Bootcamp TECH MINAs 2023

Apresentação do Desafio Final

Tema: Reclamações do Consumidor

Squad Bertha Lutz

WOMAKERS CODE

Apresentação da equipe

Alexsandra Oliveira

Bacharela em Sistemas de Informação

Beatriz Karoline

Graduanda em Engenharia

Luana Nunes

Cientista de Dados e Graduanda em Eng. Biomédica

Patrícia Alcântara

Bacharela em Gestão da Informação

Tássia Gonçalves

Graduanda em Ciência de Dados

Organização do time

- Reuniões de alinhamento;
- Comunicação por WhatsApp e e-mail;
- Divisão das questões por sorteio;
- Mentoria semanal.

Principais dificuldades

- Capacidade computacional;
- Habilidades técnicas;
- Comunicação.

Pontos positivos

- Mentoria semanal;
- Aprendizado baseado em projeto;
- Uso do ChatGPT para dúvidas.

Entendendo o cenário

Suponha que você trabalha na área de dados do Procon. Diariamente, você e sua equipe recebem dados de diversas reclamações dos consumidores. Cada reclamação, então, tem um determinado tempo para ser resolvida entre o cliente e a companhia em questão. Dado esse contexto, a sua equipe do time de operações dados históricos de reclamações de 2012 a 2016 na pasta reclamações-consumidor.

Os dados são uma amostra dos dados extraídos do Kaggle oriundos do Procon de 2012 a 2016. (Clique aqui para acessar a base de dados).

Perguntas que devem ser respondidas a área de negócios:

Parte 1 - Análise dos dados

1 - Existe alguma sazonalidade na data de abertura de uma reclamação? Ou seja, mais consumidores abrem reclamações em determinada época do ano?

2 - Qual o tempo médio de uma reclamação ativa (da abertura até a data de fechamento)?

3 -O numero de reclamações varia de acordo com a região? e de acordo com o estado? E se ponderarmos pela população média do estado?

4 - Quais as empresas que receberam mais reclamações dos consumidores? E por região e estado?

Perguntas que devem ser respondidas a área de negócios:

Parte 2 – Modelagem: Prevendo o tempo de uma reclamação ativa

O time de negócios gostaria que a sua equipe fizesse um modelo de regressão para estimar qual será o tempo médio de uma reclamação ativa.

1- Quais variáveis podem estar mais correlacionadas com o tempo de uma reclamação ativa?

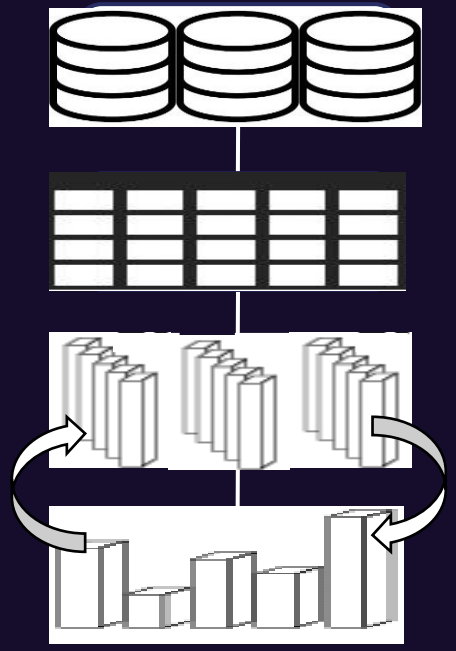
2- Construa variáveis que podem estar correlacionadas com o tempo de uma reclamação ativa a partir dos dados.

3- Analise a correlação das variáveis

4 - Construa um modelo de regressão linear em que queremos estimar o tempo de uma reclamação ativa.

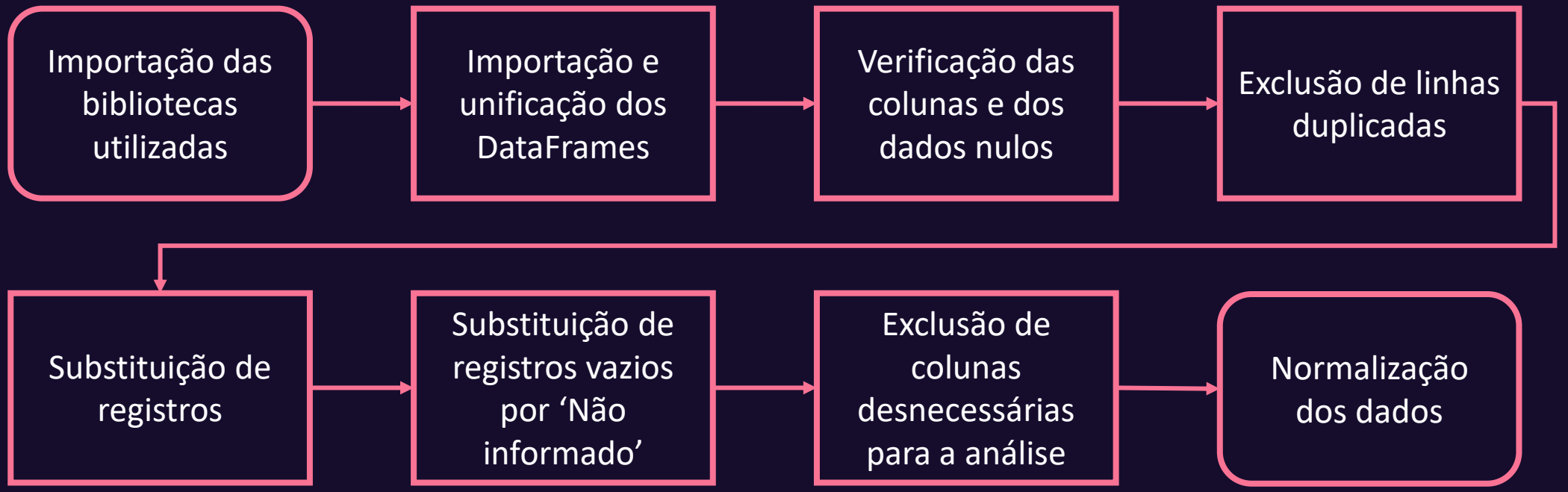
Parte 1: Análise dos dados

Fluxograma geral – Lógica utilizada



- Entendendo a base de dados
- Formatação dos DataFrames
- Análise exploratória dos dados
- Análise Visual

Tratamento da base de dados



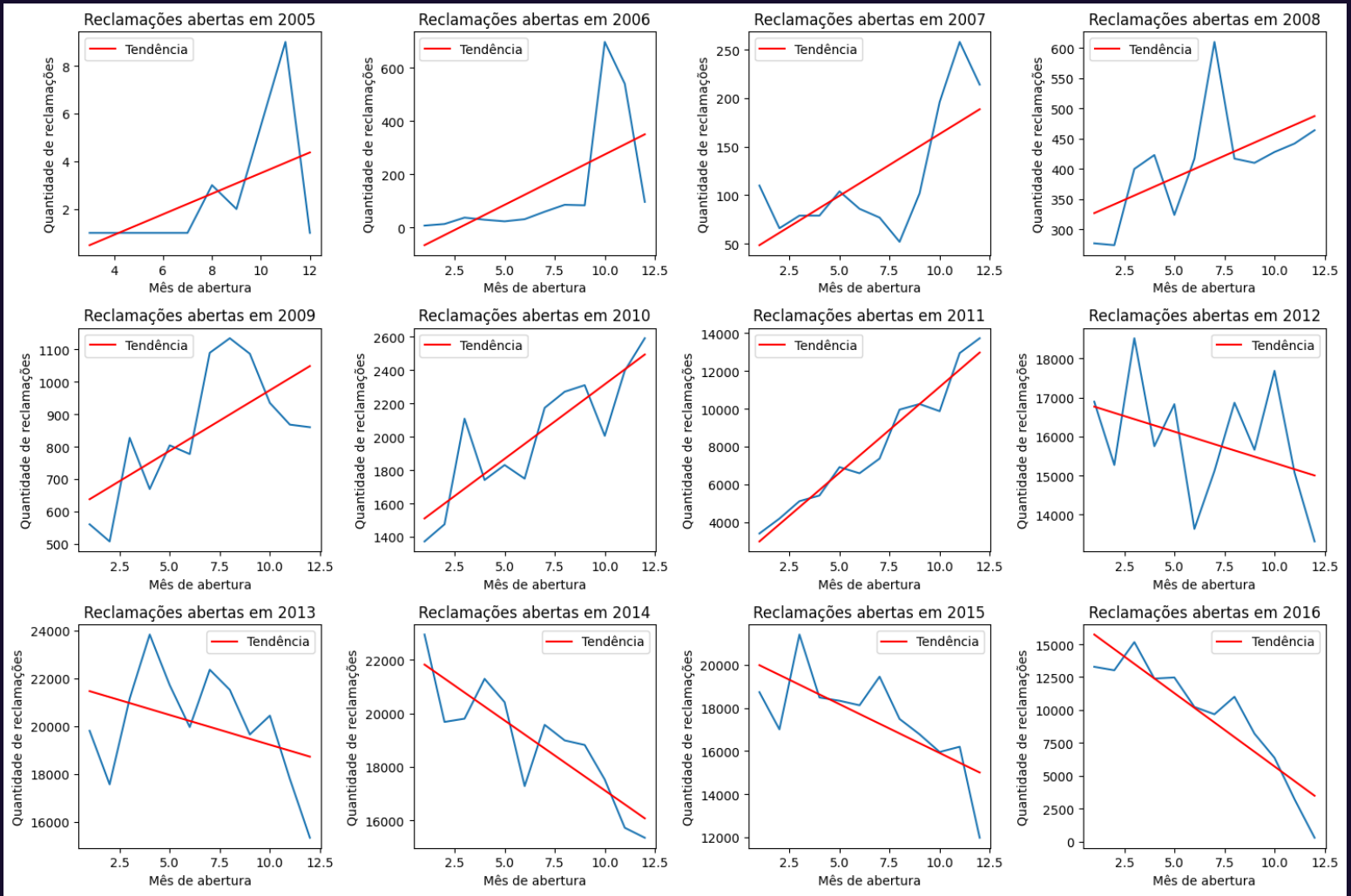
Bibliotecas utilizadas: Pandas, Numpy, Seaborn, Matplotlib, Scikit-learn, nltk, Pickle, entre outras.

Base de dados inicial: 1.206.072 registros

Base de dados final: 1.122.237 registros

Análise exploratória dos dados

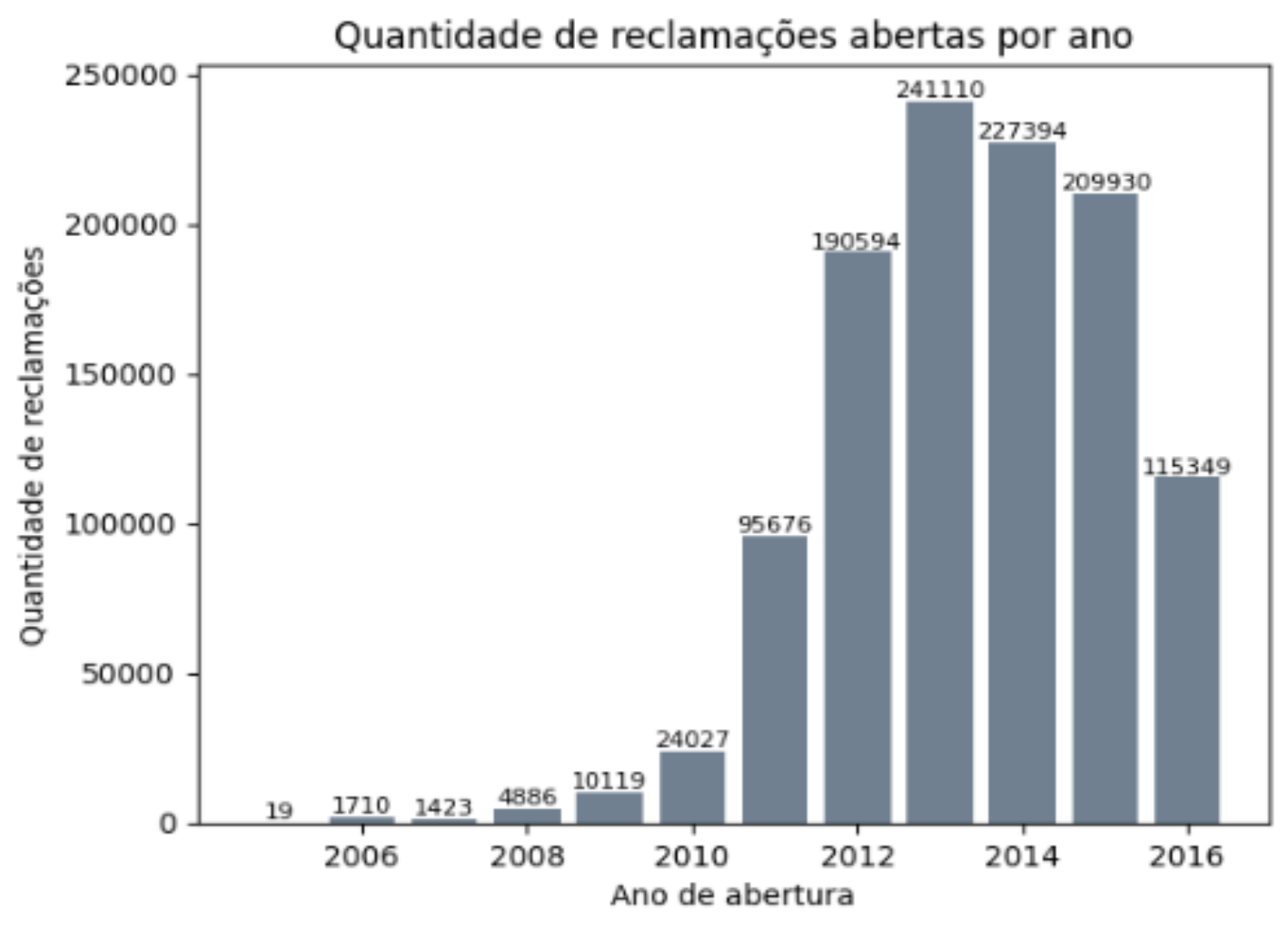
1. Existe alguma sazonalidade na data de abertura de uma reclamação? Ou seja, mais consumidores abrem reclamações em determinada época do ano?



- De 2005 e 2011 é possível notar uma tendência de crescimento;
- De 2012 e 2016 notamos uma tendência de queda;
- Muitas reclamações abertas não estão contempladas nessa análise por ter datas de arquivamento diferentes da base de dados.

Análise exploratória dos dados

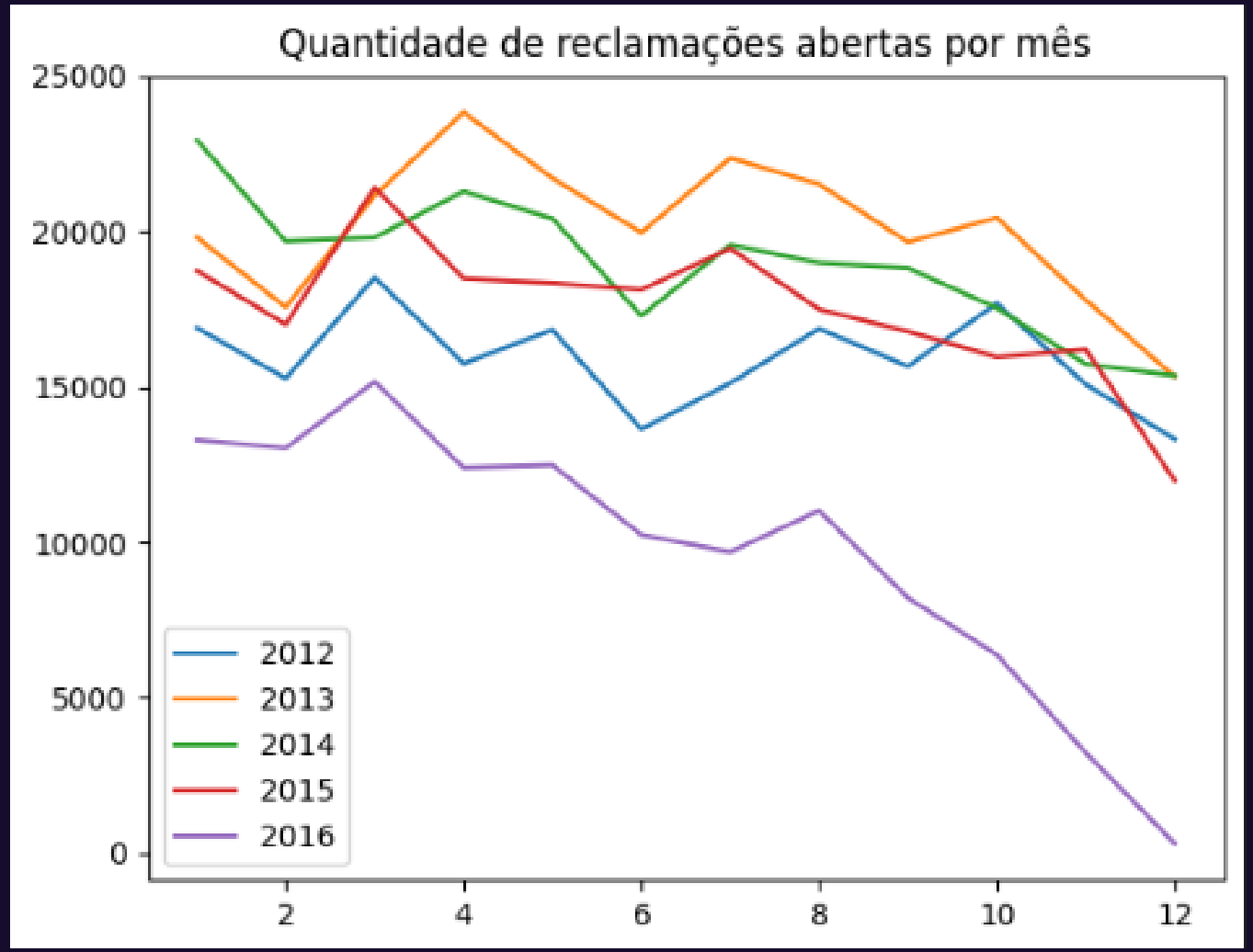
1. Existe alguma sazonalidade na data de abertura de uma reclamação? Ou seja, mais consumidores abrem reclamações em determinada época do ano?



- Discrepância quanto a quantidade de reclamações para os anos mais antigos.
- Visto que na base de dados analisada contém apenas as reclamações quanto a data de arquivamento (2012 a 2016), é possível observar:
- Reclamações abertas entre 2005 e 2011 não estão contempladas nessa análise, por terem sido arquivadas nos anos anteriores.
- Assim como há reclamações recentes que não foram fechadas, e não estão contempladas nessa base de dados.

Análise exploratória dos dados

1. Existe alguma sazonalidade na data de abertura de uma reclamação? Ou seja, mais consumidores abrem reclamações em determinada época do ano?



- Diminuição nos meses de junho e dezembro
- Aumento nos meses de março e julho.

Análise exploratória dos dados

2. Qual o tempo médio de uma reclamação ativa (da abertura até o fechamento)?

```
# Criar a coluna com o tempo da resolução
df['TempoParaFechamento'] = (df['DataArquivamento'] - df['DataAbertura']).dt.days
# Entendendo a coluna criada
df['TempoParaFechamento'].describe().round()
```

```
count    1122237.0
mean         221.0
std          337.0
min          -18.0
25%           46.0
50%           98.0
75%          226.0
max         3975.0
Name: TempoParaFechamento, dtype: float64
```

- Valores negativos;
- Arquivados no primeiro dias do mês de abertura;
- Erro humano;
- Descartar esses registros.

Análise exploratória dos dados

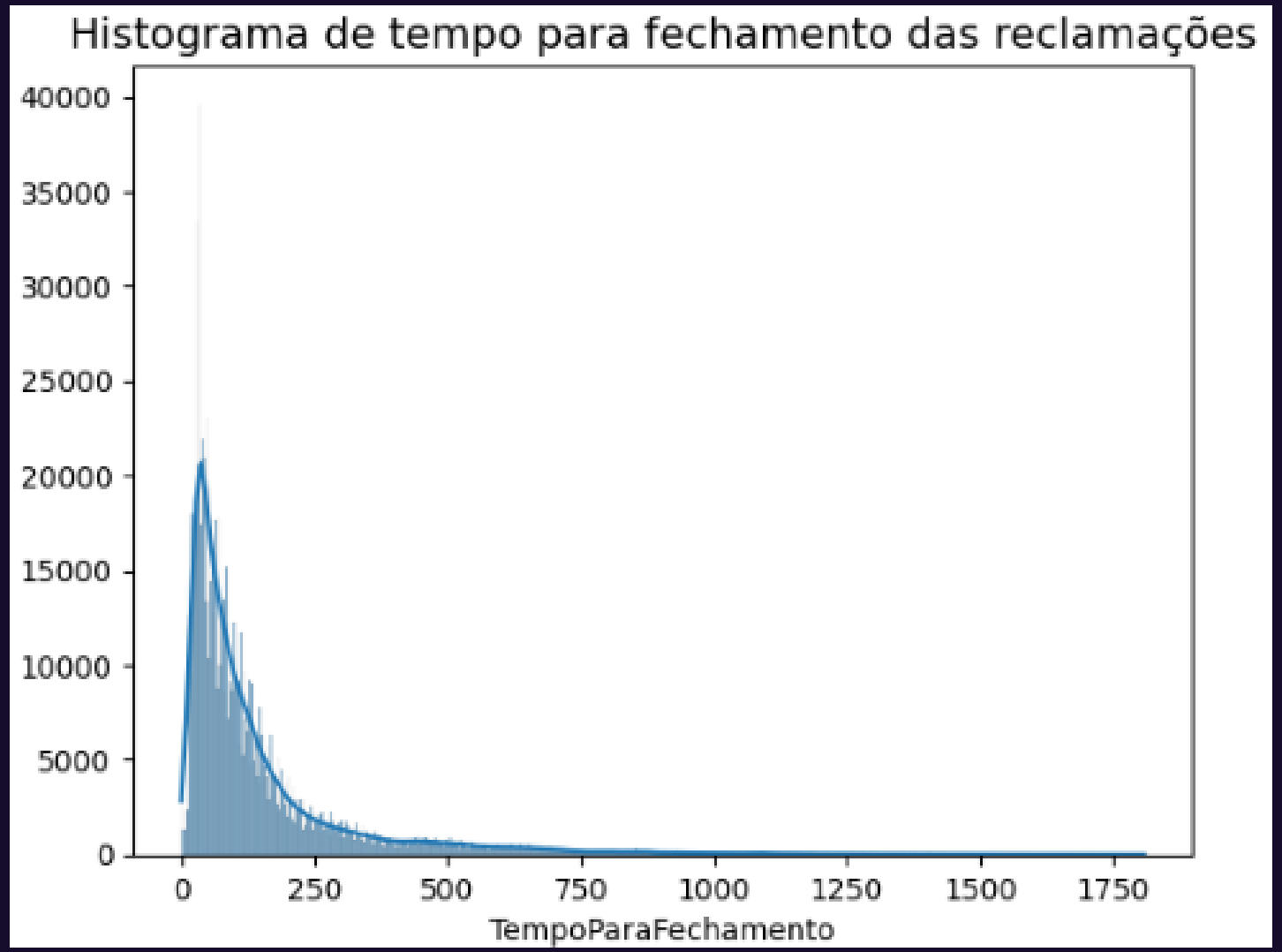
2. Qual o tempo médio de uma reclamação ativa (da abertura até o fechamento)?

	Geral	Atendidas	Não atendidas
Média	220	175	296
Média (2012 a 2016)	148	131	180
Mediana	98	85	125
Mediana (2012 a 2016)	83	76	98

Desvio Padrão	336	260	424
Desvio Padrão (2012 a 2016)	187	159	226

Análise exploratória dos dados

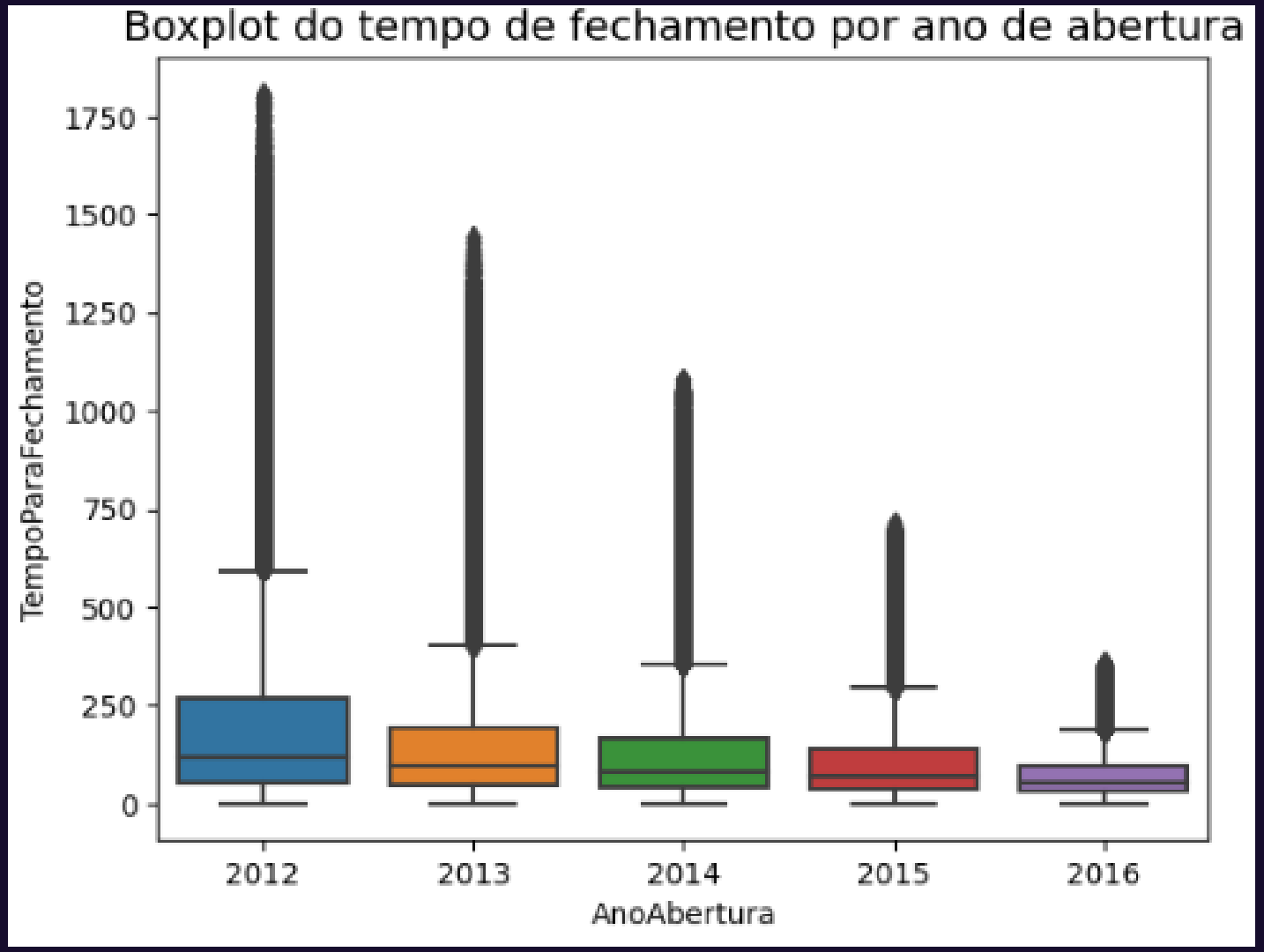
2. Qual o tempo médio de uma reclamação ativa (da abertura até o fechamento)?



- Comportamento de assimetria a direita.
- Ou seja, mesmo restringindo a análise, existe uma quantidade de reclamações com tempo para fechamento muito altos gerando essa assimetria, o que causa essa diferença grande entre a média e a mediana.

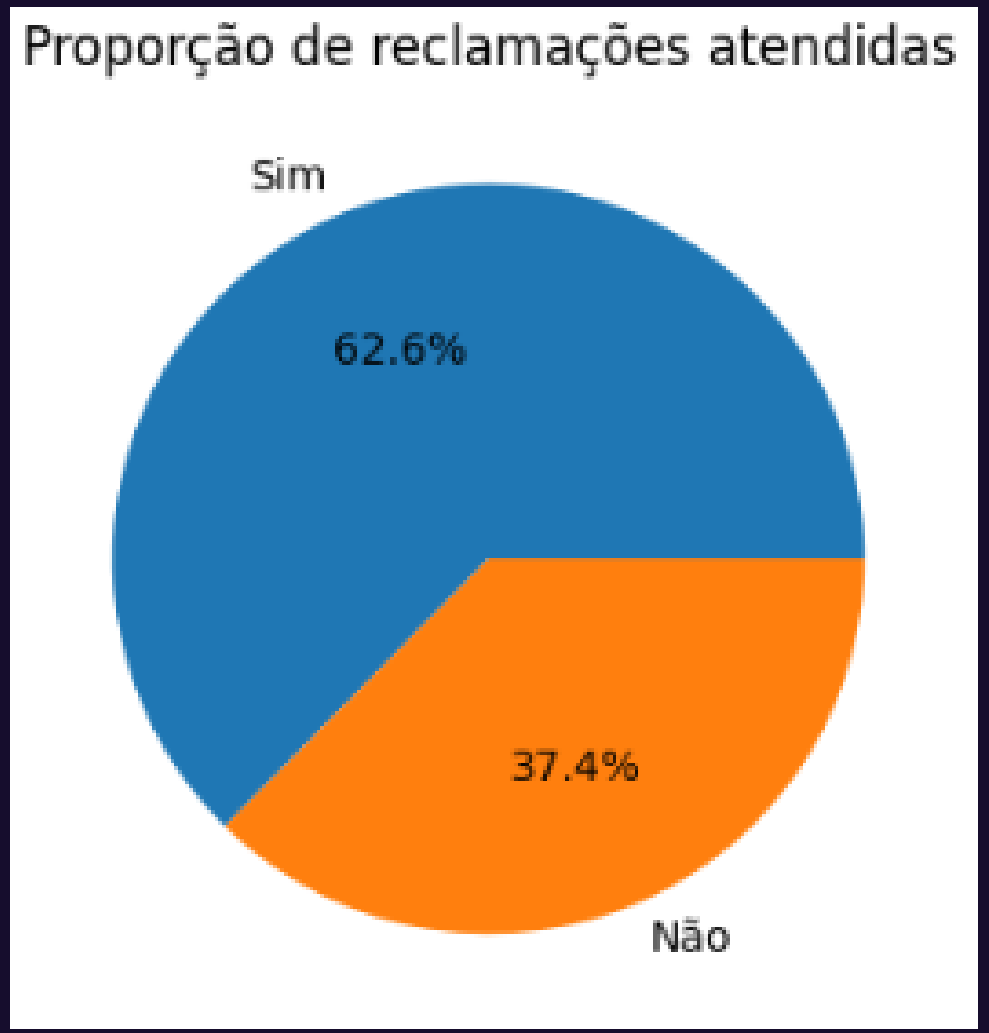
Análise exploratória dos dados

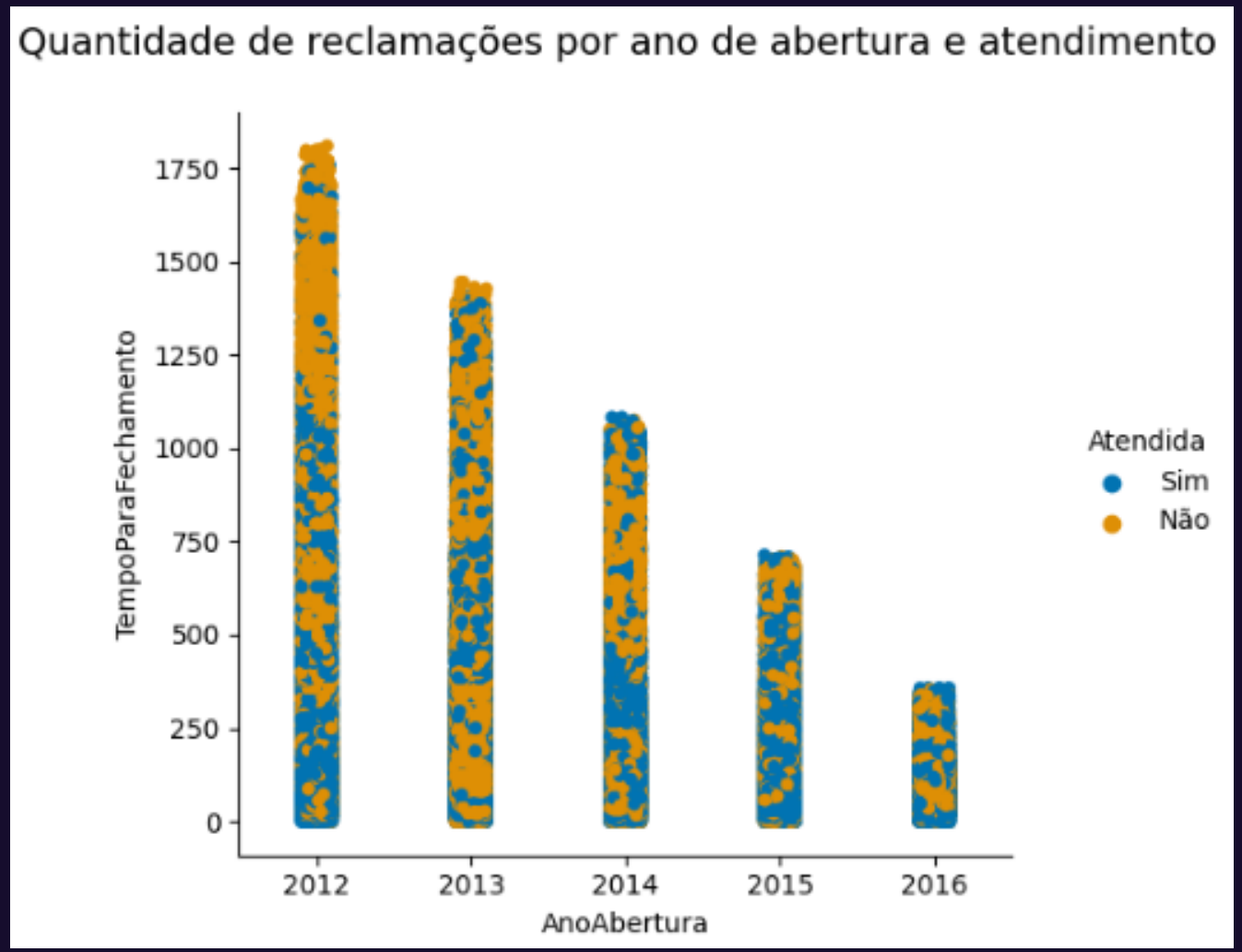
2. Qual o tempo médio de uma reclamação ativa (da abertura até o fechamento)?



- Mediana próxima, porém ao longo dos anos a variação do tempo diminuiu;
- Outliers;
- Reclamações dos anos recentes que não estão contempladas na base de dados.
- Consideramos a mediana, 98 dias, como tempo médio.

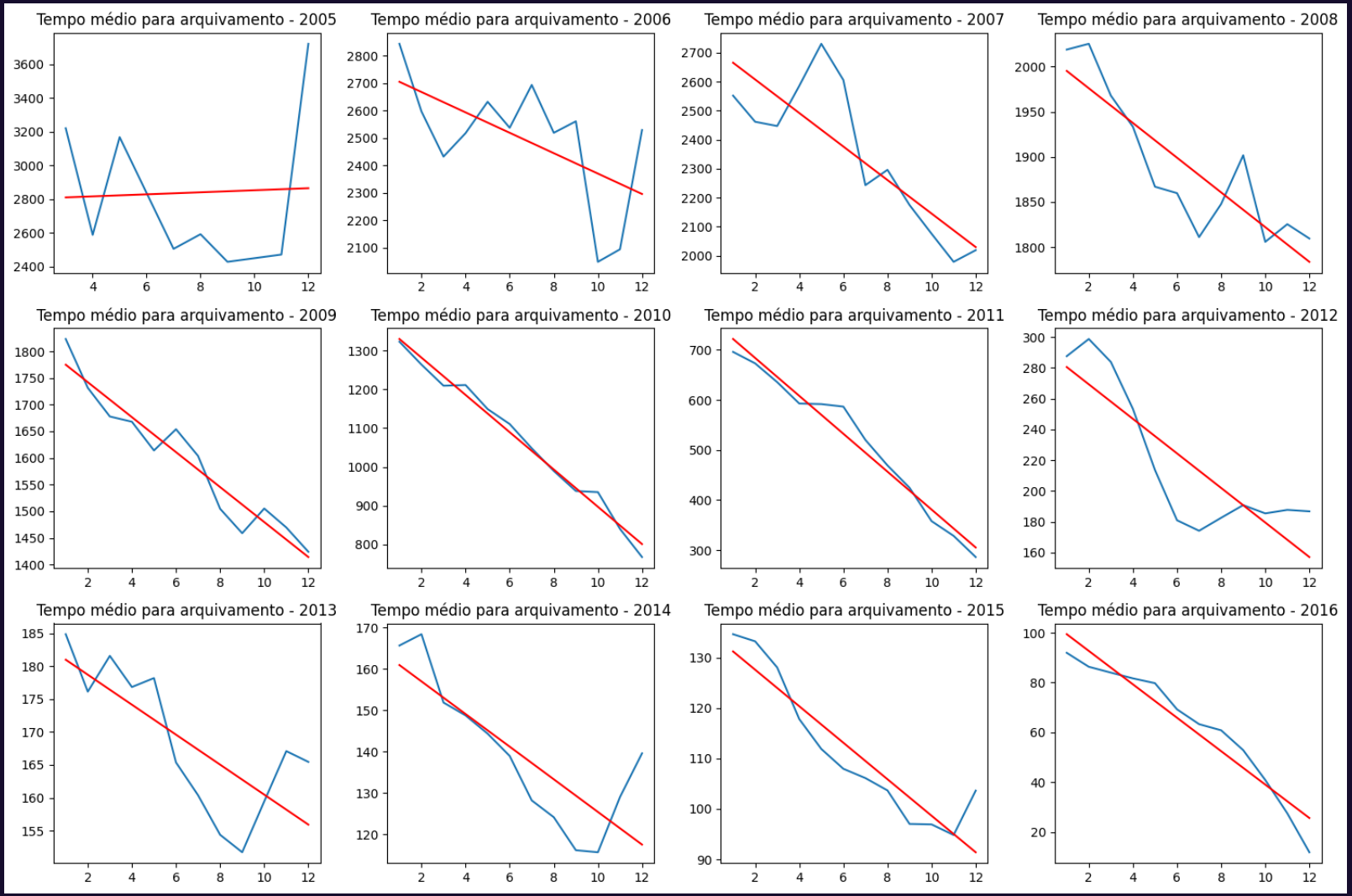
Análise exploratória dos dados





Análise exploratória dos dados

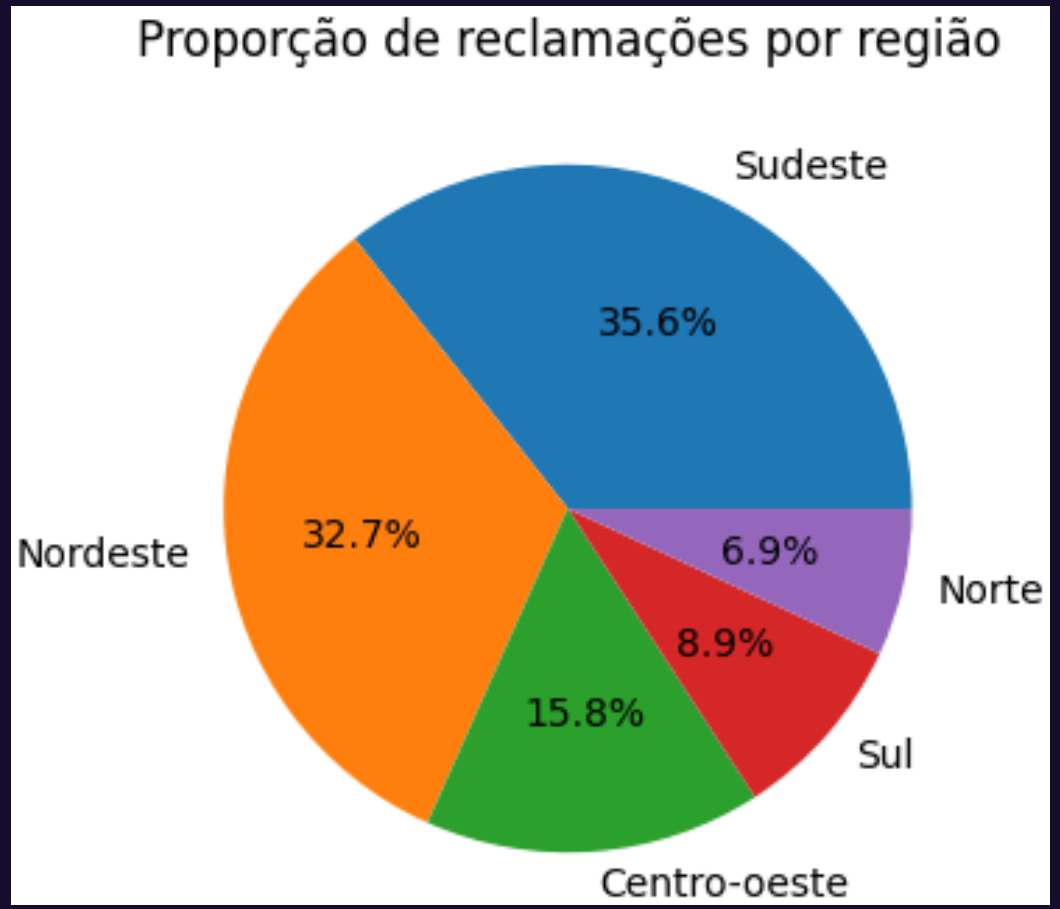
E quanto ao tempo médio para fechamentos? Será que existe sazonalidade?

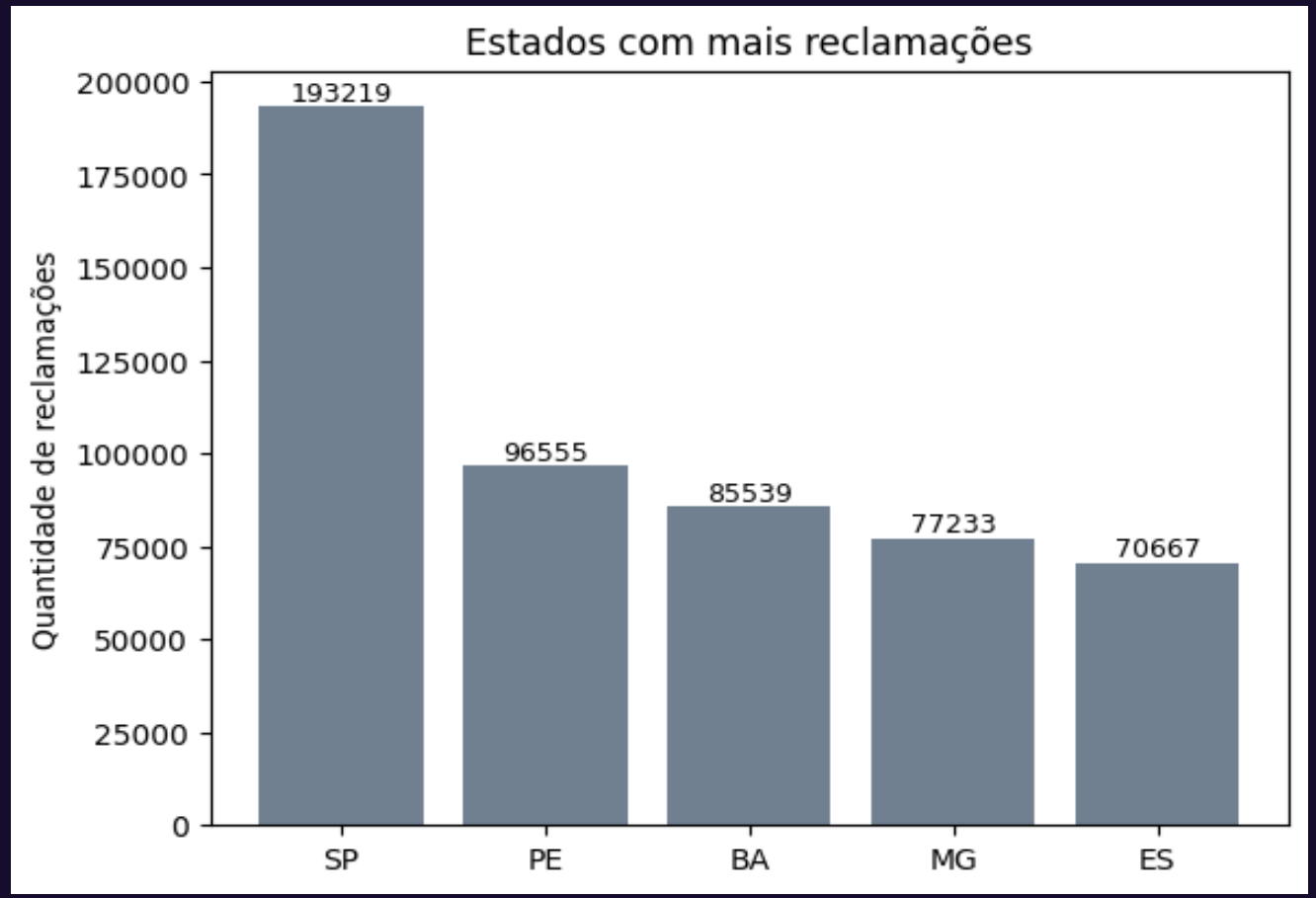


A partir dos gráficos percebe-se que reclamações abertas no início dos anos têm um tempo médio de aquivamento maior quando comparadas com as que são abertas no restante do anos.

Análise exploratória dos dados

3. a) O número de reclamações varia de acordo com a região? E de acordo com o estado?





Análise exploratória dos dados

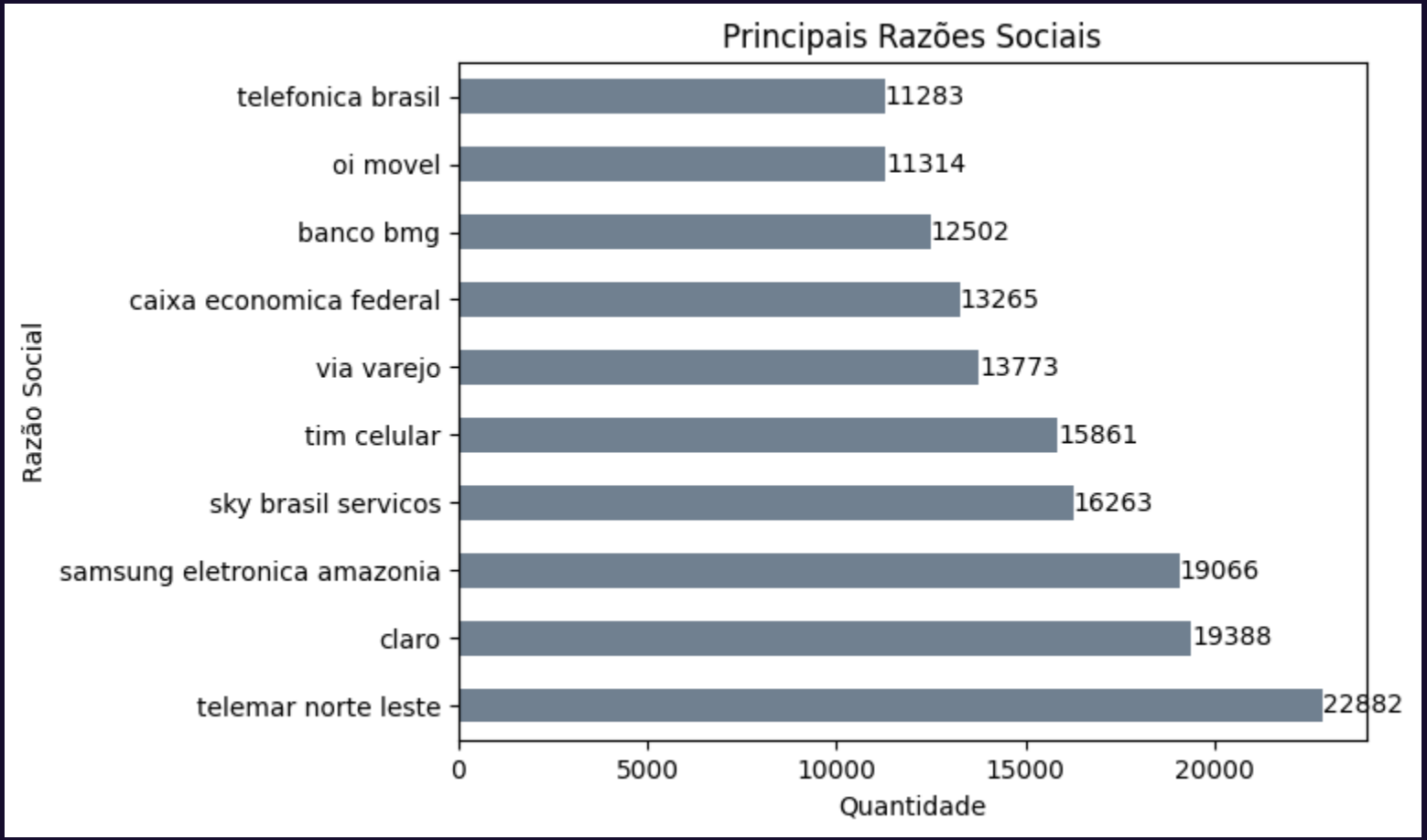
3. c) E se ponderarmos pela população média do estado?

	Regiao	Quantidade	Populacao	Reclamacoes_por_pessoa
2	Centro-oeste	180207	15660988	1.150675
1	Nordeste	368054	56915936	0.646662
0	Sudeste	399667	86356952	0.462808
4	Norte	78631	17707783	0.444048
3	Sul	95669	29439773	0.324965

	UF	Porcentagem(%)	Populacao	Reclamacoes_por_pessoa
5	MS	6.0	2682386	2.588293
12	TO	3.0	1532902	2.284947
4	ES	6.0	3973697	1.778369
9	AL	4.0	3358963	1.405612
11	MT	4.0	3305531	1.287660

Análise exploratória dos dados

4. a) Quais as empresas que receberam mais reclamações dos consumidores?



	Porcentagem (%)
telemar norte leste	2.038980
claro	1.727635
samsung eletronica amazonia	1.698942
sky brasil servicos	1.449171
tim celular	1.413349
via varejo	1.227291
caixa economica federal	1.182024
banco bmg	1.114034
oi movel	1.008173
telefonica brasil	1.005411

Análise exploratória dos dados

Outras análises sobre as empresas:

- O **CNAEs** mais reclamados foram: Bancos múltiplos e Telefonia Móvel Celular;
- O **problema** com maior tempo médio para arquivamento foi “produto sem registro/registro falso”;
- O **assunto** com maior tempo médio para arquivamento foram os relacionados a fubá/polvilho/milho;
- O **problema** com mais ocorrências foi ‘vício do produto’;
- O **assunto** com mais ocorrências quanto a telefone convencional/celular/interfone

Análise exploratória dos dados

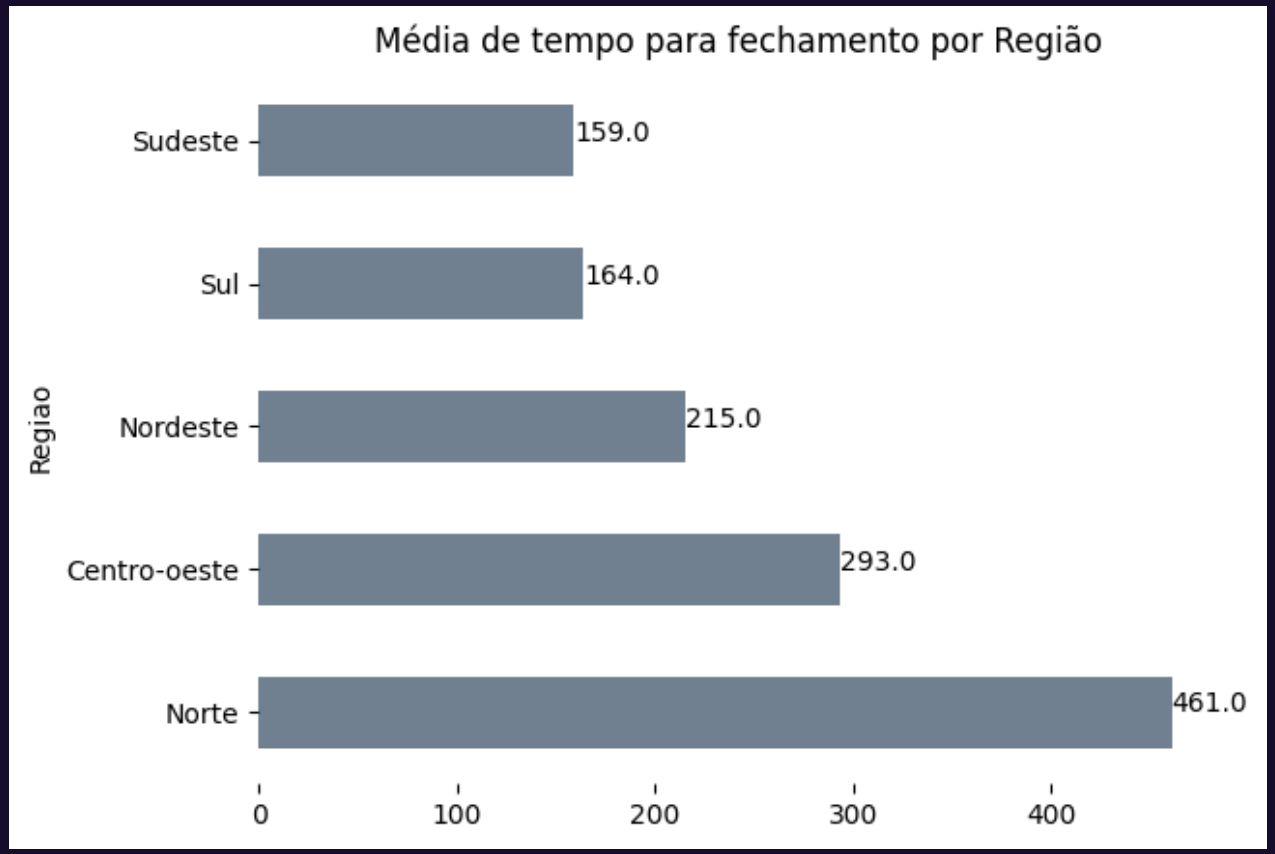
4. b) E por região? E por estado?

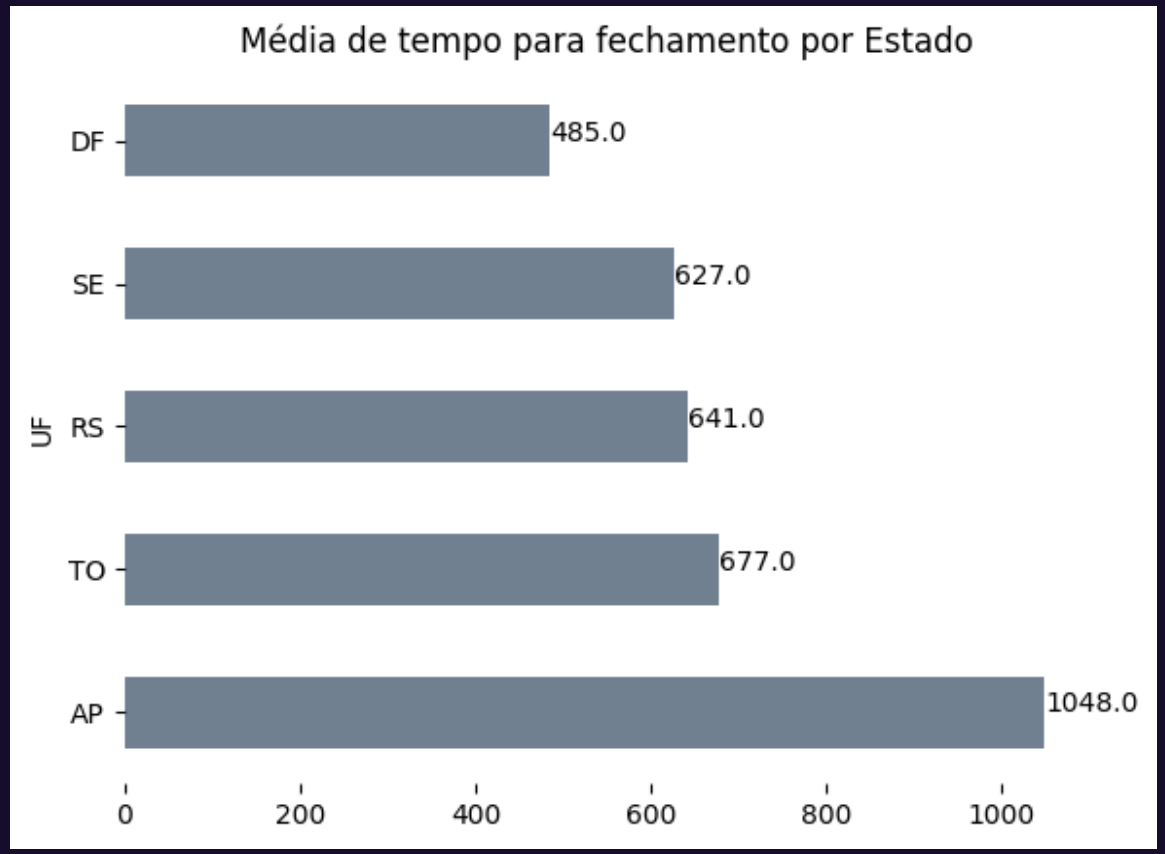
	Regiao	Empresa	Reclamacoes	Porcentagem(%)
3	Sudeste	telemar norte leste	11184	2.80
1	Nordeste	telemar norte leste	10530	2.86
0	Centro-oeste	americel	3573	1.98
4	Sul	oi	3505	3.66
2	Norte	centrais eletricas	2120	2.70

	UF	Empresa	Reclamacoes	Porcentagem(%)
24	SP	telefonica brasil	7824	4.05
10	MG	telemar norte leste	6616	8.57
18	RJ	via varejo	4330	7.40
7	ES	telemar norte leste	3954	5.60
4	BA	telemar norte leste	3895	4.55
15	PE	telemar norte leste oi fixo	3582	3.71
22	SC	oi	3505	5.90
5	CE	telemar norte leste	2363	5.41
11	MS	americel	2350	3.38
13	PA	centrais eletricas	2118	9.86

Análise exploratória dos dados

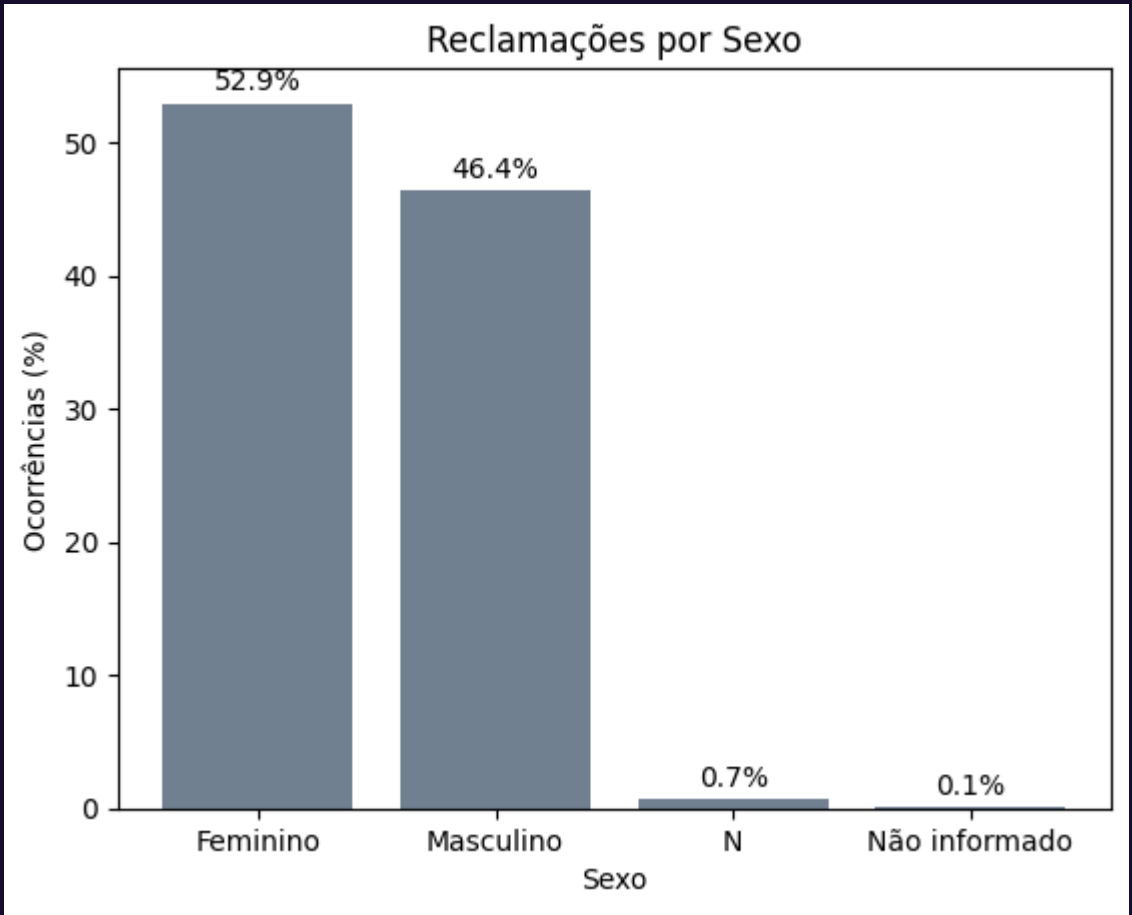
Análise quanto ao tempo para fechamento de reclamação por Região e por UF

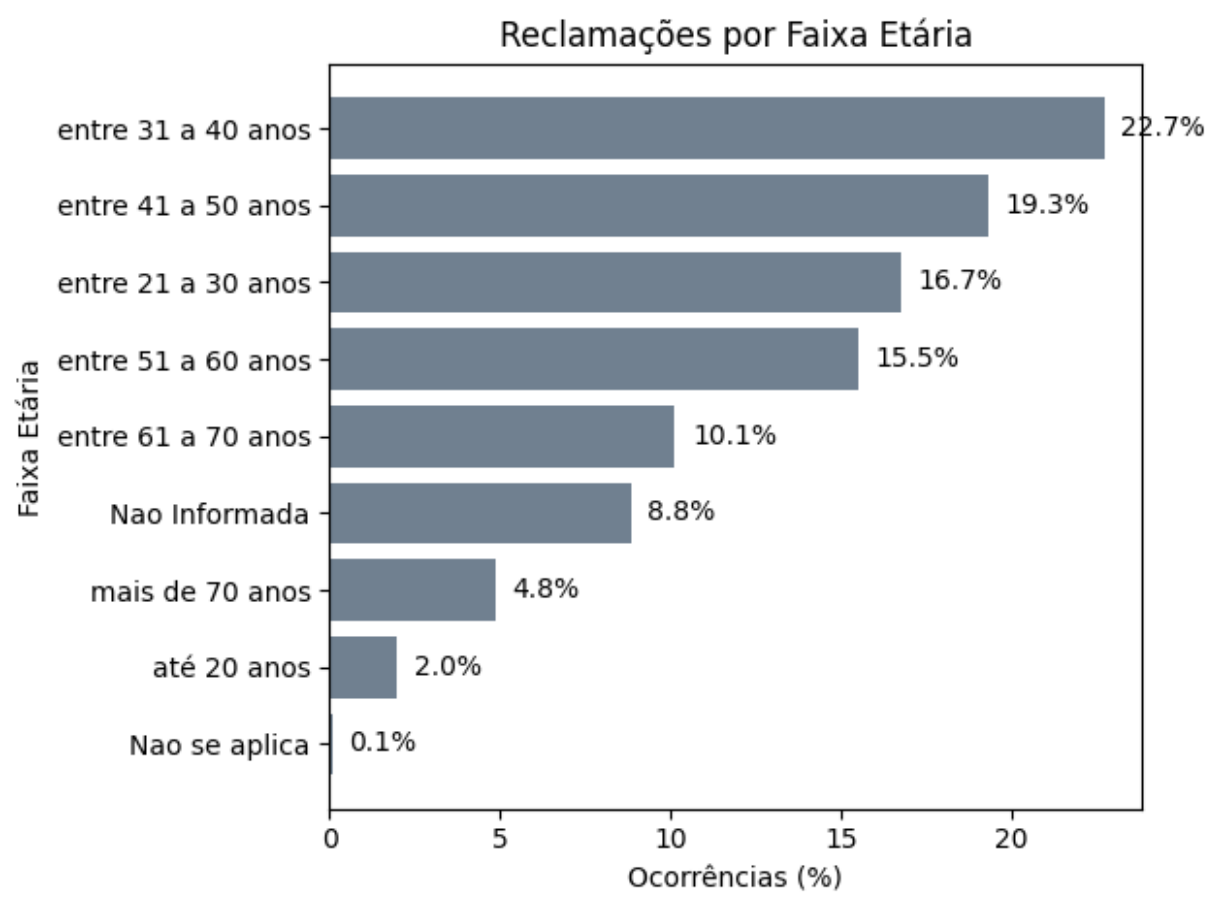




Análise exploratória dos dados

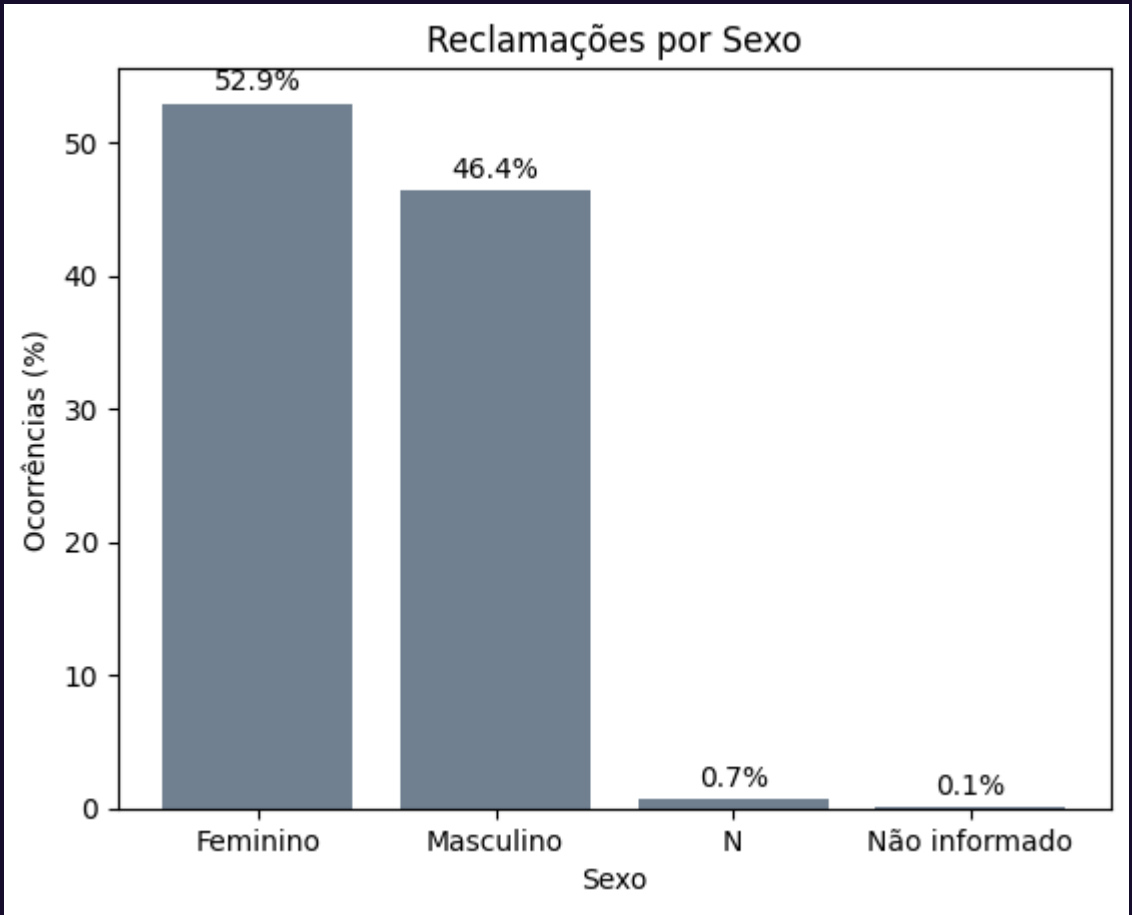
Também foi analisado quantidade de reclamações quanto ao sexo e a idade

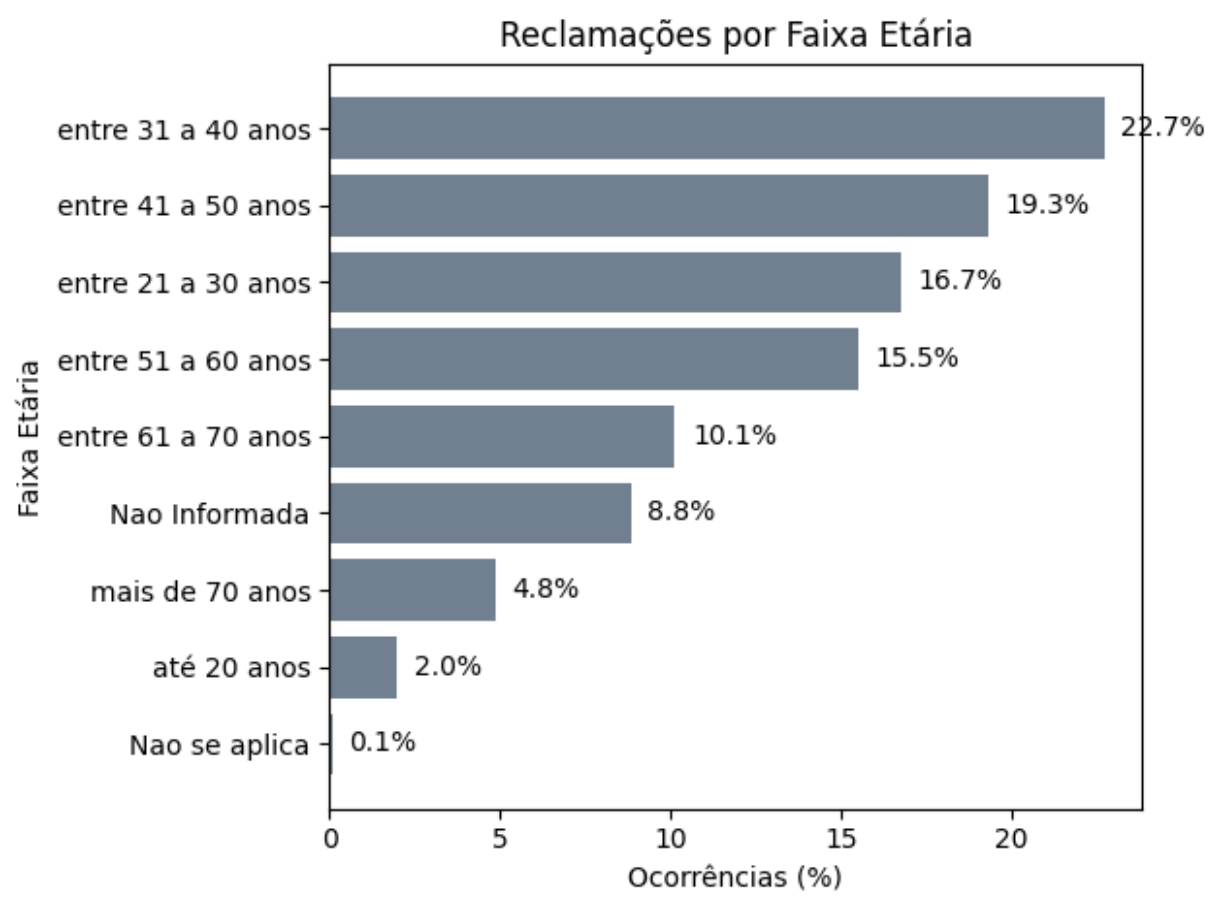




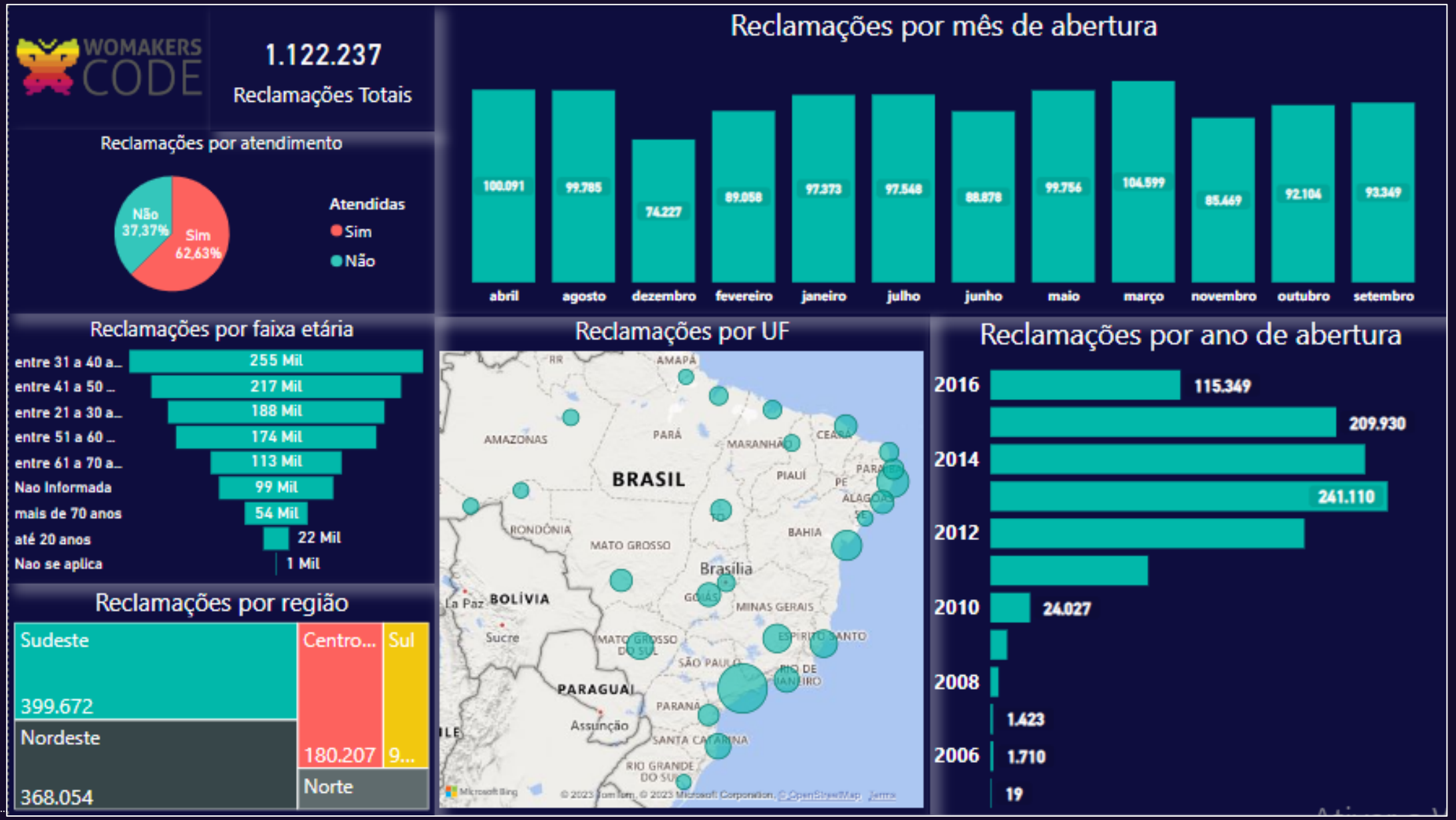
Análise exploratória dos dados

Também foi analisado quantidade de reclamações quanto ao sexo e a idade



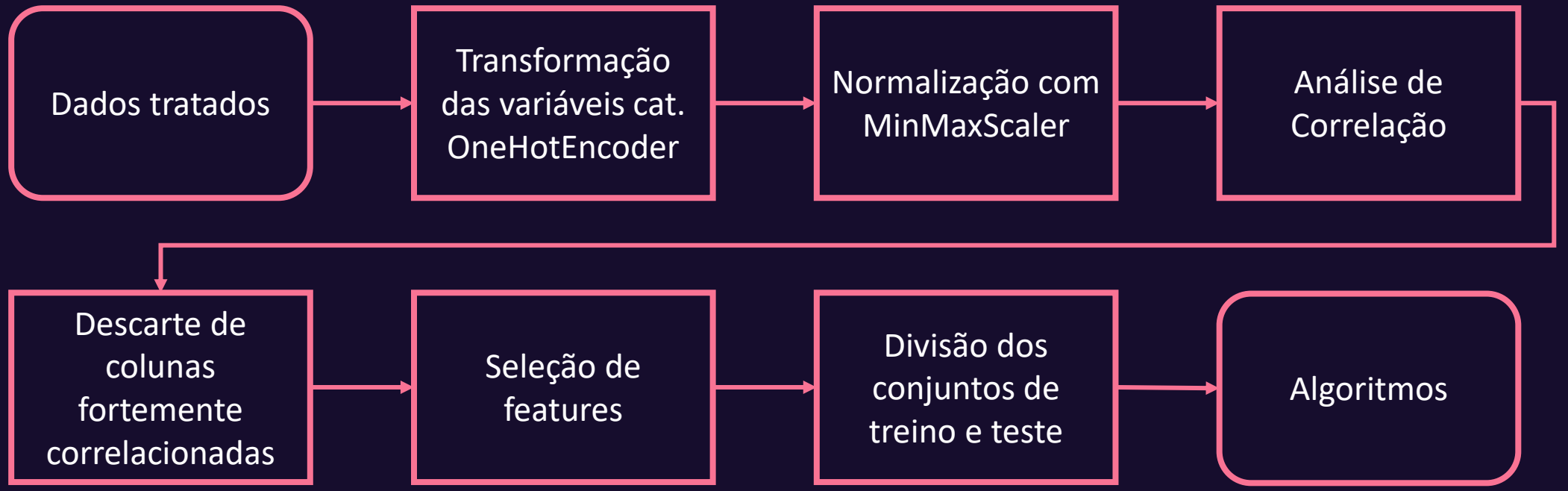


Power BI


WOMAKERS CODE
1.122.237
Reclamações Totais
Reclamações por atendimento
Não 37,37%
Sim 62,63%
Atendidas
Sim
Não
Reclamações por mês de abertura
100.091
99.785
74.227
89.058
97.373
97.548
88.878
99.756
104.599
85.469
92.104
93.349
abril
agosto
dezembro
fevereiro
janeiro
julho
junho
maio
março
novembro
outubro
setembro
Reclamações por faixa etária
entre 31 a 40 a... 255 Mil
entre 41 a 50 ... 217 Mil
entre 21 a 30 a... 188 Mil
entre 51 a 60 ... 174 Mil
entre 61 a 70 a... 113 Mil
Nao Informada 99 Mil
mais de 70 anos 54 Mil
até 20 anos 22 Mil
Nao se aplica 1 Mil
Reclamações por UF
BRASIL
Brasília
AMAZONAS
PARÁ
MARANHÃO
PIAUÍ
BAHIA
MATO GROSSO
RONDÔNIA
MINAS GERAIS
SÃO PAULO
PARANÁ
SANTA CATARINA
RIO GRANDE DO SUL
BOLÍVIA
La Paz
Sucre
PARAGUAI
Assunção
Reclamações por ano de abertura
2016 115.349
209.930
2014
241.110
2012
2010 24.027
2008
1.423
2006 1.710
19
Reclamações por região
Sudeste 399.672
Centro... 180.207
Sul 9...
Nordeste 368.054
Norte

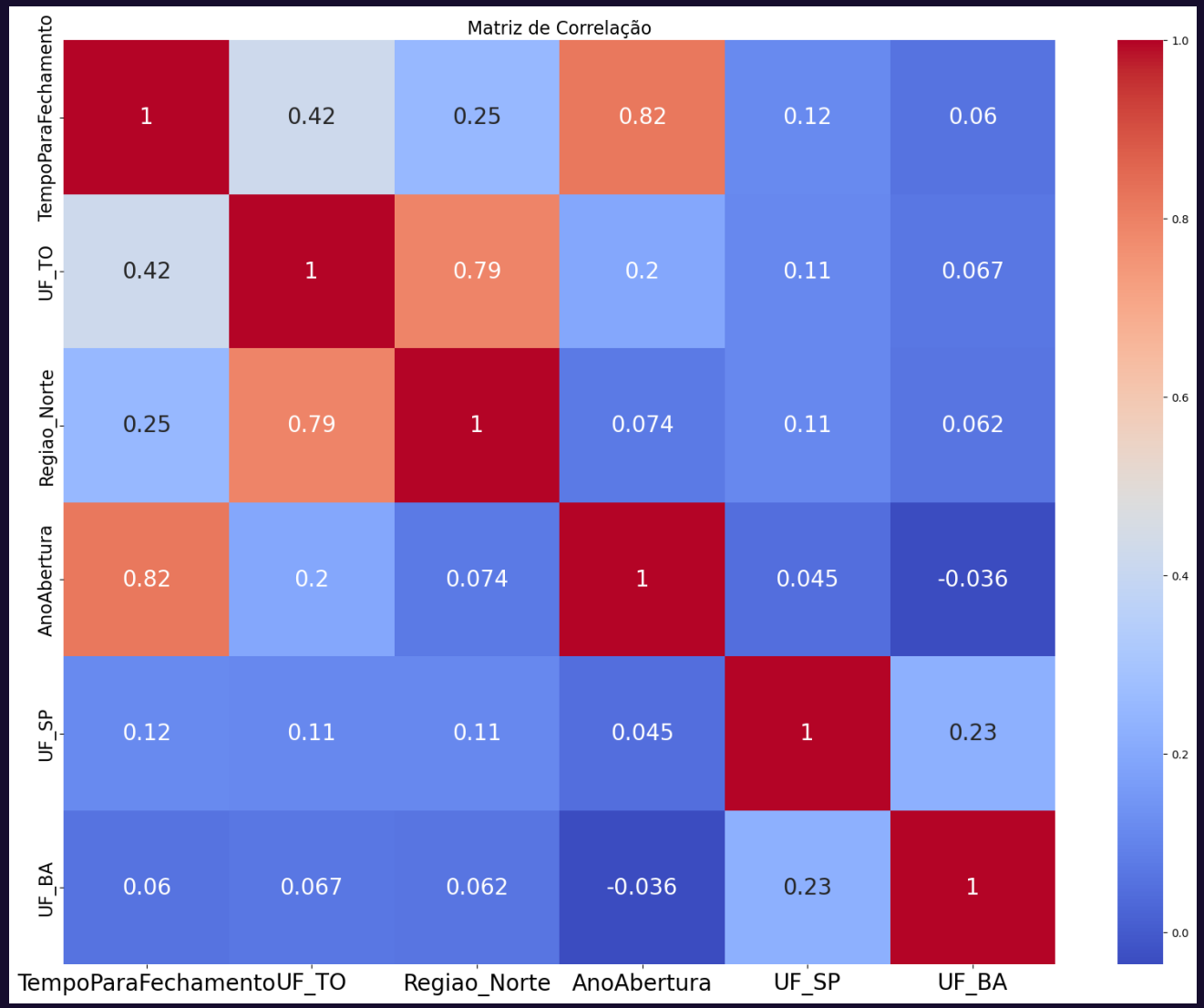
Modelagem

Pré-processamento dos dados



Modelo de Regressão Linear

Análise de correlação



Maiores correlações com o tempo para fechamento:

- AnoAbertura
- UF_TO
- Regiao_Norte
- UF_BA
- UF_SP
- UF_AP
- UF_DF

Regressão Linear

Variáveis independentes

- Ano de abertura
- Mês de abertura
- Região
- Estado
- Sexo
- Faixa etária

Variável dependente

- Tempo para fechamento

Regressão Linear

Avaliação do modelo

- Número de observações: 454245
- Método dos mínimos quadráticos
- Tipo de covariância: não robusta
- R-squared (descentralizado): 0.502
- R-squared ajustado (descentralizado): 0.502
- F-statistic: 11.180
- Prob (F-statistic): 0.00
- AIC: -1.007×10^6
- BIC: -1.007×10^6

Regressão Linear

Variáveis estatisticamente significativas

- AnoAbertura
- MêsAbertura
- UF_MG
- UF_SP
- UF_RJ
- SexoConsumidor_Masculino
- FaixaEtariaConsumidor_entre 21 a 30
- FaixaEtariaConsumidor_entre 31 a 40
- FaixaEtariaConsumidor_entre 41 a 50

Escolha do modelo

Comparação de métricas

Métricas	Regressão Linear	Decision Tree	Random Forest	XGBoost
R-squared	0,5020	0,7542	0,7600	0,7126
MSE	0,0041	0,0017	0,0017	0,0024
RMSE	0,0637	0,0418	0,0413	0,0494
MAE	0,0439	0,0232	0,0231	0,0307

Conclusões

- O Random Forest foi o modelo que obteve melhores métricas.
- Seria mais interessante que os dados fornecidos fossem baseados no ano de abertura, ao invés do ano de arquivamento.
- Talvez fosse interessante considerar as variáveis de tipos de atividade, problema e assunto para a construção do modelo de regressão.

Se conecte com a gente :)

Tássia Gonçalves

Luana Nunes

Alexsandra Oliveira

Patrícia Alcântara

“As mulheres juntas, serão uma força!”

– Bertha Lutz

WOMAKERS CODE

M.I.N.As
Mulheres em Inovação Negócios & Artes

PortoDigital

Muito obrigada!!