

# Relatório Técnico

**Bonifácio de Oliveira, Lucas Araújo, Luiz Fábio, Luiz Felipe, Rafaela Melo**

<sup>1</sup>Escola Superior de Tecnologia – Universidade do Estado do Amazonas (UEA)  
Manaus – AM – Brazil

{bldof.eng16, lsda.lic16, lfba.lic17, lfda.lic17, rmf.lic16}@uea.edu.br

**Abstract.** *This technical report describes the data analysis process for COVID-19 in Amazonas, through exploratory analysis and data visualization.*

**Resumo.** *Este relatório técnico descreve o processo de análise de dados referentes ao COVID-19 no Amazonas, por meio de análise exploratória e visualização de dados.*

## 1. Introdução

O presente relatório tem como objetivo apresentar uma análise exploratória de registros de casos pessoas com COVID-19. O banco de dados utilizado<sup>1</sup> está em constante atualização, contabilizando sempre novos registros, por isto, as análises deste relatório foram realizadas com o download do banco no dia 12/08/2020. Os códigos fontes desenvolvidos são de código aberto e estão disponíveis na plataforma Github<sup>2</sup>. As análises foram feitas através da linguagem de programação *Python*, sendo utilizadas as bibliotecas: *Pandas*, para manipular os dados; *Numpy*, para realizar alguns cálculos específicos; e *Matplotlib* e *Seaborn*, para gerar gráficos.

## 2. Visão Geral dos Casos Confirmados

O dataset apresenta 40 atributos, são eles: idade, faixa etária, sexo, bairro, classificacao, comorb\_renal, comorb\_diabetes, comorb\_imuno, comorb\_cardio, conclusao, dt\_notificacao, taxa, dt\_evolucao, raca, dt\_sintomas, criterio, sintoma\_garganta, sintoma\_dispneia, sintoma\_febre, sintoma\_tosse, sintoma\_outros, etnia, profiss\_saude, srag, se\_notificacao, distrito, bairro\_mapa, comorb\_respiratoria, comorb\_cromossomica, comorb\_hepatica, comorb\_neurologica, comorb\_hemato, comorb\_obesidade, origem, evolucao, teste\_pcr, teste\_anticorpo, teste\_antigeno, teste\_igm e teste\_igg. Em Manaus, há 39037 casos confirmados, o caso mais antigo é datado em 01/04/2020 e o mais recente em 12/08/2020.

Considerando a tarefa de análise proposta, foi realizada uma limpeza nos dados retirando atributos que não influenciam nos resultados e as linhas que não continham todos os dados preenchidos. Após a limpeza e organização restaram 11 atributos e 54782 exemplos, classificados em: 14226 casos em análise, 31942 casos descartados e 8614 casos confirmados.

Os números apresentados a partir daqui são referentes ao dataset resultante após a limpeza e organização dos dados. Do total de 8614 casos confirmados, 94.87% das pessoas se recuperaram. O número de casos no sexo feminino se sobressai ao sexo masculino,

<sup>1</sup>Disponível em: <https://covid19.manaus.am.gov.br/wp-content/uploads/Manaus.csv>

<sup>2</sup>Disponível em: [https://github.com/BonifacioDeOliveira/PP1\\_RNA\\_UEA](https://github.com/BonifacioDeOliveira/PP1_RNA_UEA)

totalizando 4715 mulheres infectadas. Com relação a idade dos indivíduos contaminados, a média é 42 anos e o desvio-padrão é 15 anos, a pessoa mais jovem a contrair a doença tem 0 anos e a mais idosa 105 anos.

O bairro com maior incidência de casos é o Cidade Nova. Os três bairros com maior número de recuperados são: Cidade Nova (371), São José Operário (321) e Flores (314). A Tabela 1 apresenta a quantidade de testes de cada tipo e a porcentagem de cada um deles com relação ao total de 49354 testes realizados. Observa-se que o teste Anticorpo foi o mais utilizado e o IGG, o menos utilizado.

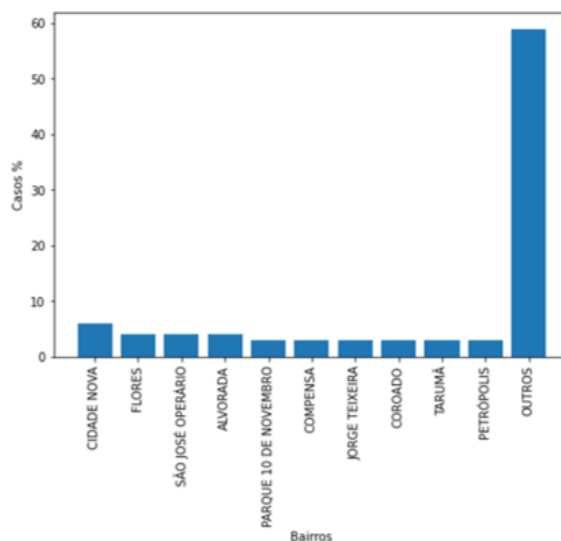
**Tabela 1. Testes Realizados**

Tipo de Teste	Quantidade	Porcentagem em relação ao total (%)
PCR	11699	23.7
Anticorpo	31656	64.14
Antígeno	5905	11.96
IGM	76	0.15
IGG	18	0.03

A taxa de letalidade calculada a partir da fração de óbitos pelo total de casos confirmados é de 5.13%. O cálculo da correlação entre a idade e o número de casos resultou em -0.2329, uma correlação negativa fraca.

### 3. Visualização de Dados

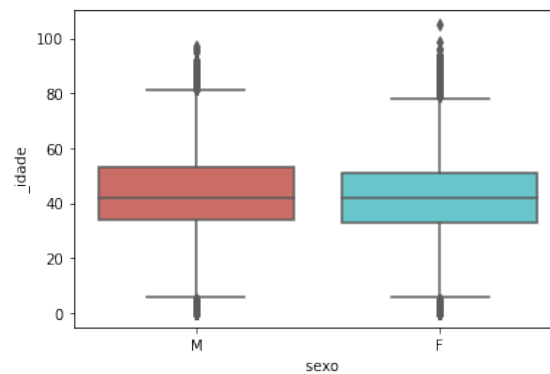
A Figura 1 apresenta um histograma informando os 10 bairros com maiores quantidades de casos registrados. A Figura 1 também indica que o bairro com maior quantidade de casos registrados foi o “Cidade Nova”. A categoria “Outros” trata da soma dos bairros remanentes.



**Figura 1. Porcentagem de casos por bairros.**

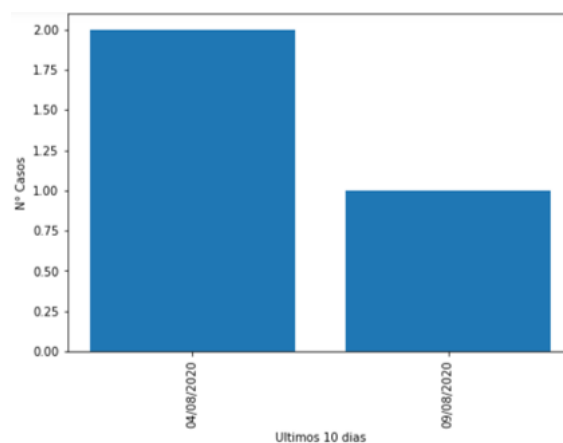
Analisando a Figura 2, representando o *boxplot* de idade com casos confirmados por sexo, é possível observar que os casos de sexo Feminino e Masculino apresentaram

uma variabilidade bem similar. Além disso ambos os sexos apresentaram valores discrepante, *outliers*, acima de 80 anos no sexo Masculino e um pouco abaixo também dos 80 anos no sexo Feminino, e ambas apresentaram *outliers* acima de 0.



**Figura 2. Boxplot da idade de casos confirmado.**

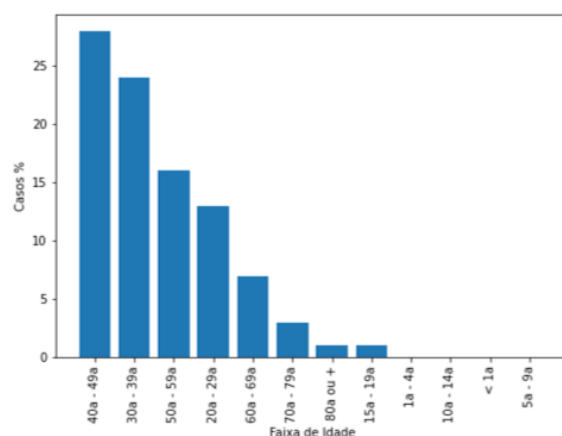
A Figura 3 é um gráfico de barras que informa a quantidade de casos entre os períodos 03/08/2020 a 12/08/2020, mas após a limpeza dos dados, apenas duas datas foram consideradas, são elas: 04/08/2020 e 09/08/2020. A Figura 3 informa que, dentre as datas, o dia 04 foi a data com mais ocorrências (com 2 casos confirmados de COVID-19).



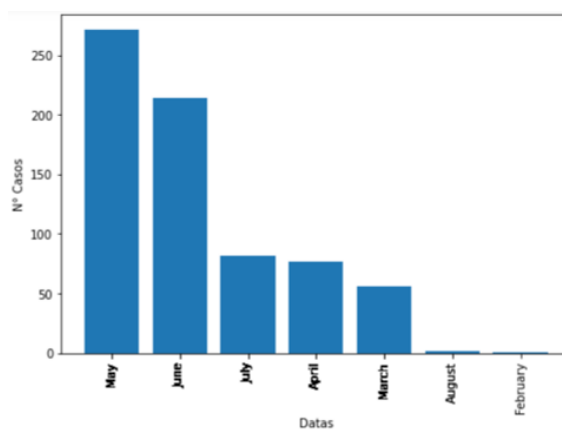
**Figura 3. Porcentagem de casos por bairros.**

Em relação aos casos que obtiveram a conclusão igual a “Recuperados” durante o período entre os dias 03/08/2020 a 12/08/2020, o data set estudado não obtinha casos marcados com recuperados durante esse período. Desta forma não foi possível gerar um gráfico similar a análise anterior pois a quantidade de casos recuperados por dia entre o período citado retornava vazio.

A Figura 4 informa a porcentagem de casos registrados por faixa etária (classificadas em décadas). Na Figura 4 as faixas etárias foram ordenadas de acordo com seus números de casos. É possível identificar que mais de 25% dos casos registrados foram de pessoas entre 40 - 49 anos de idade.

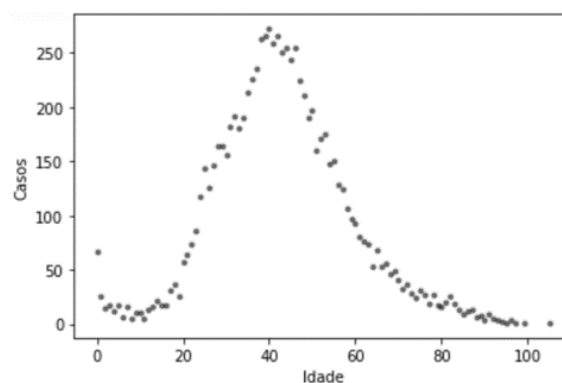


**Figura 4. Porcentagem de casos por faixas etárias.**



**Figura 5. Número de casos ao longo do tempo.**

A Figura 5 mostra as ocorrências de casos registrados ao longo do tempo (classificados por meses). A Figura 5 indica que “Maio” foi o mês com mais ocorrências registradas (mais de 250 casos).



**Figura 6. Número de casos registrados por idade.**

A Figura 6 apresentada a quantidade de casos por idade, semelhante a Figura 4 que informa as faixas etárias. A Figura 6 informa que as maiores quantidade de casos foram de pessoas com idade perto dos 40 anos. É possível observar uma tendência para

esta idade, visto que, as idades mais distantes dos 40 anos possuem menos casos.

#### **4. Tipos de Tarefas**

Considerando a grande demanda de teste para a Covid-19 e o número insuficientes de testes disponíveis, umas das formas de prevenir a contaminação de outras pessoas seria levar em consideração os sintomas dos pacientes que ainda não puderam ser testados colocando-os em observação e isolamento preventivo, onde o atributo-alvo é a Classificação, para tentar prever a possibilidade uma pessoa estar contaminada ou não. Para isso seriam levados em consideração os atributos relacionados aos sintomas que o paciente apresenta como: `_sintoma_garganta`; `_sintoma_dispneia`; `_sintoma_febre`; `_sintoma_tosse`; `_sintoma_outros`;,. Como métrica de desempenho a utilização da Revocação é a mais recomendada por se tratar de um diagnóstico médico visando prever uma pessoa possivelmente contaminada. O tipo de validação do modelo pode ser o K-Fold devido á grande quantidade de dados, a possibilidade de múltiplos testes e também a necessidade dos resultados do modelo serem o mais realístico possível, evitando diagnósticos incorretos.

O crescente número de mortos pela Covid-19 tem criado uma crise no sistema funerário, visto que muitas agências funerárias e cemitérios não estavam preparados para um cenário tão catastrófico como esse. Para evitar o problema de falta de caixões, covas etc., a regressão poderia ser utilizada para prever o número de mortos previstos para que os agentes funerários estejam preparados e todos tenham um sepultamento digno e sem frustrações. Os atributos preditores podem ser a data de notificação, data de óbito e a evolução da doença, fazendo uma projeção do crescimento do número de mortes em um determinado tempo no futuro.

Na Aprendizagem supervisionado poderíamos ter como resultados a categorização dos elementos em faixas etárias, com crianças, adolescentes, adultos e idosos, sendo possível visualizar o tamanho de cada grupo.