

# Assistente de Análise de Dados com LLM

Uma aplicação web interativa que transforma a exploração de dados

Gabriel de Antonio  
Mazetto

Mateus Pereira Alves

Disciplina: Tópicos em Ciência de  
Dados

Instituição: PUC Campinas



# Definição do Projeto

## Aplicação Web Local

Interface interativa para análise e exploração de dados sem dependências externas.

## LLM para Planejamento

API de LLM gera código e planeja tarefas usando apenas metadados não sensíveis.

## 100% Local

Todo processamento de dados do usuário ocorre em ambiente seguro e privado.



# Critérios de Aprovação

Quatro pilares garantem qualidade das funções desenvolvidas:



1

## Validação por Tentativa

Máximo 3 tentativas de geração ou correção pela IA, as Funções foram testadas com 2 datasets diferentes



3

## Generalidade

Aplicável a múltiplos contextos e datasets



2

## Não Trivialidade

Não é simples encapsulamento de funcionalidade nativa



4

## Utilidade

Avaliar se a função é útil para uma análise de dados

# Fonte dos Códigos

Documentações Oficiais

Plotly e bibliotecas especializadas

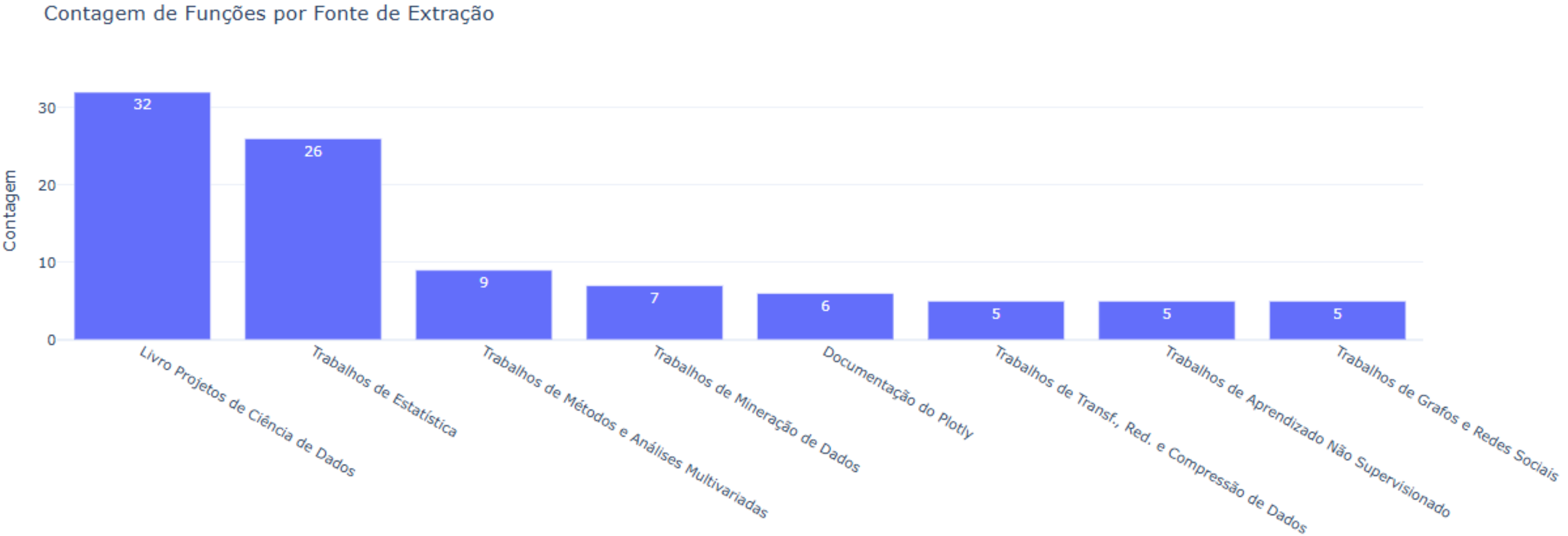
Referências Acadêmicas

"Projetos de Ciência de Dados com Python"

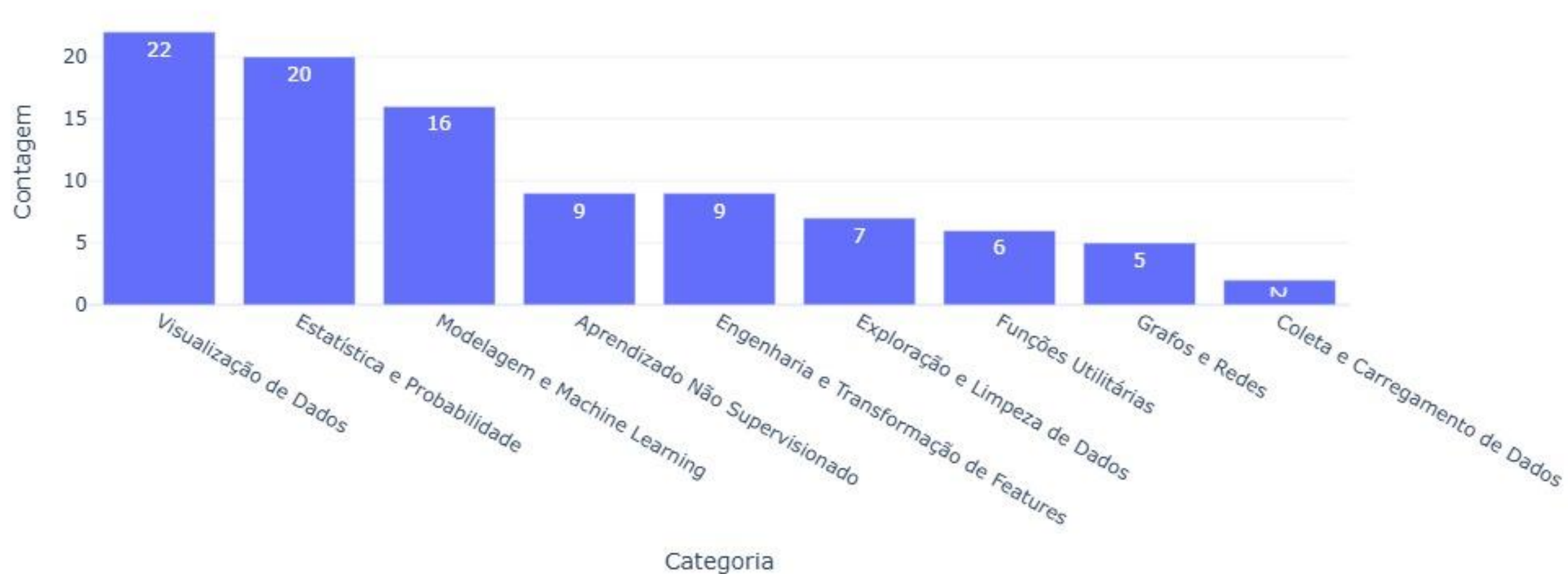
Projetos Pessoais

Códigos de projetos acadêmicos e profissionais

Contagem de Funções por Fonte de Extração

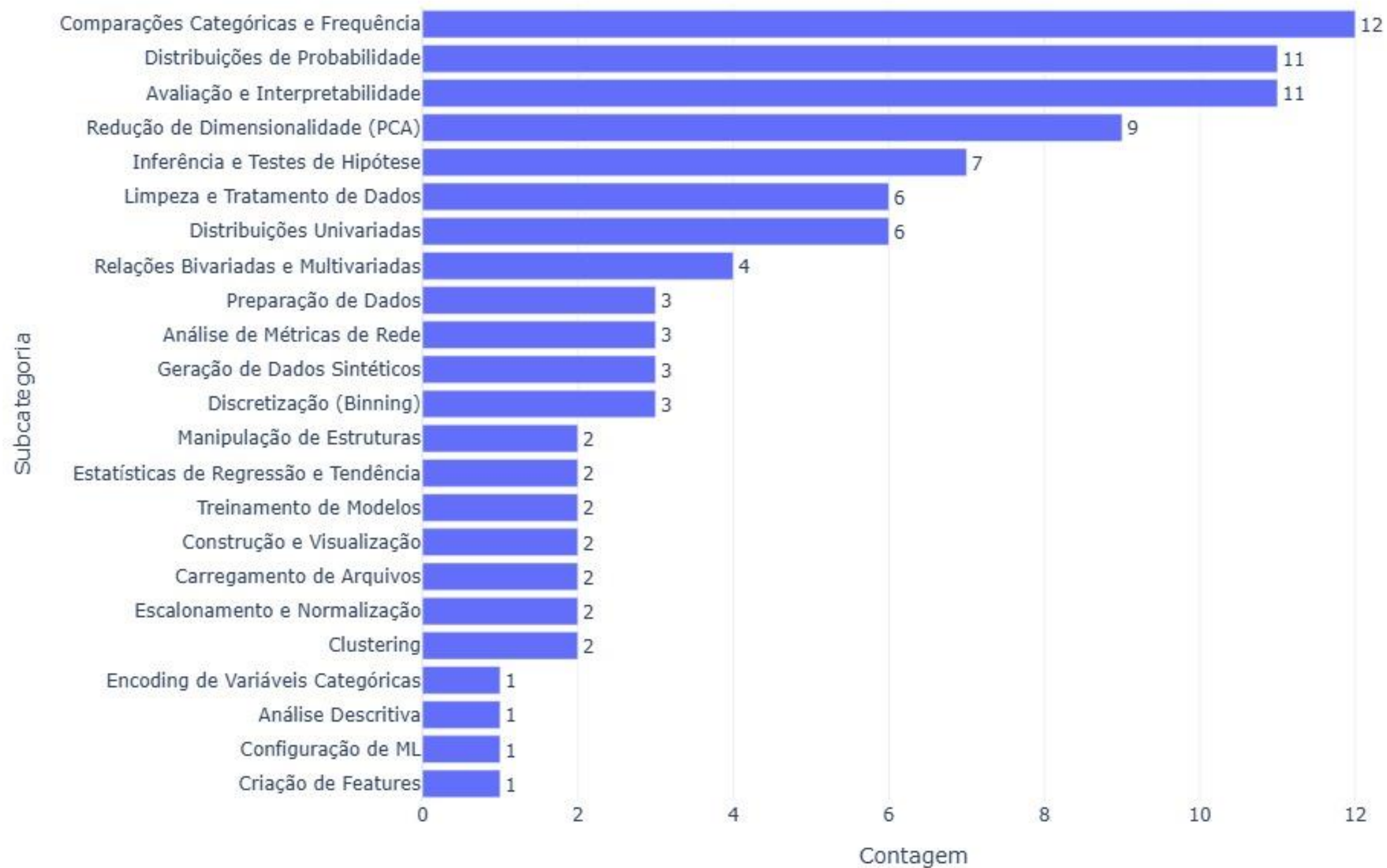


## Contagem de Funções por Categoria





## Contagem de Funções por Subcategoria



Distribuição de Funções (Categoria vs. Subcategoria)

Todas as Funções



Distribuição Geral do Tamanho das Funções (em Tokens)

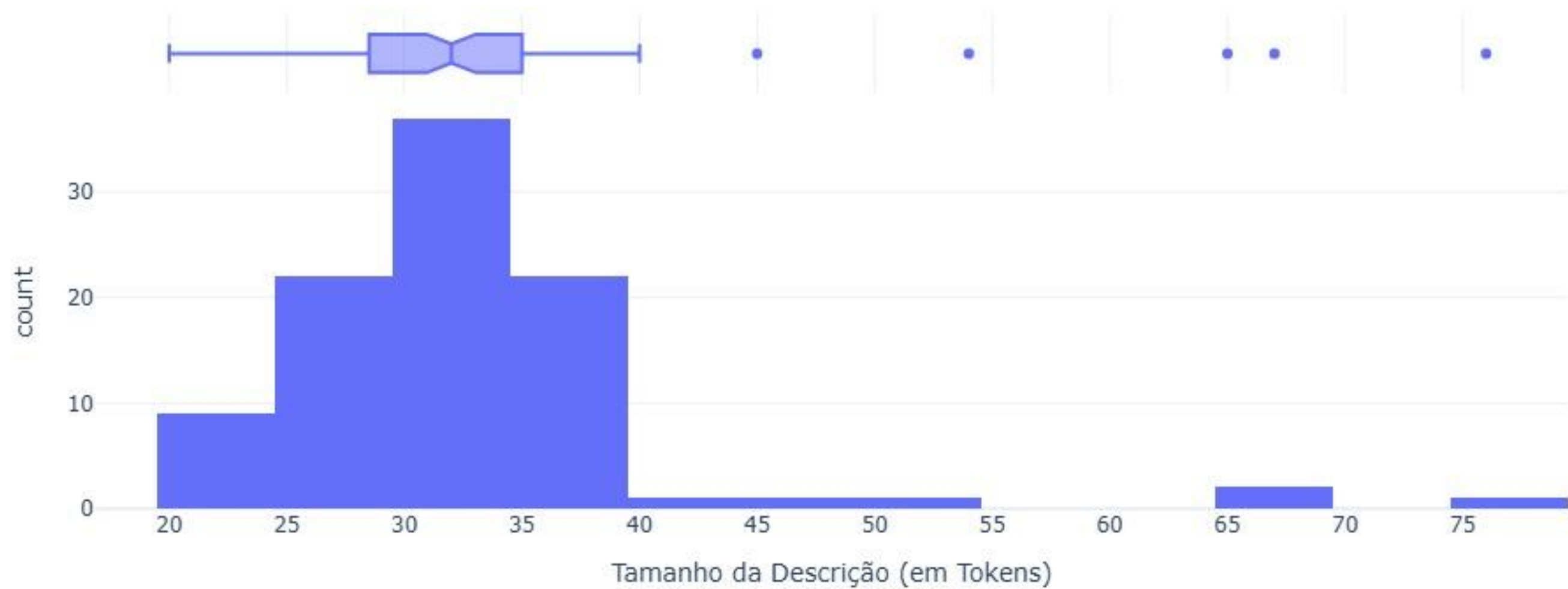




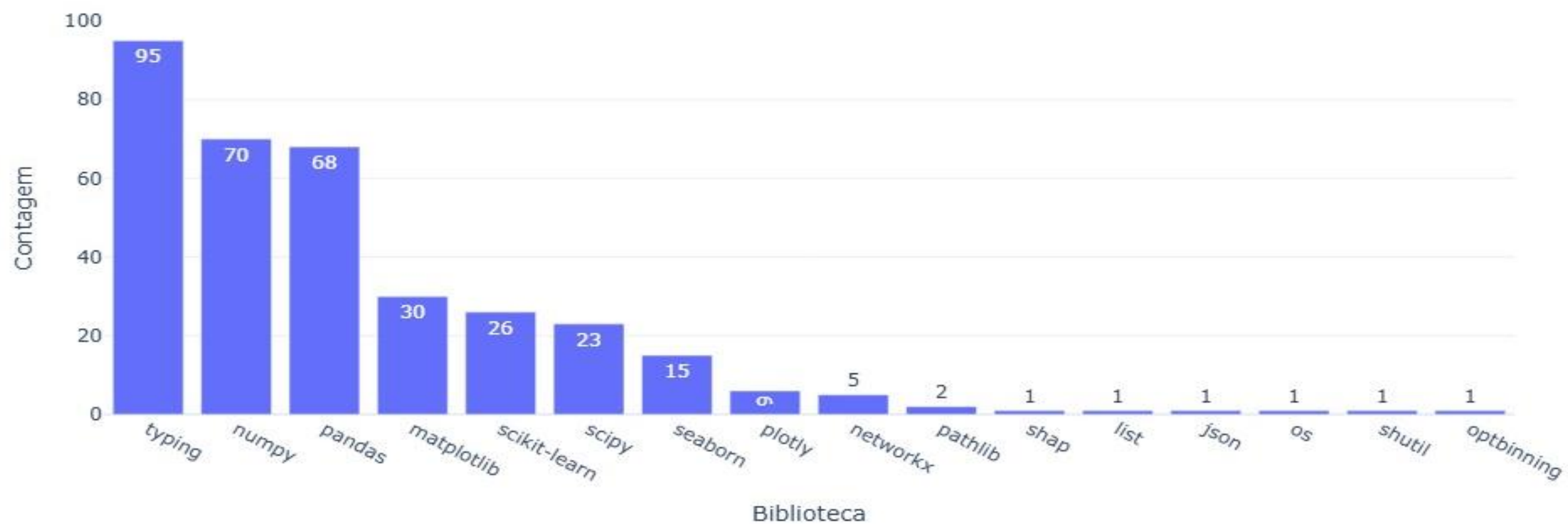
Distribuição do Tamanho das Funções por Categoria (em Tokens)



Distribuição do Tamanho das Descrições



## Bibliotecas Mais Comuns nas Funções do KB



# Estrutura da Base de Funções

```
def calculate_area(length: float, width: float) -> float:
    """Calculates the area of a rectangle.

    Args:
        length (float): The length of the rectangle.
        width (float): The width of the rectangle.

    Returns:
        float: The calculated area of the rectangle.

    Raises:
        ValueError: If either length or width is negative.
    """
    if length < 0 or width < 0:
        raise ValueError("Length and width must be non-negative.")
    return length * width
```

## Metadados

- ID único
- Título legível
- Categoria e subcategoria
- Descrição funcional
- Bibliotecas e versões

## Padrão de Código

- Tipagem 100% completa
- Docstrings Google Style
- Documentação clara
- Funções parametrizadas
- Código reutilizável