

Análise da Base de dados abertos da SiGesGuarda de Curitiba

1. Introdução

Neste trabalho proposto na disciplina de Introdução à Programação Para Ciência De Dados, do Curso de Especialização em Ciência de Dados e suas Aplicações promovido pela UTFPR, tem como objetivo a definição de um dataset de dados abertos da prefeitura de Curitiba.

O dataset escolhido foi o SiGesGuarda, que contempla todos os registros entre 2009 e 2023, contendo 429347 linhas e 35 colunas com diversas informações. [Link](#) para o dataset

Este conjunto de dados será modificado e limpo, através da linguagem de programação Python e bibliotecas Pandas e matplotlib. Após os ajustes será analisado de forma exploratória, encontrando informações de valor para tomada de decisão, como opção, das autoridades competentes.

2. Sobre o problema

Após a definição do Dataset, foi proposto pelo professor Luiz realizarmos alguns questionamentos que podem ser respondidos através da análise dos dados. Com isso podemos confirmar de fato alguns pontos ao invés da suposição.

Portanto as perguntas que serão respondidas serão:

- Quais são os tipos de registros mais comuns em Curitiba?
- Quais os dias da semana possuem mais registros?
- Quais bairros possuem mais registros?
- Qual a distribuição das ocorrências nos anos?
- Qual o período com maior frequência de registros referente à perturbação de sossego?
- Quais os dias em que há maior registros referentes à perturbação de sossego, evidenciando os períodos?

3. Análise do dataset

Os passos desta análise podem ser verificados no documento enviado anexo em HTML com o título *Trabalho Definição de Dataset Limpeza e Análise dos Dados - Emerson Adam - CDA – UTFPR*.

Esta etapa consiste em importar os dados, realizar a apropriação do contexto, entendendo melhor seu conjunto de dados, realizar a limpeza e a criação de gráficos para encontrar as respostas para as perguntas acima, definidas no tópico 2.

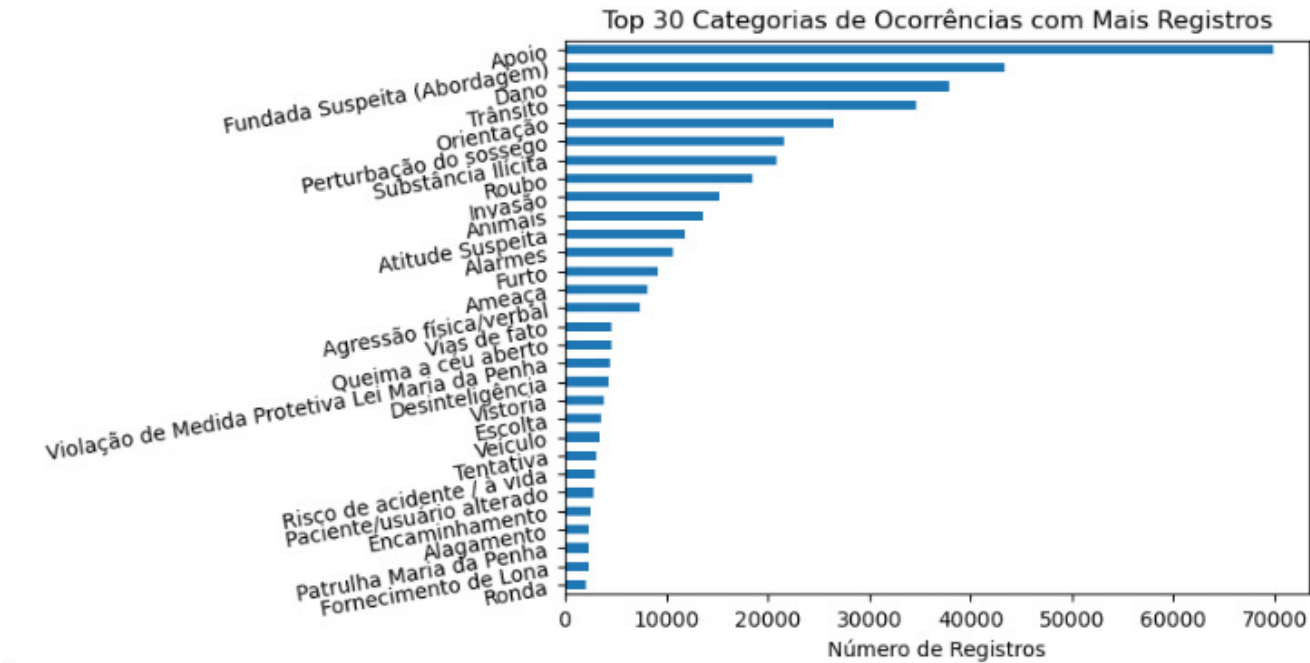
Este dataset contém um dicionário de dados, porém, como o dataset foi alterado e selecionado somente algumas colunas, será colocado neste documento somente as que foram utilizadas na análise. [Link](#) do documento original.

DICIONÁRIO DE DADOS DO DATASET SIGESGUARDA			
COLUNAS	TIPO	TAMANHO	DESCRIÇÃO DOS CAMPOS
ATENDIMENTO_BAIRRO_NOME	varchar	20	Nome do bairro em que foi realizado o atendimento
LOGRADOURO_NOME	varchar	70	Nome do logradouro
NATUREZA1_DESCRICAO	varchar	100	Descrição da primeira natureza cadastrada na ocorrência
OCORRENCIA_ANO	int	-	Ano de cadastro da ocorrência
OCORRENCIA_CODIGO	int	-	Código da ocorrência
OCORRENCIA_DATA	datetime	-	Data da ocorrência
OCORRENCIA_DIA_SEMANA	varchar	20	Dia da semana em que a ocorrência foi cadastrada
OCORRENCIA_HORA	varchar	8	Hora em que a ocorrência foi cadastrada
OCORRENCIA_MES	int	-	Mês em que a ocorrência foi cadastrada
OPERACAO_DESCRICAO	varchar	70	Nome da operação que realizará o atendimento, caso haja
ORIGEM_CHAMADO_DESCRICAO	varchar	70	Local em que se originou a chamada

Após a apropriação, podemos começar a responder as perguntas.

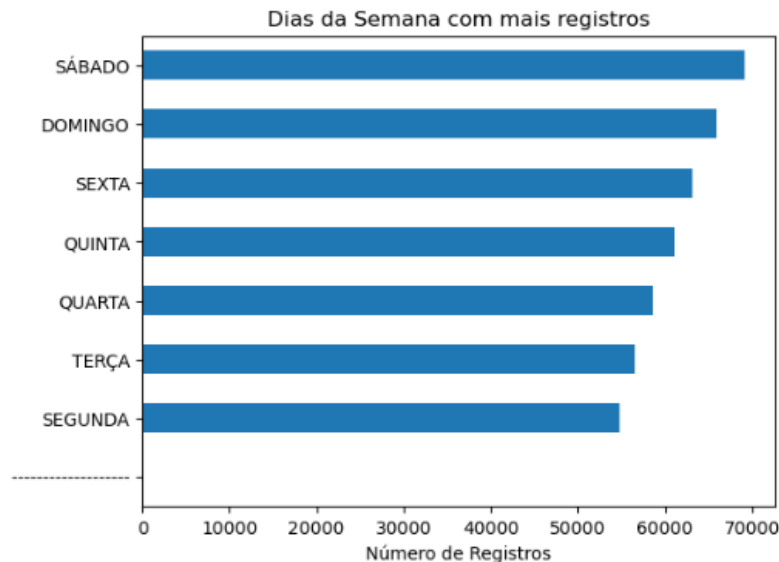
- Quais são os tipos de registros mais comuns em Curitiba?

Para responder esta pergunta, foi efetuado um gráfico com os 30 assuntos com mais registros, conforme abaixo:



- **Quais os dias da semana possuem mais registros?**

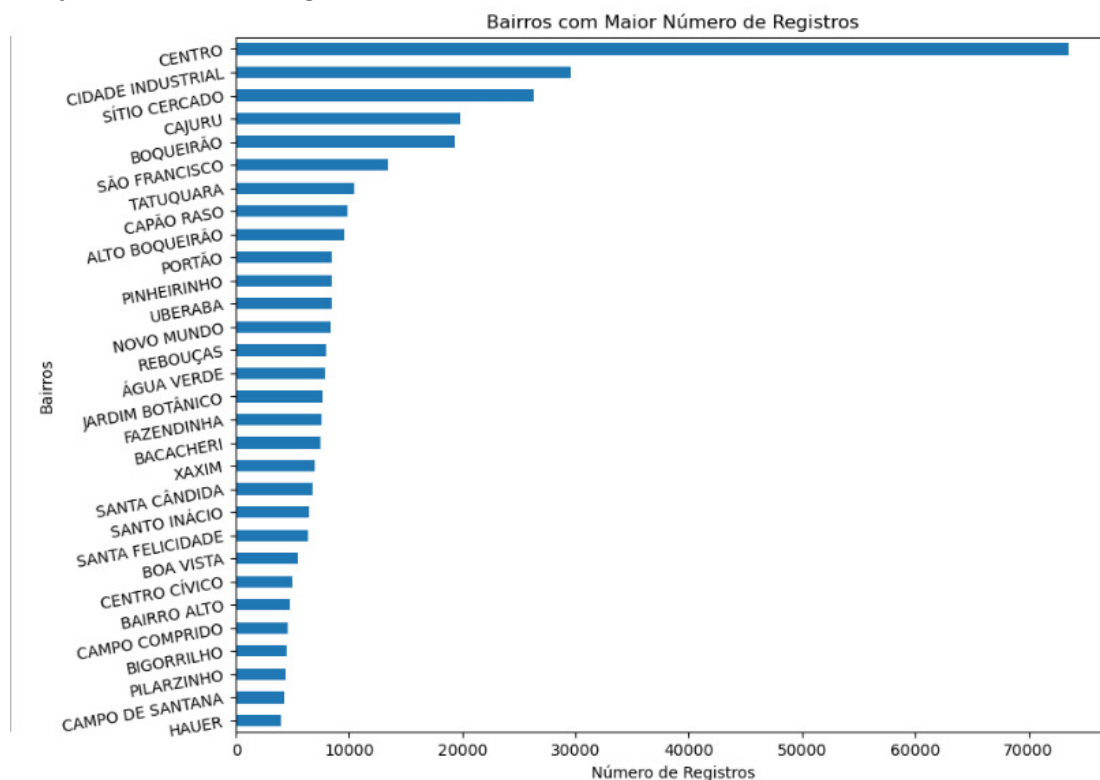
Para esta pergunta, realizamos o gráfico abaixo:



Identifica-se que os 3 dias com mais chamados são no fim de semana a partir de sexta-feira.

- **Quais bairros possuem mais registros?**

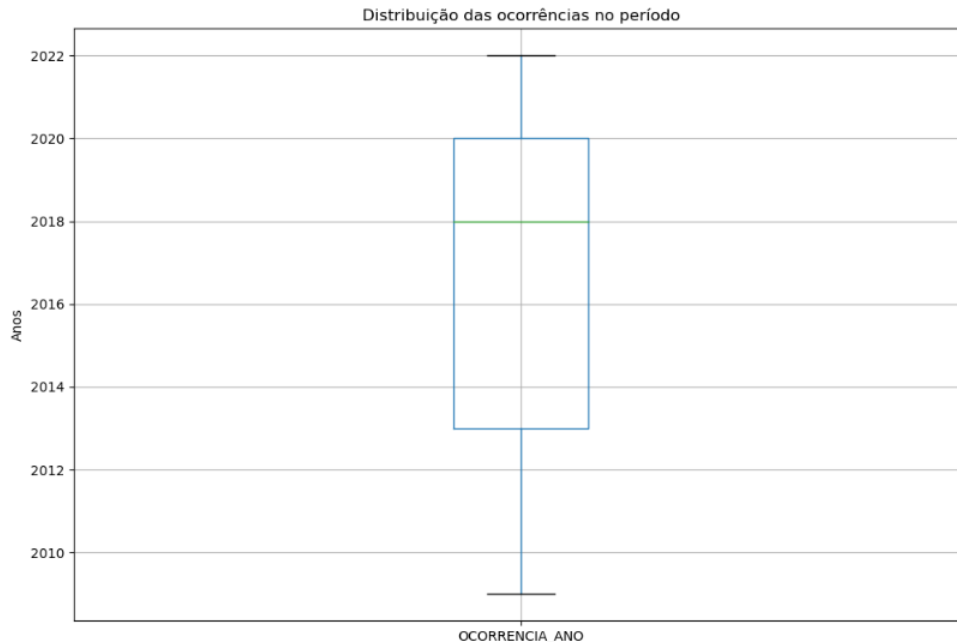
Para responder esta pergunta, o gráfico abaixo foi gerado, contendo os 30 maiores bairros em relação ao número de registros.



Nota-se que Centro, CIC e Sítio Cercado possuem uma grande diferença entre os demais.

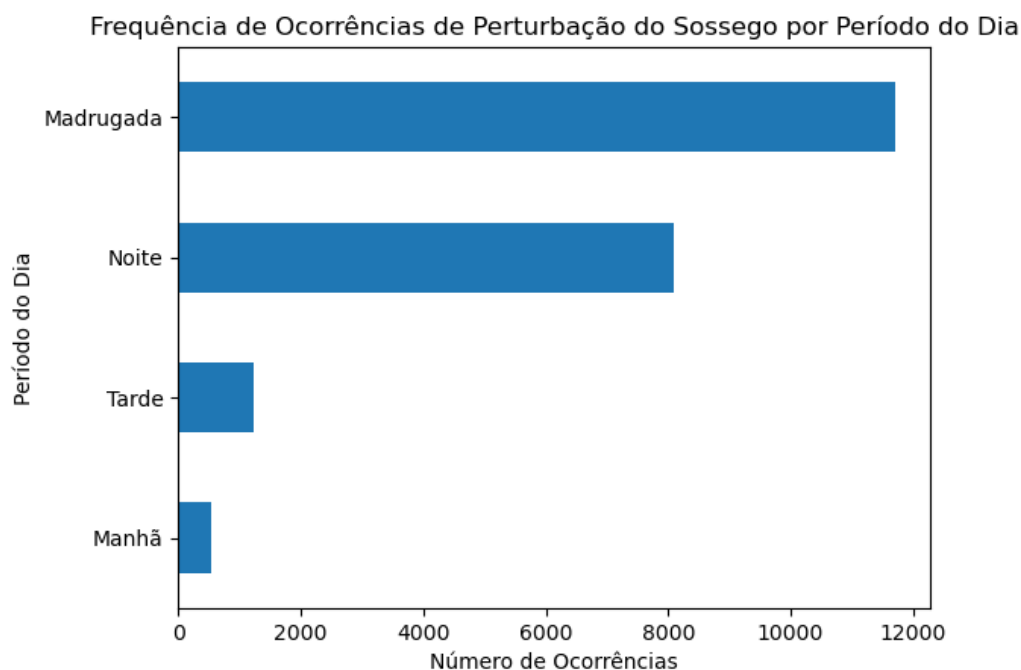
- **Qual a distribuição das ocorrências nos anos?**

Para responder esta pergunta, foi gerado o gráfico no estilo boxplot abaixo



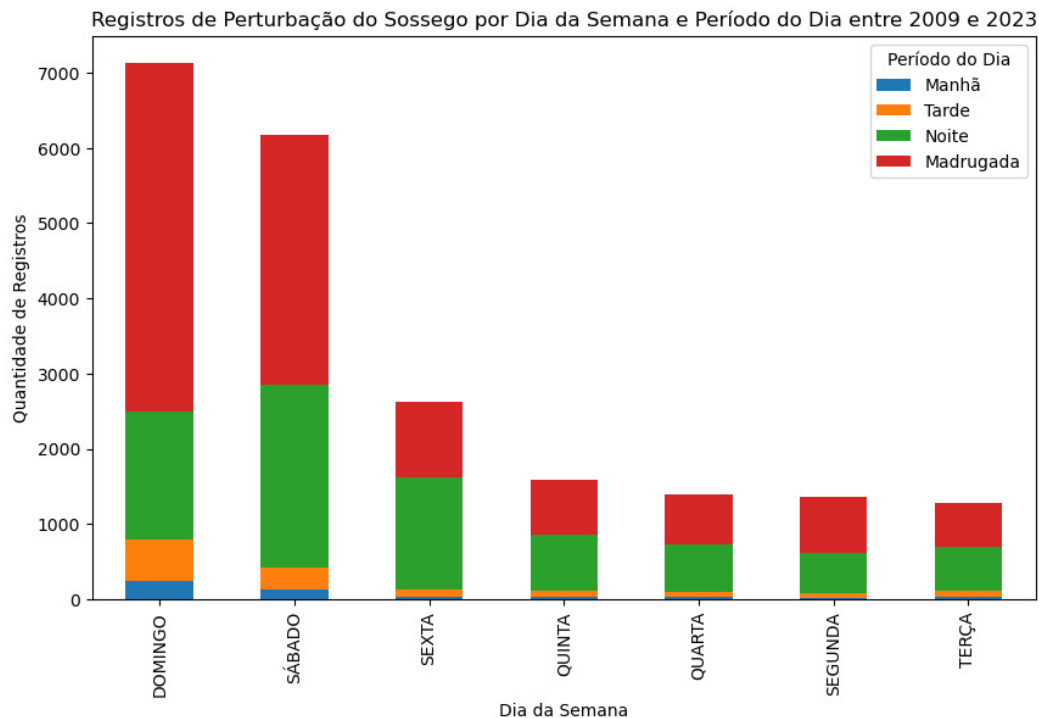
- **Qual o período com maior frequência de registros referente à perturbação de sossego?**

Para responder esta questão, geramos o gráfico de barras abaixo, separando por período de acordo com a hora da ocorrência, nota-se que de madrugada e a noite são os períodos que possuem maiores registros



- **Quais os dias em que há maior registros referentes à perturbação de sossego, evidenciando os períodos?**

Para responder esta pergunta, geramos o gráfico abaixo onde fica evidente que ao final de semana, a partir da noite de sexta-feira até o início da manhã de segunda-feira temos um aumento significativo dos registros de perturbação do sossego. Análise efetuada entre os anos de 2009 e 2023.



4. Conclusão

Após análises, podemos considerar que, especificamente na categoria perturbação de sossego, da base da SiGesGuarda, há um volume maior de registros entre a noite de sexta-feira até o início da manhã de segunda.

Portanto ações devem ser recomendadas para que esse número diminua, como por exemplo uma maior fiscalização e punição para quem esteja infringindo a lei e medidas educativas que relemem o respeito para com todos.