

# Projeto Final MC536

Grupo Shrek





# Introdução

- Cada região do mundo tem a sua cultura, sua história e sua própria culinária
- Não só entender o perfil nutricional de cada receita
- Conseguimos traçar um perfil nutricional



## Bases de dados

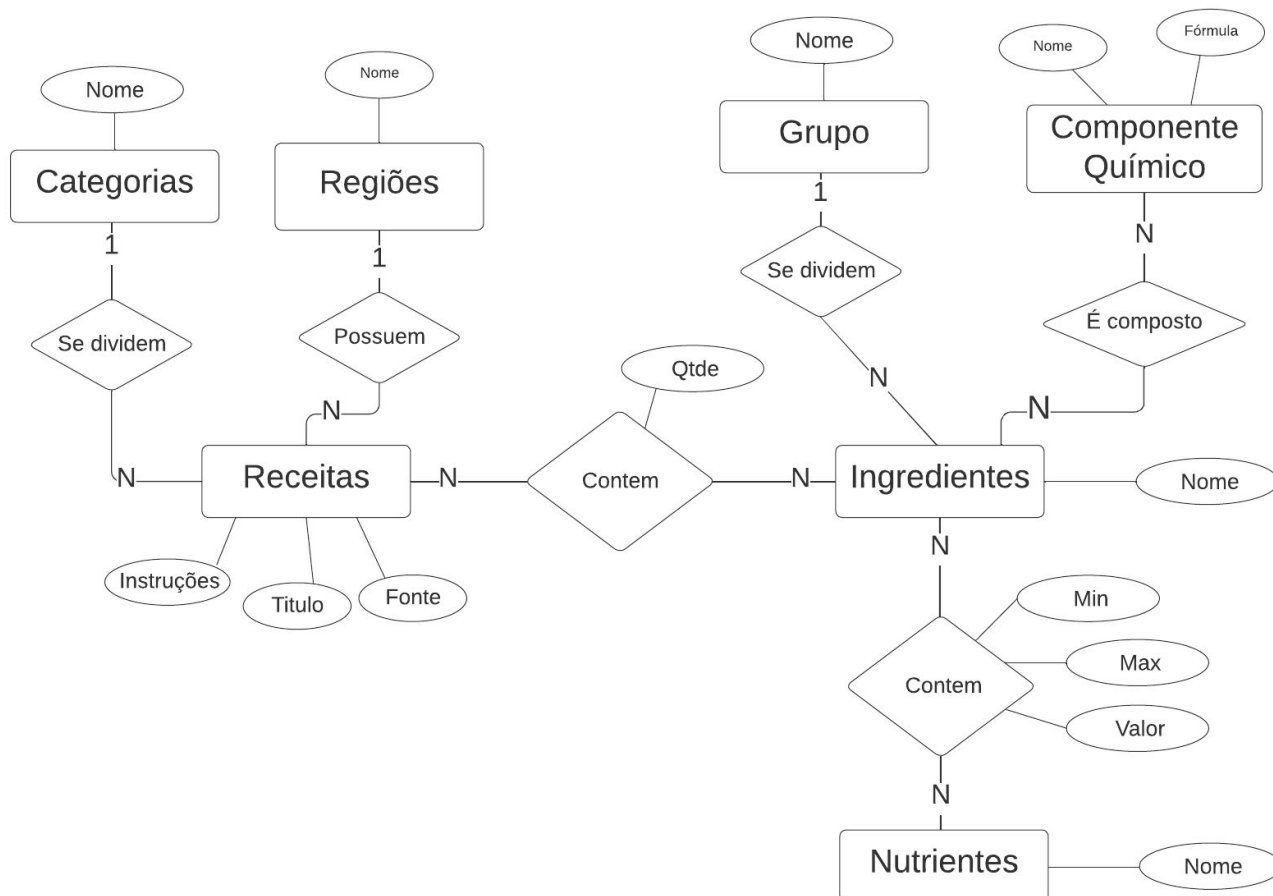


Uma base com diversos alimentos e suas composições tanto químicas como com relação aos nutrientes encontrados em cada alimento



Uma base com diversas receitas e seus ingredientes, relacionando elas com seus países/regiões de origem

# Modelo Conceitual SQL



# Modelo Lógico SQL



categoria(\_id\_, nome)

Regiao(\_id\_, nome)

Grupo(\_id\_, nome)

Receita(\_id\_, titulo, fonte, instrucoes, regiao\_id, categoria\_id)

regiao\_id chave estrangeira -> Regiao(id)

categoria\_id chave estrangeira -> Categoria(id)

Ingrediente(\_id\_, nome, grupo\_id)

grupo\_id chave estrangeira -> Grupo(id)

ReceitaIngrediente(\_receita\_id\_, \_ingrediente\_id\_, qtde)

\_receita\_id\_ chave estrangeira -> Receita(id)

\_ingrediente\_id\_ chave estrangeira -> Ingrediente(id)

# Modelo Lógico SQL



Nutriente(\_id\_, nome)

IngredienteNutriente(\_ingrediente\_id\_, \_nutriente\_id\_, valor, min, max)

\_ingrediente\_id\_ chave estrangeira -> Ingrediente(id)

\_nutriente\_id\_ chave estrangeira -> Nutriente(id)

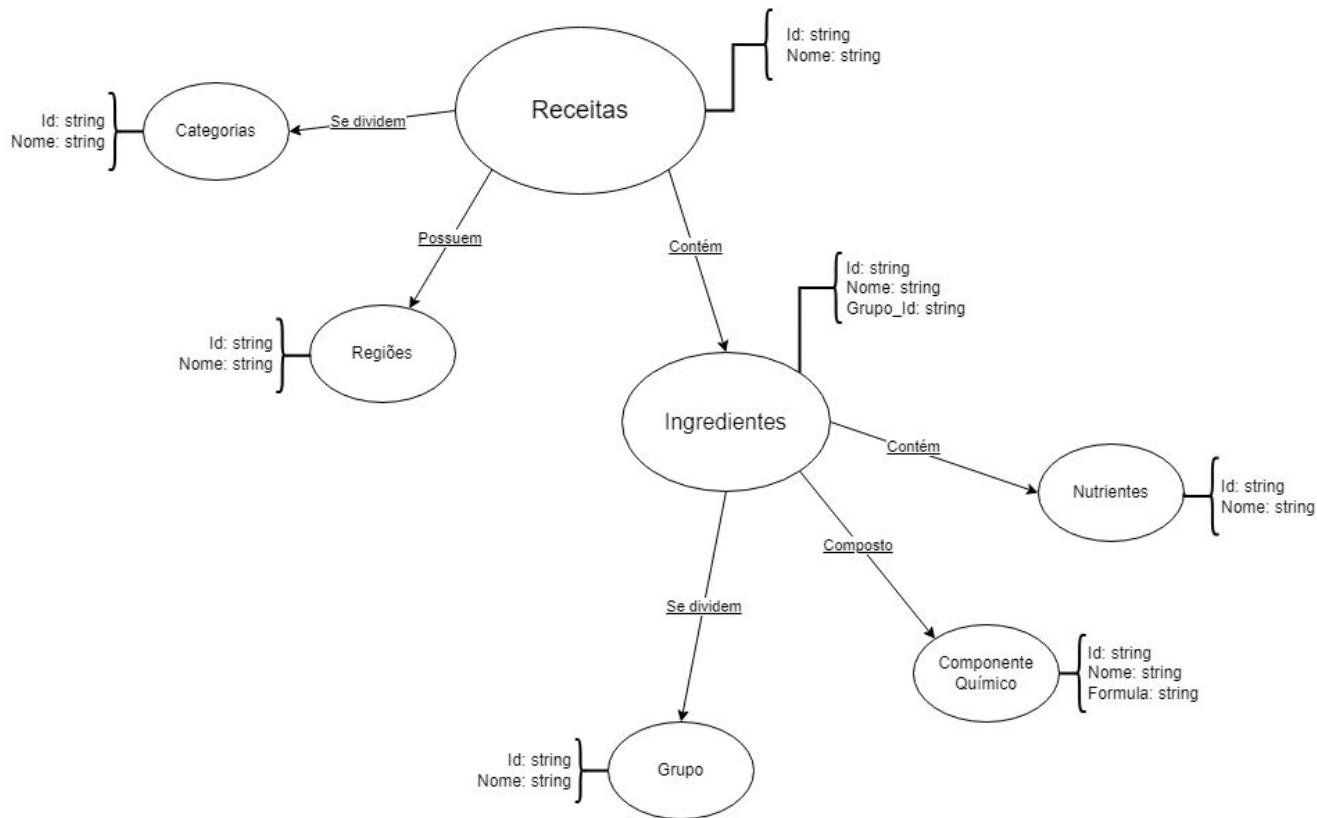
ComponenteQuimico(\_id\_, nome, formula)

IngredienteComponenteQuimico(\_ingrediente\_id\_, \_componenteQuimico\_id\_)

\_ingrediente\_id\_ chave estrangeira -> Ingrediente(id)

\_componenteQuimico\_id\_ chave estrangeira -> ComponenteQuimico(id)

# Modelagem Lógica (Grafos)





# Processo de Transformação

## Extração dos dados brutos

- TheMealDB não disponibiliza os dados por meio de csv
- Necessidade de extrair os dados para um csv para facilitar a manipulação depois
- Uso de python durante o processo





# Processo de Transformação

**Como fazer o relacionamento entre as bases?**

- As bases não usam o mesmo padrão de ID
- Comparação por nome é a resposta mais óbvia
- Evitar trabalho manual
- Ter dados consistentes
- Como resolver o problema de padrões diferentes em base?



# Processo de Transformação

## Algoritmo de proximidade de palavras

- Algoritmo visto em uma aula de Desafios de programação
- Consegue dizer a distância de uma string até outra
- Serve para sabermos se são palavras parecidas
- Conseguir usar ele para resolver nosso problema

```
def distancia_palavras(a, b):  
    n = len(a)  
    m = len(b)  
  
    t = [[[] for _ in range(n + 1)]  
          for i in range(n + 1):  
              t[i] = [0 for _ in range(m + 1)]  
  
    for i in range(n + 1):  
        t[i][0] = i  
    for j in range(m + 1):  
        t[0][j] = j  
  
    for i in range(1, n + 1):  
        for j in range(1, m + 1):  
            add = 1  
            if (a[i - 1] == b[j - 1]): add = 0  
            t[i][j] = t[i - 1][j - 1] + add  
            t[i][j] = min(t[i][j], t[i - 1][j] + 1)  
            t[i][j] = min(t[i][j], t[i][j - 1] + 1)  
  
    return t[n][m]
```



# Processo de Transformação

## Geração do dataset

- Com os dados intermediários prontos, precisamos criar as bases finais
- Uso de python para criar os scripts
- Revisão geral para tentar encontrar algum dado inconsistente
- Alguns poucos ajustes manuais mais gritantes na relação Food (FoodDB) e Ingrediente (TheMealDB)

# DataSet Publicado



título do arquivo/base	Descrição
p_categories	Tabela com todas as categorias das receitas
p_compounds	Tabela com todos os componentes químicos
p_groups	Tabela com os grupos de classificação dos ingredientes
p_ingredients	Tabela com todos os ingredientes das receitas
p_recipes	Tabela com todas as receitas
p_regions	Tabela com todas as regiões das receitas
p_nutrients	Tabela com todos os nutrientes
p_ingredients_compounds	Tabela de relação entre os ingredientes e os componentes químicos
p_ingredients_nutrients	Tabela de relação entre os ingredientes e os nutrientes
p_ingredients_recipes	Tabela de relação entre os ingredientes e as receitas

# Perguntas e Análises (SQL)

Quais os ingredientes que compõem determinadas receitas?

```
SELECT r.titulo AS receita, i.nome AS ingrediente, ri.qtde
FROM ReceitaIngrediente ri
JOIN Receita r ON ri.receita_id = r.id
JOIN Ingrediente i ON ri.ingrediente_id = i.id
WHERE r.titulo = 'Sushi';
```

# Perguntas e Análises (SQL)



Quais os ingredientes mais consumidos? E em cada região?

```
SELECT
  R.regiao_id,
  I.nome AS ingrediente,
  COUNT(*) AS total_consumido
FROM
  ReceitaIngrediente RI
  JOIN Receita R ON RI.receita_id = R.id
  JOIN Ingrediente I ON RI.ingrediente_id = I.id
GROUP BY
  R.regiao_id,
  I.nome
ORDER BY
  R.regiao_id,
  total_consumido DESC;
```

# Perguntas e Análises (SQL)

Quais os nutrientes mais consumidos em cada região?

```
SELECT r.regiao_id, n.nome AS nutriente, SUM(in.valor) AS quantidade_consumida
FROM IngredienteNutriente in
JOIN Nutriente n ON in.nutriente_id = n.id
JOIN Ingrediente i ON in.ingrediente_id = i.id
JOIN ReceitaIngrediente ri ON i.id = ri.ingrediente_id
JOIN Receita r ON ri.receita_id = r.id
GROUP BY r.regiao_id, n.nome
ORDER BY quantidade_consumida DESC;
```

# Perguntas e Análises (SQL)

Dado as receitas de determinada categoria, quais são os componentes (químico) que menos aparecem? E quais são os que mais aparecem?

```
SELECT
    cq.formula AS ComponenteQuimico,
    COUNT(icq.ingrediente_id) AS Quantidade
FROM IngredienteComponenteQuimico icq
JOIN Ingrediente i ON icq.ingrediente_id = i.id
JOIN Categoria c ON i.grupo_id = c.id
JOIN ComponenteQuimico cq ON icq.componenteQuimico_id = cq.id
WHERE c.id = 6
GROUP BY cq.formula
ORDER BY Quantidade DESC
LIMIT 10;
```

```
SELECT
    cq.formula AS ComponenteQuimico,
    COUNT(icq.ingrediente_id) AS Quantidade
FROM IngredienteComponenteQuimico icq
JOIN Ingrediente i ON icq.ingrediente_id = i.id
JOIN Categoria c ON i.grupo_id = c.id
JOIN ComponenteQuimico cq ON icq.componenteQuimico_id = cq.id
WHERE c.id = 6
GROUP BY cq.formula
ORDER BY Quantidade ASC
LIMIT 10;
```





# Perguntas e Análises (SQL)

## Não Implementadas

Perguntas
Dado um grupo específico de ingredientes, quais são os nutrientes mais/menos abundantes e quais são os componentes mais/menos abundantes? Conseguimos relacionar isso com a região de receitas que tem esses ingredientes?
Dado um conjunto de nutrientes e uma faixa de valor, quais são as regiões que tem um consumo adequado desse nutriente dentro dessa faixa de valor?



# Perguntas e Análises (Grafo)

Perguntas
1) Para o nutriente [x], e a região [y], quais são as receitas com maior presença de X.
2) Quais receitas tem mais ingredientes em comum?
3) Dado um ingrediente [x], quais nutrientes contidos nele tem valor nutricional ( $\text{Min} < \text{Valor} < \text{Max}$ )?



Gabriel Augusto Bertolino Gomes - 248287

Gabriel Freitas Pinheiro - 222339

Henrique De Lima Sousa Flores - 223911

Paulo Victor de Souza Santos - 248438

Rodrigo de Barros Camargo - 272701

