



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
UNIDADE ACADÊMICA ESPECIALIZADA EM CIÊNCIAS AGRÁRIAS  
ESCOLA AGRÍCOLA DE JUNDIAÍ  
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE  
E DESENVOLVIMENTO DE SISTEMAS

RELATÓRIO DO BANCO DE DADOS GLOBAL SHARK ATTACK: UMA  
ANÁLISE EXPLORATÓRIA DOS DADOS

MACAÍBA, RN

2025  
ELISA NASCIMENTO DOS SANTOS

**RELATÓRIO DO BANCO DE DADOS GLOBAL SHARK ATTACK: UMA  
ANÁLISE EXPLORATÓRIA DOS DADOS**

Trabalho apresentado à disciplina de Processamento Digital de Imagens do curso de graduação Tecnólogo de Análise e Desenvolvimento de Sistemas da Escola Agrícola de Jundiaí, Universidade Federal do Rio Grande do Norte, como parte da análise do banco de dados sobre Ataques Globais de Tubarões.

Orientador: Prof. Dr. Antonino Alves Feitosa Neto

MACAÍBA, RN  
2025

# Sumário

1. Introdução	4
2. O banco de dados	5
2.1 Hipóteses Iniciais	5
3. Qualidade dos Dados	7
3.1 Estatísticas básicas das variáveis	7
3.2 Análise das Distribuições	8
3.3 Tratando e identificando Outliers	10
3.4 Variáveis Independentes e Dependentes	12
3.5 Os valores faltosos se ligam ao problema?	12
3.6 Limpeza adicional	13
4. Transformações e Análise Exploratória	14
4.1 Análises de Agrupamento	15
4.1.1 Análise de Correlação, Ano e Idade	15
4.1.2 Redução de Dimensionalidade (PCA)	15
4.1.3 Baixa Dimensionalidade	16
4.1.4 K-Means usando método de Cotovelo	17
5. Análise Preditiva e Comprovação das Hipóteses	21
5.1 Qual foi o ano com mais ataques fatais?	25
5.2 Qual o país/área com mais concentração de ataques?	26
5.3 Os tubarões atacam mais quando são provocados ou não?	27
5.4 Quais são as áreas com maior número de ataques provocados e qual tem o maior número de acidentes?	27
5.5 Qual atividade marítima causa o maior número de ataques?	29
5.6 Qual a idade média entre as vítimas ao decorrer dos anos?	30
5.7 Qual horário mais perigoso para um ataque de tubarão?	30
6. Conclusões	32
Referências	33

## 1. Introdução

Ataques de tubarões acontecem todos os anos, eles ocorrem o ano todo em diversos locais do mundo, mas como sabemos de certa forma quantos ataques aconteceram em um ano em um certo local?

Felizmente o banco de dados de Mexwell nos trás as informações necessárias para respondermos essas perguntas. Assim, usando o que aprendemos até agora na matéria de Fundamentos e Técnicas em Ciência de Dados, vamos realizar análises exploratórias, explicativas e preditivas para alcançarmos valores que definem da forma mais precisa possível quantos ataques ocorreram no período de 1800 a 2023 ao redor do mundo.

Visando finalmente responder: os tubarões são realmente nossos predadores ou se eles nos mordem apenas quando provocados?

## 2. O banco de dados

Criado por Mexwell em 2023 e publicado no site Kaggle, a base relata os ataques de tubarões ao decorrer das décadas, tendo entradas não contendo anos e começando a definir a data a partir dos anos 1000. Não utilizamos algumas das variáveis do banco que não se ligavam às nossas hipóteses (variáveis específicas demais como por exemplo: nome da vítima, minuto e endereço do local de ataque).

Dentre as usadas temos o ano do ataque, a idade da vítima, a hora do ataque, o tipo do acidente, o país que ocorreu, área que aconteceu, que atividade estava fazendo antes de ser atacado, machucado causado pelo animal, se a ferida foi fatal se o ataque foi provocado ou não (ou se é incerto).

### 2.1 Hipóteses Iniciais

Primeiramente verificamos o ano com maior quantidade de ataques de tubarões em nossa base, o local desses ataques e o que aconteceu no mundo naquele ano, seja a permissão ou proibição da caça aos tubarões levando há um maior riscos de acidentes.

Em seguida, verificamos quantos ataques não são provocados, sendo realmente acidentais, voltando à questão dos tubarões atacarem ou não por instinto ou sobrevivência.

Será avaliado em que áreas ao decorrer dos anos têm maior chance de ocorrer um ataque de tubarão sem provocações do ser humano.

Verificaremos que atividades estão mais suscetíveis, nos próximos anos, a ter um ataque interrompendo a diversão, por exemplo nas olimpíadas seguintes de natação, competições de pesca ou surf, qual é a mais perigosa com os tubarões.

Qual idade ao decorrer dos anos foi mais atacada, seja os jovens, adultos ou senhores. Junto será também verificada se os ataques são mais de dia, quando se dá para ver os tubarões, ou a noite, quando eles se escondem nas águas.

Por último será verificado quais lugares são mais seguros em relação aos ataques e o que causa isso, seja a falta de dados ou algo natural como uma área de águas mais frias. Totalizando assim 7 hipóteses.

### 3. Qualidade dos Dados

Foi feita a análise das variáveis para verificar campos vazios e quais seriam melhores para a pesquisa, dividindo-as em dois tipos: Quantitativas (Ano, Idade e Hora do Ataque) e Qualitativas (Fatalidade, Tipo do Ataque, País, Área, Atividade e Espécie de tubarão que realizou o ataque)

Algumas colunas do banco original foram removidas, pois não eram de interesse para nossa análise, sendo essas:

- Nome, não estaremos analisando quem foram as vítimas.
- Localização exata (endereço), já possuímos área e país e de um modo global esses dois campos já são o suficiente.
- Nome, Ferimento e Sexo das vitimas, pois não estaremos investigando cada uma individualmente.
- PDF, href fórmula, href, case number, original order foram removidos por não adicionarem às hipóteses.

Foi feita então a limpeza dos dados numéricos e padronização das variáveis qualitativas. Reunimos o novo banco de dados tratado em seu próprio arquivo csv: “globalsharkattack\_tratado.csv” para melhor acesso e utilização.

#### 3.1 Estatísticas básicas das variáveis

Nos valores nulos de idade e ano foi feito uma média para colocar nos campos, além dos outliers serem removidos de ano mantendo o intervalo de 1800 a 2023, assim achamos que a tendência central do ano é 1986, da hora é 14h e a idade central é 24 anos, após isso para acharmos outras estatísticas básicas foi utilizado df.describe( ).

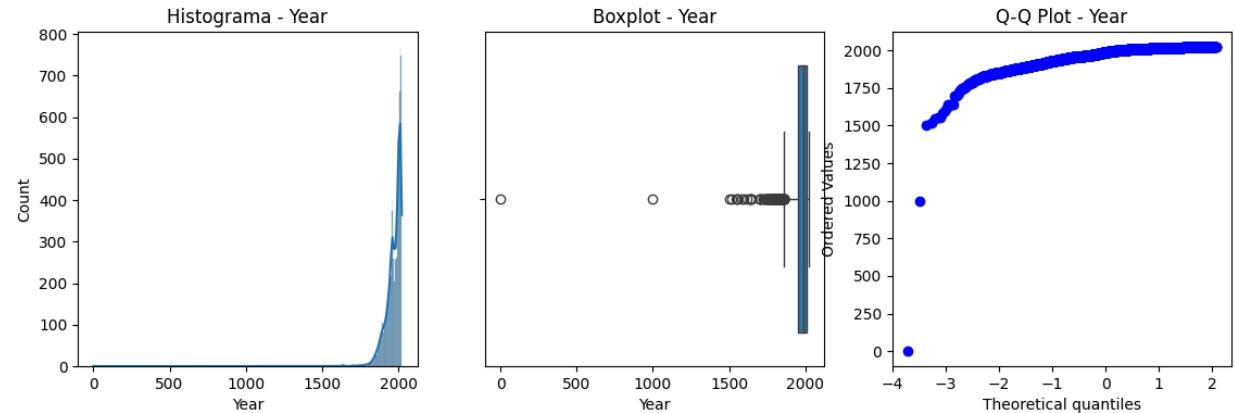
variáveis	média	moda	mediana	min	max	amplitude	std	25%	50%	75%	variância
ano	1973	2015	1986	1800	2023	223	44	1950	1986	2009	1958.37
idade	26	24	24	1	87	86	11	22	24	26	121.4
hora	13	14	14	0	23	23	0	14	14	14	6.89

Assim obtemos os valores de tendência central, dispersão e ordem. No caso do valor  $p$  para ambas as variáveis, por obtermos amostras maiores que 5000, nosso resultado foi anormal, sendo 0 para ambos.

### 3.2 Análise das Distribuições

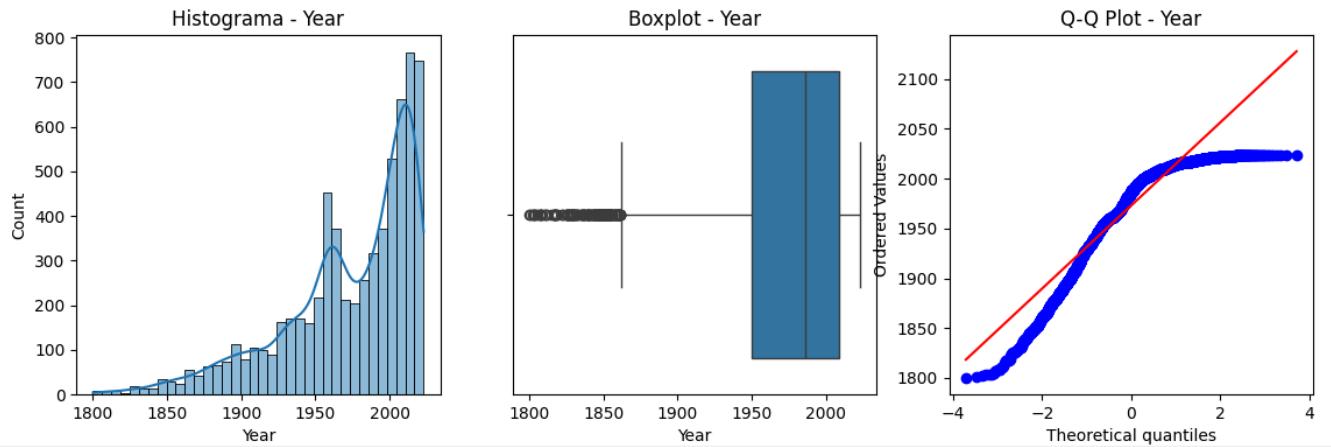
Fizemos o histograma, boxplot e Q-Q das variáveis numéricas, assim vendo os outliers de ambos.

Abaixo temos os gráficos de distribuição do ano, os outliers mais predominantes são no ano 0 e 1000, onde só possuímos uma informação de ambos, logo eles foram removidos.



Podemos observar que após aplicarmos um intervalo de 1800 à 2023 que os dados aumentam após 1900, sendo poucos antes disso, além de não possuirmos uma distribuição normal, mas sim uma assimétrica para à esquerda devido as caudas da linha.

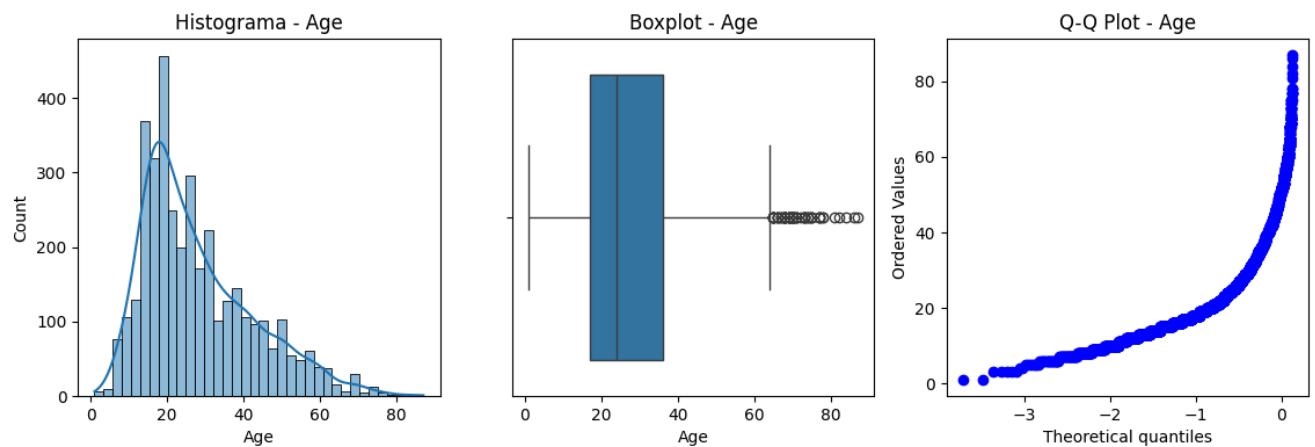
Além de mostrar que houve um aumento de casos ao decorrer dos anos e que o gráfico não é normal, precisando portanto de transformação antes de ser aplicado em algum gráfico estatístico.



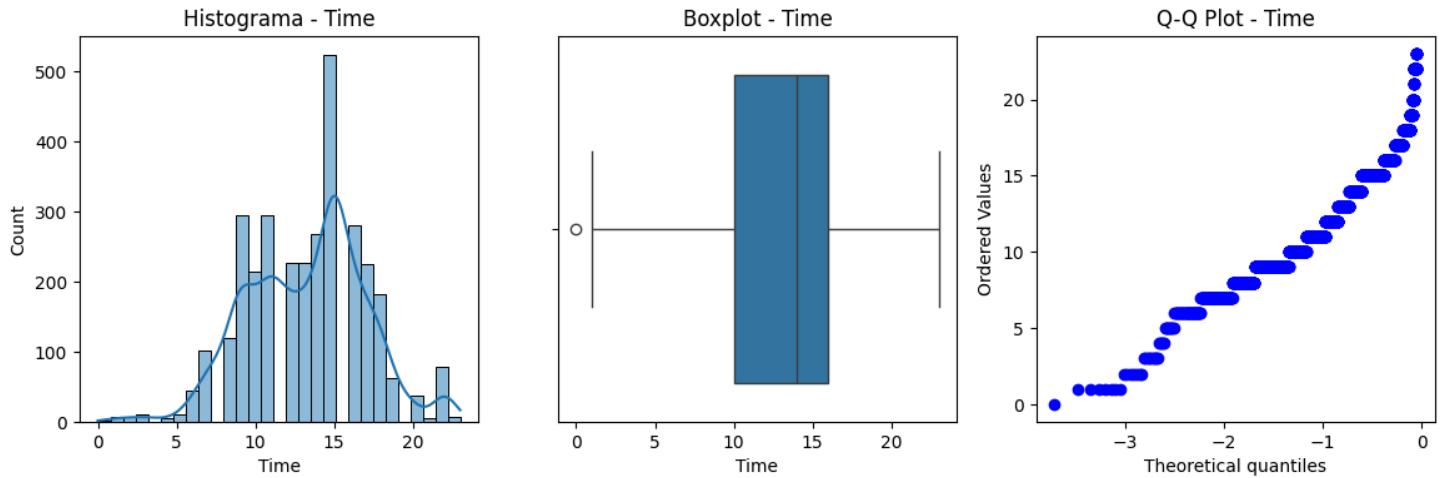
No histograma da idade não consideramos os outliers visíveis ao termos uma grande faixa etária, variando muito.

No caso do gráfico de Q-Q temos que a idade também possui grande assimetria a direita, estando quase totalmente concentrada na direita, devido aos casos de pessoas mais velhas e a enorme concentração dos ataques com pessoas entre 20 e 30 anos.

Logo, embora a maioria sejam jovens acidentados, ainda há uma quantidade maior do que a esperada de idosos.



Esses outliers das idades mais velhas não serão removidos, considerando que perderíamos o intervalo de idade.



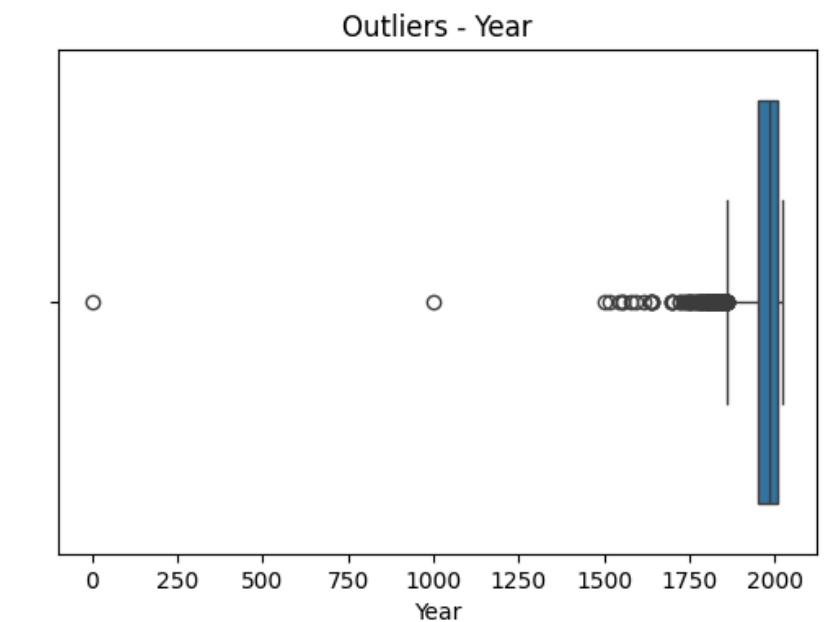
No histograma de tempo vemos uma concentração na parte da tarde dos ataques, além de uma montanha contínua de tempo, podemos verificar que os tubarões parecem atacar mais no horário do almoço e início da tarde.

Para os nossos outliers vemos que há apenas uma bolinha fora do grupo, sendo justamente os ataques que ocorreram na madrugada (0 horas da manhã).

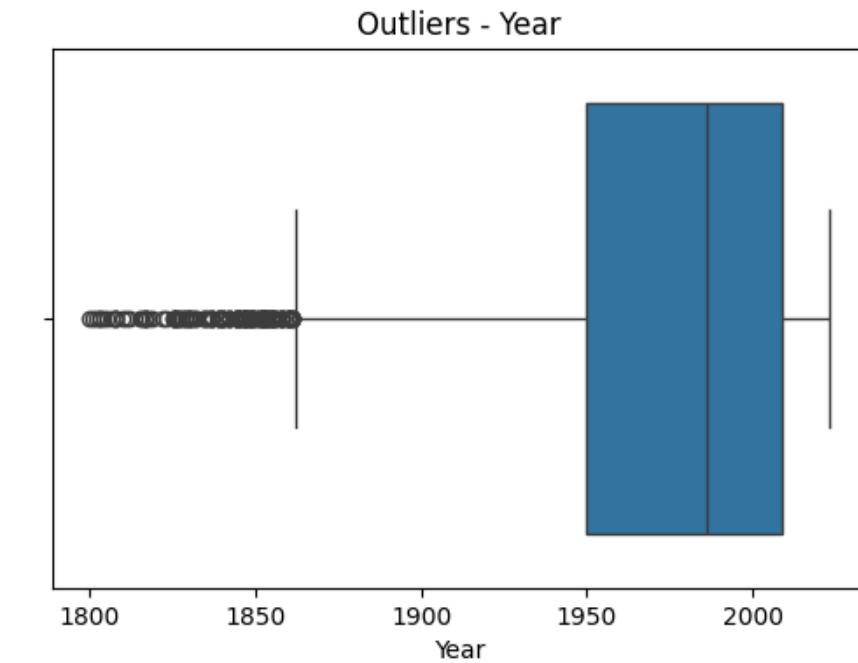
Já no gráfico q-q temos uma concentração na direita do gráfico devido a enorme quantidade de ataques a tarde.

### 3.3 Tratando e identificando Outliers

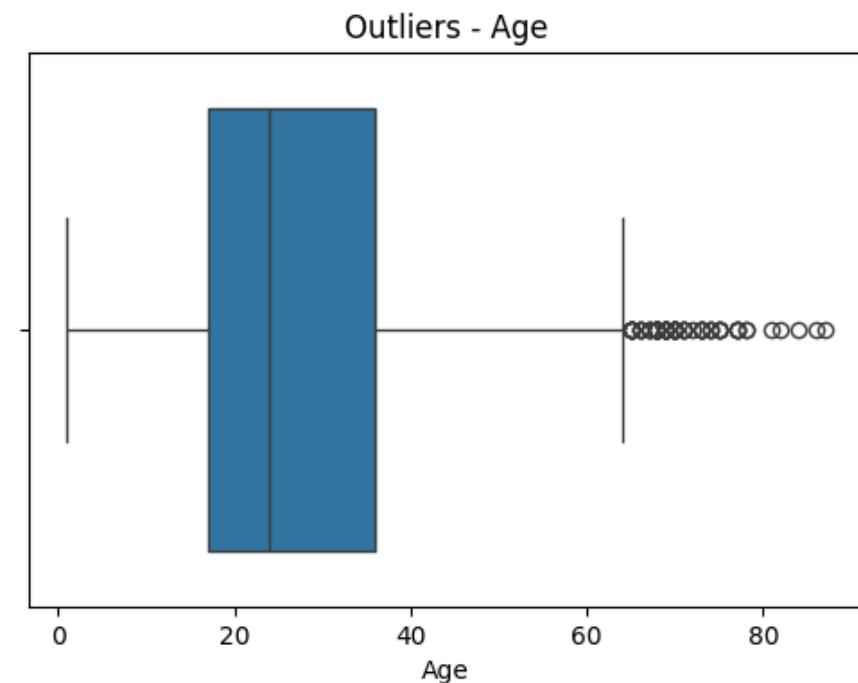
No banco inicial foram encontrados 81 outliers em ano e 73 em idade. Em ano os outliers se concentram fora do intervalo já incluído nos gráficos anteriores.



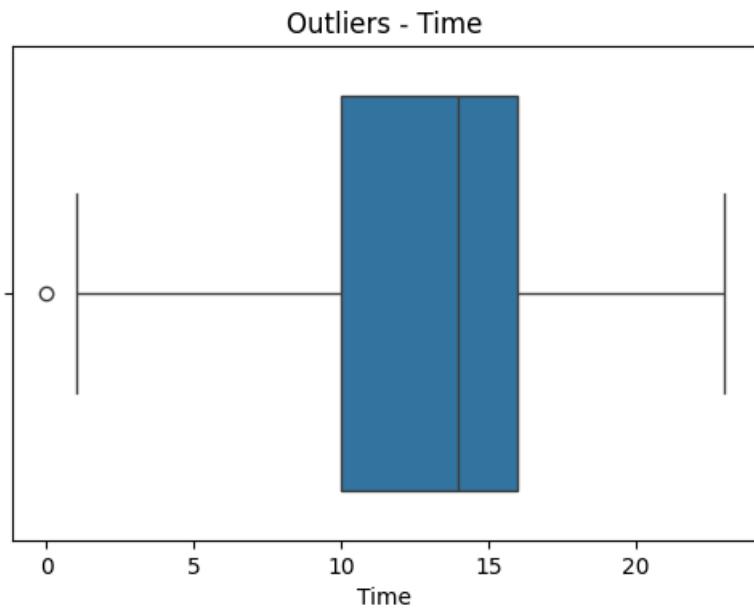
Aplicando a identificação de outliers por isolation forest no nosso banco de dados já tratado com o intervalo de 1800-2023 ainda possuímos outliers, mas esses serão mantidos de modo a continuar com nosso intervalo.



No caso da idade os outliers identificados foram justamente das pessoas mais velhas que tiveram um encontro com tubarões, nesse caso não serão removidos.



Nos nossos outliers de time temos apenas um, concentrado na esquerda, sendo justamente os ataques que ocorreram na madrugada às 0:00 horas, por isso não serão removidos, pois queremos provar se os ataques à noite superam ou não os à luz do dia e que hora os tubarões tendem a serem mais ativos ou provocados.



### 3.4 Variáveis Independentes e Dependentes

A variável independente é a variável alvo do projeto, se houve fatalidade ou não. Já as dependentes das fatalidades temos: idade, hora, sexo, tipo de acidente, ano, país, ação praticada no momento do acidente e área/local do ataque. A partir dessas colunas do banco visamos prever se um ataque é fatal devido às características do local ou da vítima.

No final temos 11 colunas principais das 21 do banco original.

### 3.5 Os valores faltosos se ligam ao problema?

Não, os campos faltosos após o tratamento desapareceram por completo, no caso foi ainda identificada 20 linhas duplicadas que foram removidas.

Coluna	Nulos Iniciais	% Nulos Iniciais	Nulos Finais
Ano	132	1,96%	0
Tipo	19	0,28%	0
País	51	0,76%	0
Área	481	7,17%	0
Atividade	586	8,72%	0
Fatalidade	0	0%	0

### 3.6 Limpeza adicional

No caso das variáveis qualitativas como fatalidade e tipo de ataque foi aplicada uma padronização, assim determinamos a fatalidade por 1 e 0, enquanto determinamos o tipo de ataque como 0 provocado, 1 não provocado e 2 questionável, para as situações onde não há como classificar com um ou outro.

No caso do tempo que possuímos a hora e os minutos fizemos um tratamento por meio de identificação de como a hora foi guardada, para caso seja diferente de 13h45 (13:45,13;45,morning, etc...) e guardamos apenas o número das horas, ou seja, em casos que o ataque foi às 2h35 no banco de dados tratado ficou 2.0.

Em caso de dados faltosos nesses campos foi colocado 0 para fatalidade (assim sendo não fatal) e no tipo foi substituído por 1, que é o caso mais ocorrente.

## 4. Transformações e Análise Exploratória

No caso dos dados faltosos restantes encontrados na etapa de tratamento das variáveis aplicamos uma regressão linear ou logística a depender de cada tipo. No caso das variáveis quantitativas foi usado o método de indução por regressão linear, e para o qualitativo foi usada a regressão logística.

Foi usado também uma métrica para medir desempenho da indução, no modelo de regressão linear foi usado  $R^2$  médio e no modelo de regressão logística acurácia. A predição foi feita levando a variável analisada como alvo e as demais do banco como preditoras, no caso do grupo numérico a predição foi feita por valor contínuo, na qualitativa foi usado valores binários para sim e não, e também foi utilizado classe múltipla para os casos com mais de duas possibilidades.

Assim, estamos medindo o quanto o modelo explica o comportamento das variáveis de idade e ano e medimos o quanto ele chegou perto do resultado.

Já na multiclasse permitimos o modelo escolher qual método seria melhor usar, para poucas variáveis foi usada a regressão logística medindo a acurácia do modelo acertar em qual dos diversos países o ataque aconteceu, por exemplo o modelo prevê que o ataque foi nos USA, mas na verdade ele foi no México, haverá uma taxa de acurácia para esses dados. Se houvesse muitas variáveis o modelo apenas substitui por moda. Assim evitando travamentos no sistema.

Mas, considerando que não tivéssemos feito o tratamento de dados faltosos na etapa de tratamento:

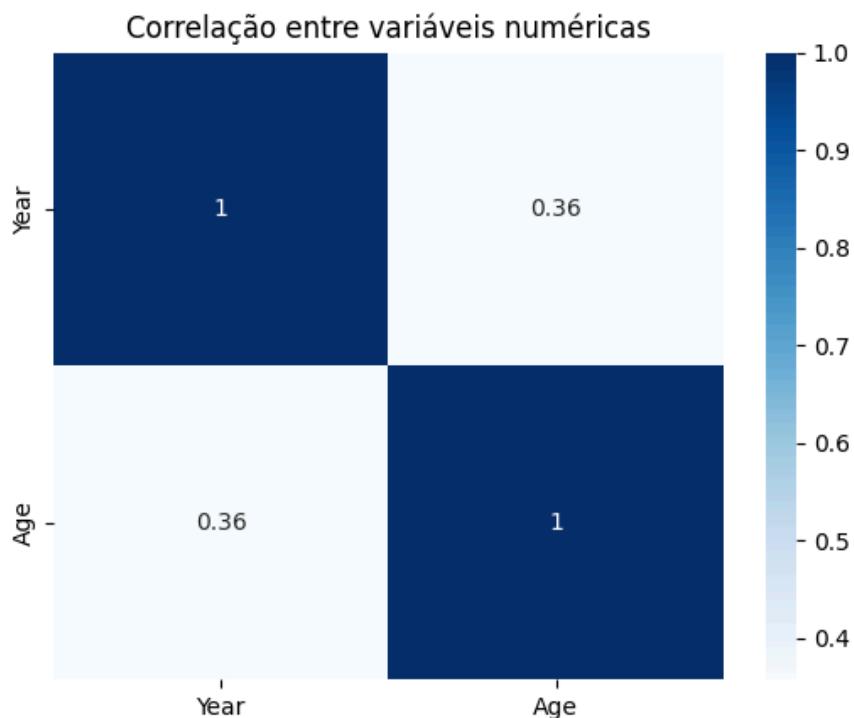
Variável	Tipo	Métrica	Valor
Year	Quantitativa	$R^2$	0.006
Age	Quantitativa	$R^2$	0.027

O restante permanece sem dados faltos. Podemos perceber que o ano e idade no banco não conseguem ser previstos pelas demais variáveis sendo muito aleatório, pois não há um padrão definido.

## 4.1 Análises de Agrupamento

### 4.1.1 Análise de Correlação, Ano e Idade

Para a correlação nosso resultado irá variar de -1 a 1, sendo -1 correlação negativa, a onde quando uma cresce a outra tende a cair, 0 pouca ou nenhuma relação e 1 para uma correlação positiva.



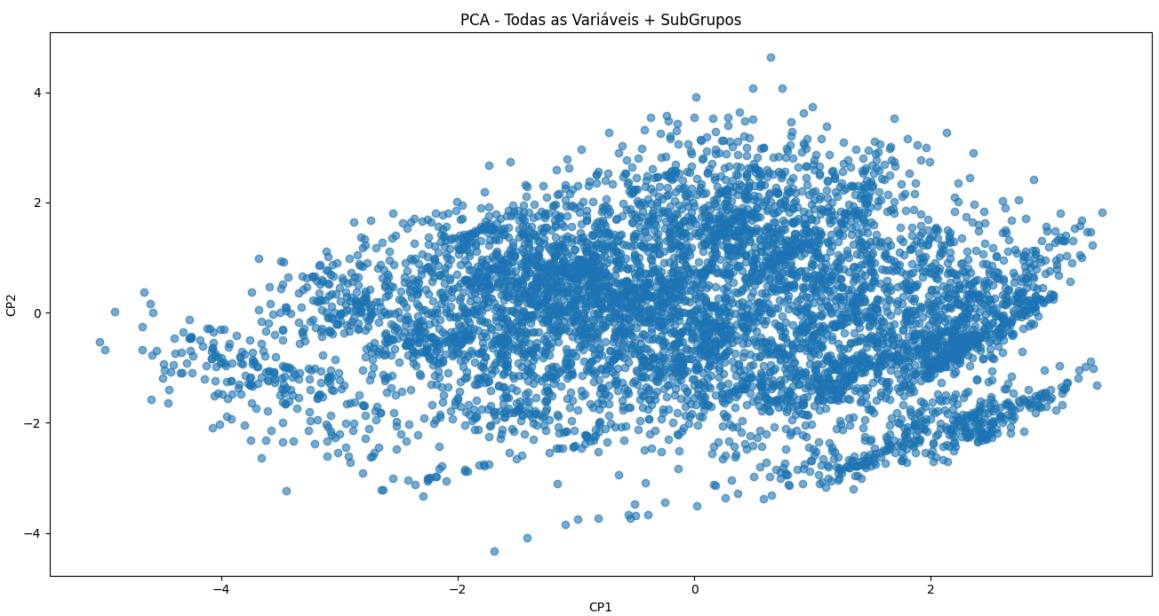
Temos uma correlação de 0.36 entre ano e idade, ou seja, eles tem quase nenhuma relação entre eles.

### 4.1.2 Redução de Dimensionalidade (PCA)

O Principal Component Analysis ou PCA busca diminuir a quantidade de colunas no banco por meio de agrupamento, remoção ou filtros como o de média.

Por meio dele, geramos novas variáveis que concentram a variação dos dados, permitindo ver grupos entre idade e ano, como por exemplo ataques mais recentes nos anos 2000 para cima terem vítimas mais novas, ou o contrário mais velhas.

Também criamos subgrupos



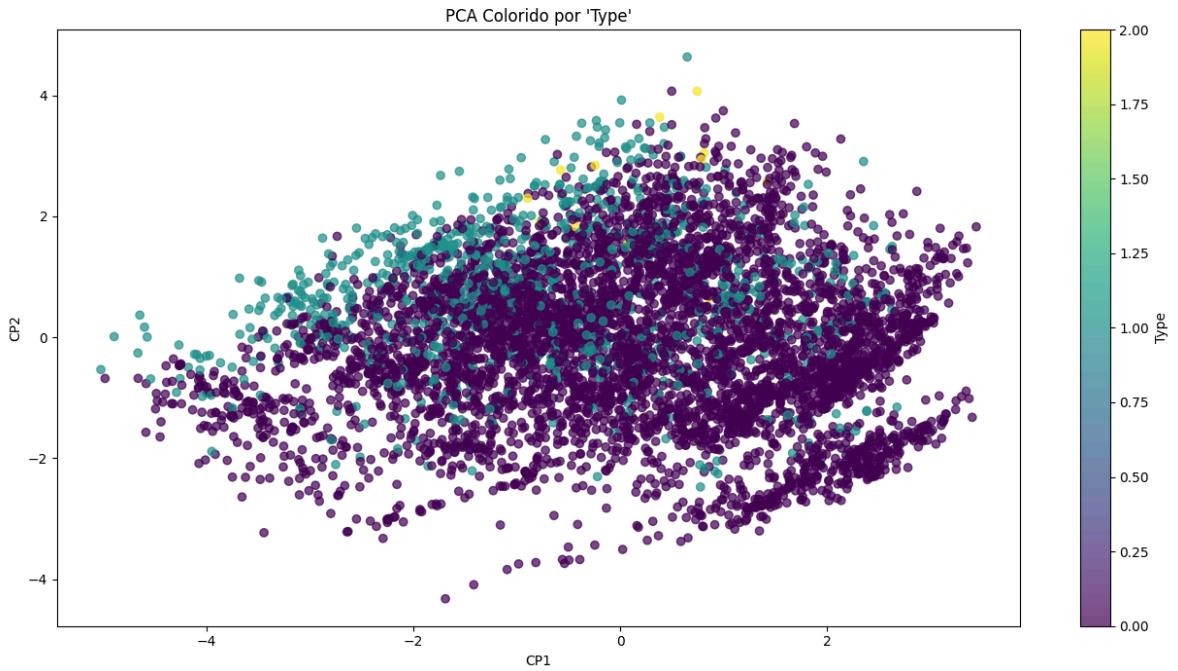
Temos o PCA com três grupos de variáveis as numéricas, categóricas e textuais que foram convertidas em números.

No gráfico abaixo podemos perceber que há uma concentração de ataques a pessoas mais velhas recentemente e um aumento dos mesmos após um certo período.

Variância explicada: [0.21345763 0.13207965]

#### 4.1.3 Baixa Dimensionalidade

Analisamos os dados em relação ao tipo de ataque ocorrido, colorindo de provocado (0), acidental (1) e questionável (2).



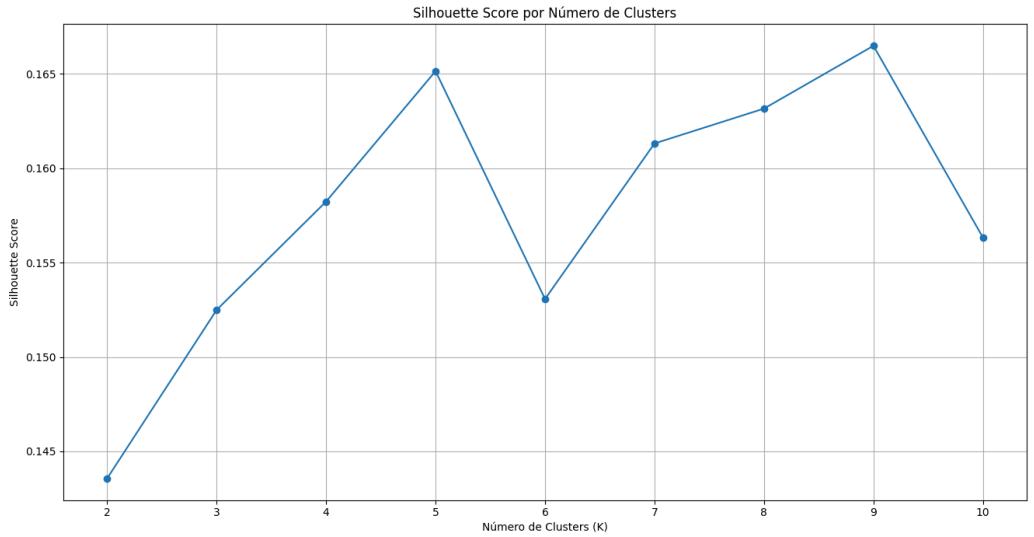
Embora não haja muita relação entre os anos, idades e tempo com o tipo do ataque, ainda podemos ver que há muitos mais casos de ataques provocados do que os realmente acidentais, especialmente entre os mais novos recentemente, mas tendo também uma concentração de acidentes com os mais velhos.

#### 4.1.4 K-Means usando método de Cotovelo

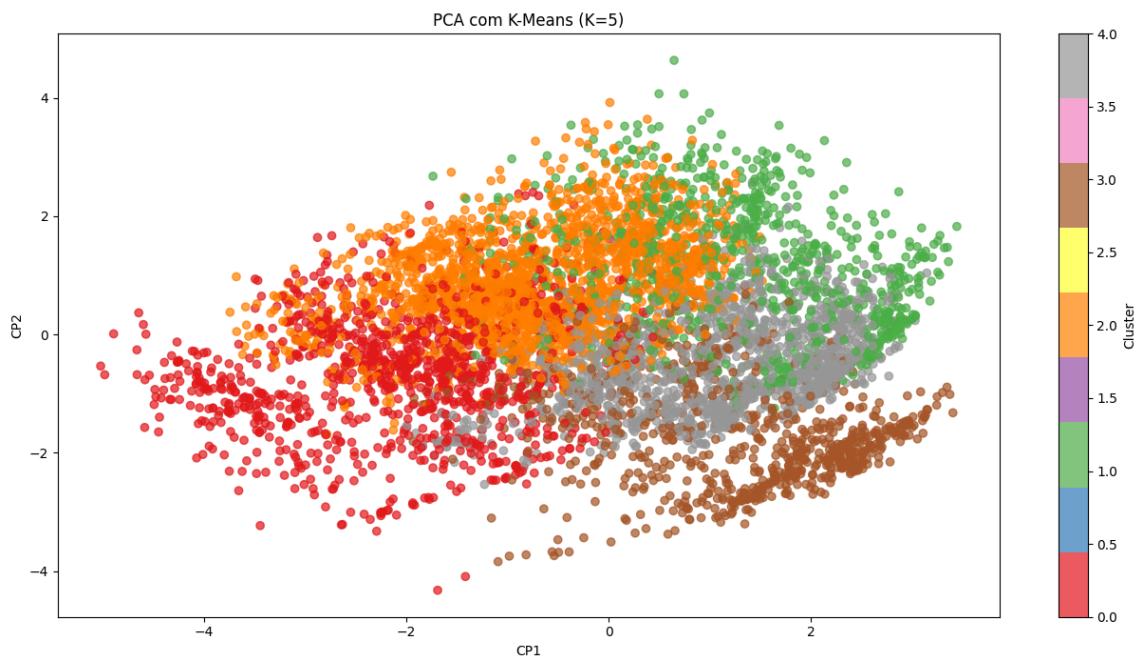
O k-means é um algoritmo que divide os dados em grupos de forma que eles mais se aproximem do centro/média, se usa o método de Cotovelo para acharmos o número ideal de grupos, rodando várias vezes e criando inúmeros grupos para ver a margem de erro/inércia entre os resultados.

Quando o gráfico de silhueta, que gera valores entre 1 e -1 para determinar a aproximação dos grupos e a distância de seus vizinhos, faça uma dobradura, o chamado cotovelo, temos o melhor grupo k, após isso não há para que adicionar mais grupos, o resultado não mudará muito.

Em nosso caso será escolhido um número de grupos ideais mantendo a simplicidade e interpretação, não precisando granular de mais.



Usando o método de silhueta temos que o melhor número de clusters, caracterizado pelo ápice do gráfico é 9, mas de modo a manter uma explicação mais limpa, mas específica usamos 6 que é a primeira queda do gráfico, poderíamos usar também 5 por ser o outro pico.



No gráfico fica notável que os grupos não são distintos, com isso as cores se misturam e sobrepõem. Mas ainda podemos determinar que:

- Verde: pessoas mais velhas com ataques variando entre provocados e acidentais nos anos mais recentes.
- Marrom: pessoas mais novas, mas ainda adultas que tiveram ataques provocados nos anos recentes.
- Cinza: Ataques na virada do milênio majoritariamente provocados por uma faixa etária mais nova.
- Vermelho: Ataques antigos e dispersos pelo mundo.
- Laranja: ataques com pessoas novas recentemente e ataques que não se encaixam com os outros.

Podemos perceber no gráfico que há uma virada entre os ataques justamente no momento que o gráfico se concentra melhor, sendo provavelmente após a virada do milênio quando os grupos vermelho e verde começam a dominar o gráfico.

Isso pode estar atrelado às proibições de caça aos tubarões, que começaram nos EUA nos anos 2000 e foram reforçadas em 2010. (*Shark Stewards*). E ao filme *Tubarão* de 1975, pois

"[...]por originar o chamado "efeito Tubarão", que supostamente inspirou "legiões de pescadores [que] se amontoaram em barcos e mataram milhares de predadores do oceano em torneios de pesca de tubarões[...]." (Wikipédia)

Esse aumento do medo por tubarões levou a caça desenfreada dos mesmos, resultando na diminuição da espécie e acidentes de pescadores menos preparados.

Com a abolição da caça temos a retornada com tudo dos animais e com isso o aumento de casos.

Tendo nossos grupos e análise podemos deduzir que os ataques a pessoas mais velhas recentemente se deve ao fato da proibição da caça aos tubarões, após quase extinção das espécies, aumento da expectativa de vida da população, uma situação mais tranquila nos estados unidos, e a poluição/aquecimento que vem atraindo cada vez mais peixes as praias e costas e com isso tubarões vem junto.

De acordo com Ifaw

*"O rápido aquecimento dos oceanos tem ativado uma improcedente imigração em massa da vida marinha. As temperaturas altas têm empurrado várias espécies marinhas, incluindo tubarões, para os pólos, especialmente nos meses do verão por procurarem por água mais frias, o que significa que eles estão se movendo para novos habitats, incluindo as costas onde eles não costumam ficar[...]"*

Com o aquecimento nós também temos a preferir as praias costeiras do que ficar em casa e quando humano e tubarão se encontram tende a dar briga.

## 5. Análise Preditiva e Comprovação das Hipóteses

Para fazermos nossa análise preditiva, ou seja, prevermos dados futuros a partir dos que já temos, escolhemos um alvo, no caso da nossa pesquisa será a coluna de fatalidade.

Assim verificamos a relação das outras variáveis com nosso alvo, descartando apenas as colunas Área e Injury, por serem muito diversificadas.

Já com as outras variáveis (tempo, ano, idade, país, atividade) foi um processo mais simples, três delas já eram numéricas e país e atividade, assim como fizemos para o PCA, foram divididas em grupos para retornar numéricos como 0,1,2,3 quando o dado se encaixava no campo.

Usamos da regressão logística, havendo assim o cálculo do logit para predição, temos um pseudo  $R^2$  de 15% de variação para a fatalidade e nosso teste de verossimilhança teve resultados baixos indicando que o modelo como um todo é significativo, ou seja, todos nossos campos são importantes para os ataques de tubarões futuros e que nenhum acontece de forma randômica.

Variável Alvo	Fatal (Y/N)
Observações	6712
Modelo	Logit
Método	MLE (Estimativa Mínima de Verossimilhança)
Tipo de Covariância	Não Robusta/ Métodos Tradicionais

Em relação a ligação de cada coluna com as fatalidades temos que:

- Os ataques estão ficando menos fatais ao decorrer dos anos, sendo que a cada 1 ano a chance de fatalidade cai cerca de 1,7% aproximadamente, podemos atribuir essa queda ao avanço da ciência e melhor observação da vida marítima por cientistas ou salva-vidas

- Quanto mais velha é a pessoa atacada pelo tubarão maior será a chance de acabar morrendo, devido aos problemas da idade.
- O tempo não muda praticamente nada na chance de ser fatal ou não o ataque. Sendo totalmente irrelevante, mas ainda necessário para nossa análise.
- O tipo de ataque também diminui a chance de ser fatal, ataques provocados tiveram mais chance de sobrevivência do que os ataques acidentais, pode-se entender que a vítima ao atacar primeiro já estava preparada para o tubarão.
- Já para os países quanto mais longe da América, ataques mais esporádicos considerando que a América tem a maior concentração de tubarões, maior é a chance de fatalidade. Isso se atribui a menor incidência de ataques e com isso a menor preparação quando um ocorre.
- Já a atividade em si praticada na hora do ataque não determina se o ataque em si será fatal.

Tendo essas informações podemos ver uma queda significativa nos ataques ao decorrer dos anos devido o avanço das pesquisas, um risco maior para pessoas mais velhas considerando a fragilidade que vem com a idade e que quando o ataque é realmente acidental maior é a chance que a vítima pereça.

Em relação a região, temos países como África e Ásia como os mais fatais quando o assunto são os ataques, no caso da África do Sul vemos essa concentração de mortes devido a área ser um destino famoso para tubarões brancos devido as colônias de focas que se amontoam na região, além de uma menor preparação e atendimento no caso de ataques (South Africa).

Já a Ásia possui uma variedade de espécies de tubarões sendo outro destino muito procurado para mergulhar entre eles, embora tenha um preparo melhor de hospitais e resgates a área também enfrenta uma batalha contra a sobrepesca para comércio de barbatanas de tubarão o que deixa os animais se sentirem ameaçados e mais suscetíveis a atacar (Lucia Malla, 2005).

	coef	std err	z	P >  z	0.025	0.975
const	30.9940	1.676	18.493	0	27.709	34.279
Ano	-0.0172	0.001	-20.199	0	-0.019	-0.016
Idade	0.0149	0.003	4.497	0	0.008	0.021
Tempo	0.0029	0.014	0.205	0.837	-0.025	0.031
Tipo	-0.6974	0.095	-7.336	0	-0.884	-0.551
País	0.2795	0.018	15.955	0	0.245	0.314
Atividade	-0.0464	0.031	-1.501	0.133	-0.107	0.014

Com isso temos que Ano, Idade, Tipo e País são as colunas que mais afetam a fatalidade.

Por meio de Cross-Validation, testar nosso banco de dados várias vezes para verificar a nossa acurácia, verificamos que o modelo acerta 81% das previsões média sendo estável devido os folds próximos, não havendo também um overfitting.

- Acurácia por Fold: [0.804, 0.810, 0.810, 0.810, 0.807]
- Acurácia Média: 0.8085

Para medir a separação das classes usamos AUC (Área sob a curva), com ele foi determinado que o modelo criado consegue separar dados desbalanceados de forma boa, mas ainda acaba havendo muitos erros.

- AUC por Fold: [0.775, 0.754, 0.759, 0.765, 0.789]
- AUC médio: 0.769

Tendo que os valores variam de 0 a 1, sendo zero um modelo que chuta a separação das classes e 1 um modelo quase impossível de existir por separar as classes com precisão.

Para os casos em que houve problemas de classificação usamos da matriz de confusão obtendo o resultado:

5159	192
------	-----

1093

268

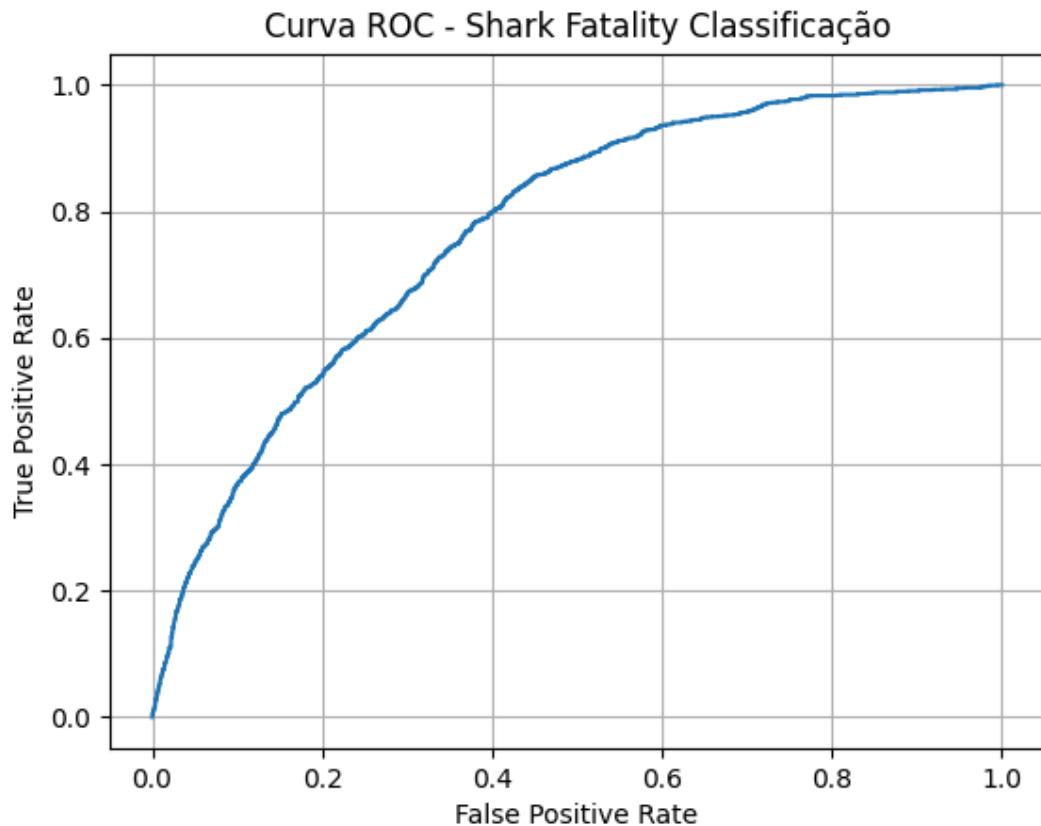
Isso nos mostra que em ataques não fatais o modelo tende a ter uma quantidade de acertos maior, já em ataques fatais ele tem uma tendência a cometer mais erros, justamente por termos poucos ataques que foram mortais (20% dos casos).

Por meio das métricas podemos ver isso melhor, onde o modelo não deixa passar um caso acidental e acaba tendo um desempenho menor nos casos fatais.

	precisão	recall	F1	suporte
fatal	0.83	0.96	0.89	5351
acidental	0.58	0.20	0.29	1361
acurácia			0.81	6712
macro avg	0.7	0.58	0.59	6712
weighted avg	0.78	0.81	0.77	6712

Por último, foi analisada a potência do modelo em separar duas classes por meio do AUC-ROC (área sob a curva característica de operação do receptor) independente de seu cutoff, a área da curva ROC que maximiza a sensibilidade e especificidade.

O resultado foi de 0.7686 sendo aceitável para fenômenos reais e tendo 77% de chance de ter uma classificação positiva. Na curvatura ROC isso fica mais evidente ao virmos que ela sobe e se aproxima do canto superior esquerdo, mostrando que a menor falsos positivos em nosso modelo e comprovando o resultado ao ter a curva mais acentuada entre 0.6 e 1.



Tendo nossos dados futuros e presentes podemos voltar às nossas hipóteses iniciais.

### 5.1 Qual foi o ano com mais ataques fatais?

O ano com maior número de ataques foi 1959 com 24 mortes do total de 93 ataques, comparando esse ano com o restante obtemos por teste de proporção 1,336 de 0,1817 assim tornando esse ano atípico.

Esses ataques ocorreram principalmente pela explosão de popularidade do surf no Havaí, justamente no ano que o Havaí virou estado dos EUA e a guerra fria estava em seu auge. Várias famílias foram morar no litoral e muitos soldados também migraram para esse novo ponto estratégico contra a USSR.

De acordo com o G1:

*"O Havaí havia se transformado em um ponto fundamental de escala e abastecimento. O arquipélago servia de plataforma para possibilitar à Marinha americana, cada vez mais poderosa, atacar alvos espanhóis nas Filipinas."*

Com o mar agitado os tubarões acabaram se aproximando mais da costa

### 5.2 Qual o país/área com mais concentração de ataques?

O país com maior número de ataques foi os Estados Unidos obtendo 2498 acidentes ao decorrer dos anos, a seguir está o top 10 países com mais ataques.

País	Ataques
EUA	2498
Austrália	1444
África do Sul	589
Nova Zelândia	142
Nova Guínea	132
Bahamas	129
Brasil	122
México	99
Itália	68
Fiji	67

Nos Estados Unidos essa concentração se deve à maior preservação dos tubarões nos anos mais recentes e a abundância de presas dos tubarões nas costas como peixes, focas, lobos-marinhos e até algumas aves, além da proibição da caça às espécies, de acordo com National Geographic:

*"Mais países, dos EUA ao Sul do Pacífico, estão aprovando e aplicando leis que proíbem a pesca de tubarões e, como resultado, os animais têm mostrado alguns sinais localizados de recuperação[...]"*

Isso se deve também ao fato que essas regiões do top 10 são principalmente de clima quente, atividades aquáticas intensas e um grande litoral.

### 5.3 Os tubarões atacam mais quando são provocados ou não?

Em nosso banco de dados temos 5539 ataques provocados, 1156 acidentais e 17 que não a como decidirmos, por meio do teste de proporção temos que os ataques provocados são 87% maiores que os provocados com z-score de 53.5 e  $p = 0$ .

Isso se dá pois o tubarão não é nosso predador natural, muitas das vezes ele morde nos confundindo com outra coisa, principalmente no surf, onde ele confunde o surfista com focas e tartarugas. De acordo com o Metrópole:

*"Apesar de ter ganhado a fama de ser um predador feroz de humanos, o oceanógrafo Ricardo de Souza Rosa explica que, geralmente, o tubarão-branco não ataca pessoas de forma intencional, já que não fazemos parte da rotina alimentar do animal."*

Essa mordida não intencional tem até um nome: "Mordida Exploratória". Nos casos dos ataques provocados temos casos mais comuns como tentar encostar em um tubarão, cercá-lo, prendê-lo em uma rede ou anzol e ficar entre ele e sua presa.

Nesses casos a verdadeira vítima é o tubarão que foi surpreendido por uma interação humana.

### 5.4 Quais são as áreas com maior número de ataques provocados e qual tem o maior número de acidentes?

A área com mais ataques não provocados foi Nova Gales do sul com 93 ataques acidentais, já a área com maior número de provocações com os tubarões foi a Flórida.

Área	Acidental
Nova Gales do Sul	93
Florida	92
California	76
Queensland	55
Kwazulu-Natal	53

Cabo Ocidental	50
Havaí	47
Austrália Ocidental	42
Austrália do Sul	32

Área	Provocado
Florida	1070
Nova Gales do Sul	420
Queensland	283
Havaí	283
California	241
Austrália Ocidental	181
Kwazulu-Natal	160
Havaí	47
Cabo Oriental	147
Cabo Ocidental	145

A Flórida é um lugar recorrente em ambas as pesquisas devido à sua concentração de atrações voltadas a tubarões, como o mergulho com eles, como mostra o site TripAdvisor, e torneios clandestinos para a caça de barbatanas, de acordo com Wild beim Wild (grupo de interesse pela vida selvagem):

*"Trata-se, essencialmente, de um esquema para ganhar dinheiro, semelhante à caça de troféus 'legal' na África", disse Raven Lynette, mergulhadora e ativista da causa dos tubarões da Califórnia. "Quem paga por essas licenças [...] pescadores que não têm conhecimento sobre a importância dos tubarões ou simplesmente pensam que eles são inimigos, [...] o que facilita aos pescadores matar e afundar espécies protegidas. Além disso, NÃO HÁ justificativa para matar tubarões[...]"*

Com isso fica claro que os ataques provocados concentrados na Flórida se dão pela pesca desenfreada e sem sentido das espécies de tubarão, além da pouca concentração da fiscalização.

## 5.5 Qual atividade marítima causa o maior número de ataques?

A atividade com mais ataques é justamente o Surfing tento 1110 ataques

Atividade	Total	Fatalidade	Proporção
Surfar	1110	0.05	0.95
Nadar	981	0.32	0.89
Pescar	487	0.09	0.69
Pesca com Lança	381	0.12	0.92
Wading	167	0.07	0.93
Banhar	160	0.42	0.92
Mergulhar	133	0.13	0.88
Mergulho Livre	126	0.15	0.96
Stand Up Paddle	113	0.12	0.99

A título de curiosidade Wading e pescar com um espeto enquanto anda na água e Stand Up Paddle é manter o equilíbrio em um prancha enquanto rema pela água.

A atividade esportiva mais segura é o Surf Skiing onde o atleta fica dentro de um caiaque longo e comprido remando pelas ondas, nesse caso a fatalidade foi indicada como 0, pois o caiaque protege a pessoa do tubarão.

Além dessa atividade se concentrar em águas doces e os tubarões são mais comuns nas salgadas. De acordo com a Wikipédia livre:

*"Enquanto a maioria dos tubarões são predominantemente marinhos, um pequeno número de espécies de tubarões se adaptaram a viver em água doce: os tubarões de rio (do gênero *Glyptis*) [...] O tubarão touro[...]"*

A atividade em si também protege o banhista no momento que ele não passa muito tempo parado esperando por ondas, facilidade para subir novamente no caiaque, a remada tende a ser mais suave e silenciosa que o bater de asas e pernas ao surfar e nadar, além de que os atletas evitam áreas mais fundas.

De acordo com surfski info:

*"Desde 2002, três atletas de surf skis foram mordidos e severamente feridos por Tubarões Brancos na False Bay, aqui em Cape Town, África do Sul."*

## 5.6 Qual a idade média entre as vítimas ao decorrer dos anos?

Ao decorrer dos anos, a idade média das vítimas se manteve em 24 anos de idade, mas nos anos mais recentes de 1990- 2023 a idade das vítimas vem crescendo aos poucos ao ponto de chegarmos a uma média de 31 anos em 2023.

Embora não seja algo que podemos com certeza apontar o porquê do crescimento de ataques a vítimas mais velhas, podemos teorizar que seja devido a maior expectativa de vida e a concentração de jovens adultos voltando a hobbies antigos como pesca, ou praticando esportes novos como surf.

Além que nessa faixa etária muitas pessoas já possuem emprego fixo e estão voltando a praticar esportes, viajar e sair para se divertir.

## 5.7 Qual horário mais perigoso para um ataque de tubarão?

O horário mais perigoso para ataques acidentais e provocados é as 14h da tarde, existem alguns fatores que comprovam nosso banco de dados

Como, de acordo com American Ocean's em uma pesquisa;

*"De acordo com o Arquivo Internacional de Ataques de Tubarão, a maioria dos ataques ocorre entre às 8h e às 18h. Isso pode ser devido ao fato de haver mais pessoas na água durante esses horários."*

Essa maior concentração de banhistas em horários de pico de atividade dos tubarões causa a concentração de horários em média às 14h da tarde. Esse período principalmente nos meses de maio à agosto é quando os tubarões estão migrando para águas mais quentes e estão mais concentrados nas orlas para comer.

De acordo com Eviroliteracy:

*"Os grandes tubarões brancos são mais ativos durante o dia, principalmente no início da manhã e no final da tarde . Esses períodos geralmente coincidem com o aumento da atividade de suas presas, como focas e leões-marinhos, que frequentam águas mais rasas perto da costa. [...]"*

Como dito anteriormente a sombra de um surfista por baixo da água é facilmente confundível com o de uma foca ou leão-marinho nadando, dessa forma o tubarão pensa que achou uma presa fácil quando na verdade é alguém querendo se divertir.

## 6. Conclusões

Por fim, por meio da análise do banco de dados podemos concluir que os ataques tiveram um notável aumento ao longo dos anos, especialmente após o banimento da caça no início dos anos 2000. Outros fatores como as mudanças ambientais e maior presença humana no mar também contribuem para a maior concentração nas costas.

Também percebemos a concentração de ataques não fatais que só acontecem pela provocação direta ao tubarão, reforçando assim que eles não nos atacam sem motivo, ou eles nos confundem com uma presa ou reagem ao ambiente.

Entre as variáveis idade, horário, país e atividade conseguimos entender melhor o perfil das ocorrências e que embora há coincidências não há um padrão forte o suficiente para haver previsões exatas para futuros ataques.

Mesmo assim, as tendências já encontradas nos ajudam a mapear riscos e entender melhor como fatores humanos e ambientais moldam os encontros.

## Referências

US Shark Fin Trade Ban, Shark Stewards, Disponível em: <https://sharkstewards.org/shark-science-education/us-shark-fin-trade-ban/>, Acesso em: 12 de novembro de 2025.

Tubarão(Filme), Wikipédia, Disponível em: [https://pt.wikipedia.org/wiki/Tubar%C3%A3o\\_\(filme\)#:%7E;text=O%20filme%20tamb%C3%A9m%20desempenhou%20um%20papel%20importante,bem%20como%20seus%20sucessos%20de%20bilheteria%20pretendidos.](https://pt.wikipedia.org/wiki/Tubar%C3%A3o_(filme)#:%7E:text=O%20filme%20tamb%C3%A9m%20desempenhou%20um%20papel%20importante,bem%20como%20seus%20sucessos%20de%20bilheteria%20pretendidos.), Acesso em: 12 de novembro de 2025

Clark, Brian Clark H., Animal Geographic, Disponível em: <https://www.nationalgeographicbrasil.com/animais/2019/06/ataques-de-tubarao-o-que-voce-precisa-saber> , Acesso em: 19 de novembro de 2025

Como EUA tomaram o Havaí para torná-lo um posto avançado de segurança nacional, Disponível em: <https://g1.globo.com/mundo/noticia/2025/03/29/como-eua-tomaram-o-havai-para-torna-lo-um-posto-avancado-de-seguranca-nacional.html> , Acesso em: 19 de novembro de 2025

MALLA, Lucia. Ser tubarão na Ásia. Disponível em: <https://luciamaalla.com/ser-tubarao-na-asia.html#:%7E;text=Mas%20na%20%C3%81sia%2C%20a%20situa%C3%A7%C3%A3o,Veja%20s%C3%B3> . Acesso em: 24 de novembro de 2025

South Africa. Mergulho com tubarões brancos. Disponível em: <https://www.southafrica.net/br/pt/travel/article/mergulho-com-tubar%C3%B5es-brancos> . Acesso em: 24 de novembro de 2025

10 melhores mergulho com tubarão em florida. TripAdvisor. Disponível em:

<https://www.tripadvisor.com.br/Attractions-g28930-Activities-c61-t195-Florida.html> . Acesso: 24 de novembro de 2025

Torneio polêmico de caça a tubarões na Flórida recompensa caçadores. Wild beim Wild. Disponível em:  
<https://wildbeimwild.com/pt/torneio-polemico-de-caca-a-tubaroes-na-florida-recompensa-cacadores/#:~:text=Torneio%20pol%C3%AAmico%20de%20ca%C3%A7a%20a%20tubar%C3%B5es%20na%20Fl%C3%BDa%20recompensa%20ca%C3%A7adores,-Equipe%20editorial%20&text=O%20pescador%20Corey%20Hexter%20compartilhou,dos%20animais%20e%20moradores%20preocupados> . Acesso em: 24 de novembro de 2025

Freshwater shark. Wikipedia. Disponível em:  
[https://en.wikipedia.org/wiki/Freshwater\\_shark#:%7E:text=Tubar%C3%B5es%20de%20%C3%A1gua%20doce%20s%C3%A3o,com%C3%A9rcio%20de%20peixes%20de%20aqu%C3%A1rio%20](https://en.wikipedia.org/wiki/Freshwater_shark#:%7E:text=Tubar%C3%B5es%20de%20%C3%A1gua%20doce%20s%C3%A3o,com%C3%A9rcio%20de%20peixes%20de%20aqu%C3%A1rio%20) . Acesso em: 24 de novembro de 2025

Shark. Surfski.info. Disponível em:  
[https://surfski.info/latest-news/tag/shark.html?start=10#:%7E:text=Since%202000%2C%20three%20surf,their%20own%20environment\)%20and%20humans](https://surfski.info/latest-news/tag/shark.html?start=10#:%7E:text=Since%202000%2C%20three%20surf,their%20own%20environment)%20and%20humans) . Acesso em: 24 de novembro de 2025

10 lugares onde ocorreram mais ataques de tubarão. American Ocean 's. Disponível em:  
[https://www.americanoceans.org/facts/where-do-shark-attacks-happen-most/?utm\\_source=chatgpt.com](https://www.americanoceans.org/facts/where-do-shark-attacks-happen-most/?utm_source=chatgpt.com) . Acesso em: 24 de novembro de 2025

Em que horário do dia os grandes tubarões brancos são mais ativos?. Eviroliteracy Team. Disponível em:

[https://enviroliteracy.org/what-time-of-day-are-great-white-sharks-most-active/?utm\\_source=chatgpt.com](https://enviroliteracy.org/what-time-of-day-are-great-white-sharks-most-active/?utm_source=chatgpt.com) . Acesso em: 24 de novembro de 2025.