

# Causas e consequências da obesidade

Filipe, João, Vasconcelos, João, Cunha, Rodrigo

---

## Abstract

Obesidade está entre as doenças mais graves do século, é um problema que além de prejudicar uma quantidade enorme de pessoas (as projeções são que até 2050 mais de a metade das pessoas do mundo seja obesas), custa autíssimo para a saúde pública e para os planos de saúde, já que está fortemente ligado a diabetes, hipertensão e outras 62 doenças.

Dessa forma buscamos entender com esse projeto as causas e consequências da obesidade, sua relação com o desenvolvimento do país, bem como sua relação com a diabetes e hipertensão, já que são as duas principais doenças relacionadas.

---

## 1. Introduction

Hoje a obesidade mata mais do que a fome, e o Brasil tem mostrado números bastante preocupantes neste tópico. Atualmente mais da metade dos brasileiros têm pelo menos sobrepeso, e cerca de 30% são obesos. Isso se torna um grande problema financeiro para planos de saúde e governo, já que a obesidade está diretamente relacionada a 64 outras doenças, dentre elas hipertensão e diabetes, duas das doenças mais caras para esses dois agentes.

Para se ter uma ideia no ano de 2017 o governo gastou R\$16bi apenas com obesidade. Dessa forma ilustra-se a importância de se obter um maior entendimento das causas desse problema, para se tomar medidas preventivas, e ainda suas consequências para possibilitar ações reativas mais inteligentes.

### 1.1. Datasets

#### 1.1.1. Países por Continentes

Base de dados de ... contendo informação do continente de 194 países.

#### 1.1.2. IDH no Mundo

Base de dados do Our World In Data contendo IDH de 189 países com dados que vão de 1990 até 2017.

#### 1.1.3. Obesidade no Mundo

Base de dados do Our World In Data contendo 8022 registros de obesidade em 191 países de 1975 até 2016.

#### *1.1.4. Densidade populacional no mundo*

Base de dados do Our World Data contendo informações sobre a densidade populacional de 164 países de 1961 até 2016, foi usado apenas o ano de 2016 por proposito de fidelidade.

#### *1.1.5. Cultural Vector*

Base de dados auxiliar do Natural Earth Data contendo código de três letras dos países assim como a sua geometria dentro de um mapa mundi. Utilizado em conjunto com outros datasets para plotagem de gráficos.

#### *1.1.6. Presença de vegetais na dieta de um país*

Base de dados do ... contendo a informação do consumo médio de vegetais de 222 países com datas variando de 1800 para cima.

#### *1.1.7. Informações clínicas de pacientes no Brasil*

Base de dados de uma pesquisa sobre obesidade do FigShare contendo informações clínicas sobre pacientes no Brasil.

## **2. Análise de Dados**

Nosso estudo inicial foi verificar a influência do IDH de um país no percentual de obesos que o mesmo tem.

### *2.1. Percentual global de obesos*

Primeiro utilizamos o dataset de obesidade no mundo, que foi coletado pela UNICEF, para verificar o crescimento do percentual global de obesos no mundo e vimos que de 1975 para 2016 o percentual subiu cerca de 14%.

### *2.2. Obesidade global em relação ao IDH*

Depois tiramos a média do IDH mundial de 1990 até 2016 e fizemos uma correlação entre a obesidade media global e o idh medio global e conseguimos observar uma fortíssima correlação entre o IDH e o nível de obesidade de um país quando se analisa a média global.

### *2.3. Obesidade em cada país*

Partimos então para uma análise mais granular com relação aos países. Tiramos a média de cada país de 1990 até 2016 e ordenamos pelo maior percentual de obesidade. Para nossa surpresa, países como EUA, Canadá e Brasil não estavam entre os TOP 5 países mais obesos do mundo e nem mesmo no TOP 10. Além disso podemos perceber todos os países presentes no TOP 10 fogem muito da média de obesidade global (que era de 20% em 2016).

#### *2.4. Obesidade em relação ao continente*

Então, decidimos fazer um boxplot para analisar a dispersão dos dados, mas antes nós verificamos a média de obesidade por continente, visto que a grande maioria dos países no TOP 10 eram ilhas da Oceania. O resultado foi que a Oceania é um dos continentes com o maior percentual de obesos no mundo, chegando a ser quase 15% maior que os outros continentes. Outra coisa que podemos perceber com os resultados foi que a Europa e a América do Sul estavam ambas na frente da América do Norte em percentual de obesos, mesmo tendo Estados Unidos e Canadá no continente.

Partimos então para analisar a dispersão dos dados e conseguimos detectar a presença de outliers, decidimos então juntar essa informação com a pesquisa do TOP 10 global de obesidade e vimos que todos os países que estavam no TOP 10 se encontram na Oceania. Visto que esse países eram outliers, resolvemos remove-los e repetir o experimento.

Conseguimos observar que apesar de ter removido os outliers dos dados, a Oceania ainda continua em primeiro lugar e que a dispersão dos dados continua com o mesmo comportamento. Isso nos leva a inferir que aqueles outliers não influenciavam tanto assim o comportamento dos dados. O fato de a oceania ter sido a primeira foi algo não esperado para os resultados por parte do grupo (tendo em mente as Américas ou a Europa como líderes). Com isso resolvemos uma análise mais detalhada da Oceania para tentar determinar um motivo demográfico para esse comportamento.

#### *2.5. Obesidade em relação a densidade populacional*

Conseguimos então observar que os países que estavam no TOP 10 eram ilhas da oceania pequenas que possuem uma densidade populacional alta. Partimos então para relacionar a densidade populacional com o percentual de obesidade de um país. A partir da nossa análise conseguimos ver que não existe uma forte correlação entre obesidade e densidade populacional, mas ainda há uma tendência negativa, o que aponta ainda uma fraca correlação inversamente proporcional. Como não achamos uma métrica demográfica razoável para explicar esse comportamento decidimos tentar correlacionar a qualidade alimentícia de cada país usando um dataset que descreve o consumo de vegetais por pessoa por país.

### **3. Aprendizado**

Para classificarmos, utilizamos 3 classificadores para testamos a eficácia básica de cada um: Gaussian Naive Bayes, K-Nearest Neighbours e Random Forest Classifier, onde mantivemos os parâmetros default de todos, exceto do Random Forest, onde variamos o número de árvores. Também fizemos o tratamento dos dados com Label Encoder e preenchemos os valores faltantes.

Para os resultados de obesidade, optamos por preencher os valores faltantes com a classe majoritária (Não), para os outros atributos, preenchemos com a média das amostras. Também descartamos atributos irrelevantes, como ID do

paciente, pois não são discriminantes com o resultado. Como os dados apresentam oversampling da classe negativa, executamos o algoritmo SMOTE para balancear os dados.

Os treinos foram feitos dividindo o conjunto de treinamento e teste : 66% treino e 33% teste. Após essa divisão, executamos o SMOTE no conjunto de treinamento para balancearmos o mesmo. Após isso apenas inserimos os conjuntos na função `.fit()` de cada classificador e, para resultados, extraímos o Score, Matriz de Confusão, Precision e Recall.

Vendo que temos a base de dados de clientes com características de cada um e o resultado de obesidade, podemos imaginar um classificador que, com os dados de um paciente, poder prever se o mesmo é propenso a ter ou não Obesidade.

#### **4. Conclusão**

Ao final da análise pudemos verificar algumas surpresas, primeiro a Oceania como sendo o local de maior densidade de obesos, imaginávamos ao início ser as Américas e a Europa. Outra surpresa foi a densidade populacional não influenciar fortemente na quantidade de obesos, já que a primeira análise dos TOP 10 países com maior índice de obesidade nos apontava para essa direção.

Pudemos concluir ainda que há uma boa possibilidade de a ingestão percapta de certos alimentos impacta fortemente no índice de obesidade de um país. Acreditamos que fazer uma análise mais detalhada de alimentos que impactam na obesidade gerará um aprendizado bastante relevante.