



UNIVERSIDADE
LUSÓFONA

Estudo de Chamadas de Emergência do INEM

Trabalho Final de Curso

Relatório Final

Nome do Aluno: Tiago van Krieken

Equipa de Orientação: Daniel Fernandes, Iolanda Velho

Trabalho Final de Curso | LEI | 30 de junho de 2023

www.ulusofona.pt

Direitos de cópia

Aplicação de *Machine Learning* para a previsão do número de chamadas recebidas do INEM, Copyright de Tiago Faria Paulo van Krieken, ULHT.

A Escola de Comunicação, Arquitetura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona de Humanidades e Tecnologias (ULHT) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Índice

Índice de Imagens	4
Índice de Equações.....	5
Índice de Tabelas.....	6
Resumo.....	7
Abstract	8
1 Identificação do Problema	9
2 Viabilidade e Pertinência.....	11
3 Benchmarking.....	12
4 Método e Planeamento	14
5 Engenharia.....	15
5.1 Recolha e Explicação de Dados	15
5.2 Análise dos Dados	16
5.3 Tecnologias e Ferramentas Utilizadas.....	21
5.3.1 <i>Python</i>	21
5.3.2 Modelos de <i>Machine Learning</i>	21
6 Solução Desenvolvida.....	22
6.1 Análise Exploratória de Dados	23
6.2 Abrangência	33
7 Resultados.....	34
7.1 Modelo de previsão do número de chamadas	34
7.2 Aplicação da Fórmula de <i>Erlang C</i>	39
8 Conclusão	40
8.1 Trabalho Futuro.....	41
8.2 Limitações	43
Bibliografia	44
Glossário.....	46

Índice de Imagens

<i>Figura 1 - Aumento do Número de Chamadas Atendidas Anualmente (fonte dos dados: inem.pt [3])</i>	10
<i>Figura 2 - Fluxo de Trabalho</i>	11
<i>Figura 18 - Calendário Gant</i>	14
<i>Figura 3 – Primeiras linhas do dataSet utilizado</i>	15
<i>Figura 4 - Tipos de Correlação de Variáveis</i>	16
<i>Figura 5 - Exemplo de Regressão Linear</i>	17
<i>Figura 6 - Número de Chamadas Recebidas</i>	23
<i>Figura 7 - Reta de Regressão do número total de Chamadas</i>	24
<i>Figura 8 - Reta de Regressão do número total de Chamadas (sem 2020/2021)</i>	25
<i>Figura 9 - Média do número de Chamadas recebidas por Mês, entre 2012 e 2022</i>	26
<i>Figura 10 - Reta de Regressão Polinomial de janeiro</i>	28
<i>Figura 11 - Reta de Regressão Polinomial de janeiro, excluindo os anos 2020 e 2021</i>	28
<i>Figura 12 - Reta de Regressão Polinomial de fevereiro</i>	29
<i>Figura 13 - Reta de Regressão Polinomial de fevereiro, excluindo os anos 2020 e 2021</i>	29
<i>Figura 14 - Reta de Regressão Polinomial de março</i>	30
<i>Figura 15 - Reta de Regressão Polinomial de março, excluindo os anos 2020 e 2021</i>	30
<i>Figura 16 - Flutuação do número de Chamadas por mês entre 2016 e 2012</i>	32
<i>Figura 17 - Associação das matérias usadas para a formulação do projeto</i>	33
<i>Figura 19 – Dataset usado para obter os modelos de Machine Learning usados</i>	34
<i>Figura 20 – Gráfico de Autocorrelação do número de chamadas por dia</i>	42
<i>Figura 21 – Gráfico de Autocorrelação do número de chamadas por semana</i>	42
<i>Figura 22 - Gráfico de Autocorrelação do número de chamadas por mês</i>	42

Índice de Equações

<i>Equação 1 – Formula Erlang C.....</i>	<i>13</i>
<i>Equação 2 – Reta de Regressão Linear.....</i>	<i>17</i>
<i>Equação 3 - Reta de Regressão Polinomial de 2º Grau</i>	<i>18</i>
<i>Equação 4 - Erro Médio Absoluto Percentual</i>	<i>19</i>
<i>Equação 5 – Correlação de Pearson</i>	<i>20</i>
<i>Equação 6 - Coeficiente de Determinação.....</i>	<i>20</i>

Índice de Tabelas

<i>Tabela 1 – Dicionário dos Dados</i>	<i>15</i>
<i>Tabela 2 - Resultados de previsão de todos os anos</i>	<i>31</i>
<i>Tabela 3 - Resultados de previsão de todos os anos, excluindo 2020 e 2021.....</i>	<i>31</i>
<i>Tabela 4 – Resultados da previsão diária do modelo de Machine Learning.....</i>	<i>36</i>
<i>Tabela 5 - Resultados da previsão semanal do modelo de Machine Learning</i>	<i>37</i>
<i>Tabela 6 - Resultados da previsão mensal do modelo de Machine Learning</i>	<i>38</i>

Resumo

Este projeto está a ser desenvolvido em conjunto com o Instituto Nacional de Emergência Médica (INEM), no sentido de utilização de modelos de *Machine Learning* para a identificação de alguma casuística das chamadas recebidas, com o propósito de prever o número de chamadas do CODU (Centro de Orientação de Doentes Urgentes) e assim sugerir melhorias na gestão de recursos do mesmo.

De acordo com o site do inem.pt [1] e 112.pt [2], o número “112” é o Número Europeu de Emergência, sendo comum para situações como saúde, incêndios, assaltos ou roubos. As chamadas efetuadas para o 112 são atendidas pela PSP e pela GNR, nas Centrais de Emergência. Caso essa chamada seja relacionada com a saúde, é encaminhada para os Centros de Orientação de Doentes Urgentes do INEM, onde é feita uma triagem através de perguntas feitas pelos especialistas, como:

- Tipo de Situação;
- Número de telemóvel;
- Localização Exata;
- Sexo e idade da pessoa que necessita de ajuda;
- A existência de qualquer situação que exija outros meios para o local, por exemplo, libertação de gases, perigo de incêndio, etc.

Após ser feita essa triagem, os operadores dos CODU desempenham um papel fundamental ao avaliar cuidadosamente a situação e fornecer orientações precisas para a população. Com base nas informações fornecidas, os operadores determinam a melhor forma de proceder, garantindo a prestação de assistência adequada e, se necessário, o envio dos meios de socorro apropriados. Desta forma, torna-se fundamental conseguir prever o número de chamadas para tomada de decisão e gestão de recursos.

Abstract

This project is being developed with the National Institute of Medical Emergency (INEM), in order to use *machine learning* models to identify some case series of incoming calls, with the purpose of predicting the number of calls per shift of the CODU (Center for Guidance of Urgent Patients) and thus suggesting improvements in the management of resources.

According to the inem.pt website [1] and 112.pt [2], the number "112" is the European Emergency Number, being common for situations such as health, fires or robberies. All calls made to the 112 are answered by PSP and GNR, in emergency centers. If this is health-related, it is referred to INEM's Urgent Patient Guidance Centers (CODU), where screening is performed through questions asked to specialists, such as:

- Type of Situation;
- Mobile number;
- Exact Location;
- Gender and age of the person requiring help;
- Existence of any situation that requires other means to the site, for example, release of gases, danger, etc.

After this screening is done, CODU operators play a key role in carefully assessing the situation and providing accurate guidance to the population. On the basis of the information provided, the operators shall determine the best way to proceed, ensuring that appropriate assistance is provided and, if necessary, that appropriate means of assistance are deployed. Knowing this, it becomes essential to be able to predict the number of calls for decision making and resource management.

1 Identificação do Problema

O Instituto Nacional de Emergência Médica (INEM) é o organismo do Ministério da Saúde responsável por coordenar o funcionamento, no território de Portugal Continental, de um Sistema Integrado de Emergência Médica (SIEM), de forma a garantir às vítimas de doença súbita a pronta e correta prestação de cuidados de saúde.

O INEM recebeu cerca de 1 370 milhões chamadas no ano 2021, ou seja, cerca de 4 mil chamadas por dia (de acordo com as estatísticas disponíveis em INEM.pt [3]). A elevada quantidade de chamadas e a conhecida falta de recursos leva à necessidade de obter a maior informação possível sobre as chamadas para uma rápida decisão por parte dos responsáveis. Com base na informação publicada em “Indicadores de Desempenho do INEM” [3], é possível prever a existência de alguns constrangimentos no CODU, que se podem dever a:

Aumento do número de chamadas anuais/mensais.

- De acordo com os dados existentes recolhidos em [3], apresentados na tabela da *Figura 1*, existiu um crescimento de 21,15% entre 2012 e 2019 do número de chamadas atendidas. É também apresentado na mesma figura a linha de tendência, cujo declive indica que o volume de chamadas atendidas tem aumentado ao longo dos anos, entre 2012 e 2019. No ano de 2020 existe uma queda do número de chamadas, ano esse afetado pela pandemia COVID-19, que poderá ter influenciado os dados. No ano de 2021 para 2022 nota-se novamente um crescimento nas chamadas atendidas, com base no número de chamadas mensais em período homólogo, prevendo-se assim, um aumento do número de chamadas atendidas até final do corrente ano e no decorrer do próximo ano.

Prevê-se que este aumento de chamadas se relacione com os seguintes fatores:

- Aumento da taxa de envelhecimento; entre os Censos de 2011 e 2021 existe um aumento significativo, pois esta população é mais sensível e utilizadora de cuidados de saúde de emergência. Ainda acordo com os Censos de 2021, há 182 idosos por cada 100 jovens no país [4];
- Alterações climáticas, com o aumento da temperatura média do planeta é previsível o aumento de doenças [5].

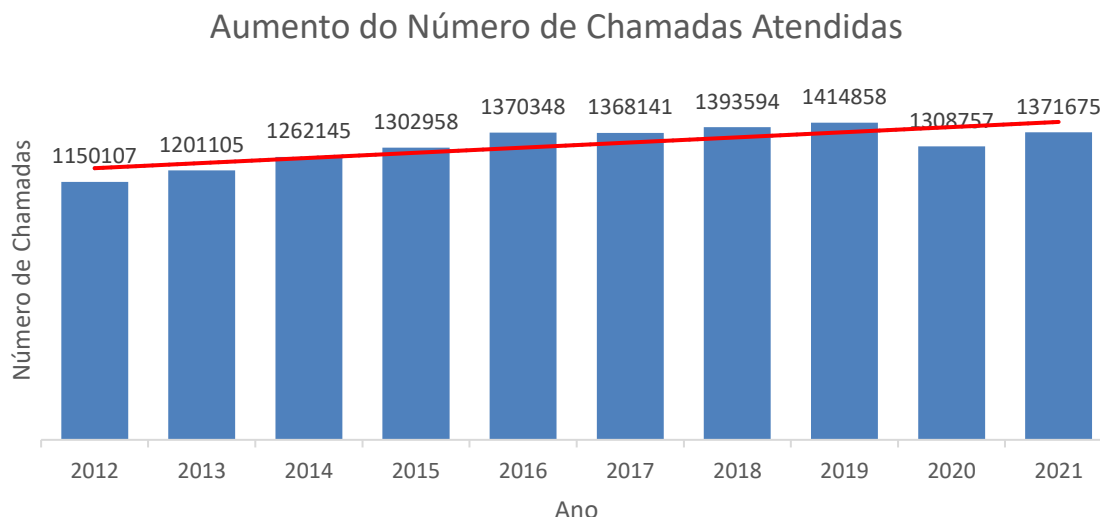


Figura 1 - Aumento do Número de Chamadas Atendidas Anualmente (fonte dos dados: inem.pt [3])

Falta de Recursos, causando problema com o tempo de resposta.

- De acordo com a notícia do Expresso, 10% das ambulâncias em Portugal estão neste momento paradas devido à falta de técnicos [6]. Com esta falta de recursos, que neste caso é a falta de recursos humanos, para além de tornar o tempo de resposta ao utente maior, vai incrementar o número de chamadas das próprias vítimas, que poderão ligar várias vezes para saberem o ponto de situação da ambulância encaminhada para as mesmas.

O objetivo deste projeto é analisar os dados, que trará conhecimento que poderá levar a uma melhoria de recursos e, consequentemente, a uma melhor resposta às necessidades da população. Resposta essa que se pretende que seja mais rápida e eficaz em caso de emergência, identificando pontos fortes e menos fortes nos dados obtidos através dessas chamadas. O projeto tem como objetivo geral aprimorar a eficiência do sistema de atendimento de emergência, proporcionando um maior segurança e conforto à população em situações de risco.

2 Viabilidade e Pertinência

Este projeto é relevante devido ao facto de ajudar a uma maior celeridade no atendimento das chamadas, podendo ter um impacto positivo no controlo e organização destas chamadas de emergência. Assim, prevê-se que a população poderá ter um melhor encaminhamento para o serviço de urgência necessário, saindo ambos, o INEM e o cidadão, beneficiados.

A viabilidade deste projeto é determinada pela sua capacidade de implementação destas medidas em qualquer momento, de forma prática e com aplicação, assegurando que este serviço seja constantemente monitorizado após a conclusão do mesmo. A garantia de uma implementação contínua e uma monitorização efetiva são elementos essenciais para assegurar a eficácia e a sustentabilidade do serviço ao longo do tempo. Além disso, o acompanhamento regular permitirá identificar possíveis oportunidades de melhoria, possibilitando ajustes e aperfeiçoamentos contínuos dos modelos para atender às necessidades em constante evolução. Por fim, este trabalho abrange estes dois pontos, pelo facto do INEM ter feito o pedido do mesmo e de estar a colaborar no seu desenvolvimento.

Este trabalho tem como propósito analisar os dados fornecidos pelo INEM, a fim de fortalecer a fundamentação, enfatizando a sua de extrema pertinência. Na *Figura 2*, é demonstrado como o fluxo de trabalho se inicia com a recolha dos dados e a análise dos mesmos, e após várias fases de desenvolvimento, chegaremos ao objetivo final de melhoria de recursos.

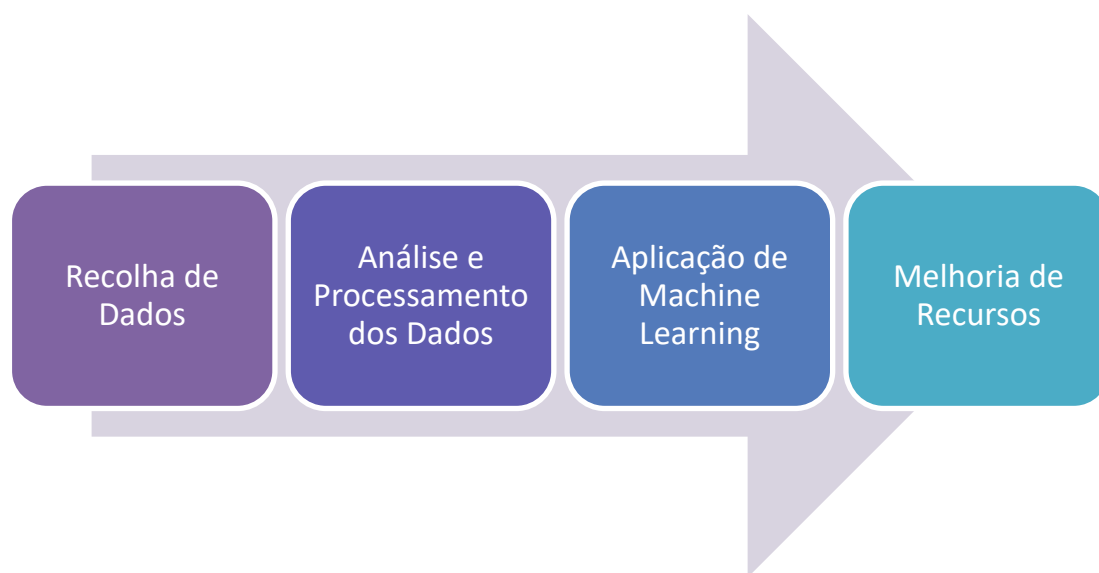


Figura 2 - Fluxo de Trabalho

3 Benchmarking

Nos dias de hoje, existem inúmeros estudos sobre *Machine Learning* e como os seus vários métodos se integram nas nossas vidas, como facilitadores [7]. *Machine Learning* utiliza aprendizagem empírica de acontecimentos passados, onde tira uma conclusão aproximada acerca do futuro desses acontecimentos, tais como, prever o número de chamadas recebidas de um determinado *call center*, entre outros. Tendo estes métodos o objetivo de prever algo, ajudam-nos também a poupar tempo e recursos, melhorando a *performance* do sistema.

O perfil de tráfego de chamadas recebidas de um *call center*, como o do INEM, tem tendência a ser bastante dinâmico. Mais precisamente, a intensidade das chamadas recebidas varia entre diferentes alturas do ano, mês, semana e dia. Por esta razão, diferentes quantidades de recursos humanos e técnicos do *call center* são necessários em diferentes períodos. Para planejar de uma forma mais otimizada os recursos do *call center*, o estudo [8] demonstra que em primeiro lugar é necessário prever o número de chamadas recebidas.

Em [8], tal como no nosso estudo que está a ser desenvolvido, pretende prever o número de chamadas recebidas de um *call center* (sendo que o INEM também recebe as suas chamadas de um *call center*), aplicando modelos de *Machine Learning*, sendo este o principal objetivo do projeto em desenvolvimento. Em [9] são aplicados modelos de *Machine Learning* como *Addictive Regression* e *Random Forest*, onde este último modelo é o único demonstrado visualmente. Este modelo é utilizado em ambas tarefas de classificação e regressão, onde várias árvores de decisão são construídas usando um subconjunto aleatório de recursos e amostras dos dados. As previsões destas árvores são então combinadas para produzirem uma previsão final. É também utilizada uma técnica que tem como objetivo melhorar a *performance* dos modelos de *Machine Learning* chamada *bagging*.

É importante ter conhecimento do número de chamadas recebidas num *call center* para saber ou calcular o número mínimo de operadores necessários para esse *call center*. O estudo [9] utiliza modelos de *Machine Learning* como ARIMA para prever o número de chamadas recebidas e, conseqüentemente, estimar o número mínimo de operadores sabendo esse número de chamadas, registadas a cada 30 minutos, utilizando também a fórmula *Erlang C*, que será explicada com mais detalhe em [10].

O estudo [9] fornece uma explicação detalhada sobre a utilização do modelo ARIMA, que, assim como a regressão linear, é um modelo utilizado para análise e previsão de séries temporais, sendo capaz de identificar padrões nos dados e utilizá-los para realizar previsões de valores futuros de forma precisa.

Para conseguirmos calcular o número de operadores necessários, teremos de utilizar a fórmula de *Erlang C*, apresentada em *Equação 1*, como é demonstrado no estudo [10] e já referido no estudo [9]. Esta fórmula é uma equação matemática que é utilizada para calcular o número de operadores que são necessários num *call center*, dado o número de chamadas e o nível de serviço que se deseja alcançar. Também permite calcular a probabilidade do tempo de espera de uma chamada (P_w), como se pode verificar em *Equação 1*, dada a intensidade de tráfego (A) e o número de operadores (N) disponíveis. Com base nos valores colocados, é possível determinar a probabilidade de, que a chamada recebida seja imediatamente atendida ou não.

$$P_w(N, A) = \frac{\frac{A^N N}{N! (N - A)}}{\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N N}{N! (N - A)}}$$

Equação 1 – Formula Erlang C

Neste momento, em Portugal não existe um estudo que analise este número de chamadas recebidas do INEM e que preveja esse número para uma melhor gestão de recursos, utilizando modelos de *Machine Learning*. Por esse motivo, este trabalho representa uma grande importância, não só para *call centers*, mas também para a população portuguesa.

4 Método e Planeamento

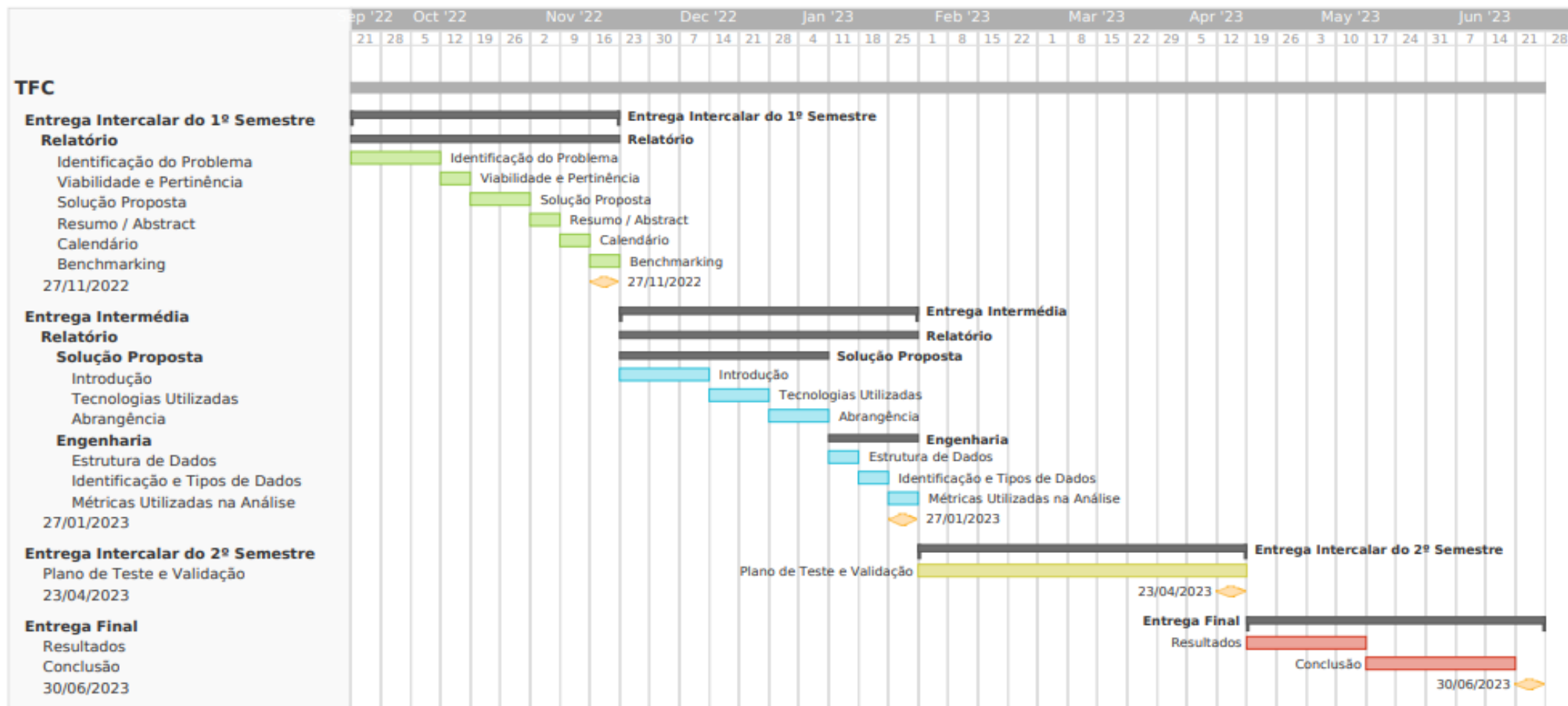


Figura 3 - Calendário Gant

5 Engenharia

5.1 Recolha e Explicação de Dados

Foram recolhidos do site oficial do INEM [3] o número total de chamadas recebidas por dia durante 10 anos, de 2012 a 2022, e organizados numa tabela num ficheiro *excel* (.xlsx). Este *dataSet*, que é um conjunto de dados organizados num formato para facilitar a análise e interpretação dos mesmos, é feito de 6 colunas e de 4 019 linhas, tendo sido construído a partir de uma recolha efetuada manualmente, e, por cada mês, exportada uma tabela com o número de chamadas efetuadas por dia desse mês. No final, foram reunidos todos os dados numa só tabela, encontrando-se o início da mesma na *Figura 4*.

Dia Semana	Dia	Mês	Ano	Data	Nº de Chamadas
Domingo	1	1	2012	01/01/2012	4538
Segunda-feira	2	1	2012	02/01/2012	3263
Terça-feira	3	1	2012	03/01/2012	3126
Quarta-feira	4	1	2012	04/01/2012	3199
Quinta-feira	5	1	2012	05/01/2012	2999
Sexta-feira	6	1	2012	06/01/2012	3354
Sábado	7	1	2012	07/01/2012	3035
Domingo	8	1	2012	08/01/2012	2896
Segunda-feira	9	1	2012	09/01/2012	3257
Terça-feira	10	1	2012	10/01/2012	3126

Figura 4 – Primeiras linhas do dataSet utilizado

De seguida, na *Tabela 1*, podemos observar a explicação de cada coluna do *dataSet* construído, apresentado na *Figura 4*.

Nome	Tipo	Significado
Dia Semana	String	Dia da semana da data referida na coluna “Data”
Dia	Int	Dia da data referida na coluna “Data”
Mês	Int	Mês da data referida na coluna “Data”
Ano	Int	Ano da data referida na coluna “Data”
Data	Date	Data, colocado por dia/mês/ano
Nº de Chamadas	Int	Número de chamadas recebidas da data referida na coluna “Data”

Tabela 1 – Dicionário dos Dados

5.2 Análise dos Dados

- **Análise Exploratória de Dados:**

Inicialmente, foi realizada uma análise exploratória de dados, que consiste em examinar as características do conjunto de dados das chamadas [11]. Esta análise foi feita por meio de métodos visuais, como gráficos e tabelas, permitindo a compreensão das fontes de dados, descobrir padrões, detetar anomalias e testar hipóteses. A análise exploratória de dados pode ser univariada, focada em uma única variável, ou multivariada, envolvendo duas ou mais variáveis simultaneamente. Esta análise utiliza técnicas gráficas, como tabelas cruzadas e histogramas, para visualizar a distribuição e frequência dos dados. No caso específico do INEM, um histograma pode ilustrar a frequência das chamadas recebidas em diferentes intervalos de tempo.

- **Análise de Correlação de Variáveis:**

Pode-se definir como correlação a medida estatística usada para avaliar se o valor de uma variável impacta no valor de outra variável, isto é, se existe uma correlação linear entre ambas as variáveis [12]. Essa correlação pode ser:

- **Correlação Positiva:** O comportamento da variável Y segue o comportamento da variável X , isto é, quando a variável Y aumenta, a variável X também aumenta, ou vice-versa, como demonstrado na Figura 5.
- **Correlação Negativa:** O comportamento da variável Y segue o comportamento oposto da variável X , isto é, quando a variável X aumenta, a variável Y diminui, como demonstrado na Figura 5.
- **Correlação Nula:** O comportamento da variável Y não segue o comportamento da variável X , como demonstrado na Figura 5.

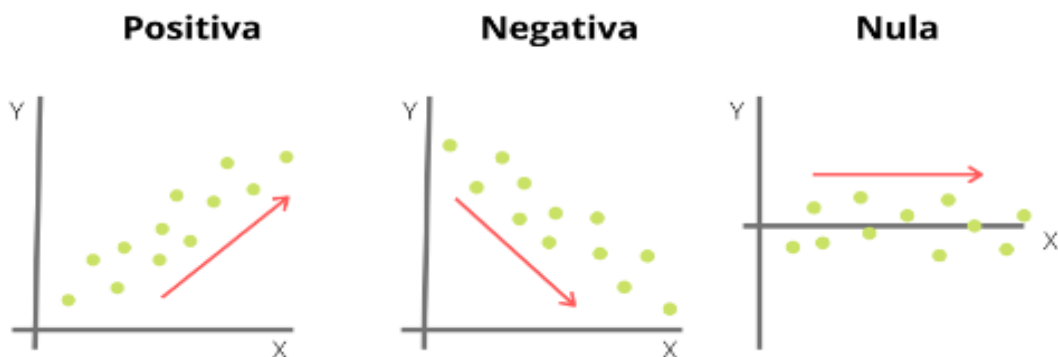


Figura 5 - Tipos de Correlação de Variáveis

- **Modelos estatísticos:**

- **Regressão Linear:**

Um dos modelos que iremos utilizar será a **Regressão Linear**. A regressão linear é definida como um algoritmo que proporciona uma relação linear entre uma ou várias variáveis independentes e uma variável dependente para prever o resultado de eventos futuros [13]. Este modelo tem como objetivo prever um valor que não se consegue concluir inicialmente, utilizando a seguinte função, demonstrada na *Figura 6*.

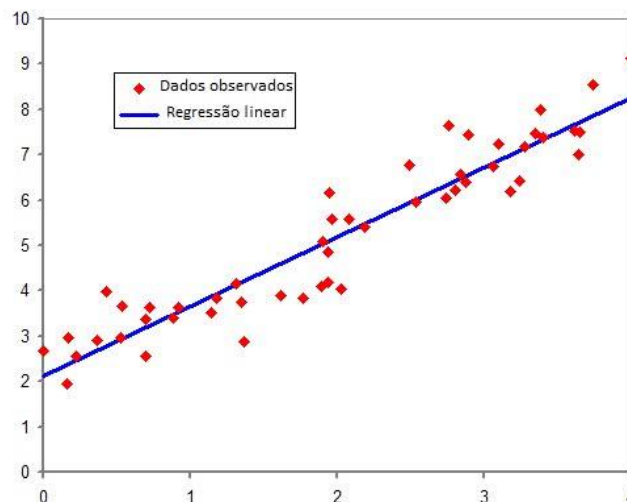


Figura 6 - Exemplo de Regressão Linear

Para calcular a reta de regressão linear, como mostrado na *Figura 6*, é necessário calcular os parâmetros da *Equação 2*:

$$y_i = \alpha + \beta x_i + \mu_i$$

Equação 2 – Reta de Regressão Linear

Onde:

- y_i : Variável dependente, que representa o que o modelo tentará prever
- α : Constante que representa a interceção da reta com o eixo vertical
- β : Representa a inclinação em relação à variável dependente
- x_i : Variável independente
- μ_i : Representa fatores residuais mais possíveis erros de medição

De acordo com o livro *Introductory Econometrics, A Modern Approach [14]*, entre o capítulo 10 e 12, explica como é utilizado ambas equações matemáticas e gráficos para o desenvolvimento de Regressão Linear com séries temporais, onde também são explicados e utilizados os *dataSets*.

Em Ciência de Dados, um *lag* refere-se a um período de tempo de um *dataSet* com séries temporais. Os *dataSets* de séries temporais, registam uma sequência de pontos de dados ao longo do tempo, onde os *lags* são usados para analisar padrões e relações entre variáveis ao longo do tempo. Um valor *lag* é o valor de uma variável em um ponto de tempo anterior e, observando a relação entre valores atuais e , podemos fazer previsões sobre valores futuros.

- Regressão Polinomial de 2º Grau:

Para fornecer um ajuste mais preciso dos dados, foi utilizado a Regressão Polinomial de 2º grau, que consiste no modelo de regressão não linear que pode ser usado para modelar relações complexas entre uma variável independente x e uma variável dependente y . Ao ajustar uma curva polinomial de 2º grau a um conjunto de dados, a relação entre x e y é modelada com a seguinte *Equação 3*:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

Equação 3 - Reta de Regressão Polinomial de 2º Grau

Onde:

- y : Variável dependente
- x : Variável independente
- β_0, β_1 e β_2 : Coeficientes a serem estimados, que representam o intercepto, o coeficiente linear e o coeficiente quadrático, respetivamente
- ε : Erro aleatório

O modelo de regressão polinomial de 2º grau é uma generalização do modelo de regressão linear simples, onde apenas uma reta é ajustada aos dados. Neste modelo, a curva ajustada é uma parábola, e, como resultado, o modelo pode ser usado para ajustar uma curva que descreve uma relação não linear entre as variáveis, podendo ser mais preciso do que o modelo de regressão linear simples em algumas situações.

- **Métricas de avaliação:**

- **Erro Médio Absoluto Percentual:**

O Erro Médio Absoluto Percentual (EMAP) é uma métrica utilizada para avaliar a precisão de um modelo de previsão. É uma medida de desempenho que expressa o erro médio em termos percentuais em relação ao valor real. Este erro é calculado como a média dos valores absolutos das diferenças percentuais entre as previsões e os valores reais, divididos pelo valor real, como demonstrado na *Equação 4*:

$$EMAP = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - p_i}{y_i} \right|$$

Equação 4 - Erro Médio Absoluto Percentual

Onde:

- y_i : Valor real observado
- p_y : Valor previsto
- N : Número total de observações

O Erro Médio Absoluto Percentual é expresso em percentagem, o que significa que é facilmente interpretável e pode ser comparado entre diferentes modelos ou previsões. Quanto menor o valor do EMAP, melhor é a precisão do modelo em relação aos dados reais. No entanto, o EMAP também pode apresentar algumas limitações, como ser mais sensível a valores atípicos e extremos.

- Coeficiente de Determinação:

O coeficiente de determinação (R^2) e o coeficiente de correlação de Pearson (r), são duas medidas estatísticas usadas para analisar a relação entre duas variáveis quantitativas.

O coeficiente de correlação de Pearson (r) mede a força e direção da relação linear entre as variáveis, variando entre -1 e 1. Um valor de r próximo de 1 indica uma correlação positiva forte entre as variáveis, enquanto um valor próximo de -1 indica uma correlação negativa forte. O valor de 0 indica ausência de correlação linear entre as variáveis.

O coeficiente de determinação (R^2), como utilizado na *Tabela 2* e *Tabela 3*, indica a proporção da variação da variável dependente que é explicada pela variação da variável independente. Varia entre 0 e 1, sendo que o valor de 0 indica que não existe nenhuma relação entre as variáveis e o valor de 1 indica uma relação perfeita entre elas. O coeficiente de determinação (R^2) é obtido através do coeficiente de correlação de Pearson (r), como mostrado na *Equação 6*. Isso significa que, quanto maior for o valor de r , maior será o valor de R^2 . Este coeficiente R^2 é uma medida importante na análise de regressão, pois indica o quanto do comportamento da variável dependente pode ser explicado pela variável independente.

Para calcular a correlação de Pearson, podemos utilizar a seguinte *Equação 5*:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$

Equação 5 – Correlação de Pearson

Onde:

- x e y : Variáveis x e y
- z_x : Desvio padrão da variável x
- z_y : Desvio padrão da variável y
- N : Número total de dados

É então calculado o coeficiente de determinação (R^2) a partir da *Equação 6*:

$$R^2 = r^2$$

Equação 6 - Coeficiente de Determinação

5.3 Tecnologias e Ferramentas Utilizadas

Para este projeto, foram utilizados modelos de *Machine Learning*, utilizando a linguagem de programação Python.

5.3.1 Python

Python é uma linguagem de programação interpretada de script, orientada a objetos. Um dos usos mais comuns do *Python* é a sua capacidade de criar e gerir estruturas de dados rapidamente, como, por exemplo, a sua biblioteca *Pandas*, cujo oferece uma infinidade de ferramentas para manipular, analisar e até mesmo representar estruturas e conjuntos de dados complexos.

Utilizando as bibliotecas da linguagem *Python*, irão ser criados *Dataframes*, que são estruturas de dados alinhados em forma de tabela, mutável em tamanho e potencialmente heterogênea.

5.3.2 Modelos de *Machine Learning*

São aplicados modelos de *Machine Learning*, que é definido como a capacidade de uma máquina imitar o comportamento humano, onde tem várias aplicações como a análise de dados e Inteligência Artificial. Estas aplicações de *Machine Learning* serão usadas para uma previsão, como o número de chamadas diárias e duração do tempo de espera das mesmas, onde o pressuposto é encontrar alguma casuística de forma a chegar a alguma conclusão.

6 Solução Desenvolvida

Neste trabalho, foi realizada uma análise detalhada dos dados de chamadas do INEM, abrangendo o período de 2012 a 2021. Além de uma análise exploratória prévia dos dados, foram aplicados modelos estatísticos e de *Machine Learning* para prever o número de chamadas recebidas. A nossa solução proposta com este estudo consiste na **identificação da casuística das chamadas recebidas**, com o propósito de prever o número de chamadas e assim sugerir melhorias na gestão de recursos do CODU.

No dia 12 de janeiro de 2023, estivemos presentes numa reunião no Instituto Nacional de Emergência Médica (INEM), para definir em concreto o objetivo deste projeto, o método de entrega dos dados e esclarecimento de mais algumas questões. Foi explicado o funcionamento do CODU e a forma como os dados são obtidos e armazenados. Após a reunião, realizou-se uma visita guiada ao centro do CODU, que nos deu uma noção detalhada de como é o seu funcionamento.

6.1 Análise Exploratória de Dados

Numa fase inicial, foi realizada uma análise estatística dos dados sem a utilização de modelos de *Machine Learning*. Foram feitas análises univariadas e multivariadas para explorar as relações entre estas e identificar possíveis padrões e tendências nos dados. Com base nesta análise exploratória, foram identificadas algumas relações interessantes entre as variáveis, as quais serão aprofundadas posteriormente com a utilização de modelos de *Machine Learning* na fase seguinte.

Número de Chamadas por ano

Como já referido no capítulo de Identificação do Problema, foi criado um gráfico, *Figura 7*, para verificar o número de chamadas por ano, entre 2012 e 2022. A partir deste gráfico, é possível observar uma tendência de crescimento no número de chamadas ao longo dos anos, com uma grande diminuição no ano 2020, sendo o ano que começou a pandemia COVID-19, e com um aumento expressivo no ano 2022, após a pandemia de COVID-19 ter terminado.

Aumento do Número de Chamadas Recebidas, excepto nos anos de pandemia (2020 e 2021)

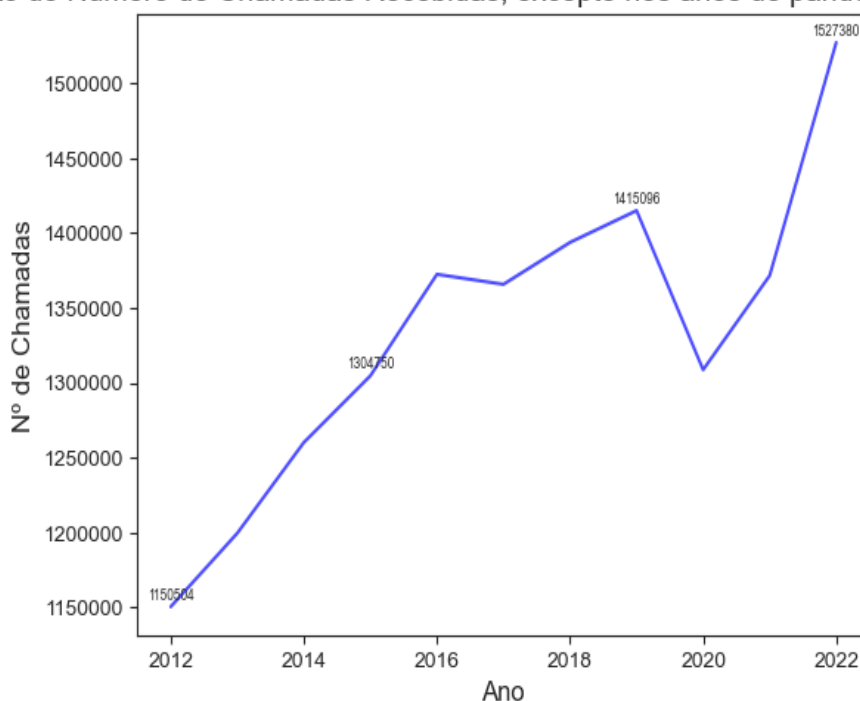


Figura 7 - Número de Chamadas Recebidas

Reta Regressão Linear do número total de Chamadas por Ano

Com base no valor acumulado das chamadas recebidas entre 2012 e 2022, foi possível calcular retas de regressão linear, incluindo todos os anos (*Figura 8*) e excluindo os anos de pandemia (*Figura 9*). Uma conclusão inicial é a tendência de aumento do número de chamadas recebidas por ano. A reta de regressão, ao considerar todos os anos, apresenta uma inclinação menor devido à redução das chamadas durante os anos de pandemia (2020 e 2021). Por outro lado, ao excluir esses anos, a inclinação aumenta. Utilizando as equações das retas de regressão, é possível prever que o número de chamadas para o ano de 2023, ao considerar os anos de pandemia, é de 1 495 129 chamadas, e no entanto, excluindo os dados dos anos 2020 e 2021, a previsão é de receber 1 576 937 chamadas..

É importante destacar que os anos de pandemia tiveram um impacto atípico e significativo na redução do número de chamadas. No entanto, em 2022, observou-se um aumento novamente, indicando que outros fatores além da pandemia também podem influenciar este número.

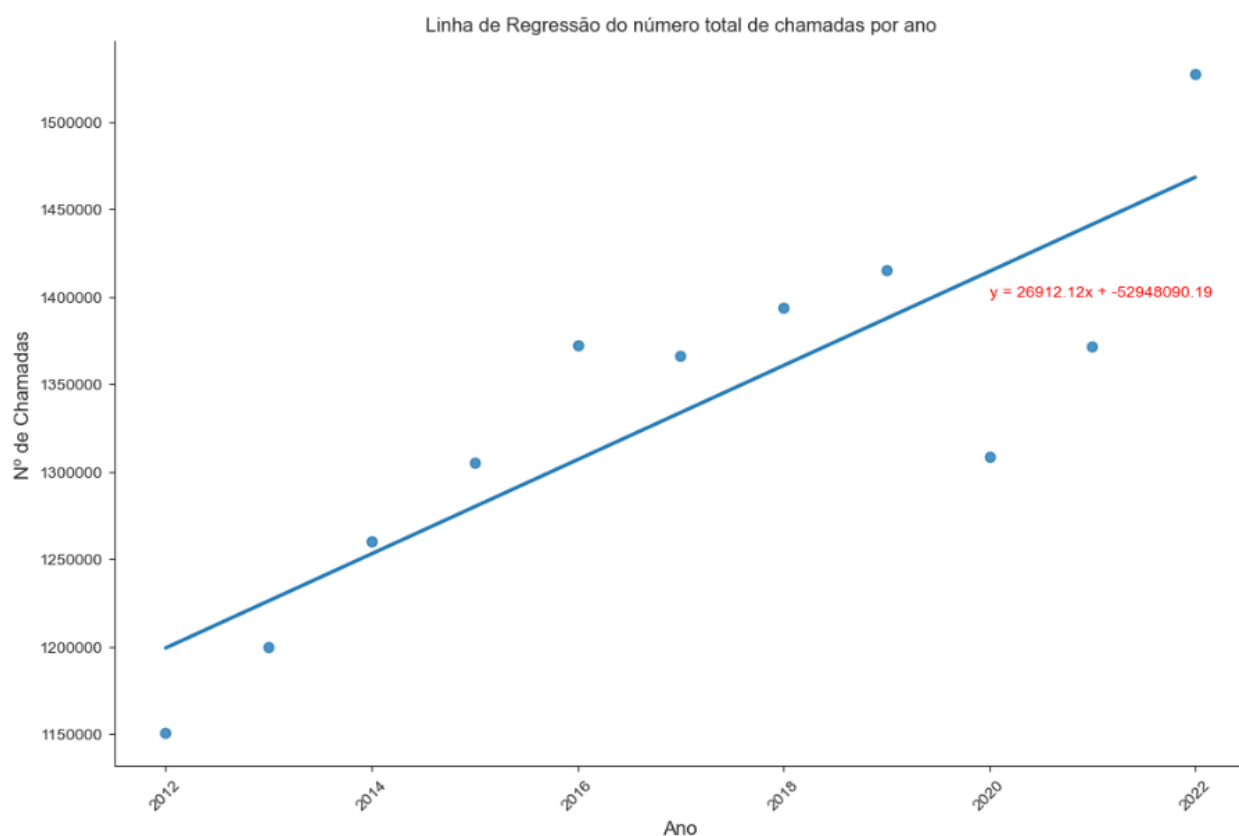


Figura 8 - Reta de Regressão do número total de Chamadas

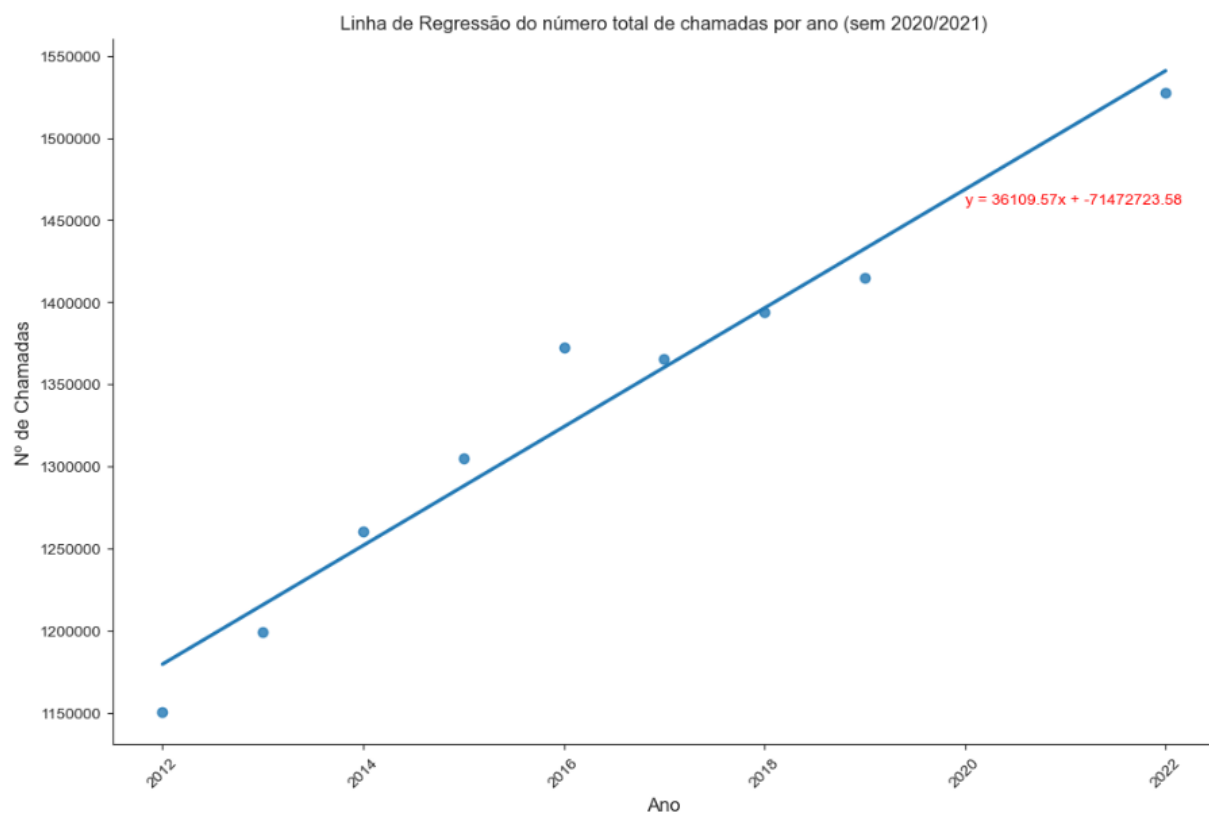


Figura 9 - Reta de Regressão do número total de Chamadas (sem 2020/2021)

Número médio de chamadas recebidas por mês entre 2012-2022

O próximo gráfico, demonstrado na *Figura 10*, foi feito para verificar a média do número de chamadas recebidas por mês ao longo de um período de 10 anos (2012-2022). O mesmo demonstra que janeiro e dezembro são os meses com um maior número de chamadas recebidas, e o mês de abril o mês com menor número de chamadas recebidas. É, no entanto, importante ressaltar que, apesar destes meses apresentarem a maior média de chamadas recebidas, não significa que sejam os meses com o maior número absoluto de chamadas, já que pode variar de acordo com diversos fatores, como feriados, eventos e condições climáticas.

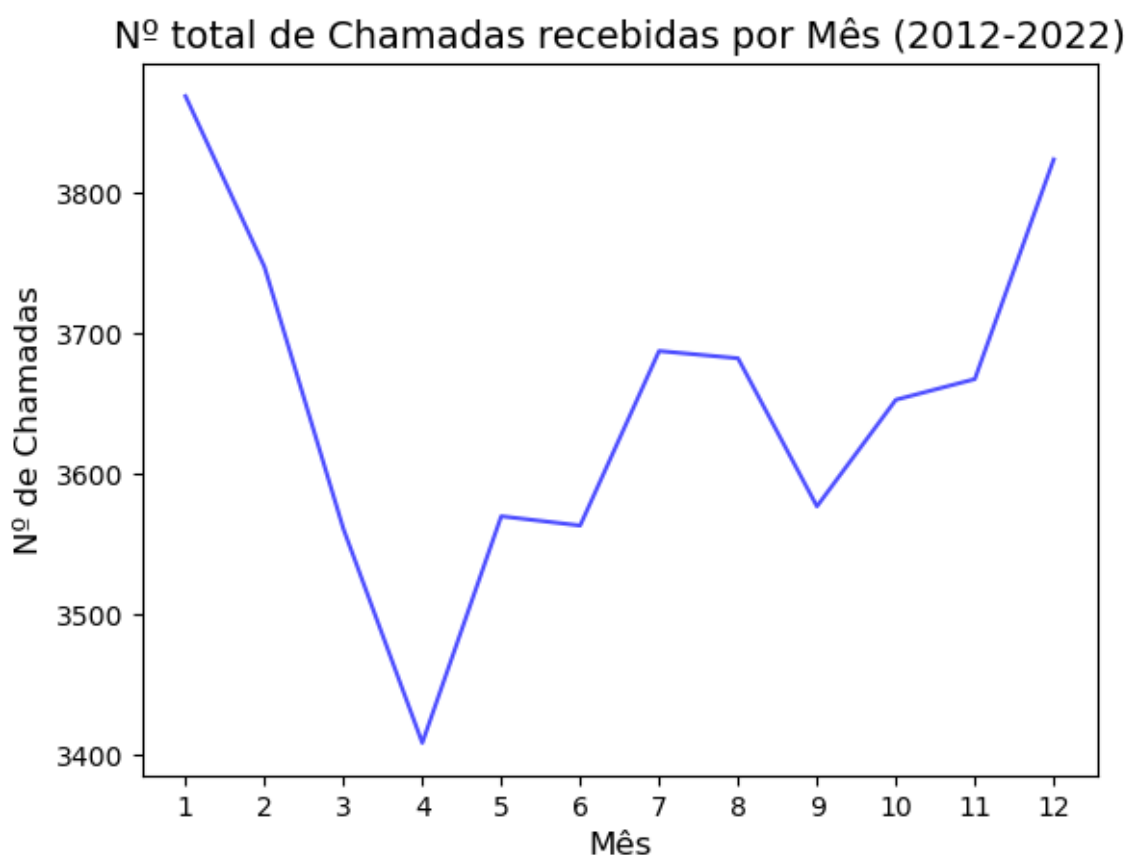


Figura 10 - Média do número de Chamadas recebidas por Mês, entre 2012 e 2022

Retas de Regressão Polinomial de 2º Grau:

Neste projeto, foram criados 6 gráficos com Retas de Regressão Polinomial de 2º Grau. Foram elaborados 2 gráficos para o mês de janeiro (com e sem os dados referentes aos anos 2020 e 2021, apresentados na *Figura 11* e *Figura 12*), 2 gráficos para o mês de fevereiro (com e sem os dados referentes aos anos 2020 e 2021, apresentados na *Figura 13* e *Figura 14*) e 2 gráficos para o mês de março (com e sem os dados referentes aos anos 2020 e 2021, apresentados na *Figura 15* e *Figura 16*).

A não inclusão dos restantes meses nos gráficos de Regressão Polinomial de 2º Grau neste projeto deve-se à constatação de que não apresentam diferenças significativas em relação aos meses selecionados (janeiro, fevereiro e março). Optou-se então por utilizar estes meses representativos, a fim de realizar uma análise mais precisa do comportamento das chamadas recebidas pelo INEM.

Esses gráficos são ferramentas valiosas para identificar possíveis padrões e tendências nos dados, bem como para prever valores futuros com base nas curvas de regressão ajustadas. Ao utilizar uma equação polinomial de segundo grau, é possível traçar uma linha curva que se adapta aos dados, permitindo visualizar a tendência geral dos mesmos. Esta abordagem gráfica proporciona uma visão mais abrangente dos dados, facilitando a compreensão das flutuações sazonais, variações de longo prazo e outros padrões que possam estar presentes.

Os gráficos com retas de Regressão Polinomial de 2º Grau são uma ferramenta valiosa neste projeto, permitindo explorar os dados, compreender suas características e obter informações relevantes para uma melhor gestão e previsão das chamadas recebidas pelo INEM.

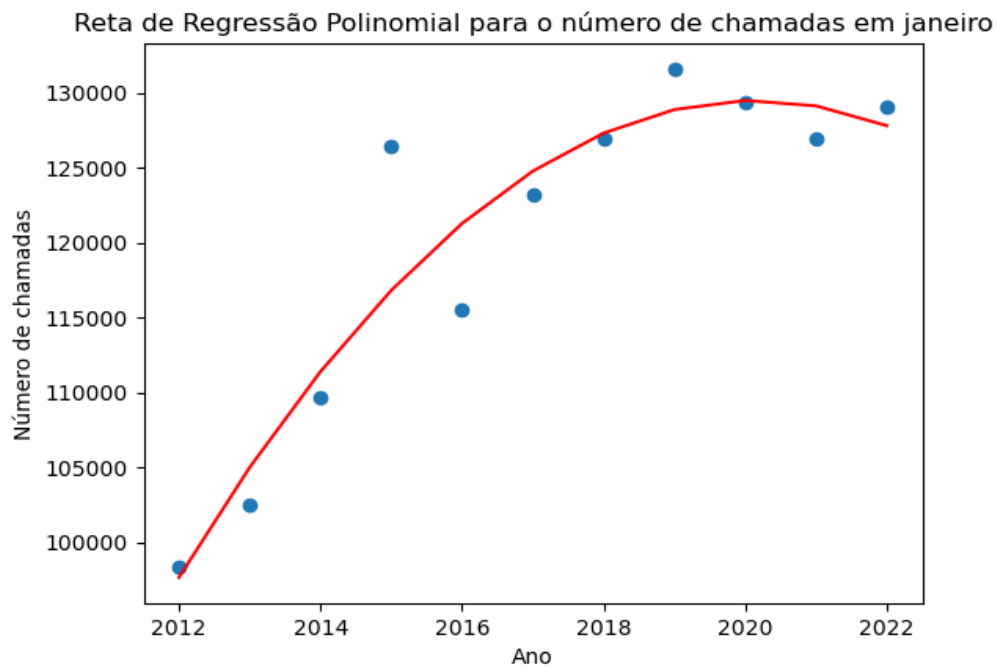


Figura 11 - Reta de Regressão Polinomial de janeiro

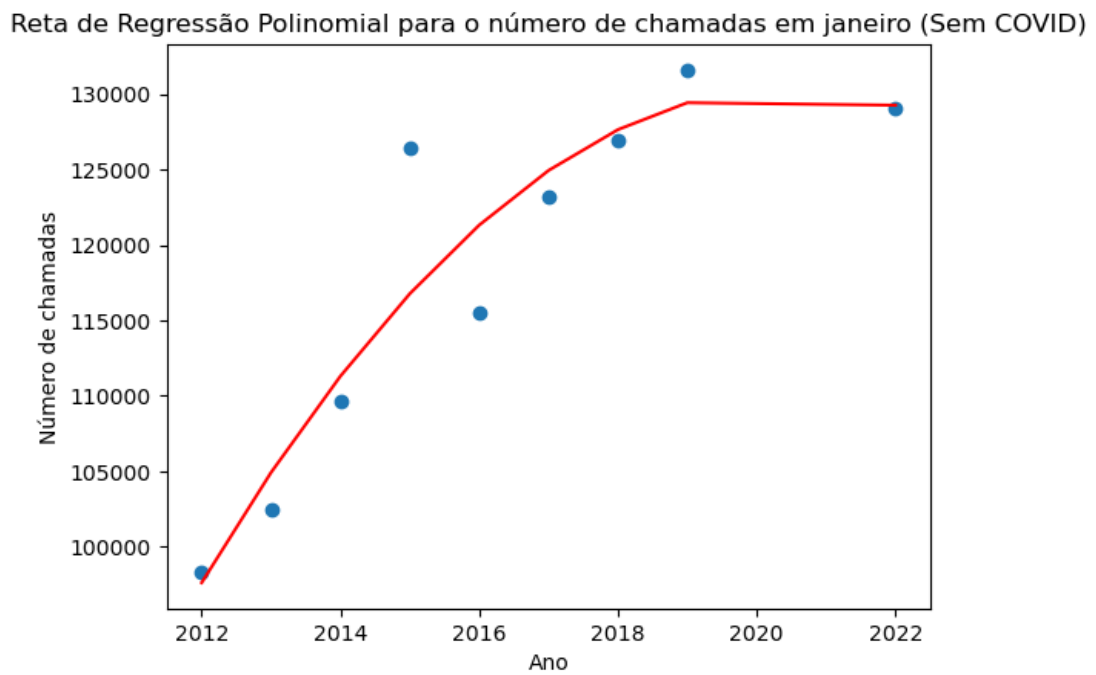


Figura 12 - Reta de Regressão Polinomial de janeiro, excluindo os anos 2020 e 2021

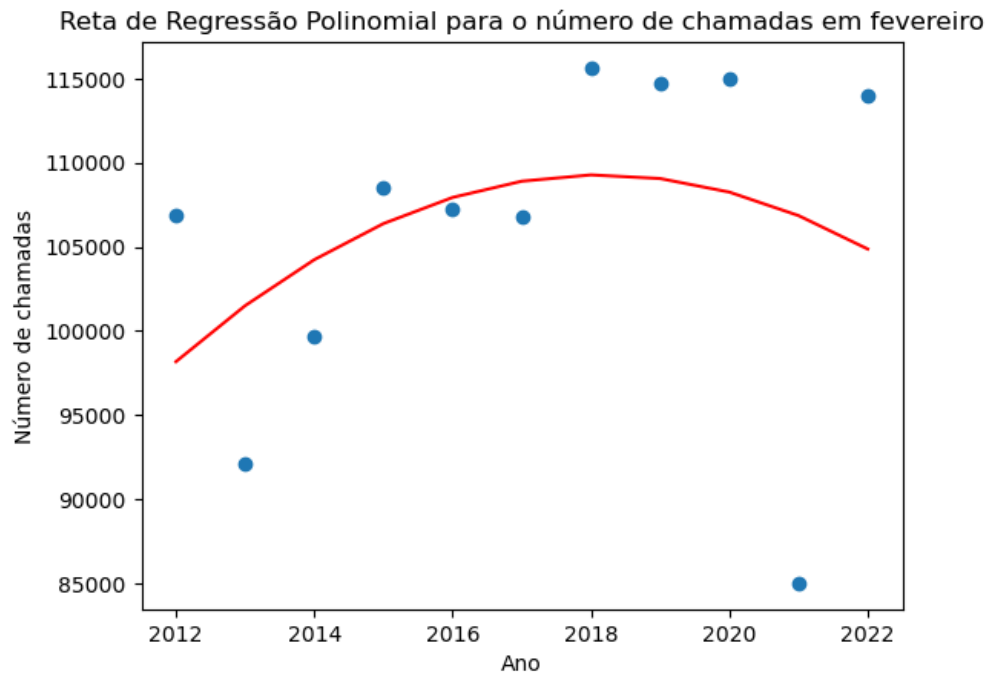


Figura 13 - Reta de Regressão Polinomial de fevereiro

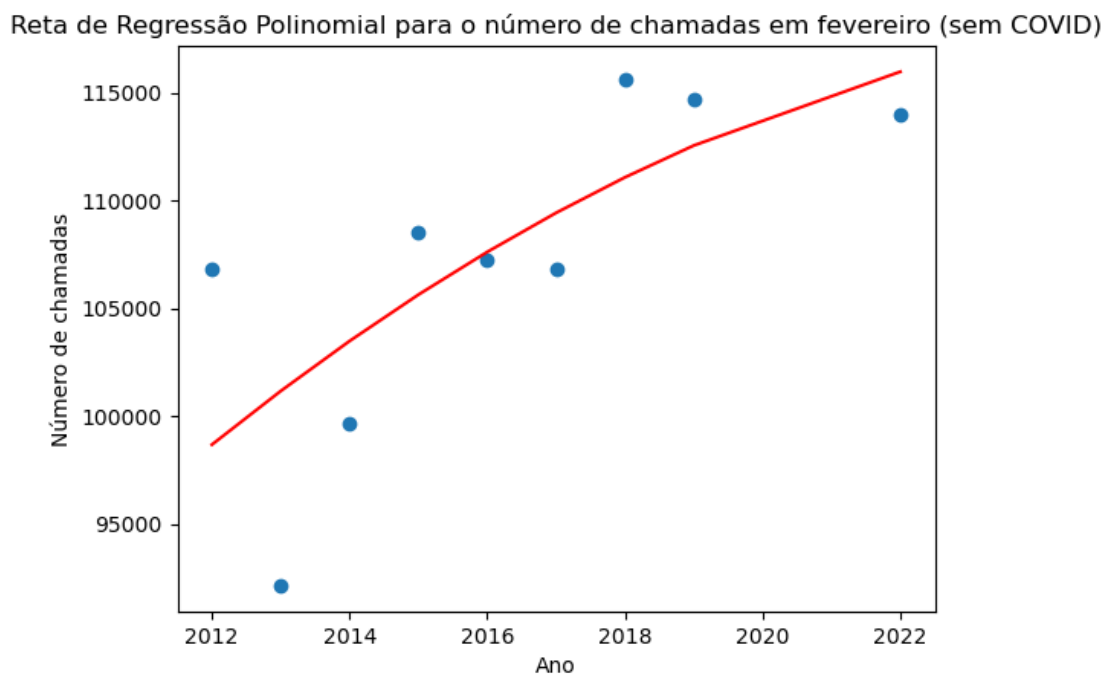


Figura 14 - Reta de Regressão Polinomial de fevereiro, excluindo os anos 2020 e 2021

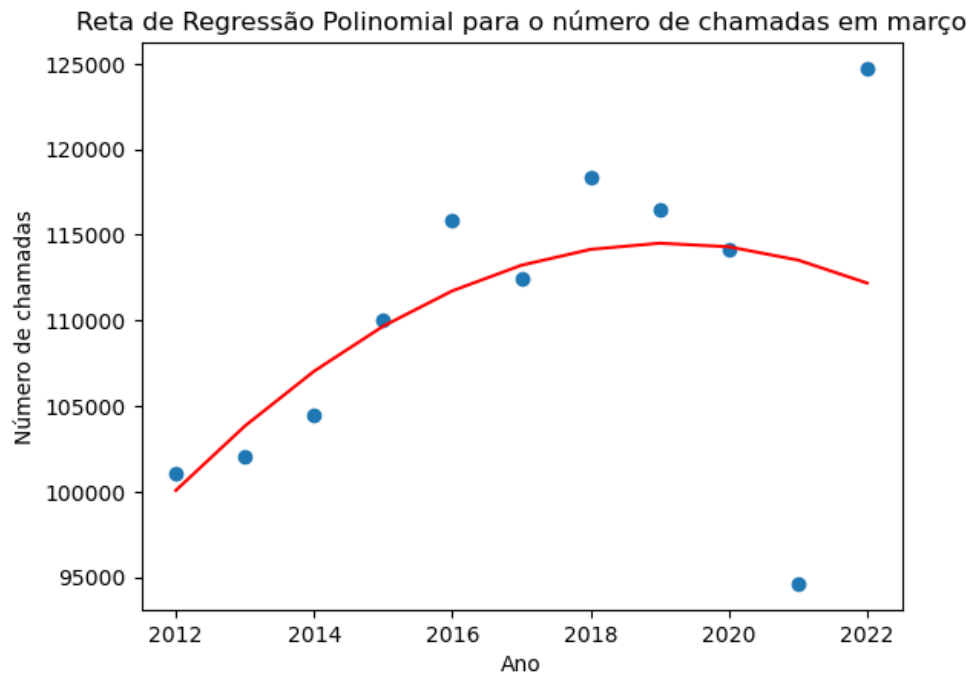


Figura 15 - Reta de Regressão Polinomial de março

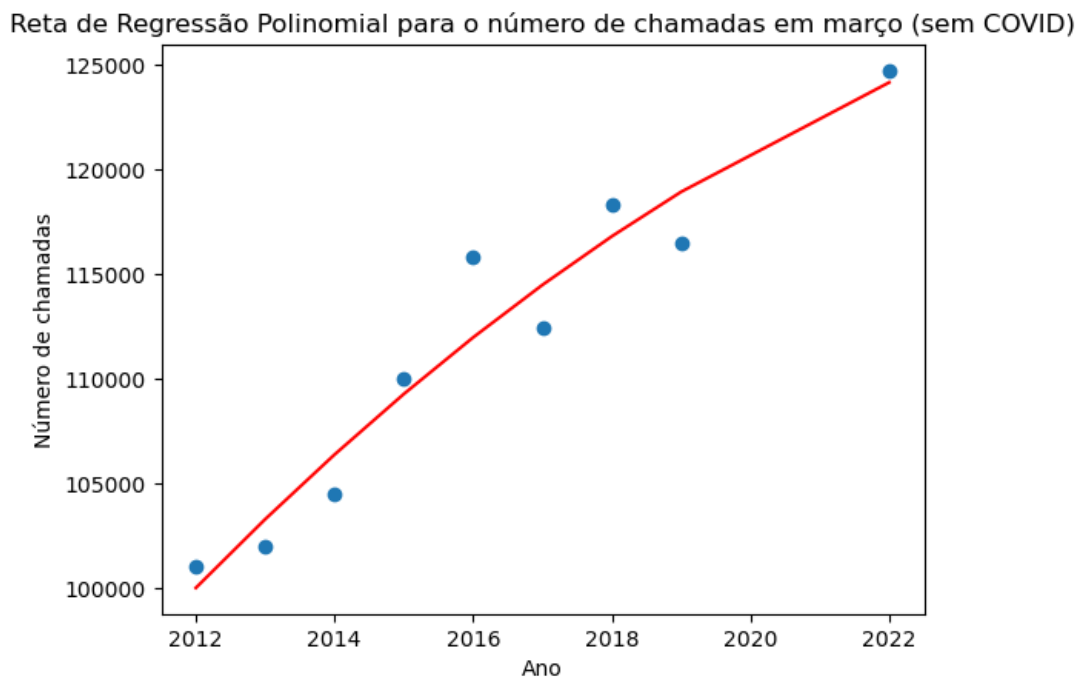


Figura 16 - Reta de Regressão Polinomial de março, excluindo os anos 2020 e 2021

Para compreendermos melhor os dados e a previsão desta reta de regressão polinomial, é apresentado, na *Tabela 2* e *Tabela 3*, as chamadas previstas para 2023 (incluindo e excluindo os dados de 2020 e 2021), o número de chamadas real, o erro médio absoluto percentual (incluindo e excluindo os dados de 2020 e 2021) em percentagem e o R^2 (incluindo e excluindo os dados de 2020 e 2021) sendo que estes foram os resultados:

Mês	Chamadas previstas para 2023 (incluindo 2020 e 2021)	Número de chamadas Real	EMAP (incluindo 2020 e 2021)	R^2 (incluindo 2020 e 2021)
Janeiro	125 508	129 474	3,06%	0.888
Fevereiro	102 297	122 026	15,53%	0.125
Março	110 268	127 016	13,19%	0.290

Tabela 2 - Resultados de previsão de todos os anos

Mês	Chamadas previstas para 2023 (excluindo 2020 e 2021)	Número de chamadas Real	EMAP (excluindo 2020 e 2021)	R^2 (excluindo 2020 e 2021)
Janeiro	127 354	129 474	1,64%	0.879
Fevereiro	116 768	122 026	4,31%	0.547
Março	125 520	127 016	1,17%	0.934

Tabela 3 - Resultados de previsão de todos os anos, excluindo 2020 e 2021

Estas tabelas apresentam os resultados da análise de regressão polinomial de 2º grau para os meses de janeiro, fevereiro e março, considerando, e não considerando, os dados dos anos de 2020 e 2021. Uma das informações apresentadas na tabela é o erro médio absoluto percentual (EMAP) onde se observa que, excluindo os anos de 2020 e 2021, o erro estimado dos valores previstos é significativamente menor em comparação com os valores que incluem esses anos. Podemos também ver que, ao excluir 2020 e 2021, o Coeficiente de Determinação (R^2) tem um melhor ajuste, sendo que, como explicado anteriormente, o Coeficiente de Determinação varia de 0 a 1, onde quanto mais elevado for o valor, melhor o ajuste do modelo aos dados.

Isso sugere que a pandemia de COVID-19 afetou significativamente a tendência do número de chamadas recebidas nesses meses, levando a uma maior imprecisão nas previsões feitas com base nos dados desses anos. Assim, a análise excluindo esses anos pode fornecer resultados mais precisos e confiáveis para entender a tendência do número de chamadas nestes meses.

Flutuação do Número de Chamadas Recebidas por mês entre 2016 e 2022

É possível analisar no gráfico exposto na *Figura 17* o número de chamadas por mês para os anos 2016, 2017, 2018, 2019 e 2022. Observa-se que há uma flutuação bastante comum entre esses anos, com um padrão de comportamento similar em vários meses. Por exemplo, é comum nestes anos haver uma descida do número de chamadas para os meses de fevereiro, abril, junho, setembro e novembro. Já para os meses de março, maio, julho, outubro e dezembro, é comum haver um aumento do número de chamadas recebidas.

Podemos concluir com este gráfico que o volume do número de chamadas recebidas no ano 2022 é muito superior aos anos anteriores, o que nos leva a tirar como conclusão que as chamadas recebidas irão aumentar nos próximos anos. Sendo que os dados dos anos 2012 a 2019 são bastante próximos, por motivos de melhor visualização, foram retirados os anos 2012 a 2015.

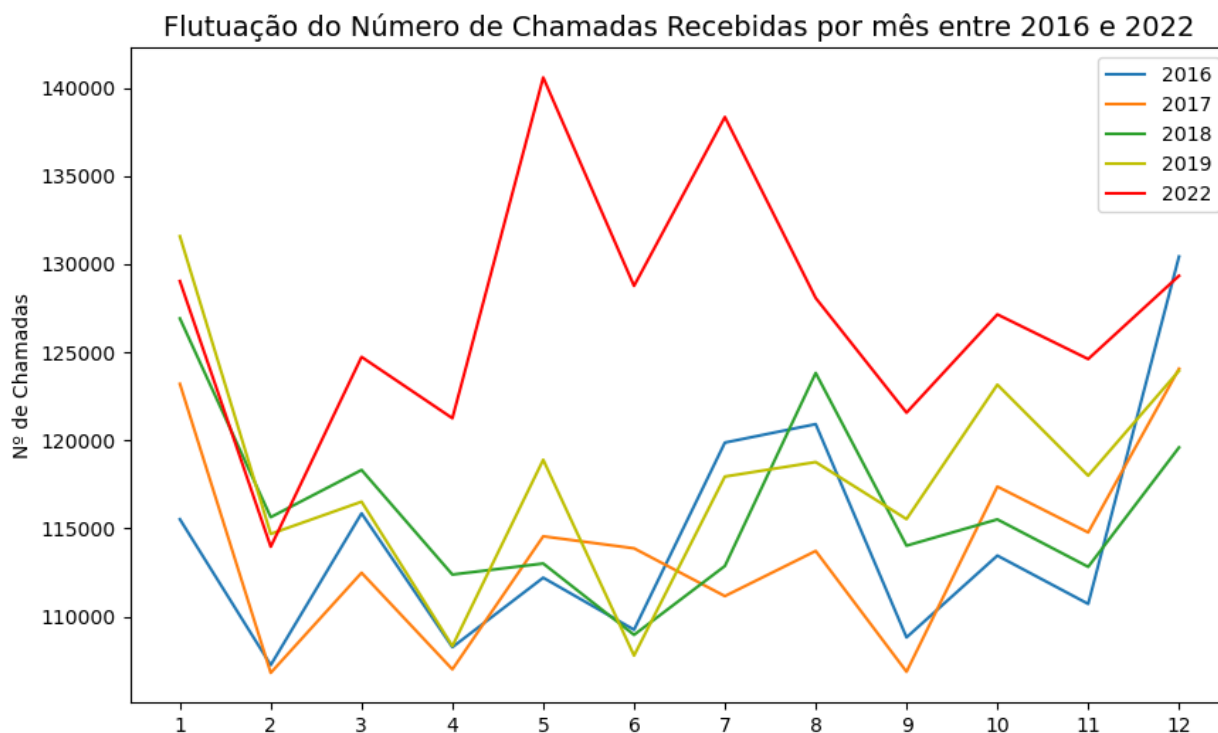


Figura 17 - Flutuação do número de Chamadas por mês entre 2016 e 2012

6.2 Abrangência

Foram utilizados conhecimentos obtidos nas seguintes disciplinas: *Data Science* (Ciência de Dados), Inteligência Artificial, Linguagens de Programação e Matemática. Cada uma destas disciplinas contribui com conhecimentos específicos, que no seu todo permitem relacionar conhecimentos e definir estratégias para o desenvolvimento e conclusão do trabalho. Será demonstrado na *Figura 18* o que será as áreas utilizadas de cada disciplina.

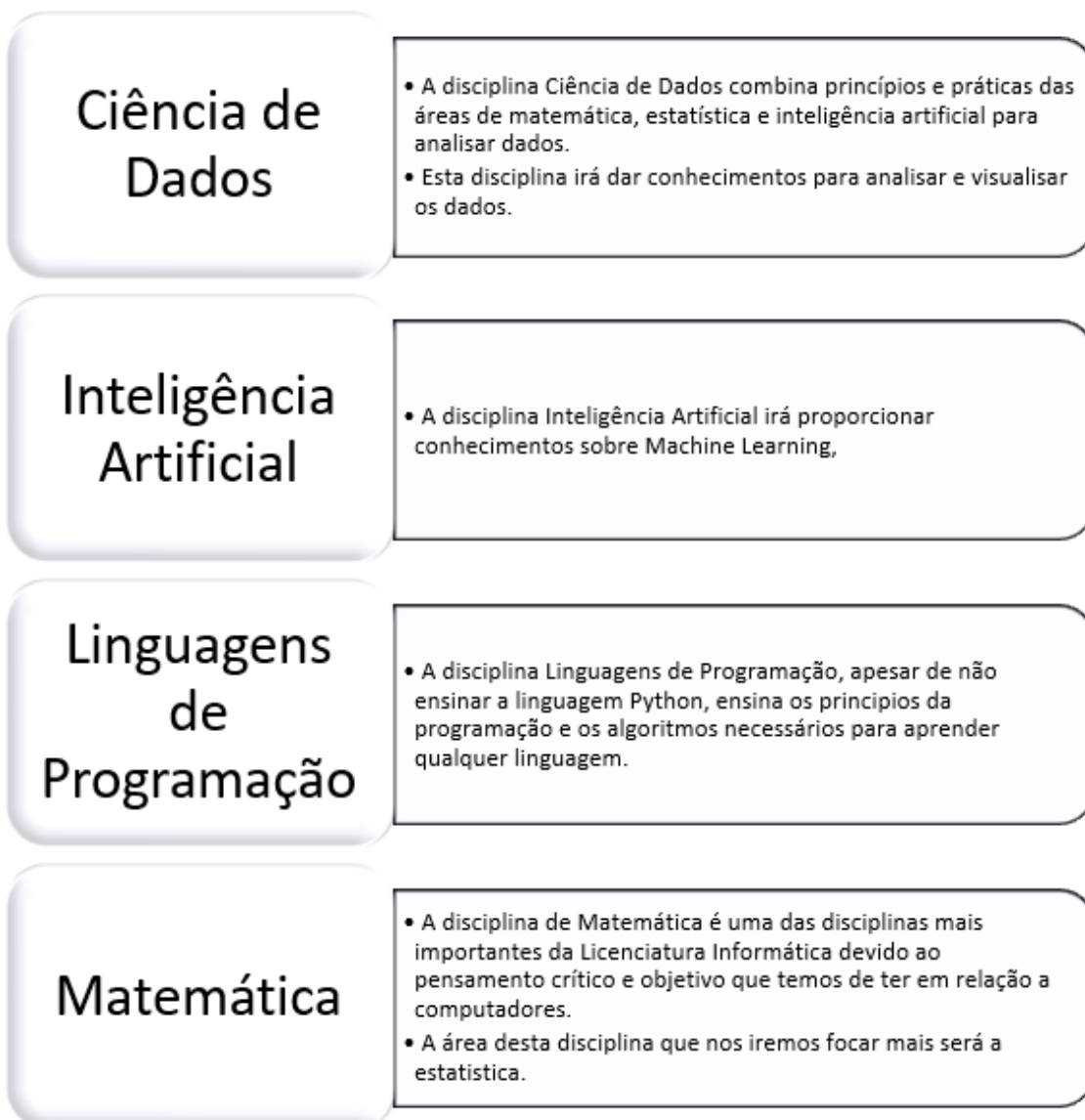


Figura 18 - Associação das matérias usadas para a formulação do projeto

7 Resultados

A utilização de *Machine Learning* tem revelado ser uma ferramenta poderosa em diversas áreas, incluindo a previsão e análise de dados. Neste capítulo, iremos explorar o campo da previsão utilizando o modelo de *Machine Learning* Regressão Linear, destacando os seus conceitos fundamentais e aplicações práticas. Com uma abordagem orientada por dados e algoritmos inteligentes, o objetivo principal consiste em desenvolver modelos precisos e fiáveis capazes de antecipar eventos e tendências com base em dados históricos, impulsionando, assim, a tomada de decisões informadas em diversos setores.

7.1 Modelo de previsão do número de chamadas

Foram desenvolvidos três modelos distintos para a previsão do número de chamadas recebidas pelo INEM: um para previsão **diária**, outro para previsão **semanal** e um terceiro para previsão **mensal**. Foi criada um *dataSet* com índices para ajudar na criação e treino dos três modelos, demonstrada na *Figura 19*. Com base no conjunto de dados abrangendo o período de 2012 a 2022 e os seus índices correspondentes, os modelos foram treinados e testados com este conjunto. Após o modelo realizar a previsão para o número de chamadas de um determinado índice, os valores reais foram incorporados ao conjunto de dados para treinar novamente o modelo e testar a próxima previsão. Este processo é repetido nos três modelos.

	Data	Nº de Chamadas	indice_dia	indice_mes	indice_semana
0	2012-01-01	4538	1	1	1
1	2012-01-02	3263	2	1	1
2	2012-01-03	3126	3	1	1
3	2012-01-04	3199	4	1	1
4	2012-01-05	2999	5	1	1
...
4013	2022-12-27	4182	3283	108	477
4014	2022-12-28	4157	3284	108	477
4015	2022-12-29	4018	3285	108	477
4016	2022-12-30	4012	3286	108	477
4017	2022-12-31	3736	3287	108	477

Figura 19 – Dataset usado para obter os modelos de Machine Learning usados

Para cada modelo, são apresentados os resultados do modelo de regressão linear de *Machine Learning*, incluindo o número previsto de chamadas recebidas, número real de chamadas recebidas, erro médio absoluto percentual (EMAP) e coeficiente de determinação (R^2). Estes valores são essenciais para avaliar o desempenho do modelo na previsão do número de chamadas recebidas pelo INEM. O EMAP fornece uma estimativa da precisão das previsões em termos percentuais e o coeficiente de determinação indica o grau de ajuste do modelo aos dados reais observados. Estes indicadores são fundamentais para avaliar a qualidade e a confiabilidade do modelo, permitindo a identificação de possíveis áreas de melhoria. Ao analisar os números previstos, reais e essas métricas de desempenho, podemos obter perspectivas relevantes para aprimorar a precisão das previsões e garantir uma abordagem mais precisa no planejamento dos recursos do INEM.

Com o objetivo de garantir uma previsão mais precisa e confiável, foi decidido excluir os dados correspondentes aos anos de 2020 e 2021 na análise do modelo. A análise prévia realizada (apresentadas na *Tabela 2* e *Tabela 3*) demonstrou que esses períodos não apresentaram um padrão consistente para a previsão do número de chamadas recebidas pelo INEM. Ao eliminar esses dados, foi possível otimizar o desempenho do modelo, a fim de melhor compreender e antecipar as tendências e os padrões emergentes para uma tomada de decisão mais precisa na gestão dos serviços de emergência médica pela parte do INEM.

É importante ressaltar que as previsões foram realizadas sem a utilização de *lags*. O modelo de regressão linear de *Machine Learning* foi desenvolvido exclusivamente com base em dados brutos, permitindo assim uma interpretação direta das previsões, facilitando a compreensão e análise dos resultados obtidos.

As datas apresentadas nas tabelas seguintes correspondem aos períodos para os quais os modelos realizaram previsões, considerando que o treino dos modelos foi até o último dia de 2022. Dessa forma, os modelos foram capazes de prever o primeiro dia, semana e mês subsequente, proporcionando uma continuidade temporal na análise e permitindo avaliar o seu desenvolvimento inicial. Essa capacidade de prever os períodos seguintes ao período de treino é fundamental para verificar a capacidade dos modelos de generalização e adaptação a novos dados em tempo real.

O objetivo principal destes modelos é fornecer previsões do número de chamadas recebidas pelo INEM numa base diária, semanal e mensal, respetivamente. Com base nos dados de dias, semanas e meses anteriores, o modelo utiliza técnicas de regressão linear para estabelecer relações entre os diferentes parâmetros e o número total de chamadas recebidas. Através dessa análise, procura-se compreender e antecipar padrões sazonais, tendências e outros fatores que possam impactar o número de chamadas recebidas pelo INEM semanalmente. Os valores obtidos encontram-se na *Tabela 4*, *Tabela 5* e *Tabela 6*.

Modelo Regressão Linear de *Machine Learning* por Dia:

O objetivo principal desse modelo é fornecer previsões do número de chamadas recebidas pelo INEM numa base diária. Os valores obtidos encontram-se na *Tabela 4*.

Dia	Número Real de Chamadas Recebidas	Número Previsto de Chamadas Recebidas	Erro Médio Absoluto Percentual (EMAP)	Coefficiente de Determinação (R^2)
Dia 1 de janeiro de 2023	4 552	4 153	8,76%	0.527
Dia 2 de janeiro de 2023	3 910	4 153	6,24%	0.528
Dia 3 de janeiro de 2023	4 601	4 153	9,72%	0.527
Dia 4 de janeiro de 2023	4 292	4 154	3,20%	0.528

Tabela 4 – Resultados da previsão diária do modelo de Machine Learning

O coeficiente de determinação (R^2) tendo a média de 0.528 indica que o modelo de regressão linear por dia explica aproximadamente 52% da variação observada nos dados. Isso significa que metade da variabilidade das chamadas recebidas pelo INEM pode ser explicada pelas características e parâmetros considerados pelo modelo. Um R^2 de 0.528 sugere uma relação moderada entre as variáveis independentes e dependentes.

Além disso, o erro médio absoluto percentual com média de 6.98% indica que, em média, as previsões do modelo têm uma diferença absoluta percentual de aproximadamente 7% em relação aos valores reais das chamadas recebidas. Este valor é uma medida da precisão das previsões em termos percentuais. Um erro médio absoluto percentual de 7% indica que, em média, as previsões estão próximas dos valores reais em 93% dos casos.

Estes resultados sugerem que o modelo possui um desempenho moderado na previsão diária do número de chamadas recebidas pelo INEM. Embora haja espaço para melhoria, uma diferença média absoluta percentual de 7% indica uma precisão aceitável nas previsões, permitindo uma estimativa razoável do número de chamadas para os serviços do INEM.

Modelo Regressão Linear de *Machine Learning* por Semana:

O objetivo principal desse modelo é fornecer previsões do número de chamadas recebidas pelo INEM numa base semanal. Os valores obtidos encontram-se na *Tabela 5*.

Semana	Número Real de Chamadas Recebidas	Número Previsto de Chamadas Recebidas	Erro Médio Absoluto Percentual (EMAP)	Coefficiente de Determinação (R^2)
Primeira semana de 2023	29 848	28 594	4,20%	0.325
Segunda semana de 2023	29 216	28 619	2,04%	0.328
Terceira semana de 2023	28 473	28 638	0,58%	0.329
Quarta semana de 2023	29 116	28 651	1,60%	0.331

Tabela 5 - Resultados da previsão semanal do modelo de Machine Learning

O coeficiente de determinação (R^2) tendo a média de 0.328 indica que o modelo de regressão linear por semana explica aproximadamente 32% da variação observada nos dados. Isso significa que uma parcela relativamente pequena da variabilidade das chamadas recebidas pelo INEM pode ser explicada pelas características e parâmetros considerados pelo modelo. Um R^2 de aproximadamente 0.3 sugere uma relação mais fraca entre as variáveis independentes e dependentes, indicando que o modelo pode não captar completamente os fatores que influenciam o número de chamadas.

Além disso, o erro médio absoluto percentual de 2.11% indica que, em média, as previsões do modelo têm uma diferença absoluta percentual de aproximadamente apenas 2% em relação aos valores reais das chamadas recebidas. Este valor é uma medida da precisão das previsões em termos percentuais. Um erro médio absoluto percentual de 2% indica uma precisão alta nas previsões, indicando que as estimativas do modelo estão próximas dos valores reais em 98% dos casos.

Estes resultados mostram que, embora o coeficiente de determinação seja relativamente baixo, o modelo apresenta uma precisão notável, com um erro médio absoluto percentual de apenas 2%.

Modelo Regressão Linear de *Machine Learning* por Mês:

O objetivo principal desse modelo é fornecer previsões do número de chamadas recebidas pelo INEM numa base mensal. Os valores obtidos encontram-se na *Tabela 6*.

Mês	Número Real de Chamadas Recebidas	Número Previsto de Chamadas Recebidas	Erro Médio Absoluto Percentual (EMAP)	Coefficiente de Determinação (R^2)
Janeiro de 2023	129 474	126 600	2,22%	0.677
Fevereiro de 2023	118 027	126 991	7,60%	0.685
Março de 2023	127 016	126 953	0,50%	0.680
Mai de 2023	118 350	127 238	7,51%	0.686

Tabela 6 - Resultados da previsão mensal do modelo de Machine Learning

O coeficiente de determinação (R^2) de 0.682 indica que o modelo de regressão linear por mês explica aproximadamente 68% da variação observada nos dados. Isso significa que uma parte considerável da variabilidade das chamadas recebidas pelo INEM pode ser explicada pelas características e parâmetros considerados pelo modelo. Um R^2 de 0.682 sugere uma relação moderadamente forte entre as variáveis independentes e dependentes, indicando que o modelo tem uma capacidade razoável de previsão.

Além disso, o erro médio absoluto percentual de 4.5% indica que, em média, as previsões do modelo têm uma diferença absoluta percentual de 4.5% em relação aos valores reais das chamadas recebidas. Este valor é uma medida da precisão das previsões em termos percentuais. Um erro médio absoluto percentual de 4.5% indica que, em média, as previsões estão próximas dos valores reais em 95.5% dos casos.

Esses resultados indicam que o modelo possui um bom desempenho na previsão diária do número de chamadas recebidas pelo INEM. Com um coeficiente de determinação de 0.68 e um erro médio absoluto percentual de 4.5%, as previsões são relativamente precisas e fornecem uma estimativa confiável da demanda por serviços de emergência médica.

7.2 Aplicação da Fórmula de *Erlang C*

Esta subsecção descreve a análise realizada com base nos dados fornecidos pelo INEM, onde foram consideradas medidas de tempo como o tempo médio de chamadas e o tempo médio de espera de atendimento. Utilizando a fórmula de *Erlang C*, foi possível determinar o número de operadores necessários para um determinado dia. Esta fórmula é amplamente utilizada para dimensionar recursos em sistemas de atendimento, tendo em consideração tanto o tempo de serviço quanto a taxa de chegada de chamadas. Ao aplicar esta fórmula aos dados do INEM, obtivemos uma estimativa do número de operadores necessários para manter um nível adequado de atendimento ao longo do dia.

Ao determinar o número de operadores necessários com base nos dados históricos e nas medidas de tempo, podemos otimizar a alocação de recursos e garantir um atendimento eficiente à população. É importante ressaltar que a aplicação da fórmula de *Erlang C* é uma abordagem inicial e que outras variáveis e considerações podem ser incorporadas posteriormente para um melhor resultado. No entanto, esta análise inicial fornece uma base para iniciar a otimização dos recursos e melhorar a eficiência operacional do INEM.

Para a aplicação da fórmula de *Erlang C*, foi utilizada uma ferramenta online especializada para realizar os cálculos necessários [15]. Esta ferramenta permitiu a inserção dos dados fornecidos pelo INEM, como o tempo médio de chamadas e o tempo médio de espera de atendimento, para obter resultados. Além das medidas de tempo mencionadas, também foram consideradas outras variáveis importantes, como a variável *shrinkage*. Esta variável é uma medida utilizada que permite ter em conta a diferença entre o número de operadores previsto e o número de quantos operadores estão disponíveis num determinado momento.

Ao aplicar a fórmula de *Erlang C* com base nos dados fornecidos pelo INEM em 27 de março de 2023, foram consideradas várias métricas importantes. Sabemos que ocorreram um total de 4379 chamadas recebidas durante um período de 24 horas, o tempo médio de espera de atendimento foi de aproximadamente 10 segundos, a duração média de cada chamada foi de cerca de 200 segundos e presumimos que 95% das chamadas foram atendidas. Com base nestas informações, o resultado da fórmula de *Erlang C* indicou que seriam necessários aproximadamente **70 operadores** para atender a procura de chamadas no período analisado.

8 Conclusão

Os modelos de regressão linear de *Machine Learning* desenvolvidos para prever o número de chamadas recebidas pelo INEM por dia, semana e mês apresentaram um desempenho satisfatório, sendo que o modelo com melhores resultados foi o modelo mensal. Os resultados obtidos sugerem que os modelos foram capazes de captar padrões e tendências nos dados de forma eficaz, permitindo uma boa estimativa do número de chamadas recebidas pelo INEM.

No entanto, é importante ressaltar que ainda há espaço para melhorias. Embora os modelos tenham apresentado um desempenho sólido, é fundamental continuar a melhorá-los e explorar outras abordagens para aumentar a precisão das previsões. Isso irá garantir que as estimativas sejam ainda mais precisas e confiáveis, contribuindo para uma melhor gestão de recursos e tomada de decisões mais assertivas pelos operadores do CODU.

Este projeto assume uma grande importância, uma vez que até o momento não existia um estudo específico que analisasse o número de chamadas recebidas pelo INEM e que permitisse prever com base em modelos de *Machine Learning*. Portanto, os resultados apresentados têm o potencial de fornecer informações valiosas, não apenas para o INEM, mas também para a população portuguesa em geral.

O calendário estabelecido foi seguido, permitindo um progresso organizado e eficiente no desenvolvimento do projeto. As etapas foram realizadas dentro dos prazos estipulados, garantindo a conclusão das tarefas de acordo com o cronograma estabelecido. O cumprimento do calendário demonstra um bom planeamento e gestão do tempo, proporcionando uma base sólida para alcançar os objetivos propostos.

Concluindo, este projeto representa um primeiro passo significativo no desenvolvimento de modelos de previsão precisos e confiáveis para o número de chamadas recebidas pelo INEM. Os resultados obtidos demonstram a viabilidade e eficácia destes modelos, ao mesmo tempo em que indicam possibilidades de melhorias futuras. Compreender e prever o número de chamadas recebidas pelo INEM de forma mais precisa é essencial para garantir uma resposta eficiente e adequada em situações de emergência, beneficiando diretamente a sociedade portuguesa como um todo.

8.1 Trabalho Futuro

É fundamental continuar a melhorar o desempenho dos três modelos distintos, explorando outras abordagens e considerando o uso de *lags*. O uso de *lags*, que são variáveis que representam os valores anteriores das chamadas recebidas, pode ajudar a captar padrões de sazonalidade, tendências temporais e efeitos de curto prazo que influenciam a procura pelos serviços de emergência médica do INEM. Ao incorporar os *lags* nos modelos, é possível ter em conta o histórico recente e passado das chamadas recebidas, fornecendo informações adicionais para uma previsão mais precisa. Esta abordagem permite que os modelos considerem a dependência temporal dos dados, melhorando a sua capacidade de captar correlações e comportamentos não lineares ao longo do tempo.

Para avaliar a viabilidade do uso de *lags*, foram realizados 3 gráficos de autocorrelação, que permitem analisar a correlação entre as chamadas recebidas em diferentes períodos, fornecendo conclusões sobre a presença de padrões de autocorrelação nos dados. Ao examinar estes gráficos, é possível identificar se existem *lags* significativos dependentes da previsão e que apresentem uma correlação forte com as chamadas recebidas em momentos anteriores. A presença de autocorrelação significativa em certos *lags* indica a dependência temporal nos dados, reforçando a importância do uso de séries temporais nos modelos de previsão.

Após analisarmos os gráficos de autocorrelação demonstrados na *Figura 20*, *Figura 21* e *Figura 22*, constatou-se que existe uma correlação significativa entre as chamadas recebidas em períodos anteriores num período diário, semanal e mensal. Estas correlações indicam uma dependência temporal nos dados, o que sugere a viabilidade do uso de *lags* nos modelos de previsão. Ao examinar os gráficos de autocorrelação, observa-se que as autocorrelações são estatisticamente significativas em diversos *lags*. Isso indica que as chamadas recebidas em dias, semanas e meses anteriores possuem uma influência estatisticamente relevante nas chamadas atuais. Portanto, é viável a utilização de *lags* para captar essas correlações e aprimorar a precisão das previsões.

Com base nesta análise, conclui-se que é viável utilizar até 250, 27 e 7 *lags* nos modelos de previsão diário, semanal e mensal, respetivamente. Esta quantidade de *lags* poderá permitir considerar um histórico adequado das chamadas recebidas, levando em conta a dependência temporal e capturando os padrões relevantes para a previsão. Esta análise destaca a necessidade e importância do uso de séries temporais na previsão mensal, uma vez que esse foi o modelo com melhores resultados.

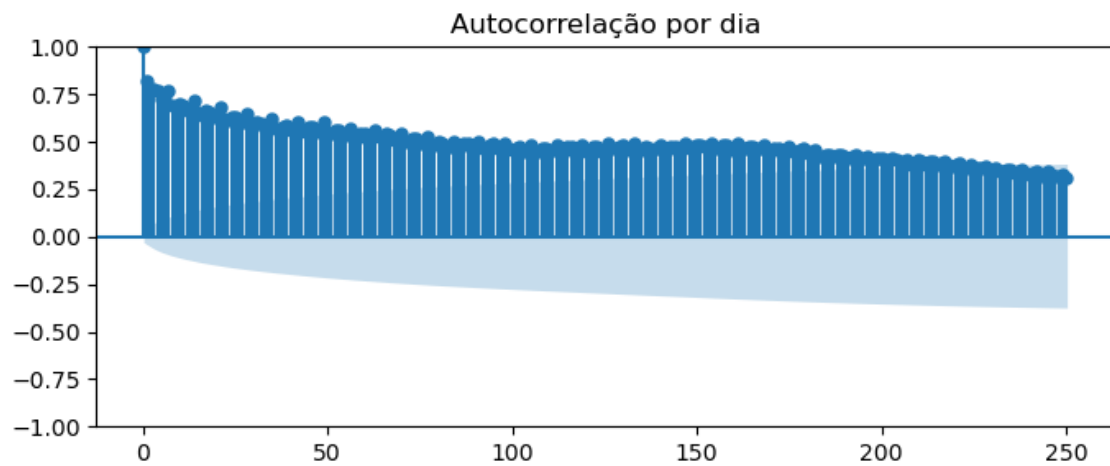


Figura 20 – Gráfico de Autocorrelação do número de chamadas por dia

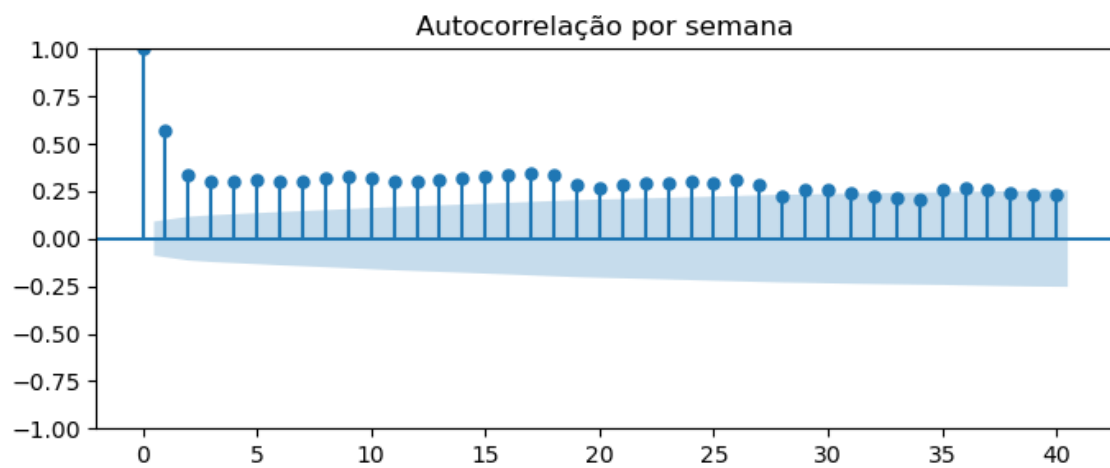


Figura 21 – Gráfico de Autocorrelação do número de chamadas por semana

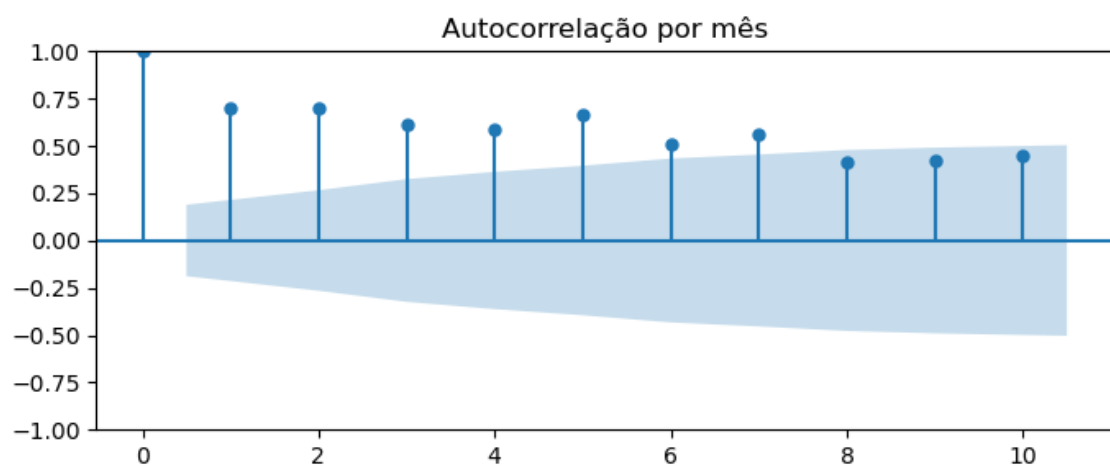


Figura 22 - Gráfico de Autocorrelação do número de chamadas por mês

8.2 Limitações

Ao realizar este projeto, foram identificadas algumas limitações que afetaram a precisão e abrangência dos resultados obtidos. É importante mencionar estas limitações para fornecer uma visão clara das restrições e possíveis áreas de melhoria no estudo. Algumas limitações como:

- **Dependência de Variáveis Externas:** As previsões e análises realizadas podem ser influenciadas por variáveis externas que estão além do controle deste projeto. Fatores econômicos, políticos, sociais ou eventos imprevistos podem afetar o comportamento das chamadas recebidas pelo INEM, e essas influências podem não ter sido totalmente consideradas nos modelos desenvolvidos.
- **Pressuposição de Estacionariedade nos Modelos:** Alguns modelos estatísticos e de previsão assumem que os padrões e relacionamentos dos dados observados no passado se mantêm constantes no futuro. No entanto, essa suposição pode não ser válida em situações em que ocorrem mudanças significativas nas condições ou tendências.
- **Limitações dos Modelos Utilizados:** Os modelos de previsão utilizados neste projeto podem ter limitações inerentes, como não captar todos os fatores relevantes que afetam as chamadas recebidas pelo INEM.

Bibliografia

- [1] INEM. “Ligue 112” SNS Serviço Nacional de Saúde. Disponível em: <https://www.inem.pt/2017/05/30/ligue-112>, Acesso em: 14 de novembro de 2022.
- [2] INEM. “Guia de Utilização do 112” Site do 112. Disponível em: http://www.112.pt/Guias/Paginas/Guia_Util_112.aspx, Acesso em: 14 de novembro de 2022.
- [3] INEM. “Indicadores de Desempenho do INEM” INEM. Disponível em: <https://extranet.inem.pt/stats/dashboard/>, Acesso em: 19 de outubro de 2022.
- [4] Ferreira, Marta. “Censos 2021 revelam que Portugal tem menos 214 mil residentes do que em 2011.” Jornal Digital Observador. Disponível em: <https://observador.pt/2021/07/28/censos-2021-revelam-que-portugal-tem-menos-214-mil-residentes-do-que-em-2011/>, Acesso em: 23 de novembro de 2022.
- [5] Zero – Associação Sistema Terrestre Sustentável. “Alterações climáticas: relatório confirma Portugal como país a necessitar de fortes medidas de adaptação climática” Zero. Disponível em: <https://zero.org/alteracoes-climaticas-relatorio-confirma-portugal-como-pais-a-necessitar-de-fortes-medidas-de-adaptacao-climatica/>, Acesso em: 12 de dezembro de 2022.
- [6] Ascensão, Joana. “Uma em cada dez ambulâncias parada por falta de técnicos, 30% abandonaram a profissão: o que se passa, afinal, no INEM?” Jornal Digital Expresso. Disponível em: <https://expresso.pt/sociedade/2022-07-31-Uma-em-cada-dez-ambulancias-parada-por-falta-de-tecnicos-30-abandonaram-a-profissao-o-que-se-passa-afinal-no-INEM--0b76107c>, Acesso em: 23 de novembro de 2022.
- [7] Kashettar, Swathi. “How Machine Learning Is Improving Our Daily Life?” Analytics Insight. Disponível em: <https://www.analyticsinsight.net/how-machine-learning-is-improving-our-daily-life/>, Acesso em: 24 de março de 2023
- [8] Bugarcic, Pavle & Jankovic, Sladjana & Mladenovic, Snezana. (2021). *Forecasting Number of Calls to the Call center Using Machine Learning*.
- [9] Hanzhang Fu, Jessie Le, Jieying Wang, Parika Gupta. *Forecasting and Optimizing Call center Staffing*.
- [10] Erik CHROMY, Tibor MISUTH, Matej KAVACKY (2011). *Erlang C Formula and its use in the Call center*

- [11] Blog da Empresa Oncase, “Entenda o que é análise exploratória de dados e sua importância para o Data Science” Oncase. Disponível em: <https://www.oncase.com.br/blog/data-driven/analise-exploratoria-de-dados/>, Acesso em: 05 de janeiro de 2023
- [12] Firmino, Daniel. “Análise de Correlação” OPUS. Disponível em: <https://www.opuspesquisa.com/blog/tecnicas/analise-de-correlacao/>, Acesso em: 08 de janeiro de 2023
- [13] Empresa IBM. “Linear Regression” IBM. Disponível em: <https://www.ibm.com/docs/en/spss-statistics/saas?topic=features-linear-regression>, Acesso em: 09 de janeiro de 2023
- [14] M. Wooldridge, Jeffrey. (2013) *Introductory Econometrics, A Modern Approach*, Chapter 10-12
- [15] Calculadora de Erlang C. “Erlang Calculator - for Call Centre Staffing” Site Call Center Tools. Disponível em: <https://www.callcentretools.com/tools/erlang-calculator/>, Acesso em: 29 de junho de 2023

Glossário

LEI:	Licenciatura em Engenharia Informática
INEM:	Instituto Nacional de Emergência Médica
CODU:	Centro de Orientações de Doentes Urgentes
SNS:	Serviço Nacional de Saúde
PSP:	Polícia de Segurança Pública
GNR:	Guarda Nacional Republicana
SIEM:	Sistema Integrado de Emergência Médica