexual violence differ across regions, reporting rates, and societal attitudes. The study also examines regional variations in awareness and education about consent, alongside socio-cultural factors that contribute to victim-blaming and the normalization of abusive behaviors in intimate relationships.

The final aim is to uncover patterns in the prevalence and reporting of violence, offering insights into the complex interaction between cultural norms and legal outcomes across Italy's diverse regions.

2.1 Research Questions

How do geographical and cultural factors across different regions of Italy influence attitudes toward gender-based violence?

---

2. What are the prevailing stereotypes about women and sexual violence across different regions in Italy, and how do these stereotypes influence public perception, legal responses, and the prevalence of violence?

3. How does the level of awareness and education about consent and sexual violence vary across Italian regions, and what impact does this have on reporting rates?

4. What socio-cultural factors contribute to victim-blaming and the normalization of abusive behaviors in intimate relationships in various geographical areas of Italy, and how do these factors shape public attitudes and response systems?

5. To what extent do regional differences in gender roles and expectations influence the prevalence and reporting of domestic and sexual violence against women across Italy?

6. Which territories exhibit the highest levels of gender stereotypes, and how do these correlate with data on sexual and domestic violence? Are there geographical areas where stereotypes persist despite lower reported rates of violence?

3. Datasets

Dataset about population: **D1**

Title of the section	**Resident population on 1st January**

Data Sources	**Popolazione residente comunale per sesso anno di nascita e stato civile.** Il sondaggio calcola la Popolazione residente comunale per sesso, anno di nascita e stato civile al 31 dicembre di ogni anno. I dati al 1° gennaio 2019 e 2020 tengono conto dei risultati del Censimento permanente della popolazione. I dati relativi agli anni 2002-2018, sono consultabili nella sezione "Popolazione Intercensuaria". **Popolazione residente comunale per sesso anno di nascita e stato civile**
Data characteristics	Data for previous years are available in the Inter censuses estimates theme. A few data for the 2012-2017 years have been subjected to a statistical review. In order to the safeguard of statistical confidentiality, for municipalities with less than 20,000 inhabitants data for marital status regarding same sex civil partner, divorced same-sex civil partner and widow/widower of same-sex civil partner are not disseminated and have been added to the married, divorced, and widowed civil status modalities respectively. Therefore, the amount of the individual municipal data, for some modalities of marital status, can be different from the sum of provincial, regional or geographical areas data.
Concepts & Classifications and other manipulations	The data on 1st January 2019 are disclosed as provisional. The final data will be released upon completion of the statistical realignment operations with the results of the permanent census of the population and housing started in October 2018.

Datasets about opinions: **D2, D3, D5**

Title of the section	**Opinions about gender roles and about violence against women - adults**
Data Sources	The survey on gender role stereotypes and the social image of violence seeks to be the tool for analysing cultural models and some of the factors influencing attitudes towards violence against women among the adult population. The used questions gather gender role stereotypes, opinions on the acceptability of violence, its permeation and its causes as well as some stereotypes about sexual violence. The spread of gender role stereotypes, on the one hand, and the attitudes towards violent behaviour, on the other, are in fact the keys to understand the cultural context in which violent relationships find their genesis and justification. As declared by the Istanbul Convention, knowing about them is essential to a better understanding of the causes of violence and monitoring them over time in order to evaluate, at least partially, how policies involving violence prevention impact the population in terms of cultural change. Istat carries out the survey on gender role stereotypes and the social image of violence as a part of a partnership agreement with the Equal Opportunities Department at the Presidency of the Council of Ministers (DEO).

	The first edition of the survey was carried out as an ad-hoc module in the Labour Force Survey (LFS) and addressed a sub-sample of respondents aged 18 to 74 during the June - November 2018 period. **Gender role stereotypes and the social image of violence**
Sampling	The sampling of the 2018 survey on stereotypes was established as a sub-sample of the sample of those responding to the Labour Force Survey during the June - November 2018 period.
Concepts & Classifications - Aggregation & consolidation	Data related to the "agreement or acceptability of at least one stereotype" refer to people who agreed strongly or somewhat with at least one of the considered stereotypes/behaviours. The rest of people did not answered or disagreed.

Dataset about victims: **D4**

Title of the section	**Victims turning to 1522 (anti-violence and stalking number)**
Data Sources	The data source is the national helpline 1522, funded by the Department for Equal Opportunity at the Presidency of the Council of Ministers to combat gender-based violence and stalking. Data are collected and stored by the telephone operators who provide a first response to the needs of users addressing the service. Data extracted from the archive of the telephone calls report socio-demographic information, the complaints, the personal consequences and on their children of the victims of violence.
Data characteristics	Data show information as recorded by the 1522 helpline operators. Missing values are due to the large number of interrupted telephone calls or, where applicable, due to missing answers by the users and victims.

3.1 Original/source datasets

- **D1_POPULATION**: This dataset contains information on the resident municipal population disaggregated by age, sex, and marital status. It was first produced in 2018 and is published by Istituto Centrale di Statistica (Italy). The data is licensed under CC BY 3.0 and can be accessed here.
- **D2_OPINIONS_VIOLENCE_GEOAREAS**: This dataset provides opinions about gender roles and violence against women, with a focus on geographical areas. It was first produced in 2018 and is published by Istituto Centrale di Statistica (Italy). The data is licensed under CC BY 3.0 and can be accessed here.
- **D3_OPINIONS_PARTNER_GEOAREAS**: This dataset covers the acceptability of intimate partner violence, disaggregated by geographical areas. First produced in 2018, it is published by Istituto Centrale di Statistica (Italy) under CC BY 3.0 and can be accessed here.
- **D4_VICTIMS**: This dataset includes information on victims who contacted the anti-violence and stalking number, disaggregated by sex, age class, and geographical areas. It was first

produced in 2018 and is published by Istituto Centrale di Statistica (Italy). The data is licensed under CC BY 3.0 and can be accessed here.

- **D5_VIOLENCE_CAUSES**: This dataset explores the causes of violence against women as perceived by adults, disaggregated by geographical areas. It was first produced in 2018 and is published by Istituto Centrale di Statistica (Italy) under CC BY 3.0 and can be accessed here.

3.2 Mashup datasets

- **MD6_VICTIMS_CAUSES_INDICATION**: This dataset analyzes violence-related data across Italian regions, including territory, causes of violence, corresponding values, and victim rates per 100,000 inhabitants. It was first produced in 2024, published by Chiara Martina and Lucrezia Pograri, and licensed under CC BY 4.0. The dataset can be accessed here.
- **MD5 K-MEANS CLUSTERS**: This dataset provides a cluster analysis of victim rates across Italian regions, categorized into Low, Medium, and High intensity clusters. It was first produced in 2024, published by Chiara Martina and Lucrezia Pograri, and licensed under CC BY 4.0. The dataset can be accessed here.
- **MD9_VICTIMS_GEO_TOTAL**: This dataset details total victim rates per 100,000 inhabitants across Italian regions, ranging from 7.90% (Molise) to 21.34% (Lazio). It was first produced in 2024, published by Chiara Martina and Lucrezia Pograri, and licensed under CC BY 4.0. The dataset can be accessed here.
- **MD7_VICTIMS_GEO_FEMALES**: This dataset presents female victim rates per 100,000 inhabitants across Italian regions, with Lazio recording the highest rate (20.7%) and Valle d'Aosta and Molise the lowest (7.16% and 7.57%, respectively). First produced in 2024, published by Chiara Martina and Lucrezia Pograri, and licensed under CC BY 4.0, it can be accessed here.
- **MD8_VICTIMS_GEO_MALES**: This dataset details male victim rates per 100,000 inhabitants across Italian regions, with Lazio reporting the highest rate (0.64%) and Calabria the lowest (0.16%). First produced in 2024, published by Chiara Martina and Lucrezia Pograri, and licensed under CC BY 4.0, it can be accessed here.

3.4 Preprocessing of data

The source datasets were processed using KNIME software, a platform for data analytics and workflow automation.

Various operations were carried out to clean, preprocess, and mash up the data, ensuring that it was structured and ready for analysis. These operations included data cleansing steps such as handling missing values, standardizing formats, and removing duplicates, as well as combining multiple data sources to create the mashup dataset. The entire workflow, which outlines each step of the data transformation process, is available for download on the project's website, providing full transparency and enabling replication of the analysis.

D1 POPULATION DATASET:

Since the census refers to the population "up to the 1st of January" of the selected year, when selecting the year variable in our dataset we did not choose directly 2018 (our time span of interest), but 2019.

D4 VICTIMS DATASET:

In this source dataset, we encountered an issue: victims value data for the Trentino-Alto Adige region were missing. To estimate the number of victims of gender-based violence in Trentino, for males, females and gender total, we used the ratio of population to victims observed in Friuli-Venezia Giulia, a region similar to Trentino in terms of population density.

From Friuli-Venezia Giulia, we know that with a population of 1.210.414, there were 132 victims for gender total. This allows us to calculate a victim-per-inhabitant rate by dividing the number of victims by the total population, resulting in approximately 0.0001523 victims per inhabitant.

Once this rate was calculated, we applied it to Trentino, which has a population of 1.074.034. By multiplying the victim rate of Friuli-Venezia Giulia by Trentino's population, we estimated approximately 116 victims for the Trentino region. The same approach has been adopted for male and female victim values, and for male victims values for Basilicata (also missing), which has been compared to Molise as population density.

$$Victims\ rate\ (per\ region)\ = \frac{Victims\ Value}{Population\ value}$$

$$Missing\ victims\ value\ estimation = victims\ rate \times population\ value\ (of\ region\ with\ missing\ data)$$

4. Analyses of the datasets

4.1 Quality analysis

In line with the **National Guidelines** ("Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico"), developed under the Data & Analytics Framework project by AgID and the Digital Transformation Team, we conducted a thorough quality assessment of our datasets to ensure their reliability and suitability for their intended purposes. Specifically, there are four main factors to look for when analysing data quality:

- **Accuracy (syntactic and semantic)**: Verifying that the data and its attributes accurately reflect the real-world values or events they represent.

- **Coherence**: Ensuring that the data is consistent and free from contradictions when compared to other related datasets within the administrative context.
- **Completeness**: Assessing whether the datasets provide exhaustive values and fully account for all related entities (sources) that contribute to the defined procedures.
- **Timeliness (or promptness of updating)**: Confirming that the data is up to date and corresponds to the relevant timeframes for the associated processes.

The results of this analysis are summarized in a table highlighting the overall quality of each dataset and identifying any areas requiring improvement.

ID Dataset	Accuracy	Coherence	Completeness	Timeliness
D1 - Population 2018	✓	✓	✓	✓
D2 - Opinions about sexual violence	✓	✓	✓	✓
D3 - Acceptability of intimate partner violence	✓	✓	✓	✓
D4 - Victims turning to 1522	✓	✓	*	✓
D5 - Indication of some causes of intimate partner violence	✓	✓	✓	✓

Figure 1: Quality analysis.

Limitations and Potential Avenues for Further Analysis

Addressing the following areas of potential further analysis could enhance the scope of the research and provide actionable insights for policymakers, advocacy groups, and community leaders working to combat violence against women in Italy.

- **Temporal limitation and impact of legislation:** The analysis of violence victims in the project is confined to data from 2018, due to the lack of more recent opinion surveys. Additionally, statistics on violence against women in Italy are only available up to 2022, with no new data published since then. Investigating the temporal development of these trends would provide valuable insights into the long-term effectiveness of intervention programs, societal changes, and the impact of new legislation. Further research could also explore how various laws aimed at combating violence against women, such as the introduction of restraining orders or stricter penalties for perpetrators, have influenced the prevalence of violence over time. A longitudinal approach could assess whether shifts in societal attitudes and legal frameworks are effectively reducing or mitigating violence, and a comparative analysis of regions that have implemented specific policies versus those that have not could offer a more nuanced understanding of the relationship between legal interventions and actual outcomes.
- **Data structure limitation:** The analysis faced constraints due to the structure of the original Istat datasets, which prevented the integration of demographic group analysis with geographical area analysis. Exploring this intersection would offer a deeper understanding

of the regional dynamics of violence against women in Italy. One potential avenue for further analysis could involve restructuring the datasets or employing advanced statistical methods, such as multivariate analysis or geographic information systems (GIS), to identify patterns and correlations between regional factors (e.g., socio-economic status, education levels, urbanization) and the incidence of violence. Additionally, the integration of qualitative data, such as interviews or surveys, could enrich the quantitative findings and offer a more comprehensive picture of the issue.

- **Socio-cultural influences and regional differences:** It would be beneficial to examine how cultural and social factors, such as religious beliefs, historical gender roles, and economic conditions, influence the prevalence and perception of violence against women across different regions. A deeper dive into the cultural norms that perpetuate victim-blaming or gender-based violence in specific areas could reveal critical insights. Regional studies comparing urban and rural environments might uncover disparities in access to support services, reporting rates, and societal attitudes towards victims.
- **Intersectionality of gender-based violence:** Future research could further investigate the intersectionality of violence against women, particularly how factors such as race, class, and sexual orientation interact with gender to shape experiences of violence. Exploring how marginalized groups, such as immigrant women or LGBTQ+ individuals, experience violence differently could provide a more nuanced understanding of the issue and inform targeted intervention strategies.

4.2 Legal analysis

The legal analysis of the source datasets is essential to ensure the **long-term sustainability** of the production process and of the publication of datasets and to guarantee a of both the data production process and the publication of datasets, while also guaranteeing a balanced service that aligns with public responsibilities and respects **individual rights.**

This analysis was carried out using a reference checklist consisting of a series of binary questions regarding the topics of **privacy issues**, **IPR policy**, **licenses**, **limitations on public access**, **economic conditions**, and **temporal aspects**.

A key aspect of the legal analysis is determining the appropriate publication license for the newly created mashup datasets. This decision must align with the licenses of the source datasets, which in our case, were all released under the **CC BY 3.0**. To guide this process, we utilized the Licensing Assistant tool provided by the European Commission. After evaluating the options, we opted to publish all nine mashup datasets under the CC BY 4.0 license.

The table below summarizes the original licenses of the source datasets, and the final publication license applied to the mashup datasets:

ID	Dataset	Original licenses	Final license
MD6	Victims causes indication	CC BY 3.0, CC BY 3.0, CC BY 3.0	CC BY 4.0
MD7	Victims geo fermales	CC BY 3.0, CC BY 3.0, CC BY 3.0	CC BY 4.0
MD8	Victims geo males	CC BY 3.0, CC BY 3.0, CC BY 3.0	CC BY 4.0
MD9	Victims geo total	CC BY 3.0, CC BY 3.0, CC BY 3.0	CC BY 4.0

Figure 2: publication license applied to the mashup datasets.

See more in the section "**Legal Analysis**" of the project's website.

4.3 Ethical analysis

For the ethical analysis of our project's data, we applied the **Data Ethics Principles and Guidelines** and the **Odi Project's** detailed **framework** for assessing the ethical aspects of our data processing.

Both the source and mashup datasets for our project are exclusively derived from the Italian National Institute of Statistics (ISTAT). Therefore, we first focused on evaluating the fairness of ISTAT's data collection and management practices. Following that, we established clear ethical guidelines to ensure responsible handling of the datasets throughout our project's lifecycle.

Data Ethics Principles

The "Open Voices" project aims to analyse the socio-cultural dynamics that influence violence against women in Italy, therefore the ethical approach in data management is fundamental to ensure that the rights of victims are respected, that no stereotypes or discrimination are fostered, and that the analysis contributes positively to raising awareness on such a sensitive issue.

Human being at the center: ISTAT's policy is deeply aligned with both ethical standards and legislative principles. The organization prioritizes the dissemination of statistical information to promote awareness of Italy's social and economic conditions. It also strives to enhance public decision-making by providing clear, accessible statistical data. In addition, ISTAT conducts research to continually refine statistical methodologies and improve Italy's statistical literacy.

- **Equality:** In the context of the "Open Voices" project, the main objective is to analyze the socio-cultural dynamics that influence violence against women in Italy, Therefore, the collection of data concerns very sensitive issues such as gender violence and its perception within society. ISTAT data on key equality issues, such as gender discrimination and domestic violence.
- **Transparency:** ISTAT ensures transparency in data management by providing comprehensive documentation. This documentation covers the data collection methods, clarifies the use of specific terms and definitions, and outlines policies and licenses that safeguard against misinterpretation of the data.

- **Accountability:** ISTAT's quality assurance procedures align with European frameworks, specifically the European Statistics Code of Practice. This adherence strengthens both accountability and governance within the national statistical system and aligns it with European standards.
- **Individual data protection:** ISTAT ensures that its datasets are anonymized in compliance with legal requirements. The organization's practices adhere to strict confidentiality standards, ensuring that the privacy of respondents is always respected. Their approach to handling sensitive data complies with European data protection laws (e.g., Regulation (EU) 2016/679 and Legislative Decree No. 322/1989).

Ethical concerns and their management

Although ISTAT adheres to ethical principles in data collection and management, the team has paid particular attention to the ethical management of source information, given the great sensitivity of the content dealt with in our project, which concerns gender-based violence. Data on sexual violence, the acceptability of partner violence and access to support services (such as 1522) are sensitive issues and ethical concerns in their treatment have been addressed in the following ways:

- **Data Integrity and Privacy:** To ensure data integrity and privacy, the values from the source datasets were aggregated and presented as percentages, avoiding any correlation with real individuals.
- **Protection of Vulnerable Groups:** Certain sensitive data were intentionally omitted to avoid the risk of discriminatory behavior. The project aimed to protect women experiencing violence, minimizing the risk of stigmatization.
- **Avoiding Generalizations and Misinterpretations:** The objective of the project was to identify potential patterns in the dynamics of gender-based violence, not to make inferences or generalizations. In our results and conclusions documentation, we emphasize that any observed patterns in the data should not be generalized due to inconsistencies in the data and the absence of other potentially relevant socio-economic factors.

All relevant documentation regarding the data processing for the creation of mashup datasets and visualizations is provided in the GitHub repository of the project.

4.4 Technical analysis

All source datasets have been assessed following the **metadata model established by AGID**, which categorizes metadata quality into four levels. This classification is based on two key factors: the strength of the data-metadata relationship and the level of detail provided.

Note: Further details and reconstructed metadata for the source datasets are available in the metadata analysis table below.

ID	Provenience	Format	Metadata	URI	License
D1	I.Stat	.csv	Level 4: An SDMX structured file is downloadable with a strong data-metadata bond and a datum-level detail of description. They are machine readable.	2018Population	CC BY 3.0
D2	IstatData	.csv	Level 4: An SDMX-structured file is available for download, featuring a strong connection between data and metadata, with detailed descriptions at the datum level. These files are machine-readable. Level 2: Additional metadata, offering clear information about sources and methodologies, is provided on a separate webpage, accessible via a sidebar menu.	OpinionsViolenceGeoAreas	CC BY 3.0
D3	IstatData	.csv	Level 4: An SDMX-structured file is available for download, featuring a strong connection between data and metadata, with detailed descriptions at the datum level. These files are machine-readable. Level 2: Additional metadata, offering clear information about sources and methodologies, is provided on a separate webpage, accessible via a sidebar menu. >webpage, accessible through a sidebar menu	OpinionsPartnerGeoAreas	CC BY 3.0
D4	I.Stat	.csv	Level 4: An SDMX structured file is downloadable with a strong data-metadata bond and a datum-level detail of description. They are machine readable.	Victims	CC BY 3.0
D5	I.Stat	.csv	Level 4: An SDMX structured file is downloadable with a strong data-metadata bond and a datum-level detail of description. They are machine readable. >webpage, accessible through a sidebar menu	ViolenceCauses	CC BY 3.0

Figure 3: metadata analysis table.

RDF Metadata Assertion of the Datasets

All generated mashup datasets have been meticulously documented using metadata, adhering to the guidelines outlined in the **DCAT-AP_IT** standard. This choice aligns with the **Italian Agency for Digitalization (AGID)** directives for the valorization of public information heritage.

Given that these datasets represent data of significant national interest and are derived from Istat, an Italian public research institution, the adoption of **DCAT-AP_IT** (2016), was deemed most appropriate.
Although this national standard is based on the earlier **DCAT v1.0** and introduces stricter constraints compared to the more flexible DCAT-AP 2.0, it was chosen because an increasing number of Italian Public Administrations are aligning with DCAT-AP_IT, ensuring better compatibility within the national ecosystem. Furthermore, adhering to DCAT-AP_IT allows our datasets to align with detailed national guidelines, thereby facilitating seamless **interoperability and harmonization** with other public data resources in Italy.

Semantic Enrichment of Datasets

The metadata for the source datasets were primarily derived from the original data sources. When metadata was incomplete or unavailable, additional information was inferred and

supplemented following the same principles applied to the mashup datasets. For instance, themes were assigned to source datasets based on recognized European Authority standards. To enhance the semantic description of datasets, we have utilized several key ontologies, including **DCAT, DCTERMS, PROV, FOAF, ADMS, SKOS, CC**, and **DCAT-AP IT**. These ontologies provide a structured framework for describing datasets and their metadata, ensuring interoperability and adherence to **Linked Open Data (LOD) standards**. This approach facilitates the discoverability, understandability, and reuse of datasets across diverse contexts.

The **DCAT (Data Catalog Vocabulary) ontology** plays a central role, serving to describe datasets and data catalogs and publish them on the Web. Key properties such as dcat:dataset, associate datasets with a catalog, while dcat:theme categorizes datasets by their topics, and dcat:distribution specifies the available data formats. For example, datasets can be linked to themes like population statistics or societal issues using dcat:theme. Learn more about DCAT Version 3 at W3C DCAT.

DCTERMS (Dublin Core Terms) is used to describe general metadata, including dcterms:title for dataset titles and descriptions, and dcterms:accessRights to indicate accessibility. This inclusion ensures the use of a widely recognized standard for metadata description. More information about DCTERMS can be found at Dublin Core Terms.

ADMS (Asset Description Metadata Schema) is included to complement DCAT by describing assets like datasets and services, particularly in government and public administration contexts. ADMS plays an important role in the asset management landscape. For further details, consult the ADMS Specification.

CC (Creative Commons) provides a vocabulary for licensing datasets, such as cc:license, ensuring clear and standardized license attribution. Visit Creative Commons specifications for further insights.

Finally, **DCAT-AP IT (DCAT Application Profile for Italy)** is referenced through dcatapit: to align with Italian government open data standards. It extends DCAT to meet specific national requirements. For more information, check out DCAT-AP IT.

FAIR principles

During the project development, we aimed to align our efforts with the **FAIR principles** established by the **GO FAIR Initiative**. These principles, developed by a consortium of scientists and organizations, provide **guidelines** to ensure digital assets are **Findability**, **Accessibility**, **Interoperability**, **Reusability**, with a strong emphasis on machine-actionability.

See more in the section "**Technical Analysis**" of the project's website.

5. Visualizations and results

Various graphic representations of the data facilitate a deeper exploration of the project's topic, allowing for the analysis of potential correlations between key factors across different regions. Together, these visualizations present the data in an engaging and accessible manner, supporting a comprehensive analysis of the topic.

The visualizations have been created using two libraries: **Leaflet.js** for the map and **Plotly.js** for the bar charts and bubble chart. Below are the types of visualizations used:

- **Choropleth Map:** Choropleth map serves as the most effective method to visually display regional variations in our data, highlighting trends and correlations across geographical areas. The victims' values are presented as numbers, while regional clusters are generated through data normalization by population, applying the k-means algorithm to the rate of victims per 100,000 inhabitants. This map allows us to observe general trends in population and violence victimization rates. Higher rates are seen in regions like Lazio, Campania, and Abruzzo, while the lowest rates, often due to lower population density, are found in Molise. Other regions, such as Valle d'Aosta, Trentino Alto Adige, and Molise, report lower victimization rates, with most regions falling in between these extremes.

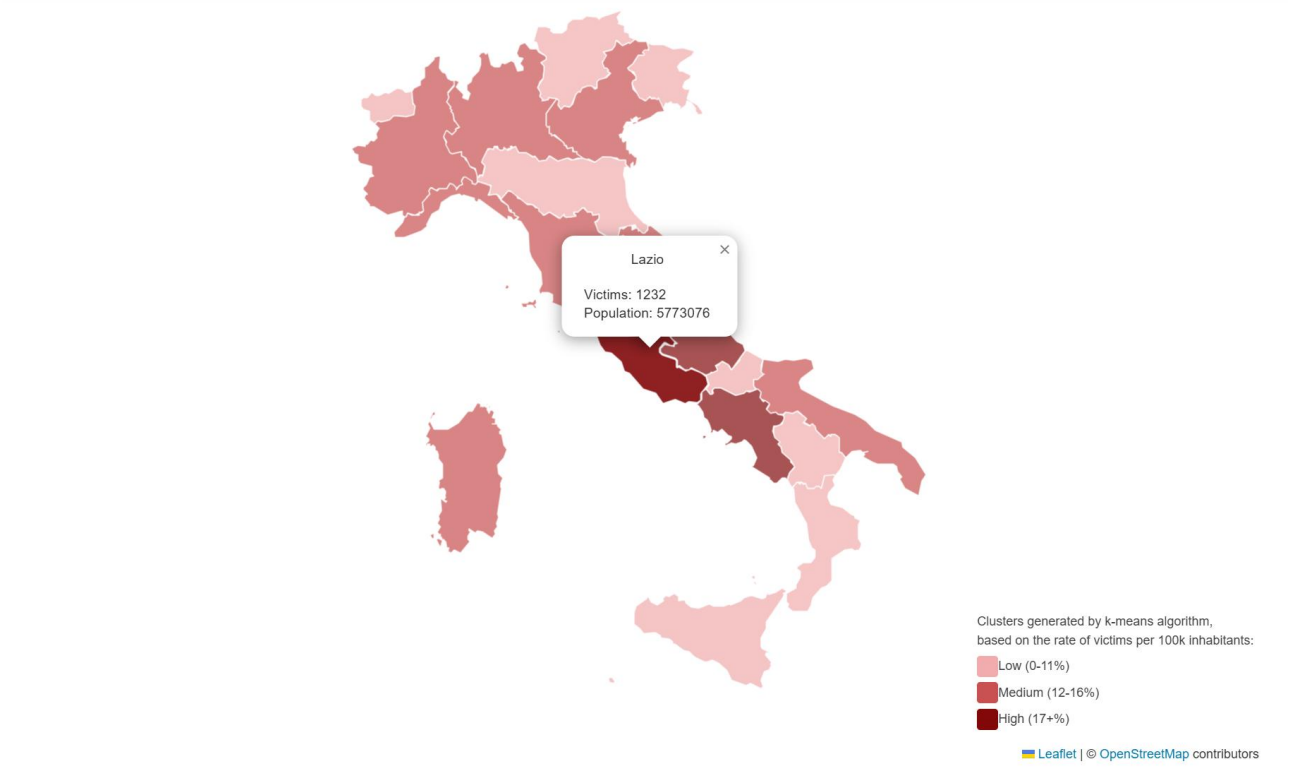

Figure 4: choropleth map of Italian regions.

- **Bar Charts:** Bar charts provide an immediate, clear comparison of values for each of the variables, offering a straightforward way to assess differences across regions.

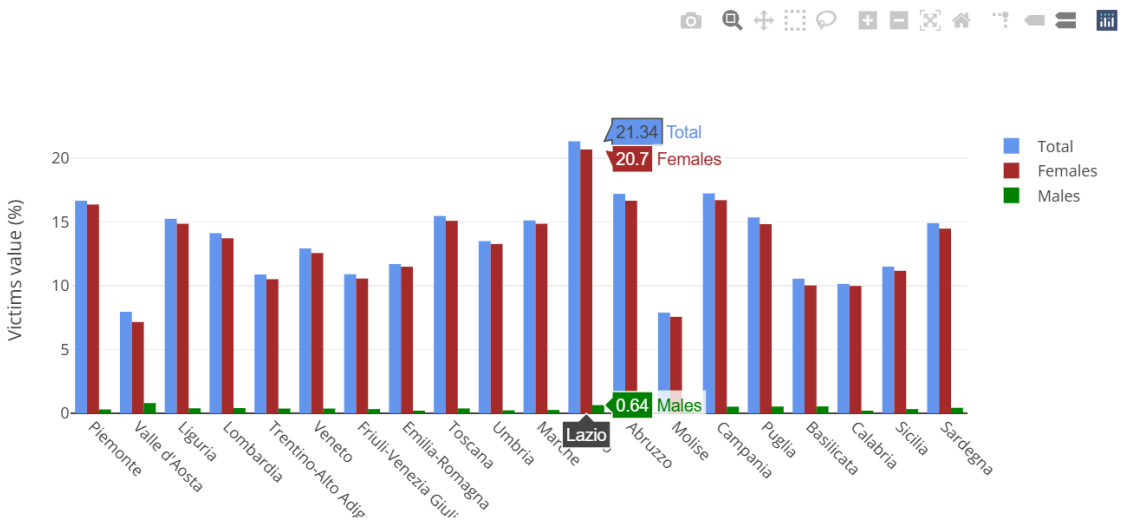

Figure 5: bar chart showing victims rate per 100k inhabitants across regions.

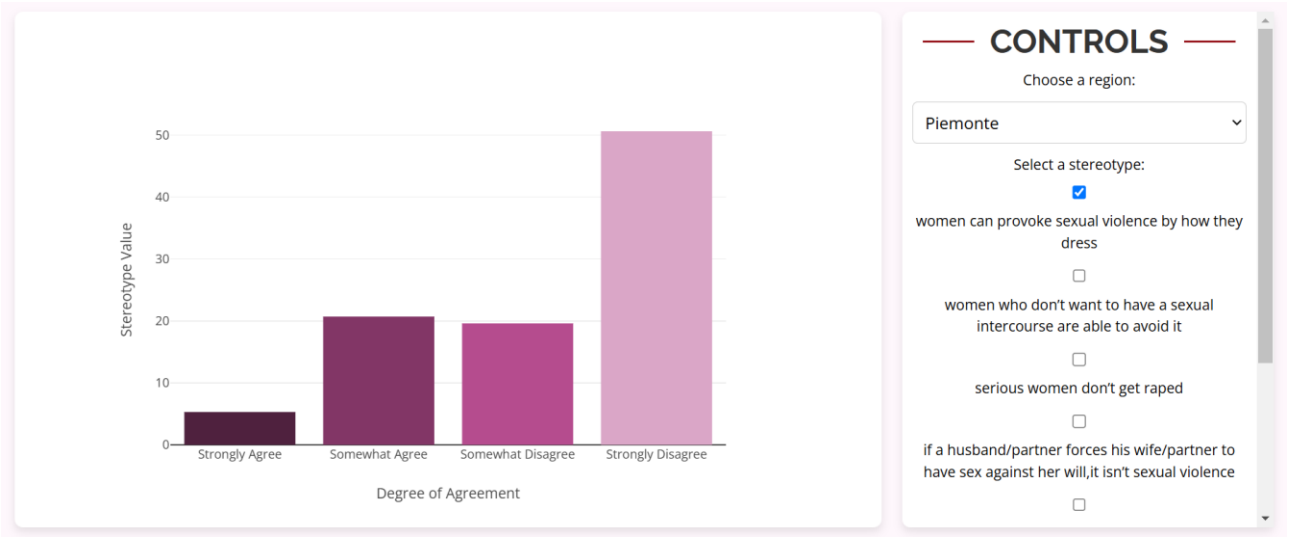

Figure 6: Opinions about sexual violence.

This chart explores the point of view of Italian adults on gender roles and sexual violence, broken down by region. The bar graph shows the degree of agreement with various stereotypes. The data show how different regions of Italy respond to these beliefs, highlighting cultural and social variations in perceptions of violence against women.

Results:

Regional differences in perceptions of sexual violence in Italy are evident. This thesis highlights the wide regional differences in perceptions of sexual violence in Italy. Northern regions, such as Tuscany and Lombardy, show a greater awareness and strong disagreement about misconceptions about sexual violence, as the idea that women can provoke violence with their clothing or that violence within marriage is not considered to be such.

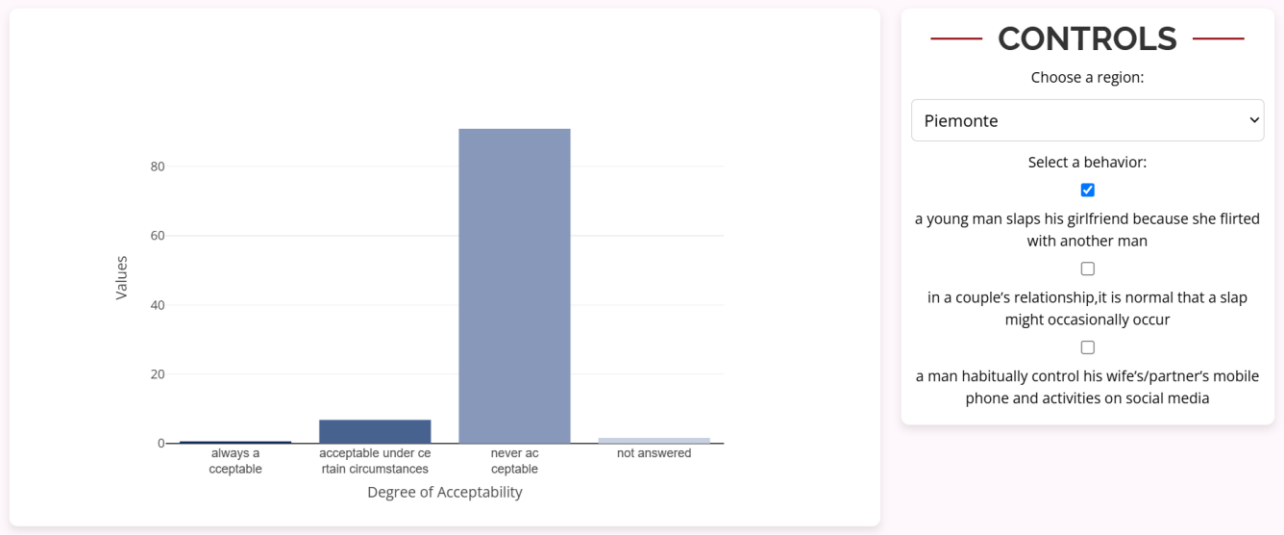

Figure 7: Acceptability of intimate partner violence.

This chart examines the acceptability of intimate partner violence in different regions of Italy. The bar graph shows the acceptability of various violent behaviors. The data shows the percentage of people who consider these behaviors acceptable or not, broken down by region, contributing to an understanding of cultural differences in opinion about domestic violence.

Results:

In contrast, in regions such as Calabria, Puglia or Liguria, more permissive and stereotyped views persist, with significant percentages of people justifying or minimizing sexual violence. These findings suggest the importance of targeted educational interventions to raise public awareness and reduce the gap in regional perceptions, in order to promote a culture of respect and awareness about gender violence. The differences between regions indicate a significant variation in the perception of acceptability of violence. Regions such as Lazio, Sardinia and Piedmont show a clear opposition to couple violence, while in regions like Abruzzo and Basilicata there are still worrying margins of tolerance. In some regions, conditional or partial acceptance of these behaviors suggests that gender stereotypes and cultural norms permissive towards control or violence may persist.

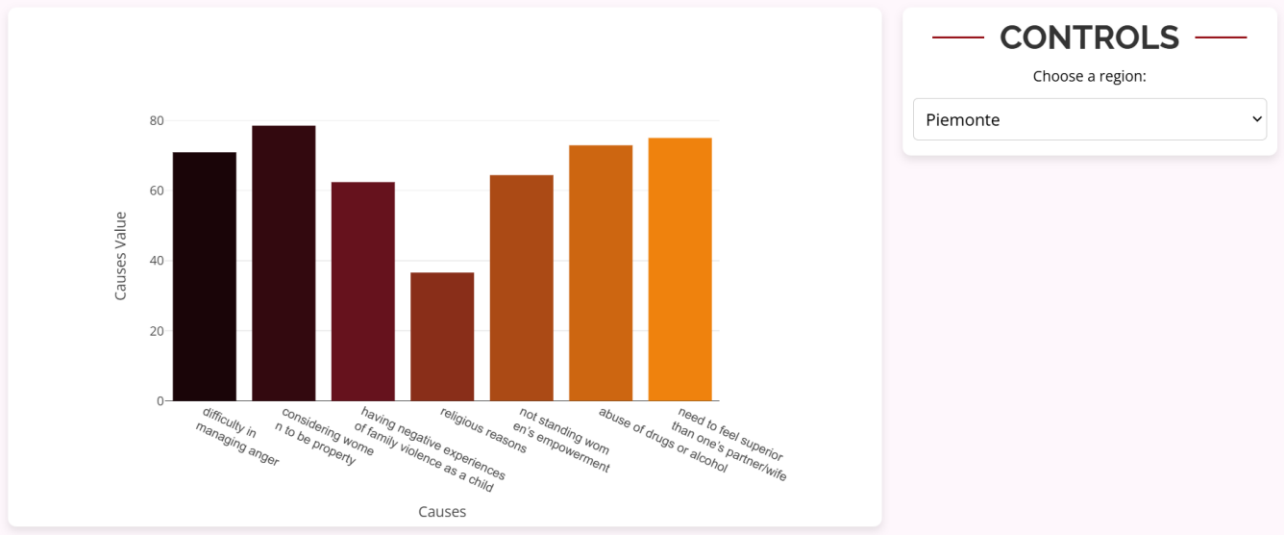

Figure 8: Indication of some causes of intimate partner violence.

This chart explores the perceived causes of violence in intimate relationships in different regions of Italy. The data show how often each factor is identified as a cause of violence, contributing to a deeper understanding of the cultural and social roots of this phenomenon.

Results:

The data show a marked regional disparity in the factors associated with gender-based violence in Italy. Regions such as Friuli Venezia Giulia and Emilia Romagna emerge with the highest rates in several categories, such as women's perception as property and the need for superiority towards their partners, indicating a persistent rooting of gender stereotypes. Factors such as drug or alcohol abuse, prevalent in Sardinia, and childhood experiences of domestic violence, widespread in Basilicata, underline the importance of addressing the root causes of violence, including addiction and intergenerational transmission. Although some regions, such as Piedmont and Sardinia, show a clear opposition to violence, the presence of tolerant attitudes in other areas highlights the need for targeted interventions. Finally, it is crucial to promote cultural change through educational campaigns, strengthen psychological support and intervene on addictions to reduce regional disparities and prevent gender violence in a sustainable and effective way.

- **Bubble Chart:** The bubble chart complements the analysis by adding a dynamic layer of visualization, further aiding in understanding the relationships between variables. This chart compares causes of violence across Italian regions, highlighting their relationship with victimization rates. Key factors include anger management difficulties, viewing women as property, childhood family violence, religious beliefs, opposition to women's empowerment, substance abuse, and the need for dominance in relationships. The analysis reveals a strong correlation between the belief that women are property and high victimization rates, particularly in Lombardia and Friuli-Venezia Giulia. Substance abuse also shows a strong link to violence, especially in Trentino Alto Adige and Friuli-Venezia

Giulia. Regions like Lazio have a higher victim count, indicating the widespread impact of these causes. The intensity of each cause varies by region, with anger management issues being prevalent but less impactful in areas like Veneto and Marche. The need for dominance remains a consistent factor across regions. Overall, the data underscores the complexity of violence causes, suggesting that regional interventions should address specific contributing factors.

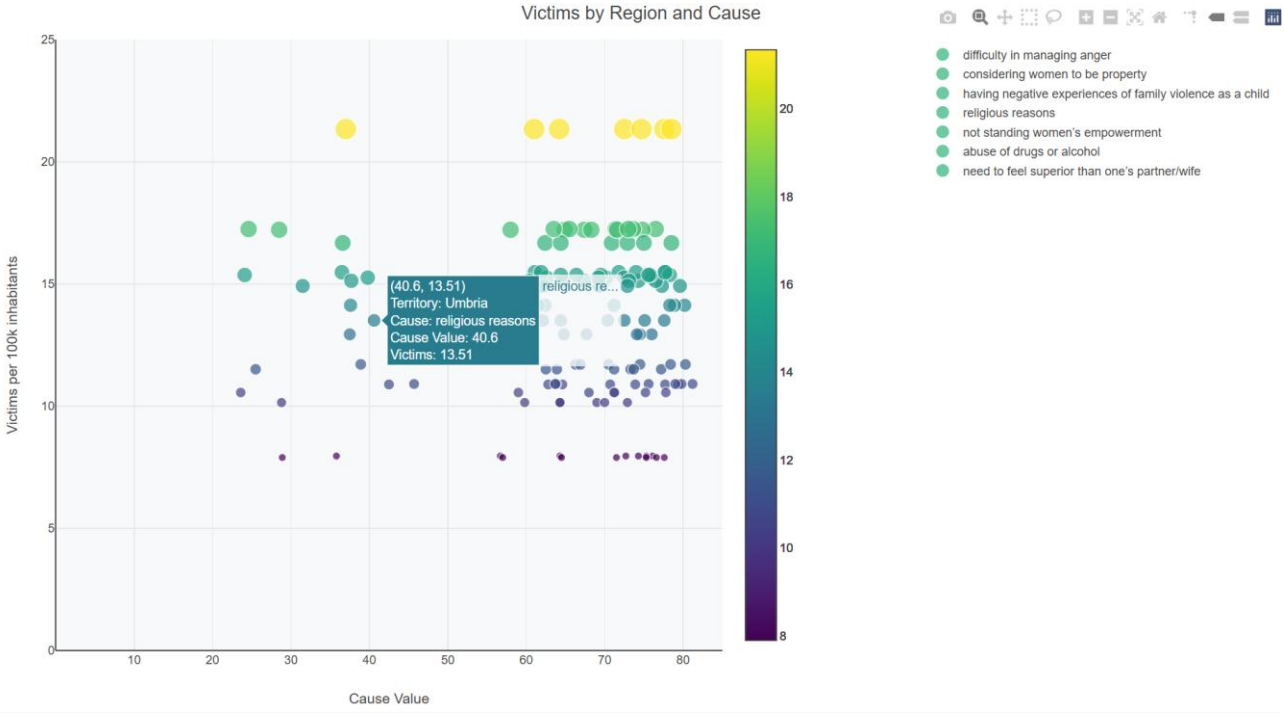

Figure 9: victims by regions and cause of violence.

6. Sustainability of the project

The source datasets used for this project are provided exclusively by the **Italian National Institute of Statistics (Istat)**, which maintains them in its various databases. However, due to the ongoing transition of Istat's content to IstatData, the URIs provided in this project may eventually become obsolete.

Open Voices is the **final project** developed for the *Open Access and Digital Ethics* course (a.y. 2024/2025) within the Digital Humanities and Digital Knowledge Master's Degree (University of Bologna). As such, it is **not actively maintained** and **will not be updated** in the future.

7. RDF metadata assertion

All generated mashup datasets have been meticulously documented using metadata, adhering to the guidelines outlined in the DCAT-AP_IT standard. This choice aligns with the **Italian Agency for Digitalization (AGID)** directives for the valorization of public information heritage.

Given that these datasets represent data of significant national interest and are derived from Istat, an Italian public research institution, the adoption of **DCAT-AP_IT** (2016), was deemed most appropriate. Although this national standard is based on the earlier **DCAT v1.0** and introduces stricter constraints compared to the more flexible DCAT-AP 3.0, it was chosen because an increasing number of Italian Public Administrations are aligning with DCAT-AP_IT, ensuring better compatibility within the national ecosystem. Furthermore, adhering to DCAT-AP_IT allows our datasets to align with detailed national guidelines, thereby facilitating seamless **interoperability and harmonization** with other public data resources in Italy.

The metadata for the source datasets were primarily derived from the original data sources. When metadata was incomplete or unavailable, additional information was inferred and supplemented following the same principles applied to the mashup datasets. For instance, themes were assigned to source datasets based on recognized European Authority standards. To enhance the semantic description of datasets, we have utilized several key ontologies, including **DCAT, DCTERMS, PROV, FOAF, ADMS, SKOS, CC**, and **DCAT-AP IT**. These ontologies provide a structured framework for describing datasets and their metadata, ensuring interoperability and adherence to **Linked Open Data (LOD) standards**. This approach facilitates the discoverability, understandability, and reuse of datasets across diverse contexts.

The **DCAT (Data Catalog Vocabulary) ontology** plays a central role, serving to describe datasets and data catalogs and publish them on the Web. Key properties such as dcat:dataset, associate datasets with a catalog, while dcat:theme categorizes datasets by their topics, and dcat:distribution specifies the available data formats. For example, datasets can be linked to themes like population statistics or societal issues using dcat:theme. Learn more about DCAT Version 3 at W3C DCAT.

DCTERMS (Dublin Core Terms) is used to describe general metadata, including dcterms:title for dataset titles and descriptions, and dcterms:accessRights to indicate accessibility. This inclusion ensures the use of a widely recognized standard for metadata description. More information about DCTERMS can be found at Dublin Core Terms.

ADMS (Asset Description Metadata Schema) is included to complement DCAT by describing assets like datasets and services, particularly in government and public administration contexts. ADMS plays an important role in the asset management landscape. For further details, consult the ADMS Specification.

CC (Creative Commons) provides a vocabulary for licensing datasets, such as cc:license, ensuring clear and standardized license attribution. Visit Creative Commons specifications for further insights.

Finally, **DCAT-AP IT (DCAT Application Profile for Italy)** is referenced through dcatapit: to align with Italian government open data standards. It extends DCAT to meet specific national requirements. For more information, check out DCAT-AP IT.

8. Team and statement of responsibility

The project has been developed by:

- Chiara Martina – chiara.martina@studio.unibo.it
- Lucrezia Pograri – lucrezia.pograri@studio.unibo.it

9. Licenses and credits

Images and icons

All the icons are taken from ICONS8. They are available for unrestricted commercial and noncommercial use without permission or fee (CC0)

Image on homepage by the freelance illustrator Hanna Barczyk (c) 2024

Web template

The website of the project is built on the HTML5 template "Vesperr" by BootstrapMade and released under MIT

Source Datasets

Creative Commons Attribution 3.0 Unported (CC BY 3.0)

Mashup Datasets

Creative Commons Attribution 4.0 International (CC BY 4.0)

Softwares used

Leaflet.js: Copyright (c) 2010-2023, Volodymyr Agafonkin Copyright (c) 2010-2011, CloudMade - All rights reserved (BSD 2-Clause "Simplified" License)

Plotly.js: Copyright (c) 2021 Plotly, Inc - All rights reserved (MIT License)

KNIME: Copyright (c) 2007 Free Software Foundation, Inc. (General Public License (GPL), Version 3)

Index

Index of Figures