**Título do Trabalho**

**Afonso H. de Lima Santos1, Carla Monik G. do Nascimento1, Rafael da S. Ribeiro1, Bruno G. Araújo1**

1Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), Campus Canguaretama, BR-101, CEP: 59190-000, Km 160, S/N, Areia Branca, Canguaretama - RN, Brasil

h.afonso@escolar.ifrn.edu.br, carla.gomes@escolar.ifrn.edu.br, r.rafael@escolar.ifrn.edu.br, bruno.gomes@ifrn.edu.br

**Resumo**

A cada ano ocorre uma mudança nas principais raças encontradas em clubes de cães, visando entender melhor esse comportamento essa análise pode ajudar a conhecer as raças populares em determinados anos pode ajudar os clubes se preparam para receber esses cães permitindo que eles adequar sua estrutura física para melhor atender as peculiaridades e necessidades de cada raça com base nos dados analisados*.*

Palavras-chave: Cães. Clubes de Cães. Análise de dados. Dataset. *Python.*

# 1. Introdução

Clubes para cães são grupos organizados para donos de cães. Esses clubes oferecem uma variedade de atividades, incluindo treinamento de obediência, competições de habilidades, passeios em grupo e eventos sociais. Além disso, muitos clubes para cães também atuam como uma rede de apoio para seus membros.

Para isso é necessário oferecer informações e conhecer quais as raças mais populares em determinados anos e assim os clubes se preparem com informações e conselhos sobre cuidados com cães e outros assuntos relacionados a cães. Participar de um clube para cães pode ser uma ótima maneira de socializar seu cão, fazer novos amigos e aprender mais sobre seu animal de estimação.

Visando atender essa necessidade de conhecer quais cães são mais populares em cada um dos anos, veio a necessidade de utilização de análise de dados para acompanhar o comportamento da popularidade de cada raça nos principais clubes em determinados anos.

# 2. Objetivo Geral

O objetivo geral do presente artigo consiste em entender quais as principais raças de cães são mais populares em determinados locais.

# 3. Referencial Teórico

## 3.1. Análise de Dados com Python

*Python* é uma linguagem de programação de propósito geral, de alto nível e interpretada. Possui sintaxe clara e fácil de aprender, tornando-a amplamente utilizada em aplicações de software como desenvolvimento *web*, ciência de dados, inteligência artificial, entre outros. Além da sua sintaxe amigável, *Python* conta com uma vasta gama de bibliotecas pré-construídas, que oferecem recursos adicionais e facilitam o desenvolvimento de aplicações. Estas bibliotecas, testadas e utilizadas por muitos desenvolvedores, são uma grande vantagem pois economizam tempo e garantem a qualidade do código, já que são menos propensas a erros do que o código escrito por um único desenvolvedor.

E para complementar ainda mais as vantagens do uso do *Python*, vale destacar a utilização do *Jupyter Notebook*, que é uma ferramenta de desenvolvimento interativo e de análise de dados em *Python*. O *Jupyter* permite escrever, executar e documentar o código *Python* de forma clara e organizada, tornando-o uma escolha popular entre os cientistas de dados e outros profissionais que trabalham com análise de dados e visualização. Além disso, o *Jupyter* permite compartilhar facilmente o código e os resultados da análise com outros membros da equipe ou com a comunidade, tornando-o uma ferramenta colaborativa e acessível para desenvolvedores de todos os níveis.

Em resumo, o uso do *Python* juntamente com as suas bibliotecas e o *Jupyter Notebook* torna o desenvolvimento de aplicações mais rápido, eficiente e acessível, ampliando as capacidades da linguagem e possibilitando a criação de soluções avançadas para vários setores.

Outro ponto importante é que as bibliotecas têm uma ampla comunidade de usuários e desenvolvedores que podem fornecer suporte e ajuda. Entre as principais bibliotecas está a *NumPy* que é uma biblioteca de processamento numérico em *Python*, projetada para trabalhar com *arrays* multidimensionais de dados numéricos. É uma das bibliotecas mais importantes para a computação científica e análise de dados, oferecendo recursos avançados para operações matemáticas, estatísticas e lógicas em *arrays*, e Pandas é uma biblioteca de análise de dados em *Python* que fornece ferramentas eficientes e fáceis de usar para manipulação, processamento e análise de dados tabulares. A biblioteca oferece recursos para lidar com dados ausentes, trabalhar com séries temporais e realizar agregações e manipulações de dados.

Portanto, análise de dados é o processo de examinar, limpando, transformando e modelando dados para descobrir informações úteis, concluir tendências e tomar decisões baseadas em dados. É uma combinação de técnicas estatísticas, matemáticas e computacionais para transformar dados brutos em insights significativos e aplicáveis. A análise de dados pode ser aplicada em muitas áreas, como marketing, finanças, saúde, setor público, entre outras. A análise de dados envolve coletar dados, prepará-los para análise, realizar análises exploratórias e modelagem, e comunicar os resultados de forma clara e significativa. O objetivo final da análise de dados é fornecer informações valiosas que possam ser usadas para tomar decisões informadas e melhorar resultados em uma organização ou indústria.

# 3. Metodologia

O método de pesquisa utilizado neste artigo é a pesquisa aplicada, que busca solucionar problemas práticos através da combinação de teoria e prática. O objetivo desta pesquisa foi entender o registro de raças de cães e suas posições em clubes em determinados anos, a fim de fornecer soluções úteis e práticas para questões reais relacionadas aos clubes.

A pesquisa foi quantitativa, realizada com uma metodologia estruturada que incluiu a identificação do problema, coleta de dados, análise de um dataset e apresentação de resultados. O resultado final foi fornecer informações valiosas sobre os clubes de cães, para auxiliar na tomada de decisões.

# 4. Resultados

## 4.1. Preparação e Exploração dos dados

Para realizar a exploração dos dados foi utilizado uma base de dados pública sobre clubes de cachorros, uma base de dados é uma coleção organizada de informações armazenadas em um sistema informático. Ela permite aos usuários armazenar, acessar e gerenciar informações de forma eficiente.

Depois de escolher a base de dados, foi utilizada a linguagem *python*, em conjunto com algumas ferramentas como o pandas, matplotlib e *Jupyter*, que são ferramentas essenciais para se trabalhar com *Python*.

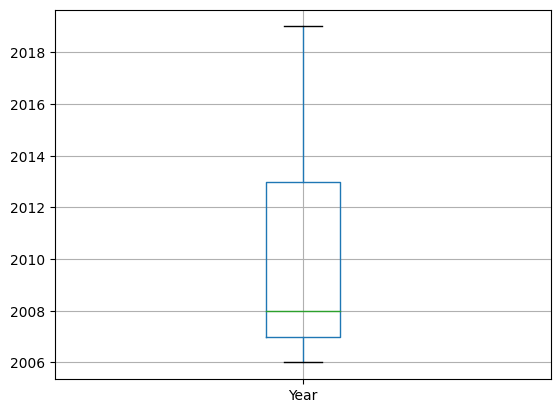
## 4.2. Análises

Análise é o processo de examinar dados para identificar padrões e tendências. É uma ferramenta valiosa para tomar decisões informadas, melhorar a eficiência e entender o desempenho. Nesse artigo, foram realizadas cinco análises, cujos resultados foram descritos.

A primeira análise foi para identificação e tratamento de valores ausentes/nulos. O resultado mostrou que o *dataframe* em análise não tinha valores nulos, mas na coluna "*Registrations*" foram detectados visualmente caracteres "-", " " e "." que dificultavam a análise. Para corrigir os dados, foram utilizados alguns métodos, que podem ser encontrados no repositório do projeto. Foram substituídos os valores "-" da coluna "*Registrations*" por "0", removemos o espaçamento e o ponto, e, por fim, mudaram o tipo de dado para inteiro. Assim, conseguimos realizar uma análise mais completa.

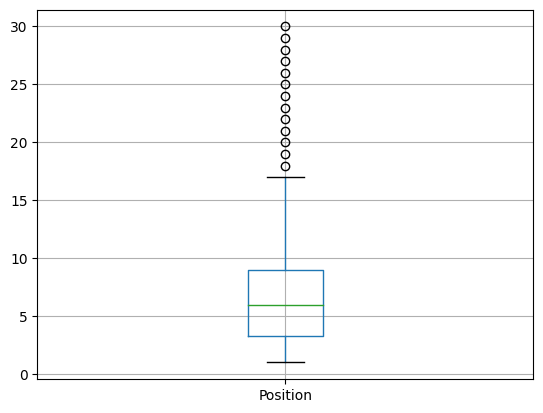
Uma segunda análise foi realizada para identificar e tratar valores atípicos, conhecidos como *outliers*. Eles são valores fora do padrão em relação ao restante dos dados. Para identificá-los, foi utilizado o método de "*boxplot*". Nele, os *outliers* são representados como pontos fora dos limites superiores e inferiores. Os resultados dos gráficos das colunas "Year", "Position" e "Registrations" não apresentaram valores anômalos em relação ao restante dos dados. Nessa análise, procuramos pelos *outliers* em "Year", "Position" e "Registrations", respectivamente, e os gráficos obtidos para cada caso podem ser vistos a seguir.

**Figura 1. *outliers* na coluna ano**



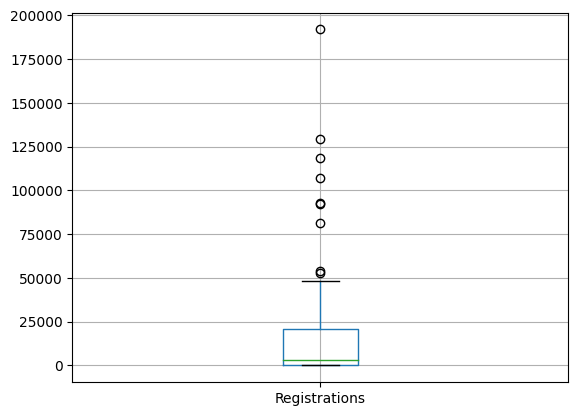
Fonte: Autoria própria (2023)

**Figura 2 . *outliers* na coluna posição**



Fonte: Autoria própria (2023)

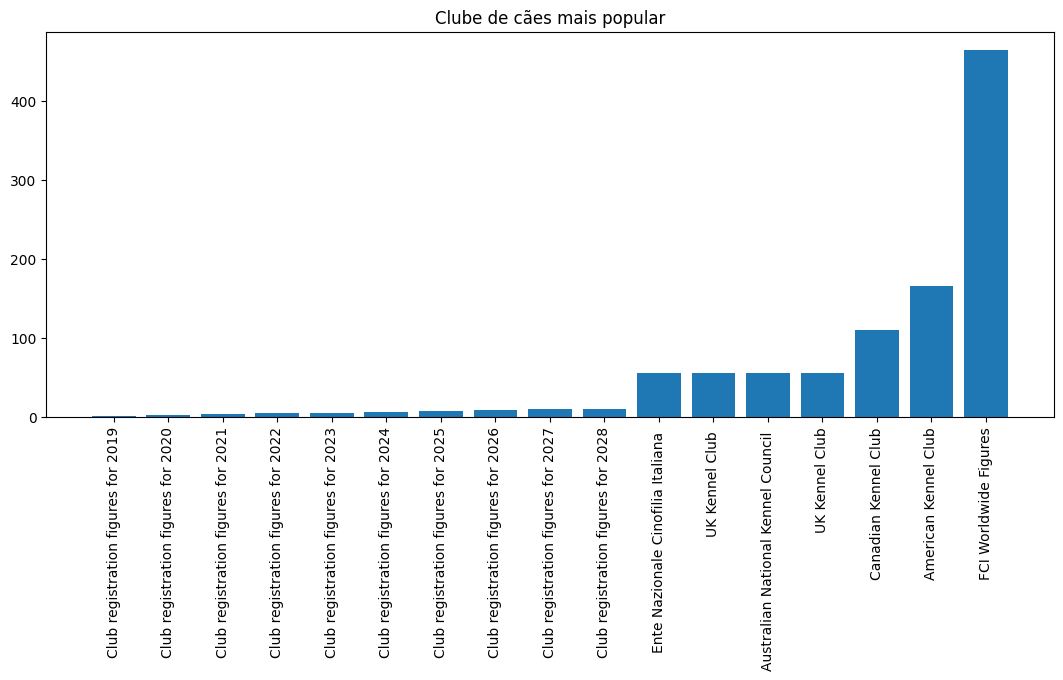
**Figura 3 . *outliers* na coluna registro**

****

Fonte: Autoria própria (2023)

Foi realizado um Tratamento em formato de dados para deixá-los adequados para a utilização. Como foi utilizado apenas um *dataframe* não foi necessário realizar a junção para verificar se vários *datasets* identificar chaves primárias e realizar a junção (*join*), outro ponto que não se fez necessário foi a *feature engineering.*  Por se tratar de um *dataframe* pequeno não foi necessário criar novos dados para substituir os originais. Foram realizadas 5 análises descritivas, a primeira análise foi para entender qual é o Clube de cães mais popular (Figura 4), como resultado tivemos o seguinte gráfico.

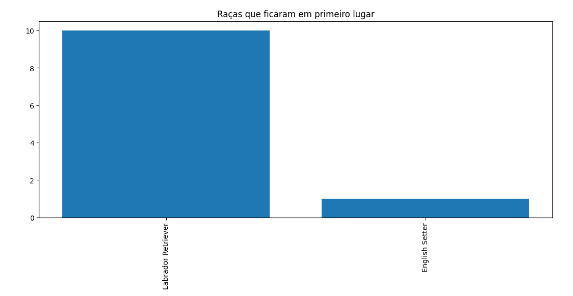
**Figura 4 . Clube de cães mais popular**



Fonte: Autoria própria (2023)

Sobre as raças que ficaram em primeiro lugar ao longo dos anos, concluiu-se que é o Labrador Retriever e English Setter ocupam os primeiro lugares, como podemos ver no gráfico (Figura 5).

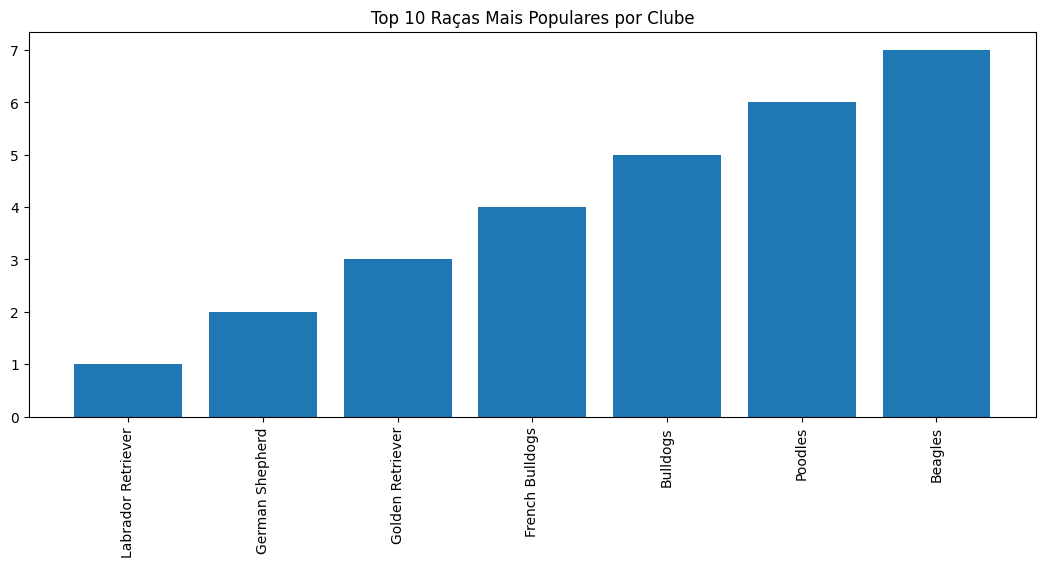
**Figura 5 . Raças que ficaram em primeiro lugar**

****

Fonte: Autoria própria (2023)

Ao realizar uma busca sobre a raça de cão mais popular em 10 clubes de cães? (Figura 6) gerar o gráfico nota-se que apenas 7 raças de cães apareceram, mas devido alguns clubes apresentarem a mesma posição para respectivas raças, ou seja, alguns clubes obtiveram as mesmas posições de cães mais populares.

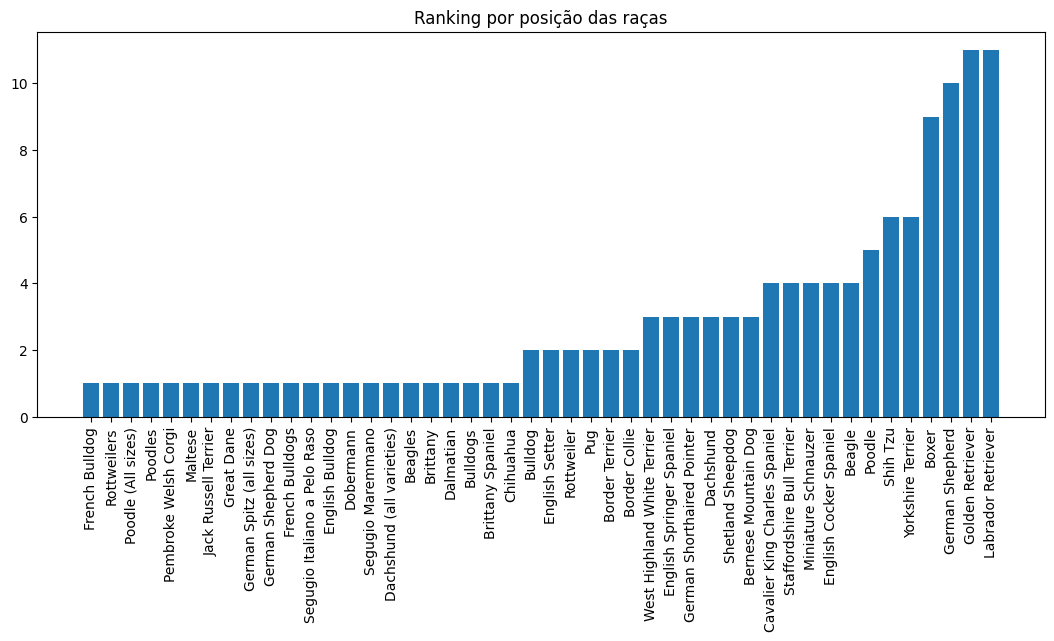
**Figura 6 .Raça de cão mais popular em 10 clubes de cães**



Fonte: Autoria própria (2023)

Outra relação importante que conseguimos obter foi qual é a relação entre a raça e a posição (Figura 7), obtivemos como resultado que raças pequenas estão em primeiro lugar.

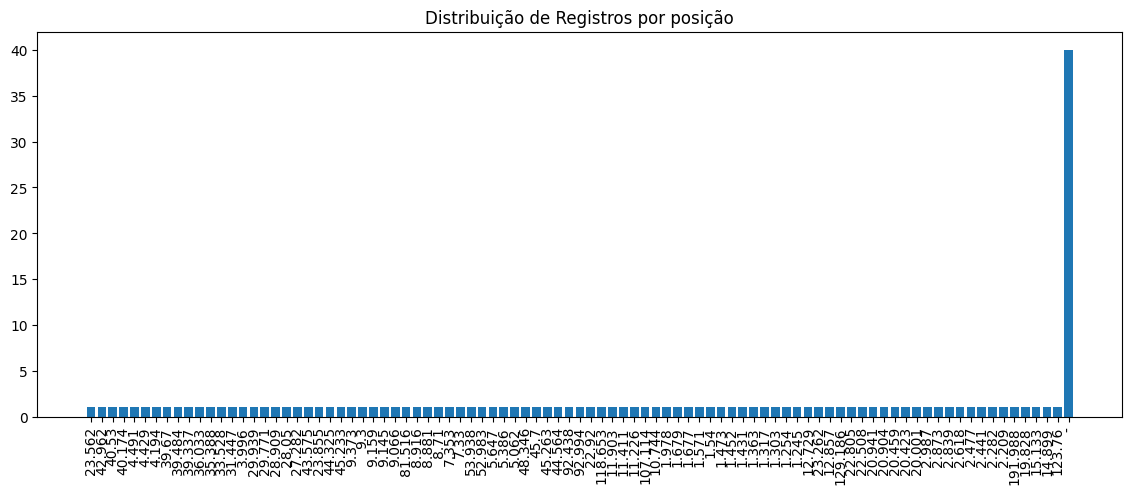
**Figura 7. Relação raça posição**



Fonte: Autoria própria (2023)

Sobre a distribuição de registros por posição (Figura 8) obtivemos os seguintes resultados que podem ser visualizados na imagem a seguir

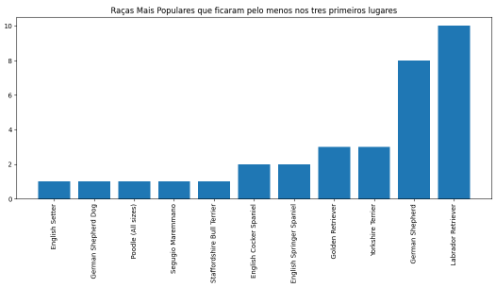
**Figura 8. Qual é a distribuição de registros por posição?**

****

Fonte: Autoria própria (2023)

Foi realizada uma nova análise onde revelou que as raças de cães mais populares que se mantiveram entre as três primeiras posições foram Labrador Retriever, German Shepherd e Yorkshire Terrier juntamente com o Golden Retriever ocupando a mesma posição no ranking. Esses resultados (Figura 9) mostram a consistência do apreço dos proprietários de animais pelas raças mencionadas.

**Figura 9. Raças Mais Populares que Ficaram pelo Menos nos Três Primeiros Lugares**

****

Fonte: Autoria Própria (2023)

**5. Conclusão**

Após a realização da análise de dados no dataset do Kaggle, foi possível concluir que as principais raças em primeira posição no ranking são raças de menor porte, mas as raças de maior porte são mais numerosas nos clubes. Esses resultados são importantes porque assim os clubes podem se preparar para receber essas raças de cães.

Além disso, durante o projeto, foi possível ver a importância da linguagem *Python*, que ajudou devido à sua capacidade de processar grandes quantidades de informações e ter uma sintaxe simples e fácil de aprender. Além disso, há uma grande comunidade de desenvolvedores que criam bibliotecas e ferramentas de análise de dados, tornando o processo mais fácil e acessível.

Em resumo, os resultados alcançados estimularam a continuidade desta análise, aprofundando a compreensão da relação entre os resultados encontrados nos clubes e os concursos de beleza para cães.

# 

# Referências

JUPYTER NOTEBOOK. serviços da Web para computação interativa. Disponível em:

<<https://jupyter.org/>> Acesso em: 24/01/2023

KAGLE. dataset para análise de dados. Disponível em:

<<https://www.kaggle.com>/> Acesso em: 24/01/2023

MATPLOTLIB. biblioteca abrangente para criar visualizações estáticas. Disponível em

<<https://matplotlib.org/>> Acesso em: 27/01/2023

PYTHON. linguagem de programação de alto nível. Disponível em:

<<https://www.python.org/>> Acesso em: 31/01/2023