

Equipe: Cleidiane Souza, Isabela Marinho, Leonardo Melo, Milena Cordeiro e Mônica Rivas.

Especialização em Data Science & Analytics - Senai Cimatec

Relatório

Introdução

A **Expectativa de Vida** é uma medida essencial para compreender a saúde e o bem-estar de uma população. Ela representa o número médio de anos que os indivíduos de um país podem esperar viver, com base nas condições de vida existentes no momento do nascimento. A **Expectativa de Vida** é influenciada por uma variedade de fatores, incluindo educação, saúde, saneamento básico, segurança no trabalho e índices de violência. Esses fatores desempenham um papel crucial na determinação da qualidade de vida e na formulação de políticas públicas.

Este relatório tem como objetivo apresentar o resultado da análise dos dados relacionados à expectativa de vida e fatores de saúde para 193 países, com base no conjunto de dados coletados no site de repositório de dados da Organização Mundial da Saúde (OMS), juntamente com os dados econômicos correspondentes obtidos do site das Nações Unidas. A análise desses dados nos ajudará a compreender melhor os fatores que influenciam na **Expectativa de Vida** dos países analisados bem como as tendências e padrões relacionados à saúde global.

O conjunto de dados inclui informações sobre a **Expectativa de Vida** e uma variedade de fatores de saúde, como taxa de mortalidade infantil, taxa de mortalidade de adultos e incidência de doenças infecciosas. Além disso, também foram coletados dados econômicos, como o PIB per capita e o índice de desenvolvimento humano (IDH).

A análise desses dados permitirá uma compreensão e identificação mais aprofundada das tendências e padrões globais de saúde, bem como da relação entre saúde e desenvolvimento econômico, para a criação de um modelo de regressão linear que preveja quais fatores influenciam na **Expectativa de Vida** dos países analisados.

Esses insights podem ser fundamentais para a formulação de políticas eficazes que visem melhorar a qualidade de vida da população, reduzir as desigualdades e promover um desenvolvimento sustentável.

Ao longo deste relatório, serão apresentados os principais resultados da análise de dados, destacando correlações e tendências identificadas nos indicadores de saúde e econômicos. Além disso, serão oferecidas recomendações estratégicas para ações governamentais, com base nas descobertas obtidas a partir dessa análise abrangente.

É importante ressaltar que as informações contidas neste relatório são fundamentadas em dados reais e representam uma valiosa fonte de conhecimento para apoiar a tomada de decisões informadas por parte dos gestores dos países.

Metodologia

Inicialmente, analisou-se os dados fornecidos sobre os países e suas respectivas expectativas de vida, bem como as outras variáveis relacionadas. Para tanto, realizou-se a análise exploratória dos dados que, dentre outros, identificou a ausência de muitos dados e a existência de valores atípicos, ou seja, valores fora do padrão de comportamento dos demais para o mesmo atributo no conjunto. Sendo realizado os devidos tratamentos para sanar os problemas supramencionados.

Cabe salientar que boa parte do tempo foi utilizada para examinar e tratar os dados, uma melhor etapa de coleta pode subsidiar menor tempo gasto com tratamento e maior esforço em encontrar um melhor modelo preditivo.

Na sequência, construímos alguns modelos de regressão e por fim, selecionamos o modelo que apresentou o melhor desempenho nos testes estatísticos, com base em sua capacidade de prever a expectativa de vida. Esse modelo nos permitirá responder às perguntas propostas a seguir.

Adicionalmente iniciamos a construção de um simulador de **Expectativa de Vida** com base no melhor modelo escolhido. Essa iniciativa visa comprovar as relações apresentadas entre os dados e futuramente embasar as decisões de governos mais prováveis de atingirem seus objetivos.

Outra potencialidade desse estudo, foi o teste de outros modelos de regressão que mostraram-se promissores como substitutos do atual, porém necessitando mais testes para afirmar sua potencial superioridade.

Análise dos Dados

O conjunto de dados apresentado possui 2.938 registros com dados de 193 países entre os anos 2000 e 2015, contendo os seguintes indicadores:

- **Situação:** Estado Desenvolvido ou Em Desenvolvimento;
- **Expectativa de Vida:** Expectativa de Vida na idade;
- **Mortalidade Adultos:** Taxas de mortalidade adulta de ambos os sexos (probabilidade de morrer entre 15 e 60 anos por 1.000 habitantes);
- **Mortalidade Infantil:** Número de óbitos infantis por 1.000 habitantes;
- **Álcool:** Álcool, consumo registrado per capita (15+) (em litros de álcool puro);
- **Percentual de Despesas:** Gastos com saúde como porcentagem do Produto Interno Bruto per Capita (%);

- **Hepatite B:** Cobertura vacinal contra Hepatite B (HepB) em crianças de 1 ano (%);
- **Sarampo:** Sarampo - número de casos notificados por 1.000 habitantes;
- **IMC:** Índice de Massa Corporal Médio de toda a população;
- **Mortes Menor de 5 anos:** Número de mortes de menores de cinco anos por 1.000 habitantes;
- **Poliomielite:** Cobertura vacinal contra pólio (Pol3) em crianças de 1 ano (%);
- **Gastos Totais:** Gastos do governo geral com saúde como porcentagem do gasto total do governo (%);
- **Difteria:** Cobertura vacinal contra Difteria, tétano e coqueluche (DTP3) em crianças de 1 ano de idade (%);
- **HIV/AIDS:** Mortes por 1.000 nascidos vivos HIV/AIDS (0-4 anos);
- **PIB:** Produto Interno Bruto per capita (em USD);
- **População:** População do país;
- **Desnutrição 10-19 anos:** Prevalência de magreza em crianças e adolescentes de 10 a 19 anos (%);
- **Desnutrição 5-9 anos:** Prevalência de magreza em crianças de 5 a 9 anos (%);
- **Composição Renda:** Índice de Desenvolvimento Humano em termos de composição de renda dos recursos (índice variando de 0 a 1);
- **Escolaridade:** Número de anos de escolaridade (anos).

Dos 2.938 registros, 2.426 (82,57%) referem-se a países desenvolvidos e 512 (17,43%) a países em desenvolvimento.

Para cada ano de referência existem dados de 183 países, a única exceção é o ano de 2013 que possui dados de 193 países.

Analisando os dados, observamos a presença significativa de dados ausentes nos seguintes indicadores:

População: 652 valores ausentes (22.19%);

Hepatite B: 553 valores ausentes (18.82%);

PIB: 448 valores ausentes (15.25%);

Gastos Totais: 226 valores ausentes (7.69%);

Álcool: 194 valores ausentes (6.60%);

Composição Renda: 167 valores ausentes (5.68%);

Escolaridade: 163 valores ausentes (5.55%);

IMC, Desnutrição 5-9 anos e Desnutrição 10-19 anos: 34 valores ausentes (1.16%);

Poliomielite e Difteria: 19 valores ausentes (0.65%);

Expectativa de Vida e Mortalidade Adultos: 10 valores ausentes (0.34%).

Esses registros com dados ausentes foram excluídos da análise.

Constatamos que alguns indicadores, cuja unidade de medida é em 1.000 habitantes, possuíam registros com valores acima desse limite, configurando irregularidade na informação, como segue:

Sarampo: 525 registros;

Mortalidade infantil: 13 registros;

Mortes Menor de 5 anos: 16 registros.

Esses registros com valores acima de 1.000 foram excluídos da análise.

Identificamos também que o indicador **IMC** apresentava irregularidades nos dados.

O **IMC** é calculado dividindo o peso do indivíduo (massa) pela sua altura ao quadrado, sendo a massa em quilogramas (kg) e a altura em metros (m). Pela tabela abaixo, é possível identificar que **IMC** maior que 40kg/m² já configura obesidade grau III.

Existem 1.139 registros com **IMC** acima de 50kg/m² (aproximadamente 40% dos dados originais).

Tabela do IMC

Classificação	IMC(kg/m ²)
Muito abaixo do peso	Menor que 18,5
Normal	18,5 - 24,9
Acima do peso	25 - 29,9
Obesidade grau I	30 - 34,9
Obesidade grau II	35 - 39,9
Obesidade grau III	Maior que 40

Devido a grande quantidade de registros com valores exorbitantes, optamos em desconsiderar esse indicador da nossa análise.

Após todos os ajustes realizados, o conjunto de dados foi reduzido para 1.693 registros.

Diante da urgência na obtenção desta análise, optamos por simplificar o pré-processamento dos dados, porém na próxima versão, que já está em andamento, iremos preencher os dados ausentes dos indicadores **População** e **PIB**, e enriquecer o conjunto de dados agregando mais um indicador, **Densidade Demográfica**, com base nos dados fornecidos pelas Nações Unidas (<https://population.un.org/wpp/Download/Standard/CSV/> e <https://data.un.org/Data.aspx?d=SNAAMA&f=grID%3A101%3BcurrID%3AUSD%3BpcFlag%3A1>).

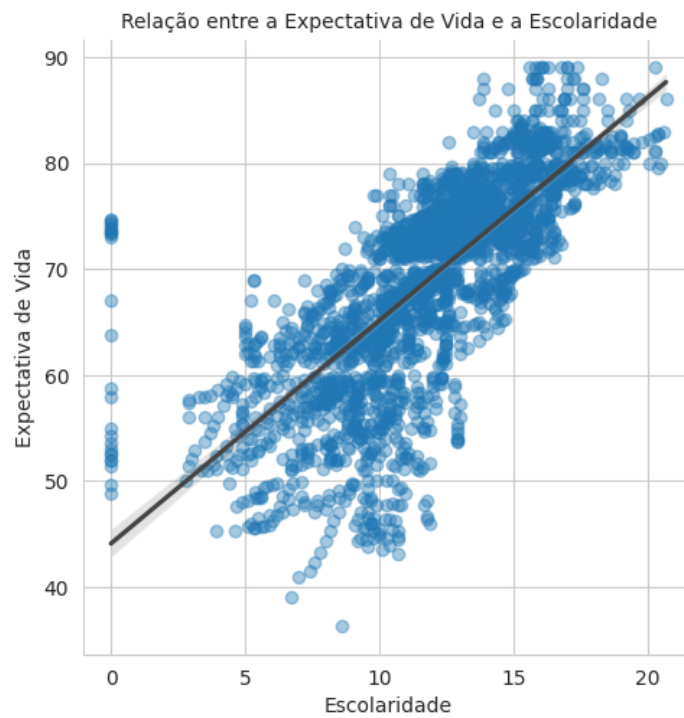
Perguntas-chave

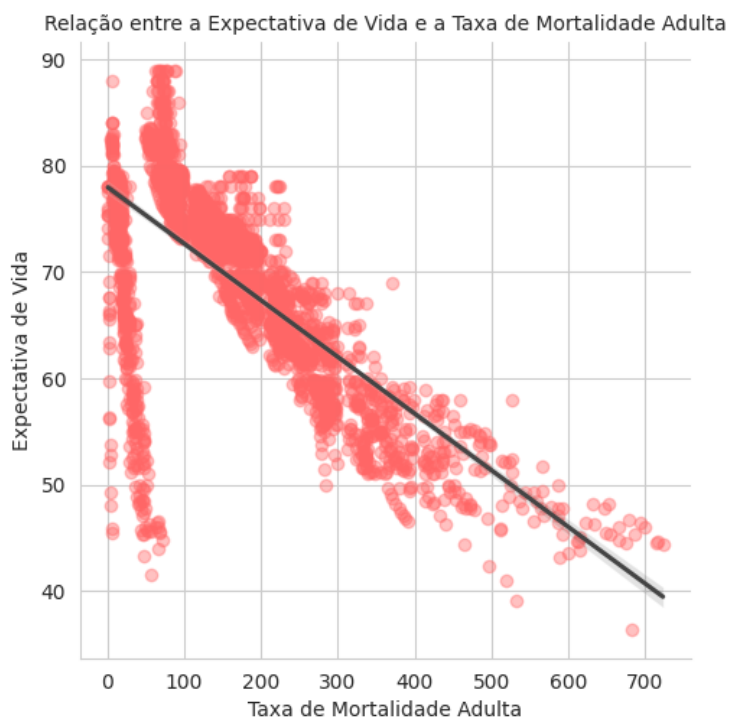
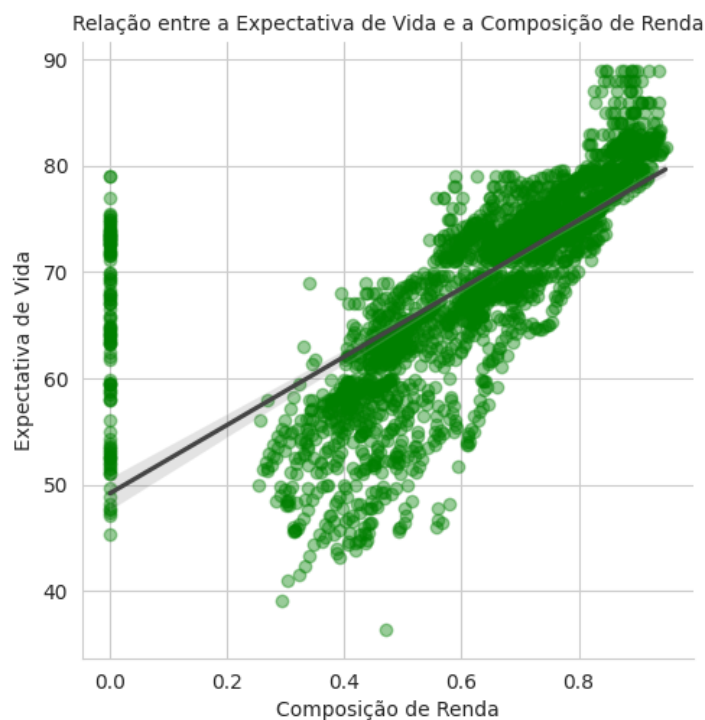
1. Quais são as variáveis que realmente afetam a Expectativa de Vida?

Avaliando o grau de correlação entre as variáveis independentes e a variável que se quer prever, a **Expectativa de Vida**, concluiu-se que os atributos **Escolaridade**, **Composição Renda** e **Mortalidade Adultos** isoladamente são os que mais afetam a **Expectativa de Vida**, vide gráficos de dispersão abaixo.

Por outro lado, a combinação de atributos que tiveram impacto estatisticamente significativo na **Expectativa de Vida**, foram: **Escolaridade**, **Composição Renda**, **Mortalidade Infantil**, **Mortalidade Adultos**, **Mortes Menor de 5 anos**, **Difteria**, **HIV/AIDS**, **População**, **Desnutrição 10-19 anos** e **Situação**.

Sendo criados os modelos de regressão que prevêm a Expectativa de Vida baseados nesses atributos.





2. Um país com Expectativa de Vida menor (<65) deve aumentar seus investimentos em saúde a fim de melhorar sua Expectativa de Vida média?

Dentre os atributos fornecidos, escolheu-se as variáveis **Percentual de Despesas** e **Gastos Totais** para avaliar se o aumento de investimentos com saúde, dentro da faixa etária definida, implicaria em uma melhoria na Expectativa de Vida. Sendo que **Percentual**

de Despesas corresponde a gastos com saúde como porcentagem do Produto Interno Bruto per capita (%) e **Gastos Totais** corresponde a gastos do governo geral com saúde como porcentagem do gasto total do governo (%).

Analisando isoladamente as variáveis, nenhuma das duas possui uma correlação significativa com a **Expectativa de Vida**. De fato, **Percentual de Despesas** e **Gastos Totais** apresentaram baixas correlações de 0,01 e -0,16, respectivamente, em relação à **Expectativa de Vida**.

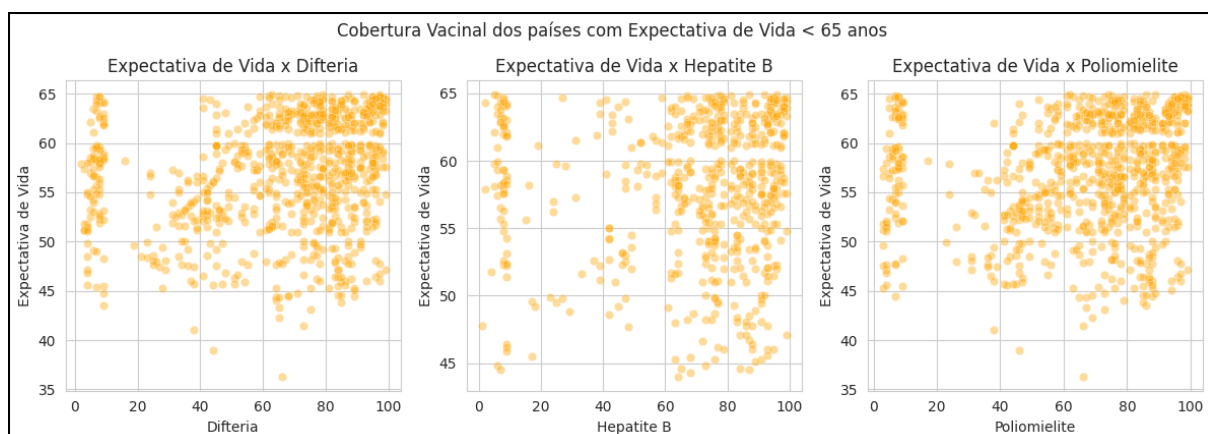
Analisando o modelo de regressão linear, essas variáveis não tiveram um desempenho relevante na explicação da variação da **Expectativa de Vida** e não foram consideradas como preditoras significativas, ou seja são "variáveis não significativas".

Diante do exposto, considerando o modelo e o conjunto de dados analisados, não podemos afirmar que um país com **Expectativa de Vida** menor deva aumentar seus investimentos em saúde.

Tendo em vista as supracitadas correlações não significativas dos gastos com saúde e que a cobertura vacinal pode ser considerada um investimento em saúde, também selecionou-se as variáveis **Hepatite B**, **Poliomielite** e **Difteria**

A variável **Hepatite B** possui um correlação baixa com a **Expectativa de Vida**, além disso, após análise exploratória foi decidido excluí-la por possuir dados ausentes. A **Poliomielite** e a **Difteria** isoladamente possuem correlação moderada com a **Expectativa de Vida**, vide gráficos de dispersão abaixo.

No entanto, ao combinar os dados da cobertura vacinal com os demais atributos, gerando o modelo de regressão linear múltiplo, apenas a **Difteria** foi considerada uma variável significativa, impactando na explicação da **Expectativa de Vida**. Sendo assim podemos considerar que investimentos em saúde, que proporcionam uma maior cobertura vacinal da **Difteria**, contribuem para aumentar a **Expectativa de Vida** de um país.



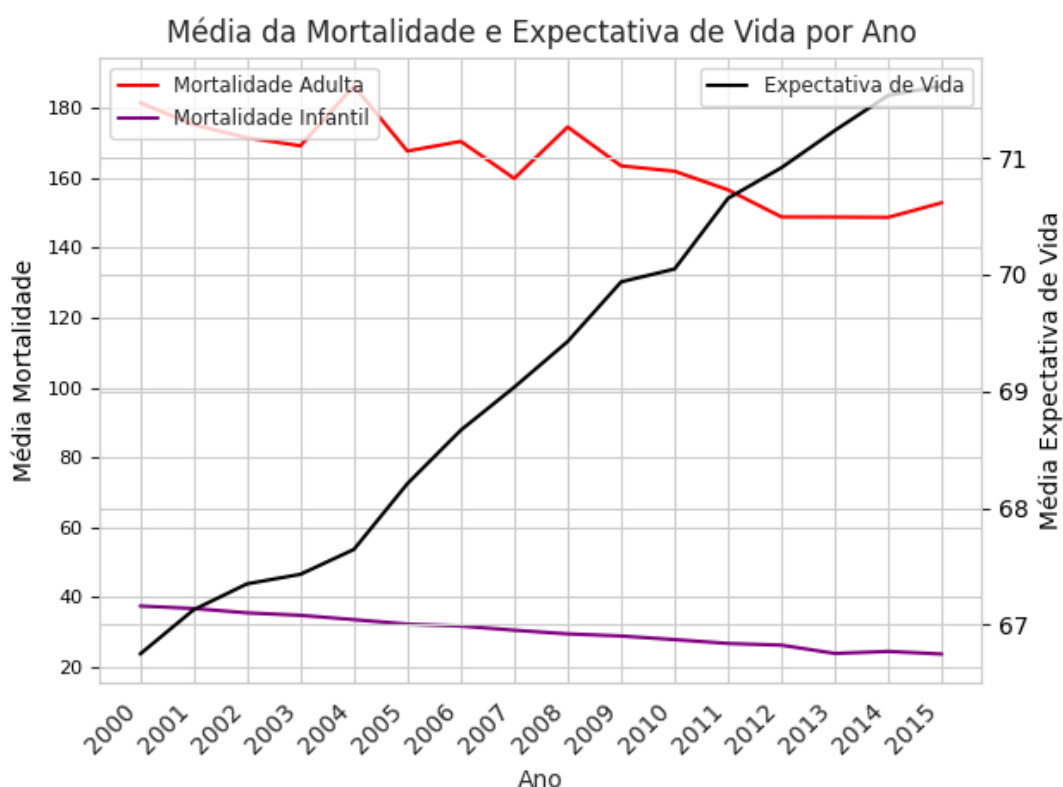
3. Como as taxas de mortalidade infantil e de adultos afetam a Expectativa de Vida?

A análise revelou uma relação significativa e negativa entre a taxa **Mortalidade Adultos** e a **Expectativa de Vida**. Quanto menor a taxa **Mortalidade Adultos**, maior tende a ser a **Expectativa de Vida** média de um país. Isso indica que reduzir a mortalidade em adultos é um fator chave para melhorar a expectativa de vida.

No caso da taxa **Mortalidade Infantil**, embora tenha sido observada uma relação negativa com a **Expectativa de Vida**, essa associação é menos evidente em comparação com a mortalidade em adultos. Apesar disso, a variável **Mortalidade Infantil** teve importância no modelo de regressão linear, sendo considerada um fator com impacto estatisticamente significativo na determinação da **Expectativa de Vida**.

Esses resultados ressaltam a importância de políticas e intervenções voltadas para a redução da mortalidade em adultos, uma vez que essa variável tem um impacto mais significativo na expectativa de vida. No entanto, não devemos subestimar a relevância da redução da mortalidade infantil, já que ambos os aspectos têm influência na saúde e no bem-estar geral de uma população.

Obs.: Mesmo analisando a **Mortalidade Infantil** pela **Situação** do país a sua correlação com a **Expectativa de Vida** é baixa em ambos, "Desenvolvido" e "Em desenvolvimento".



4. A Expectativa de Vida tem correlação positiva ou negativa com hábitos alimentares, estilo de vida, exercícios, fumo, consumo de álcool etc.?

Analisando o consumo registrado per capita de álcool puro, observa-se uma correlação moderada positiva com a **Expectativa de Vida**, e tal variável não teve significância para o modelo de regressão linear.

A correlação positiva entre **Álcool** e a **Expectativa de Vida** pode parecer surpreendente, pois é amplamente conhecido que o consumo excessivo de álcool pode ter efeitos negativos na saúde. No entanto, é importante lembrar que a correlação não implica causalidade direta e há outros fatores envolvidos nessa relação que podem explicar essa correlação positiva, por exemplo:

1. Consumo moderado de álcool: Estudos sugerem que o consumo moderado de álcool (em quantidades controladas) pode ter benefícios para a saúde. Por exemplo, o consumo moderado do vinho tinto e os efeitos cardiovasculares benéficos.
2. Fatores socioeconômicos e culturais: Em alguns países ou regiões, o consumo de álcool pode ser parte de tradições sociais, e essas áreas podem ter melhores condições socioeconômicas e acesso a cuidados de saúde.
3. Variação na qualidade e tipo de álcool consumido: A qualidade e o tipo de álcool consumido podem variar em diferentes países e regiões. Por exemplo, algumas culturas consomem principalmente bebidas alcoólicas com menor teor alcoólico, como vinho, que podem ter efeitos mais benéficos à saúde em comparação com bebidas destiladas com alto teor alcoólico.

Em relação aos hábitos alimentares, têm-se os atributos **Desnutrição 5-9 anos** e **Desnutrição 10-19 anos**, que refletem, de alguma forma, neste quesito. Contudo, só teríamos esses dados para a faixa etária entre 5 e 19 anos. Assim, foi observada uma correlação moderada negativa desses atributos em relação à **Expectativa de Vida**, o que implicaria em quanto maior o nível de desnutrição, menor a expectativa de vida. Ressalta-se que, em uma análise combinada de atributos, somente a variável **Desnutrição 10-19 anos** foi significativa para o modelo de regressão linear múltipla.

O Índice de Massa Corporal Médio, **IMC**, é uma medida de referência internacional reconhecida pela OMS (Organização Mundial da Saúde) para avaliar o peso de uma pessoa e pode dar indícios sobre a saúde dos indivíduos. Como citado anteriormente, identificamos irregularidades nos valores do referido dado e para garantir a confiabilidade do modelo, decidimos excluir essa informação. Inclusive, identificamos que a correlação entre o **IMC** e a **Expectativa de Vida** era moderada positiva, o que confirma a presença de valores inconsistentes, já que o esperado seria uma correlação forte ou moderada negativa.

No entanto, é importante ressaltar que os outros fatores mencionados na pergunta, como estilo de vida, exercícios, fumo, não estão explicitamente presentes no conjunto de dados fornecido. Para avaliar a correlação entre esses fatores e a **Expectativa de Vida**, seria necessário obter dados adicionais relacionados a esses aspectos.

5. Qual é o impacto da escolaridade na Expectativa de Vida dos humanos?

O impacto da **Escolaridade** na **Expectativa de Vida** foi analisado e constatou-se que a **Escolaridade** é a variável que possui maior significância individual para a **Expectativa de Vida**, explicando 55,5% da variabilidade. A análise revelou uma correlação forte positiva entre a **Escolaridade** e a **Expectativa de Vida**, indicando que países com níveis mais altos de escolaridade tendem a ter uma expectativa de vida mais elevada.

Isso sugere que a educação desempenha um papel crucial na promoção da saúde e no aumento da longevidade das populações, pois existem várias razões pelas quais a escolaridade pode ter um impacto positivo na **Expectativa de Vida**, tais como: informações e conhecimentos sobre hábitos e comportamentos saudáveis, melhores oportunidades de emprego que pode resultar em melhores condições de vida, maior probabilidade na tomada de decisões assertivas decorrente do desenvolvimento de raciocínio lógico, dentre outras.



6. A Expectativa de Vida tem uma relação positiva ou negativa com o consumo de álcool?

Como já descrito na resposta da 4ª questão, analisando o consumo registrado per capita de álcool puro, observa-se uma correlação moderada positiva com a **Expectativa de Vida**, e tal variável não teve significância para o modelo de regressão linear.

A correlação positiva entre **Álcool** e a **Expectativa de Vida** pode parecer surpreendente, pois é amplamente conhecido que o consumo excessivo de álcool pode ter efeitos negativos na saúde. No entanto, é importante lembrar que a correlação não implica

causalidade direta e há outros fatores envolvidos nessa relação que podem explicar essa correlação positiva, por exemplo:

1. Consumo moderado de álcool: Estudos sugerem que o consumo moderado de álcool (em quantidades controladas) pode ter benefícios para a saúde. Por exemplo, o consumo moderado do vinho tinto e os efeitos cardiovasculares benéficos.
2. Fatores socioeconômicos e culturais: Em alguns países ou regiões, o consumo de álcool pode ser parte de tradições sociais, e essas áreas podem ter melhores condições socioeconômicas e acesso a cuidados de saúde.
3. Variação na qualidade e tipo de álcool consumido: A qualidade e o tipo de álcool consumido podem variar em diferentes países e regiões. Por exemplo, algumas culturas consomem principalmente bebidas alcoólicas com menor teor alcoólico, como vinho, que podem ter efeitos mais benéficos à saúde em comparação com bebidas destiladas com alto teor alcoólico.

7. Países densamente povoados tendem a ter menor Expectativa de Vida?

O conjunto de dados não apresenta variável sobre a densidade populacional, então para responder à pergunta utilizaremos a variável **População**.

A **População** possui relação fraca com a **Expectativa de Vida**, mas apesar disso é uma das variáveis que tiveram significância para o modelo.

Embora a relação direta entre a **População** e a **Expectativa de Vida** possa ser fraca, a população pode desempenhar um papel indireto na saúde e na expectativa de vida por meio de outros fatores associados, pois, países com populações maiores podem enfrentar desafios relacionados à distribuição de recursos, infraestrutura de saúde e acesso a serviços de qualidade, o que pode influenciar na expectativa de vida.

É importante considerar que a influência da **População** na **Expectativa de Vida** pode variar de acordo com as características específicas de cada país, visto que em países com sistemas de saúde robustos e recursos adequados, a relação entre a **População** e a **Expectativa de Vida** pode ser menos significativa em comparação com países com sistemas de saúde mais sobrecarregados.

8. Qual é o impacto da cobertura de imunização na Expectativa de Vida?

No conjunto de dados temos três variáveis relacionadas à cobertura vacinal: **Hepatite B**, **Poliomielite** e **Difteria**.

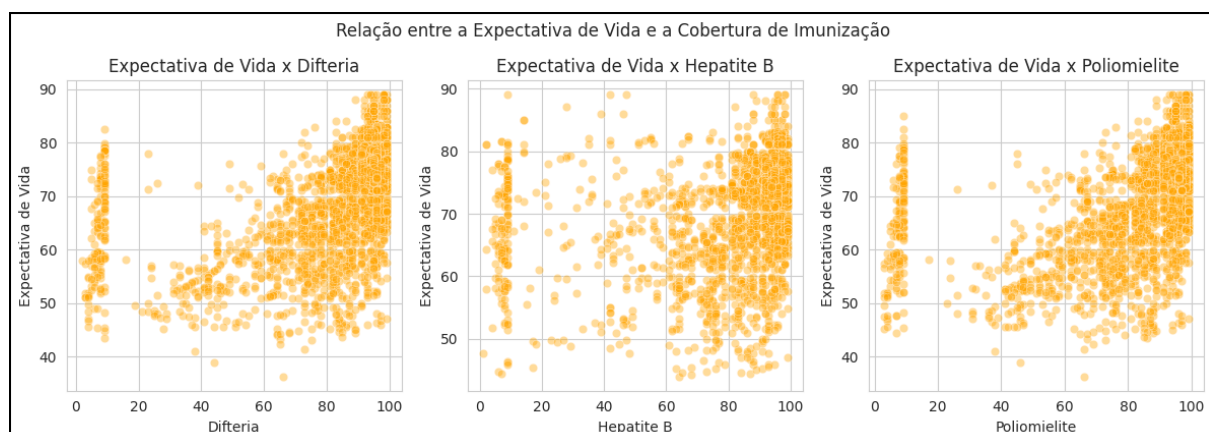
A variável **Hepatite B** possui uma correlação baixa com a **Expectativa de Vida**, além disso, após análise exploratória foi decidido excluí-la por possuir muitos dados ausentes. A **Poliomielite** e a **Difteria** possuem correlação moderada com a **Expectativa de Vida**, quanto maior a cobertura vacinal maior a expectativa de vida do país.

Já no modelo de regressão linear, apenas a **Difteria** foi considerada uma variável significativa, impactando na explicação da **Expectativa de Vida**. Sendo assim, podemos considerar que investimentos em saúde, que proporcionam uma maior cobertura vacinal da **Difteria**, contribuem para aumentar a expectativa de vida de um país.

Para determinar o impacto da cobertura de imunização na **Expectativa de Vida** com base no conjunto de dados fornecido, seria necessário realizar uma análise mais detalhada dos dados. No entanto, de maneira geral, a cobertura de imunização tem um impacto positivo na **Expectativa de Vida** e na saúde populacional. A imunização efetiva contra doenças infecciosas pode prevenir a ocorrência de epidemias, reduzir a morbidade e a mortalidade, além de melhorar a saúde geral das populações.

A imunização pode proteger contra uma ampla gama de doenças infecciosas, como sarampo, poliomielite, difteria, coqueluche, entre outras. Ao prevenir a ocorrência dessas doenças, a cobertura de imunização contribui para a redução da mortalidade e morbidade associadas a elas. Além disso, a imunização é especialmente importante para proteção de grupos vulneráveis, como crianças, idosos e pessoas com sistema imunológico comprometido, pois, ao garantir altas taxas de cobertura de imunização nesses grupos, é possível reduzir os riscos de complicações e óbitos relacionados a doenças preveníveis. Desempenha um papel fundamental na redução da mortalidade infantil. Vacinas contra doenças como o sarampo, a pneumonia e a diarreia, por exemplo, têm impacto direto na diminuição das taxas de mortalidade em crianças. Ela também, não apenas reduz a incidência de doenças, mas também ajuda a evitar sequelas e complicações associadas a elas.

No entanto, é importante ressaltar que o impacto da cobertura de imunização na **Expectativa de Vida** pode variar de acordo com o contexto específico de cada país e com as doenças prevalentes em determinadas regiões.



Conclusão

Em virtude das análises apresentadas, os dados revelaram alguns insights acerca das relações significativas entre as variáveis independentes e a **Expectativa de Vida**. Além disso, criou-se o modelo preditivo de regressão linear, resultante do processo de aprendizado sobre os dados fornecidos e da avaliação dos indicadores estatísticos de performance, o que nos permitiu selecionar o melhor modelo.

Entre as variáveis analisadas, destacaram-se a **Escolaridade**, a **Composição Renda**, a taxa de **Mortalidade Adultos**, o índice **HIV/AIDS**, a taxa de **Mortalidade Infantil**, a taxa de **Mortes Menor de 5 anos**, a cobertura vacinal contra **Difteria**, a **Desnutrição** em diferentes faixas etárias, a **População** e a **Situação** do país. Essas variáveis demonstraram ter uma correlação significativa com a **Expectativa de Vida**, influenciando positiva ou negativamente esse indicador.

A **Escolaridade**, **Composição Renda** e taxa de **Mortalidade Adultos** se destacaram em relação às demais variáveis, evidenciando a importância de abordá-las ao formular políticas públicas voltadas para a saúde e o bem-estar das populações. Investimentos em educação, acesso equitativo a cuidados de saúde, medidas de prevenção de doenças, combate à desnutrição, vacinação e melhorias nas condições socioeconômicas são fundamentais para promover a saúde e aumentar a expectativa de vida das populações.

Portanto, esses resultados podem fornecer informações valiosas para orientar políticas públicas, entre outras, e intervenções em países com o objetivo de melhorar a **Expectativa de Vida** média. Porém, reforçamos a necessidade de melhorar a qualidade dos dados dos indicadores fornecidos.