AnalySUS

### Equipe **Outliers**

### Slides <https://docs.google.com/presentation/d/1SamhW__Wnulvbk0g4U8GcphSlAH1DOG1/edit?usp=sharing&ouid=111397257354245207001&rtpof=true&sd=true>

Formada por:

| Igor Duarte | <https://github.com/igorduartt> | [linkedin.com/in/igor-duarte-54478ba3](http://linkedin.com/in/igor-duarte-54478ba3) |
| --- | --- | --- |
| Rafael P. dos Santos | <https://github.com/Rafae1PS> |  |
| Josué dos Santos | <https://github.com/JosueSantos> | <https://www.linkedin.com/in/js-josu%C3%A9-santos/> |
| Vanessa Sharine | <https://github.com/VanSharine> | <https://www.linkedin.com/in/vanessa-sharine-c-63a95a1b2/> |
| Madson Dias |  |  |
| Aislan |  |  |

### SUS

O Sistema Único de Saúde (SUS) é um dos maiores e mais complexos sistemas de saúde pública do mundo, abrangendo desde o simples atendimento para avaliação da pressão arterial, por meio da Atenção Primária, até o transplante de órgãos, garantindo acesso integral, universal e gratuito para toda a população do país.

A Atenção Primária à Saúde (APS) é o primeiro nível de atenção em saúde e se caracteriza por um conjunto de ações de saúde, no âmbito individual e coletivo, que abrange a promoção e a proteção da saúde, a prevenção de agravos, o diagnóstico, o tratamento, a reabilitação, a redução de danos e a manutenção da saúde com o objetivo de desenvolver uma atenção integral que impacte positivamente na situação de saúde das coletividades. Trata-se da principal porta de entrada do SUS e do centro de comunicação com toda a Rede de Atenção dos SUS.

O foco na atenção primária se torna um dos pontos principais na logística de recursos e atendimento. Já que evitando que enfermidades venham a se agravar, estaremos evitando a ocupação de leitos de internações, aliviando todo o sistema em cadeia.

### Objetivos

Observando o poder computacional de auxílio na tomada de decisão, este projeto visa analisar os dados referentes às internações do SUS, levantando informações pertinentes para a melhoria no atendimento na atenção primária.

Utilizando-se de métodos preditivos de análise de dados é pretendido realizar um apontamento prévio com poucas informações iniciais no momento do atendimento primário, podendo sugerir uma maior prioridade em casos de maior probabilidade de uma futura internação.

### Dados

Para a análise foi utilizado dois datasets principais:

O primeiro ( **internacoes\_fort** ) possui todos os registros de internações de pessoas que residem/residiam em Fortaleza entre os anos de 2014 a 2019 de acordo com o Sistema de Informações Hospitalares do SUS do Departamento de Informática do SUS (SIH-SUS/DATASUS).

Para a criação do segundo Dataset ( **df\_icsap** ) foi utilizado um pacote em linguagem R que automatiza a classificação e descrição das Condições Sensíveis à Atenção Primária (CSAP) segundo a Lista Brasileira de CSAP que foi definida em 2008 através de portaria ministerial (Nº 221, DE 17 DE ABRIL DE 2008).

Para o auxílio dos dois datasets selecionamos alguns datasets secundários:

Utilizamos dados da prefeitura de Fortaleza disponíveis em <https://simda.sms.fortaleza.ce.gov.br/simda/populacao/faixa>, no site é possível exportar um arquivo com dados relacionados à população de cada bairro por faixa etária.

Criamos dois datasets ( **bairros** ) (Contém todos os ceps extraídos e seus bairros respectivamente e ( **pop\_bairros** ) contendo toda a população por bairros de acordo com o censo de 2010, para que seja possível calcular a taxa de internações por 100 habitantes.

Foi utilizado também a listagem da CID-10, que é a 10ª revisão da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde, uma lista de classificação médica da Organização Mundial da Saúde.

| **Dataset** | **Registros** | **Variáveis** |
| --- | --- | --- |
| internacoes\_fort | 867.018 | 113 |
| df\_ic\_icsap | 557.908 | 17 |
| bairros | 11.188 | 4 |
| pop\_bairros | 120 | 13 |
| cid | 14.230 | 4 |

### Formato da Análise Exploratória

A análise está dividida em duas partes: **1.** Limpeza e Mesclagem de dados; **2.** Descrição e análise utilizando os dados já focados na atenção primária.

Para orientar a exploração foi utilizado 11 perguntas guias:

| 1. Qual é o número de registros da base de dados selecionada? |
| --- |
| 2. Quais as idades dos pacientes de maneira geral? |
| 3. Quantos dias em média um paciente passou internado de acordo com o tipo de leito? |
| 4. Quem tem mais registros de internações? Homens ou mulheres? |
| 5. Qual o ranking de atendimentos de acordo com o hospital executante? |
| 6. Qual causa foi mais recorrente durante o período analisado? |
| 7. Qual dia da semana tem o maior fluxo médio de pacientes? |
| 8. Qual o tempo médio de permanência por hospital? |
| 9. Quais bairros tiveram mais registros de internações? |
| 10. Qual a quantidade de internações de acordo com a faixa etária dos pacientes? |
| 11. Quais os valores das AIHs de internações? |

Inicialmente não possuíamos um dataset filtrado apenas com as internações de Fortaleza. Era o arquivo ( **df\_internacoes** ) com 1,59 GB, 113 colunas e 2.956.343 registros de todo o Estado do Ceará. Foram removidas as linhas desnecessárias para o objetivo da nossa análise, e filtradas pelos pacientes que residiam em Fortaleza.

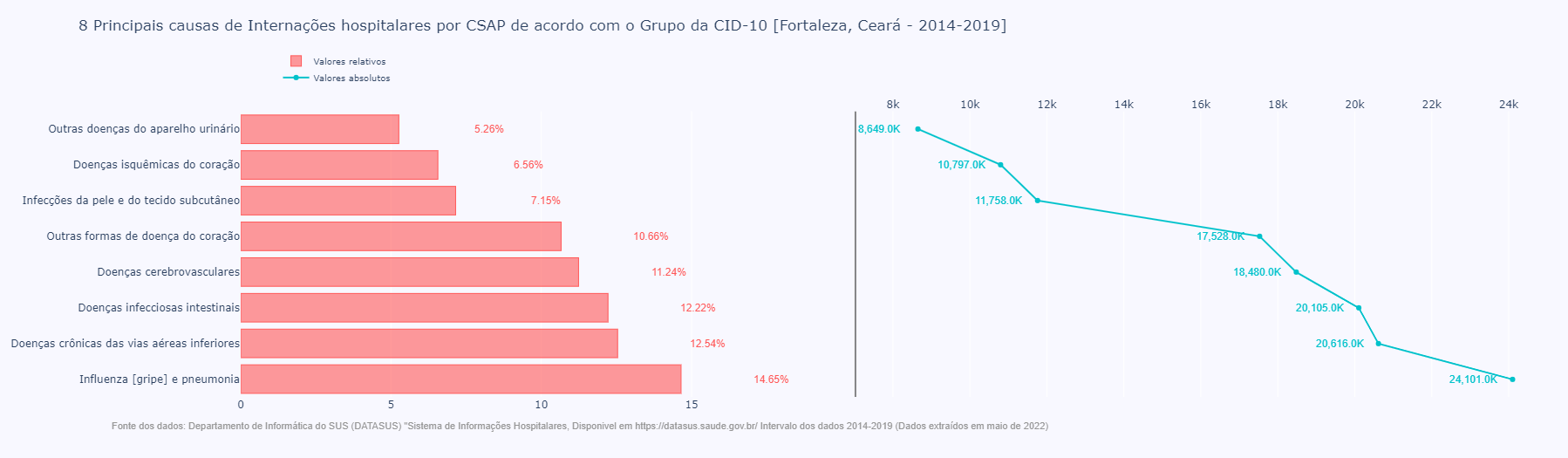
Também foi utilizado para a construção do segundo dataset o indicador ICSAP - Internações por Condições Sensíveis à Atenção Primária. Que é um indicador de avaliação de saúde baseado em um conjunto de enfermidades que, se tratadas de forma eficaz na Atenção Primária e em tempo oportuno, reduzem o risco de internações hospitalares. A utilização deste indicador permite que os gestores consigam melhorar o planejamento e a gestão dos serviços de saúde, já que possibilita avaliar o desempenho da Atenção Primária.

Com o dataset filtrado e o ICSAP, foram mesclados juntamente com as outras bases auxiliares. Após, foram tratados os dados para um formato padronizado obtendo um dataset com diversas variáveis.

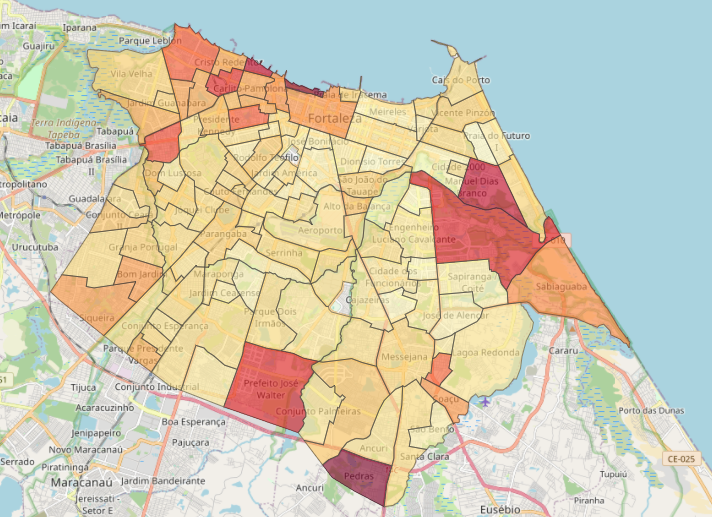
Para responder as perguntas feitas antes da análise, foram selecionadas 27 variáveis com o intuito de tornar o Dataframe mais parcimonioso. Foram elas:

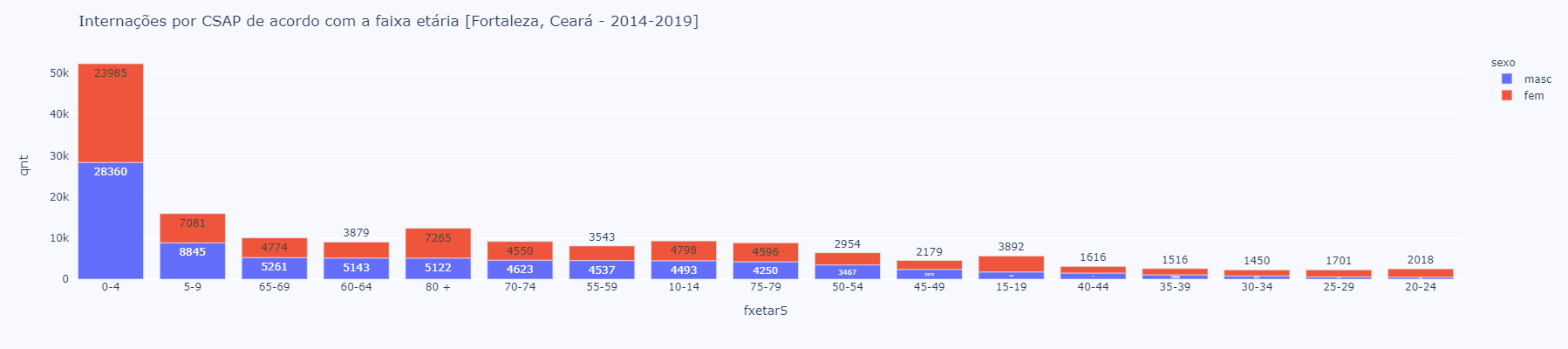
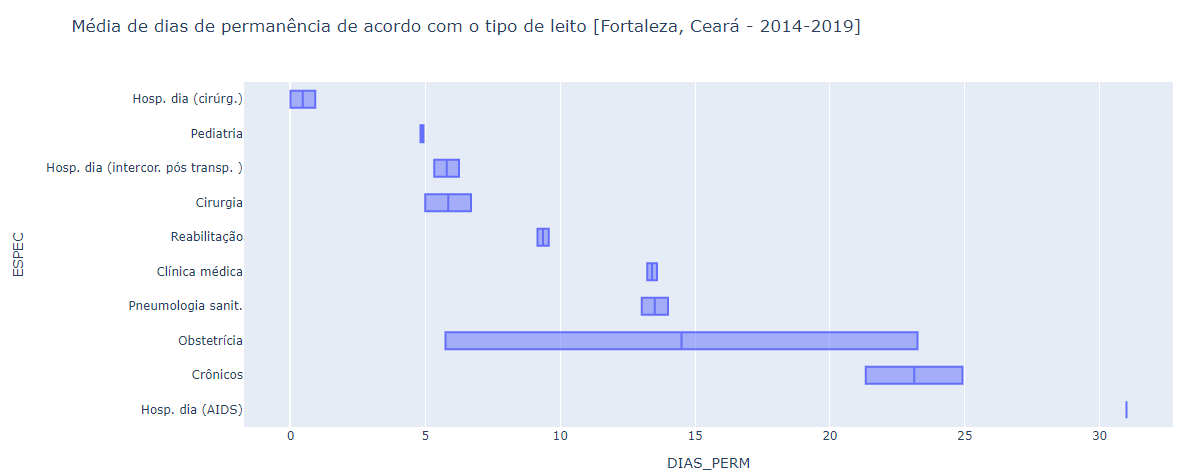
| ANO\_CMPT | O ano de registro da internação |
| --- | --- |
| MES\_CMPT | O mês de registro |
| ESPEC | Tipo do leito de internação |
| N\_AIH | Código da autorização de internação hospitalar |
| CEP | Código postal de residência do paciente |
| MUN\_RES | Município de residência do paciente |
| SEXO | Sexo do(a) paciente |
| IDADE | Idade do(a) paciente |
| PROC\_REA | Procedimento principal que foi realizado durante a internação |
| VAL\_TOT | Valor total da internação |
| DT\_INTER | Data da internação do(a) paciente |
| DT\_SAIDA | Data da alta |
| DIAG\_PRINC | Diagnóstico principal (de acordo com a CID-10) |
| MORTE | Indica se houve ou não a morte do paciente |
| CNES | Código do estabelecimento que realizou a internação |
| DIAS\_PERM | Dias de permanência |
| MUNIC\_MOV | Município onde o paciente foi internado |
| ETNIA | Etnia do(a) paciente |
| COD\_IDADE | Código da idade do paciente (Dias, meses ou anos) |
| IDH-B | Índice de desenvolvimento humano dos bairros (Disponível no site da prefeitura de fortaleza |
| pop\_total | População total e por bairro de acordo com o censo de 2010 |
| regional | Regional do bairro (1 a 12) |
| bairro | Bairro de residência do paciente |
| bairro\_formatado | Nome do bairro de acordo com site da prefeitura |

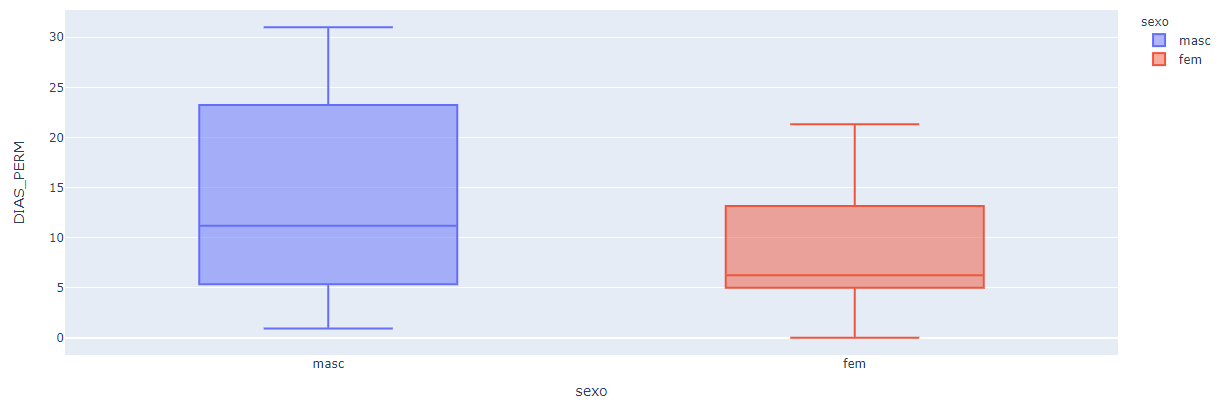
### Apontamentos



* 8 Grupos de diagnósticos resumem cerca de 80% do total de internações por condições sensíveis à Atenção Primária no território fortalezense.
* Bairros com menor IDH-B tendem a ter maiores taxas de internações, mas essa explicação não pode ser generalizada em todos os casos, por isso, é importante avaliar com quais variáveis as taxas se associam melhor. Moura Brasil possui a maior taxa por 1000 habitantes - cerca de 531 pessoas. (Acreditamos que pode ser resultado de erros de registros de CEPS, já que o bairro possui uma população relativamente menor em relação a outros bairros.

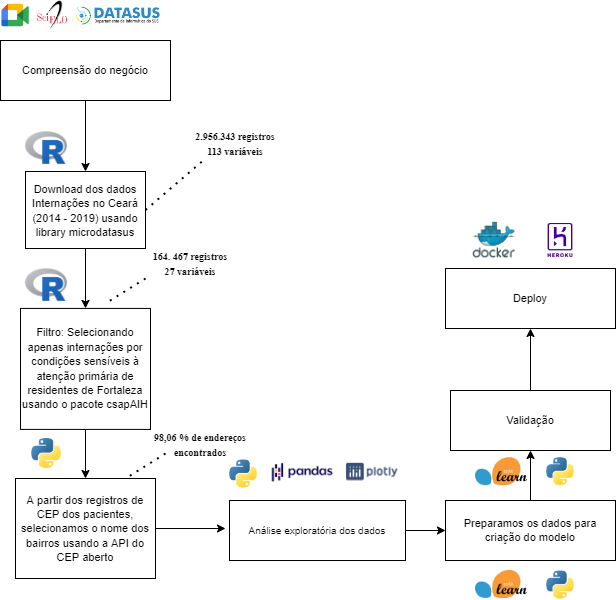


* Crianças de até 9 anos e idosos de 60 a 80 anos são os mais afetados com as doenças de causas evitáveis, uma das nossas hipóteses é de que essa população possui menores condições físicas e biológicas para resistir a doenças, além de ter um menor suporte de Atenção Primária.
* A partir de 2017 o número de internações volta a crescer em ambos os sexos.
* É possível observar que em diferentes sazonalidades alguns picos se repetem, demonstrando um certo padrão.
* Leitos do tipo obstétricos possuem uma maior variabilidade no número de internações e uma das maiores medianas. O grande número de partos realizados anualmente contribui de maneira significativa com a maior probabilidade de desenvolvimento de doenças relacionadas ao parto.
* Homens passam mais tempo internados por condições sensíveis do que as mulheres em Fortaleza.



Link Colab com gráficos interativos:

<https://colab.research.google.com/drive/16tCm4-0dpT3wk83A1FGxhSDL45NDCNFB?usp=sharing>



<https://scikit-learn.org/stable/auto_examples/applications/plot_cyclical_feature_engineering.html>

………………………………..

# AnalySUS

O Sistema Único de Saúde (SUS) é um dos maiores e mais complexos sistemas de saúde pública do mundo, abrangendo desde o simples atendimento para avaliação da pressão arterial, por meio da Atenção Primária, até o transplante de órgãos, garantindo acesso integral, universal e gratuito para toda a população do país.

A Atenção Primária à Saúde (APS) é o primeiro nível de atenção em saúde e se caracteriza por um conjunto de ações de saúde, no âmbito individual e coletivo, que abrange a promoção e a proteção da saúde, a prevenção de agravos, o diagnóstico, o tratamento, a reabilitação, a redução de danos e a manutenção da saúde com o objetivo de desenvolver uma atenção integral que impacte positivamente na situação de saúde das coletividades. Trata-se da principal porta de entrada do SUS e do centro de comunicação com toda a Rede de Atenção dos SUS.

O foco na atenção primária se torna um dos pontos principais na logística de recursos e atendimento. Já que evitando que enfermidades venham a se agravar, estaremos evitando a ocupação de leitos de internações, aliviando todo o sistema em cadeia.

## Objetivos

Observando o poder computacional de auxílio na tomada de decisão, este projeto visa analisar os dados referentes às internações do SUS, levantando informações pertinentes para a melhoria no atendimento na atenção primária.

Utilizando-se de métodos preditivos de análise de dados é pretendido realizar um apontamento prévio com poucas informações iniciais no momento do atendimento primário, podendo sugerir uma maior prioridade em casos de maior probabilidade de uma futura internação.

* Captar dados da base oficial do Governo
  + Selecionar as bases necessárias
  + Identificar bases secundárias para complementar a informação
* Realizar a análise exploratória
  + Identificar variáveis e descrevê las
  + Tratar valores faltantes
  + Realizar comparações
  + Gerar gráficos e anotações
* Criar um modelo de Regressão capaz de predizer as internações de dois meses adiante
  + Preparar os dados
  + Selecionar possíveis modelos
  + Selecionar o modelo mais eficiente

## Conteúdo

Notebooks