

# Extração e Análise de Dados Médicos no Brasil

1<sup>st</sup> Matheus Henrique Alves  
Engenharia da Computação  
CEFET-MG  
Divinópolis, Brasil  
matheu98@hotmail.com

**Resumo**—O aumento do número de médicos no Brasil nos últimos anos tem contribuído para o crescimento da área da saúde, mas ainda existem desafios na escolha de um profissional qualificado. Este trabalho tem como objetivo consolidar informações públicas sobre médicos em uma base de dados estruturada, com foco no tempo de registro no Conselho Regional de Medicina (CRM) como um indicador de experiência profissional. Através da coleta de dados do portal do Conselho Federal de Medicina (CFM), seguido de tratamento e análise, busca-se desenvolver um sistema de recomendação que facilite a escolha de médicos, considerando fatores como especialidade, tempo de CRM e localização geográfica.

**Index Terms**—dados médicos, análise de dados, CRM, experiência profissional

## INTRODUÇÃO

A área da saúde no Brasil tem experimentado um crescimento expressivo no número de profissionais médicos nos últimos anos. Segundo a *Demografia Médica no Brasil 2023* [1], o número de registros de médicos com título de especialidade aumentou 84% na última década, projetando que o país alcançará a marca de 1 milhão de médicos até 2035.

Apesar desse aumento no quantitativo de médicos, a escolha de um profissional qualificado continua sendo um desafio para os pacientes. A diversidade de especializações, a distribuição desigual dos profissionais pelo território nacional e as disparidades salariais entre gêneros são fatores que complicam a decisão dos pacientes na seleção de um médico adequado [1]. A falta de informações centralizadas e acessíveis sobre a experiência e a qualidade dos serviços prestados pelos médicos contribui para essa incerteza, levando os pacientes a confiarem em critérios subjetivos ou recomendações limitadas.

Uma das formas de estimar a experiência de um médico é por meio do tempo de registro no Conselho Regional de Medicina (CRM), que reflete, de maneira indireta, sua trajetória profissional. Com o avanço das tecnologias de coleta e análise de dados, torna-se viável desenvolver ferramentas que auxiliem na identificação de profissionais com maior experiência, considerando parâmetros como especialidade, localização e tempo de registro.

## OBJETIVO

O objetivo deste trabalho é consolidar informações públicas disponibilizadas pelo Conselho Federal de Medicina (CFM) em uma base de dados estruturada, possibilitando a análise do tempo de registro no Conselho Regional de Medicina (CRM) como indicador da experiência profissional. Busca-se explorar

os dados coletados como base para o desenvolvimento futuro de um sistema de recomendação que auxilie na identificação de médicos qualificados, considerando parâmetros como tempo de CRM, especialidade e localização.

## METODOLOGIA

A metodologia será estruturada em três etapas principais: coleta de dados, tratamento e armazenamento das informações, e análise visual por meio de um dashboard interativo. Cada etapa será projetada para assegurar a organização e a qualidade dos dados, com foco na utilização de tecnologias acessíveis e robustas.

### A. Coleta de Dados

Os dados serão extraídos do portal oficial do Conselho Federal de Medicina (CFM) [2], que disponibiliza informações públicas sobre médicos registrados no Brasil. O portal oferece uma busca paginada, onde cada página exibe 10 médicos em formato de cartões contendo as informações relevantes, além de filtros como especialidade, situação do CRM, nome, número do CRM, UF e cidade.

Para garantir a qualidade e relevância das informações, serão aplicados dois filtros principais. O primeiro filtro é a restrição à unidade federativa (UF) de Minas Gerais (MG). O segundo filtro restringirá a coleta aos médicos com registro ativo no Conselho Regional de Medicina (CRM).

A coleta será realizada por cidade, com o filtro de cidade sendo aplicado em cada iteração da coleta, juntamente com os critérios de estado e situação do CRM. A navegação entre as páginas será feita por meio dos botões de navegação. Para evitar bloqueios automáticos por parte do site, a biblioteca `time` será utilizada para introduzir intervalos entre as requisições.

As informações a serem coletadas incluirão:

- Nome do médico;
- Número do CRM e tempo de registro;
- UF e cidade de atuação;
- Especialidade.

As bibliotecas necessárias para a realização da coleta incluem a biblioteca `Selenium` [3], que será utilizada para automação da navegação na web, e a biblioteca `time`, que será empregada para introduzir intervalos entre as requisições, ajudando a evitar bloqueios por parte do site.

Os dados capturados serão armazenados inicialmente em arquivos no formato CSV, e, após a coleta, serão submetidos

a um processo de tratamento para remoção de inconsistências e transformação, possibilitando uma análise precisa e eficaz.

## B. Tratamento de Dados

Os arquivos CSV resultantes da coleta serão consolidados em uma única base de dados utilizando a biblioteca *Pandas* [4]. Nesta etapa, serão realizados vários processos essenciais para garantir a qualidade e a consistência dos dados. Primeiramente, será realizada a limpeza dos dados, que inclui a remoção de duplicatas, o preenchimento de valores ausentes e a padronização dos formatos de texto.

Além disso, será realizada a extração de informações utilizando expressões regulares (por meio da biblioteca *re*) para padronizar o tempo de registro do CRM a partir de datas. A estruturação dos dados também será realizada, organizando-os em colunas padronizadas para facilitar tanto a análise quanto a integração com as ferramentas de visualização.

## C. Análise e Visualização

A base consolidada será importada para o Power BI [5], onde será desenvolvido um dashboard interativo para a visualização e análise dos dados. Esse dashboard fornecerá uma representação clara de métricas e percepções relacionados à distribuição dos médicos por especialidade e tempo de registro, à concentração geográfica dos profissionais no estado de Minas Gerais e às comparações entre cidades e regiões.

As visualizações permitirão uma exploração detalhada das informações coletadas, facilitando a identificação de padrões e tendências, que poderão ser utilizados para o desenvolvimento de ferramentas mais avançadas, aprimorando a compreensão dos dados e suportando decisões informadas.

## RESULTADOS ESPERADOS

Os resultados esperados deste trabalho incluem:

- 1) **Base de Dados Estruturada:** A criação de uma base de dados consolidada e bem estruturada, contendo informações sobre médicos registrados no Conselho Regional de Medicina (CRM), com dados como nome, especialidade, tempo de registro e localização geográfica. Esta base de dados permitirá análises aprofundadas e poderá ser utilizada para o desenvolvimento de futuras ferramentas.
- 2) **Ferramenta de Visualização:** O desenvolvimento de um dashboard interativo utilizando Power BI, que possibilitará a visualização e análise das informações coletadas. Espera-se que a ferramenta facilite a identificação de padrões na distribuição de médicos, como a concentração de profissionais por especialidade e região, além de destacar os médicos com maior tempo de experiência.
- 3) **Impacto no Processo de Decisão:** Dados que melhore a tomada de decisão dos pacientes na escolha de médicos, sendo dados confiáveis e detalhados. Espera-se que o Dashboard contribua para uma escolha mais informada e consciente, promovendo um atendimento de saúde de melhor qualidade.

Esses resultados poderão ser avaliados por meio de testes de usabilidade e feedback de usuários.

## CONCLUSÃO

O presente trabalho tem como objetivo suprir a lacuna existente na busca por médicos no Brasil, propondo uma solução que centraliza informações relevantes e amplamente disseminadas, tornando-as acessíveis em uma única plataforma. A proposta é a criação de um sistema de recomendação que, além de reunir dados importantes sobre os profissionais da saúde, permita ao usuário uma análise mais criteriosa e fundamentada ao selecionar um médico.

A implementação desse sistema não só facilita a busca por médicos, mas também otimiza a tomada de decisão dos pacientes, proporcionando-lhes um critério adicional relevante para a escolha de um médico. A centralização dos dados e a personalização da busca, com base em informações detalhadas e atualizadas, são aspectos fundamentais que garantem a qualidade e a eficiência do processo de recomendação. Com isso, espera-se que a ferramenta contribua para uma escolha mais consciente e informada, promovendo melhores resultados tanto para os pacientes quanto para os profissionais de saúde.

## TRABALHOS FUTUROS

Como perspectiva para trabalhos futuros, pretende-se desenvolver uma plataforma de recomendação de médicos, onde, além de sugerir profissionais com base em critérios objetivos, os pacientes poderão avaliar os médicos com os quais realizaram consultas. Este sistema de feedback permitirá que os usuários compartilhem suas experiências, oferecendo uma visão mais completa e colaborativa sobre os profissionais de saúde, o que contribuirá para uma escolha mais informada.

Além disso, o plano é integrar funcionalidades para agendamento de consultas diretamente pela aplicação, permitindo que os pacientes marquem suas consultas de forma prática e eficiente. Esse agendamento será vinculado ao sistema de recomendação, garantindo uma experiência fluida e personalizada para os usuários. Para viabilizar financeiramente a plataforma, a intenção é adotar um modelo de monetização por meio de parcerias afiliadas com médicos. Assim, cada consulta marcada através do aplicativo gerará uma comissão para a plataforma, criando uma fonte de receita sustentável.

## REFERÊNCIAS

- [1] SCHEFFER, M. et al. *Demografia Médica no Brasil 2023*. Conselho Federal de Medicina, 2023. Disponível em: <https://portal.cfm.org.br>. Acesso em: 05 dez. 2024.
- [2] Conselho Federal de Medicina (CFM). *Consulta a médicos*. Disponível em: <https://portal.cfm.org.br/>. Acesso em: 05 dez. 2024.
- [3] Selenium Software. *Selenium: WebDriver*. Disponível em: <https://www.selenium.dev/>. Acesso em: 05 dez. 2024.
- [4] WES McKINNEY. *Pandas: Powerful Python data analysis toolkit*. Disponível em: <https://pandas.pydata.org/>. Acesso em: 05 dez. 2024.
- [5] Microsoft. *Power BI*. Disponível em: <https://powerbi.microsoft.com/>. Acesso em: 05 dez. 2024.