

# **PROJETO DA DISCIPLINA PROGRAMAÇÃO I**

## **Violência Doméstica em Pernambuco (2015-2022)**

**PROFESSORES:** Filipe Calegário  
Sergio Soares

**ALUNOS:** Gabriel Arnaud (gamf)  
Hítalo Nascimento (hban)  
Ingrid Freire (iafsl)  
Otávio Augusto (oacn)  
Rayza Laurindo (rld)  
Renata Santana (ras13)

Recife, novembro de 2022



<b>1. A ORGANIZAÇÃO DO CÓDIGO</b>	<b>3</b>
• Estrutura de diretórios	3
• Instalação das dependências	3
• Pré processamento dos dados	4
• Análise dos dados	4
• Resumo dos resultados	4
<b>2. FERRAMENTAS UTILIZADAS</b>	<b>5</b>
• Para construção do código	5
• Para análise dos dados	5
• Para organização da equipe	6
<b>3. A DIVISÃO DE TRABALHO DENTRO DO GRUPO</b>	<b>6</b>
<b>ANÁLISES</b>	<b>6</b>
<b>4. CONCEITOS QUE FORAM APRESENTADOS DURANTE A DISCIPLINA E UTILIZADOS NO PROJETO</b>	<b>7</b>
• Algoritmos e conceitos fundamentais em Python	7
• Operadores lógicos e comandos condicionais	7
• Estrutura de dados	7
• Funções	8
• Módulos e Pacotes	8
<b>5. OS DESAFIOS E ERROS ENFRENTADOS NO DECORRER DO PROJETO E AS LIÇÕES APRENDIDAS</b>	<b>9</b>
• Qual foi o maior erro cometido durante o projeto? Como vocês lidaram com ele?	9
• Qual foi o maior desafio enfrentado durante o projeto? Como vocês lidaram com ele?	9
• Quais as lições aprendidas durante o projeto?	10



Esse trabalho de visualização de dados, cujo título é Violência Doméstica em Pernambuco (PE), foi realizado utilizando os dados brutos disponibilizados pela Secretaria de Defesa Social (SDS) do estado de Pernambuco sobre violência doméstica, apurados entre janeiro de 2015 a agosto de 2022. Nesse projeto tivemos como objetivo explorar a visualização dos dados de modo analítico, onde dividimos em subconjuntos os dados de interesse para melhor análise e compreensão.

Para a visualização dos dados desse dataset, foi necessário realizar o tratamento dos dados, processar e filtrar as informações mais importante desse tema.

Link do repositório do projeto completo: [Repositório GitHub](#)

Link com os dados usados nas análises: [Dados Brutos](#)

## 1. A ORGANIZAÇÃO DO CÓDIGO

- Estrutura de diretórios

```
|— Data
|   |— MICRODADOS_DE_VIOLENCIA_DOMESTICA_JAN_2015_A_AGO_2022.xlsx
|— README.md
|— analyses
|   |— teste_violencia.ipynb
```

- Instalação das dependências

Logo no início do arquivo **teste\_violencia.ipynb** encontramos os comandos para instalação das bibliotecas usados durante o projeto, essas libs são obrigatórias para que o projeto funcione corretamente. Além disto, no próprio README.md do repositório do GitHub encontramos as orientações para instalação e execução do projeto.



- Pré processamento dos dados

O passo seguinte a instalação e importação das bibliotecas é a realização do pré processamento dos dados do dataset escolhido. Nessa etapa, é feita uma breve verificação dos tipos colunares, uma análise dos valores colunares, uma verificação de valores faltantes (onde o valor equivale a Null) e por fim o tratamento destes dados analisados tais como a normalização dos dados de idade e criação de uma nova coluna intitulada “MÊS DO FATO” com o intuito de facilitar no estudo dos dados e geração dos gráficos.

- Análise dos dados

Nesta etapa o objetivo foi fazer uso dos dados pré processados na etapa anterior e realizar alguns estudos. Dentre estes:

- análise exploratória inicial;
- análise temporal;
- análise com foco onde a natureza dos casos de violência é *ameaça por violência*;
- análise dos casos de violência na região metropolitana;
- análise dos casos de violência na região do agreste;
- análise dos casos de violência na capital;

- Resumo dos resultados

Neste estudo foi identificada a predominância de casos com as seguintes características: sexo feminino, faixa etária de 35-64 anos, natureza “Ameaça por violência doméstica/familiar” e total de envolvidos 1. Ademais, o ano de 2019 apresentou maior número de ocorrências, seguido de 2020 e 2021, respectivamente. Quanto aos meses, o mês de janeiro apresentou maior incidência de casos, seguido dos meses de março e agosto, respectivamente. A região Metropolitana apresentou maior número de ocorrências, seguida da região Agreste. A cidade do Recife, compreendida nos dados como a região Capital, apresentou a nível de município o maior número de casos. Portanto, concluímos que a violência doméstica, infelizmente, atinge na grande maioria dos casos as mulheres, sendo



essa uma problemática com extrema urgência em solucionar a fim de tornar a sociedade mais justa e igualitária.

## 2. FERRAMENTAS UTILIZADAS

- Para construção do código
  - **Git e GitHub:** o Git foi essencial para ajudar na gestão do repositório e no versionamento do código enquanto a equipe trabalhou na construção de suas respectivas análises. Como repositório remoto, escolhemos trabalhar com o GitHub uma vez que a maioria da equipe já tinha familiaridade com a ferramenta.
  - **Visual Studio Code:** IDE escolhida pela equipe. Optamos por trabalhar com VS Code por se tratar de uma IDE de fácil aprendizagem e disponibilizar diversas extensões que facilitam muito o dia a dia do desenvolvimento.
  - **GitLens:** extensão do VS Code que facilita a interação do usuário com o Git e GitHub, através dela pode-se realizar commits, gerenciar o staging e fazer push das alterações feitas no código por meio de uma interface gráfica de fácil uso.
  - **GitGraph:** extensão do VS Code que também interage com o Git e GitHub, contudo seu propósito é exibir em gráfico de árvores as modificações realizadas no código bem como quem as fez. Facilitou para que a equipe pudesse acompanhar as modificações feitas no repositório remoto pelos outros desenvolvedores.
- Para análise dos dados
  - **Jupyter Notebook:** embora existesse a possibilidade de fazer todo o projeto da análise dos dados usando Python puro optamos por utilizar o Jupyter Notebook pela vantagem que este trás na exibição do código juntando tanto análise quanto resultado e exibindo estes em um fluxo lógico bem ilustrado.
  - **Pandas:** a biblioteca Pandas é um requisito essencial para manipulação e análise de dados com Python. Além de toda facilidade que ela proporciona, sua documentação está muito bem organizada, o que facilita muito a consulta dos colaboradores da equipe durante o desenvolvimento.
  - **Matplotlib:** a biblioteca Matplotlib foi escolhida para geração dos gráficos das análises. O fator decisivo para seu uso foi, além da facilidade de uso, a fácil integração com a biblioteca Pandas.

**UNIVERSIDADE FEDERAL DE PERNAMBUCO-UFPE**  
**CENTRO DE INFORMÁTICA**



- Para organização da equipe
  - **Scrum:** aproveitando dos checkpoints realizados nas aulas, encaixamos na nossa metodologia de trabalho um modelo adaptado de Scrum. Cada semana foi um ciclo de atividades.
    - **Sprint 1 03/10 - 09/10:** Leitura sobre a proposta do projeto da disciplina e determinação / especificação do projeto;
    - **Sprint 2 10/10 - 16/10:** Estudo sobre tratamento de dados e bibliotecas Pandas e Matplotlib, encontro de horizontalização de conhecimento sobre essas tecnologias e realização do pré processamento dos dados;
    - **Sprint 3 17/10 - 23/10:** Estudo sobre Git e GitHub, encontro de horizontalização de conhecimento sobre essas tecnologias e início do desenvolvimento das análises;
    - **Sprint 4 24/10 - 30/10:** Foco no desenvolvimento das análises e suporte entre os membros da equipe para tirar dúvidas;
    - **Sprint 5 31/10 - 06/11:** Finalização das análises e preparação do relatório e apresentação.
  - **Whatsapp:** A fim de concentrar toda a comunicação da equipe escolhemos o Whatsapp como ferramenta de comunicação assíncrona.
  - **Google Meet:** Para comunicação síncrona optamos por usar o Google Meet tanto para os encontros de horizontalização com equipe quanto para dar suporte e tirar dúvidas.
  - **Google Drive:** Todos os encontros de horizontalização foram gravados e os vídeos foram armazenados para uso da equipe em uma pasta compartilhada.

### 3. A DIVISÃO DE TRABALHO DENTRO DO GRUPO

ANÁLISES	
Gabriel	Pré-processamentos dos dados, Análise Temporal
Hítalo	Análise por “Natureza”
Ingrid	Análise da região Metropolitana
Otávio	Análise da região Agreste
Rayza	Análise da Capital



Renata	Análise por “Natureza”
--------	------------------------

#### **4. CONCEITOS QUE FORAM APRESENTADOS DURANTE A DISCIPLINA E UTILIZADOS NO PROJETO**

- Algoritmos e conceitos fundamentais em Python

Por se tratar de um Jupyter Notebook toda a arquitetura do projeto é um algoritmo. Em aula vimos que um algoritmo é um conjunto ordenado de passos executáveis, não ambíguos, definindo um processo que tem um término. Durante o Notebook é exatamente isso que ocorre. O algoritmo dita um processo de tratamento, criação de informação e retirada de gráficos de um dataset. Para tal foram usados, também, conceitos fundamentais da linguagem de programação Python nesse processo, tais como declaração de variáveis, atribuição e manipulação dos valores armazenados.

- Operadores lógicos e comandos condicionais

O uso de operadores lógicos e comandos condicionais é fundamental para que o código tenha o mínimo de poder de decisão. “O que fazer quando isso ou aquilo acontece?”. Desta maneira, o uso destes recursos foi indispensável.

```
if sexo == 'DESCONHECIDO' or sexo == 'NAO INFORMADO':  
    sexo = 'DESCONHECIDO/NAO INFORMADO'  
else:  
    sexo = sexo
```

- Estrutura de dados

Por termos utilizado a biblioteca Pandas, a principal estrutura de dados que usamos foi o DataFrame, se trata de uma estrutura de dados bidirecional semelhante a uma matriz. Com ela, o tratamento e manipulação de dados tabelares é extremamente facilitado.

**UNIVERSIDADE FEDERAL DE PERNAMBUCO-UFPE**  
**CENTRO DE INFORMÁTICA**



df								
	MUNICÍPIO DO FATO	REGIAO GEOGRÁFICA	NATUREZA	DATA DO FATO	ANO	SEXO	IDADE SENASP	TOTAL DE ENVOLVIDOS
0	AFOGADOS DA INGAZEIRA	SERTÃO	INJÚRIA POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2015-01-01	2015	FEMININO	6) 35-64	1
1	AFRANIO	SERTÃO	AMEAÇA POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2015-01-01	2015	FEMININO	5) 30-34	1
2	AFRANIO	SERTÃO	LESÃO CORPORAL POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2015-01-01	2015	FEMININO	4) 25-29	1
3	AGUAS BELAS	AGRESTE	ESTUPRO DE VULNERÁVEL POR VIOLÊNCIA DOMÉSTICA/...	2015-01-01	2015	FEMININO	2) 12-17	1
4	ALAGOINHA	AGRESTE	AMEAÇA POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2015-01-01	2015	FEMININO	6) 35-64	1
...	...	...	...	...	...	...	...	...
253698	SAO JOSE DO BELMONTE	SERTÃO	MAUS TRATOS POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2022-08-31	2022	FEMININO	(vazio)	1
253699	SAO JOSE DO EGITO	SERTÃO	AMEAÇA POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2022-08-31	2022	FEMININO	3) 18-24	1
253700	SURUBIM	AGRESTE	PERTURBAÇÃO DO SOSSEGO POR VIOLÊNCIA DOMÉSTICA...	2022-08-31	2022	FEMININO	6) 35-64	1
253701	VERDEJANTE	SERTÃO	AMEAÇA POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2022-08-31	2022	FEMININO	6) 35-64	1
253702	VERDEJANTE	SERTÃO	LESÃO CORPORAL POR VIOLÊNCIA DOMÉSTICA/FAMILIAR	2022-08-31	2022	FEMININO	3) 18-24	1

- Funções

Usadas para ajustar dados do dataset, o uso de funções foi indispensável. Uma vez que precisamos repetir o mesmo tipo de tratamento para todas as linhas da nossa tabela, a forma mais fácil de fazê-lo foi por meio do uso de funções.

```
def ajuste_sexo (sexo):
    if sexo == 'DESCONHECIDO' or sexo == 'NAO INFORMADO':
        sexo = 'DESCONHECIDO/NAO INFORMADO'
    else:
        sexo = sexo
    return sexo
```

```
def ajuste_idades(idade):
    if idade == '(vazio)':
        idade = 'Não informado'
    else:
        idade = idade[3:]
    return idade
```

```
def ano_mes (data):
    data = str(data)[:7]
    return data
```

- Módulos e Pacotes

O uso do pacote Pandas foi recorrente durante todo o Notebook. Para facilitar o uso em várias situações diferentes adotamos o nome reduzido “pd”.

```
import pandas as pd
```

```
import statsmodels.tsa.api as tsa
```





## **5. OS DESAFIOS E ERROS ENFRENTADOS NO DECORRER DO PROJETO E AS LIÇÕES APRENDIDAS.**

- Qual foi o maior erro cometido durante o projeto? Como vocês lidaram com ele?

Inicialmente, o grupo passou por algumas dificuldades básicas, pois havia um certo desnível quanto ao conhecimento de cada integrante em relação à programação em si, à importação e manuseio de outras bibliotecas, do domínio da ferramenta Git e, até mesmo, na compatibilidade da agenda de todos os participantes para as reuniões do grupo.

Ao se analisar os erros, a maioria deles foram de fato mínimos e tiveram um impacto irrisório no final do trabalho. Um fator que dificultou o início do trabalho foi a questão de encontrar um horário em comum para as reuniões. Diante desse cenário, foram sugeridos horários distintos, sendo escolhido o horário com maior participação dos integrantes da equipe. Em seguida, dividimos o grupo em três duplas, nas quais os participantes tivessem níveis diferentes de experiência e através de trocas pudessem proporcionar ao grupo um equilíbrio de conhecimento.

- Qual foi o maior desafio enfrentado durante o projeto? Como vocês lidaram com ele?

Quanto aos desafios, a maior dificuldade para os membros foi, de fato, o domínio do Git. Por certo, apenas dois dos membros tinham uma maior experiência e domínio da ferramenta, o que ocasionou em uma necessidade de abordar as funcionalidades do Git com os outros integrantes. A solução encontrada foi a inserção de uma aula de uso da ferramenta git no cronograma de atividades, a qual foi ministrada então pelos integrantes com mais experiência. Também foi ministrada uma aula básica sobre o uso da biblioteca Pandas, a qual foi essencial para que cada integrante conseguisse entregar a sua parte no projeto.



- Quais as lições aprendidas durante o projeto?

Para nós como um grupo, as principais lições aprendidas foram a importância do trabalho em equipe, da divisão de tarefas e da comunicação entre os integrantes, de modo que a barreira da falta de experiência foi contornada através da divisão em duplas, que foram divididas levando-se em consideração o conhecimento de cada um na ferramenta utilizada. Desse modo, foi imprescindível a categorização básica dos membros para essa estratégia, a fim de criar um equilíbrio na equipe e proporcionar a todos o suporte necessário para que as tarefas fossem cumpridas e os conceitos fossem aprendidos. A realização deste trabalho foi, não apenas uma oportunidade de conhecer melhor as ferramentas do dia a dia na programação, como Git, VS code e outras, mas também uma grande experiência em gerenciamento de projeto, de pessoas e na divisão perfeita para um trabalho em equipe.