

# Regional recipes and their costs

André Santos Rocha	235887
Gustavo Henrique Luiz Merlo	171401
José Felipe Theodoro	219081
Luiz Felipe Corradini Rego Costa	230613
Mariano Cho	230797
Pedro da Rosa Pinheiro	231081

# Motivação

- Analisar as dietas e alimentação de forma econômica e social.
- Averiguar o perfil nutricional de uma região.
- Avaliar relações entre cultura local, alimentação e perfil nutricional.
- Verificar o preço de uma receita em uma determinada região.
- Quais ingredientes são mais populares no panorama culinário mundial.





# Bases de Dados Utilizadas

- Food Prices for Nutrition: Contém preços de comidas de 186 países  
<https://databank.worldbank.org/source/food-prices-for-nutrition/Type/TABLE/preview/on#>
- Culinary DB: Repositório de diversas receitas de 22 regiões do mundo  
<https://cosylab.iiitd.edu.in/culinarydb/#databasedescription>
- FooDB: Provê informações de micronutrientes, macronutrientes e constituintes de vários alimentos  
<https://foodb.ca/>
- World Development Indicators (DataBank): contém o PIB per capita de todos os países  
<https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=#>



# Modelo Lógico

IngredientesCDB(NomeIngrediente, \_IDIngrediente\_, Category)

ReceitasCDB(NomeReceita, \_IDReceita\_, SubRegiao)

BaseCDB(\_NomeIngrediente\_, \_NomeReceita\_)

NomeIngrediente chave estrangeira -> IngredientesCDB(NomeIngrediente)

NomeReceita chave estrangeira -> ReceitasCDB(NomeReceita)

FoodPriceDB(\_Pais\_, 2017[YR2017], \_SeriesName\_)

Countries(\_SubRegiao\_, \_Pais\_)

SubRegiao chave estrangeira -> ReceitasCDB(SubRegiao)

Pais chave estrangeira -> FoodPriceDB(Pais)



# Modelo Lógico

FooDB(FoodGroup, \_FoodName\_)

SubgruposFoodDBPrice(\_SeriesName\_, \_FoodGroup\_)

SeriesName chave estrangeira -> FoodPriceDB(SeriesName)

FoodGroup chave estrangeira -> FooDB(FoodGroup)

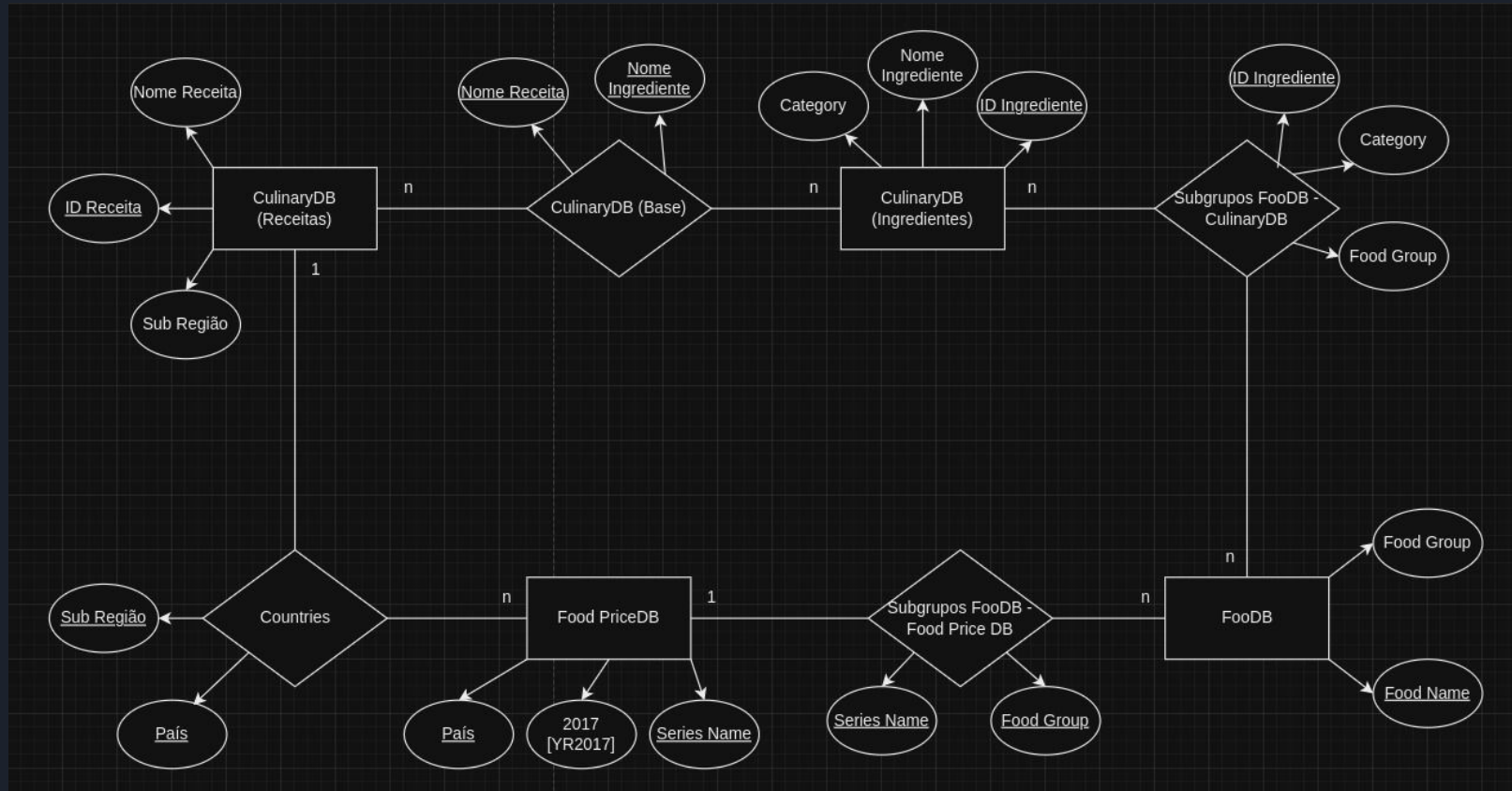
SubgruposFooDBCDB(\_IDIngrediente\_, Category, FoodGroup)

IDIngrediente chave estrangeira -> IngredientesCDB(IDIngrediente)

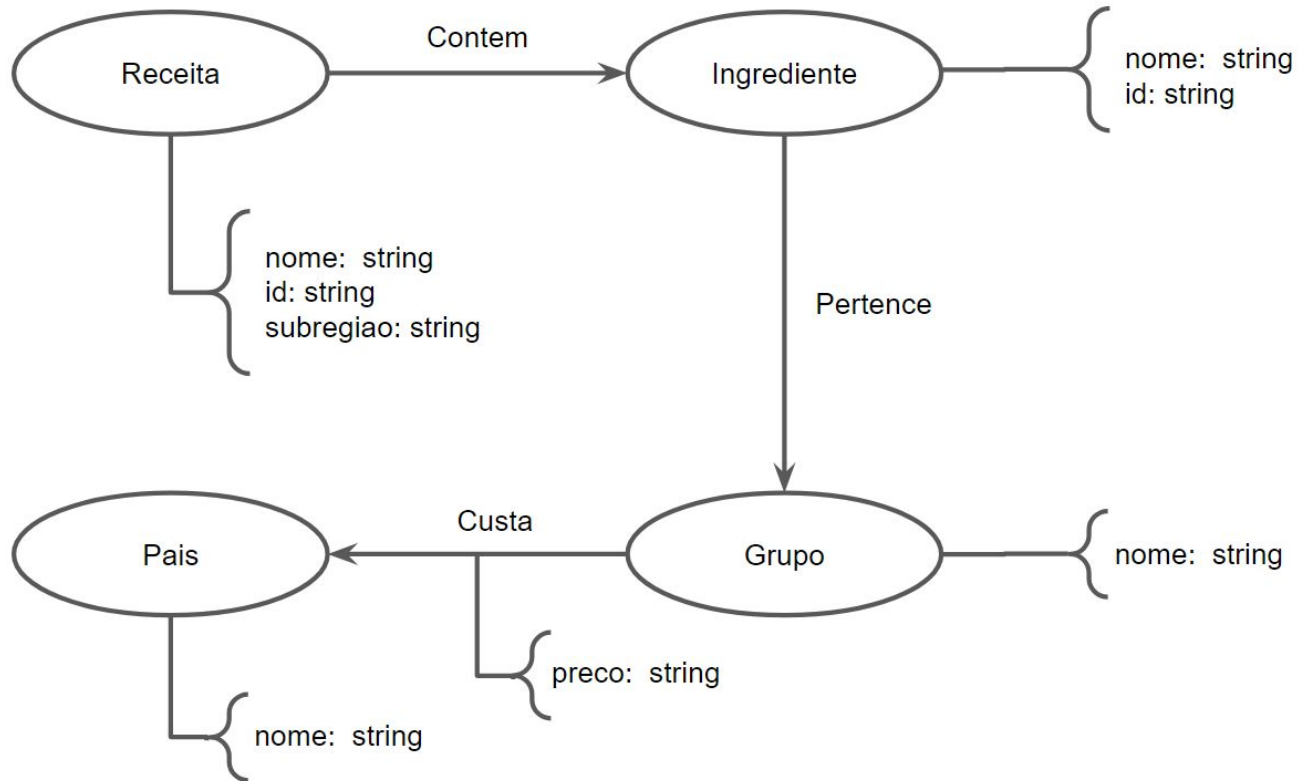
Category chave estrangeira -> IngredientesCDB(Category)

FoodGroup chave estrangeira -> FooDB(FoodGroup)

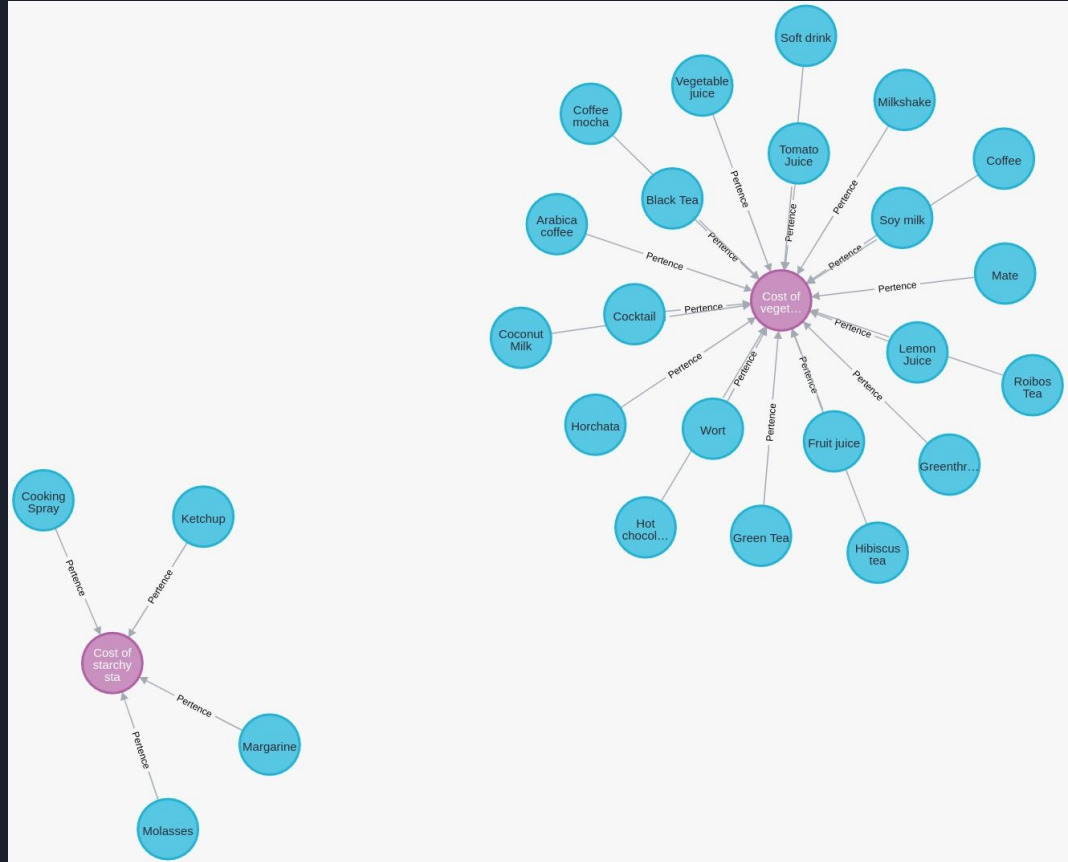
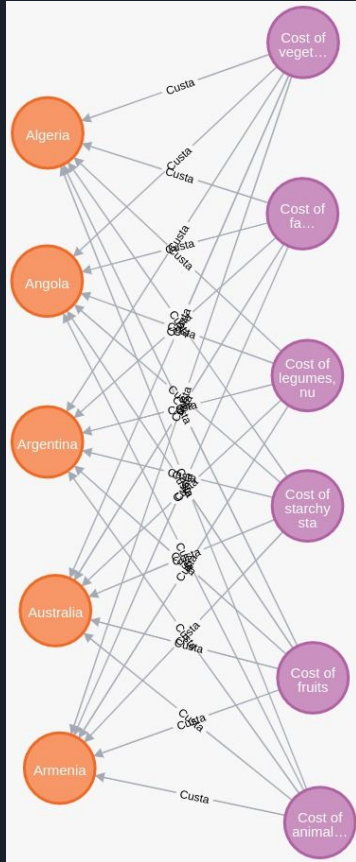
# Modelo Conceitual



# Modelo Lógico de Grafos



# Grafos no Neo4j







# Operações de Preparo

## Relação ingrediente e grupo alimentar

26	Pan dulce	pan dulce	935	Bakery	Cereals and cereal products
27	Raisin bread	#bread-raisin#	936	Bakery	Cereals and cereal products
28	Wonton wrapper	wonton wrapper	937	Bakery	Cereals and cereal products
29	Chocolate mousse	chocolate-mousse	945	Bakery	Cereals and cereal products
30	Fudge	fudge	958	Bakery	Cereals and cereal products
31	Candy bar	candy bar	959	Bakery	Cereals and cereal products
32	Egg	egg	0	Meat	Animal foods
33	Beef	beef; steak; veal; sirloin	270	Meat	Animal foods
34	Beef Processed	beef-processed	271	Meat	Animal foods
35	Chicken	chicken	272	Meat	Animal foods
36	Ham	ham	274	Meat	Animal foods
37	Lamb	lamb; keema	275	Meat	Animal foods

# Operações de Preparo

Relação ingrediente e grupo alimentar

```
CREATE TABLE ingredients_group (  
  name VARCHAR(50),  
  synonyms VARCHAR(200),  
  id VARCHAR(4),  
  category VARCHAR(50)  
) AS SELECT  
  "Aliased Ingredient Name" AS name,  
  "Ingredient Synonyms" AS synonyms,  
  "Entity ID" AS id,  
  "CATEGORY" AS category  
FROM CSVREAD('../data/raw/02_Ingredients.csv');
```

```
ALTER TABLE ingredients_group  
ADD food_group VARCHAR(200) NULL;
```

```
UPDATE ingredients_group  
SET food_group='Beverages'  
WHERE category='Beverage Alcoholic'
```

```
UPDATE ingredients_group  
SET food_group='Vegetables'  
WHERE category='Vegetable'
```

```
UPDATE ingredients_group  
SET food_group='Herbs and Spices'  
WHERE category='Spice'
```

# Operações de Preparo

Relação país e subregião

1	Country	World Region
2	Algeria	Africa
3	Angola	Africa
4	Argentina	South America
5	Armenia	Middle East
6	Australia	Australia & New Zealand
7	Austria	DACH Countries
8	Bahamas	Caribbean
9	Bahrain	Middle East
10	Bangladesh	Indian Subcontinent
11	Belarus	Eastern Europe
12	Belgium	DACH Countries
13	Benin	Africa

Relação de grupo alimentar entre tabelas

1	FOOD_GROUP	SERIES_NAME
2	Fats and oils	Cost of oils and fats
3	Herbs and Spices	Cost of vegetables
4	Gourds	Cost of fruits
5	Aquatic foods	Cost of animal-source foods
6	Teas	Cost of vegetables
7	Cereals and cereal products	Cost of starchy staples
8	Baking goods	Cost of starchy staples
9	Nuts	Cost of legumes, nuts and seeds
10	Milk and milk products	Cost of animal-source foods
11	Vegetables	Cost of vegetables
12	Fruits	Cost of fruits
13	Pulses	Cost of legumes, nuts and seeds
14	Coffee and coffee products	Cost of legumes, nuts and seeds
15	Animal foods	Cost of animal-source foods
16	Cocoa and cocoa products	Cost of legumes, nuts and seeds
17	Eggs	Cost of animal-source foods

```
CREATE TABLE Foodb (  
  id VARCHAR(5),  
  common_name VARCHAR(50),  
  scientific_name VARCHAR(50),  
  food_group VARCHAR(35),  
  food_subgroup VARCHAR(35),  
  -- food types é uma opção, pois aí conseguiríamos  
) AS SELECT  
  id, name, name_scientific, food_group, food_subgroup  
FROM CSVREAD('../data/raw/Food.csv');
```

```
ALTER TABLE Foodb  
ADD Series_name VARCHAR(200) NULL;
```

```
-- Atualiza para o segundo Serial Name  
UPDATE foodb  
SET Series_name = 'Cost of fruits'  
WHERE Food_group IN ('Fruits', 'Gourds');
```

# Resultados finais

Valor da receita em cada país

index	RECIPE_ID	TITLE	COUNTRY	PRICE
0	25512	"16 Bean" Pasta E Fagioli	Italy	8.536000000000000
1	22761	"16 Bean" Pasta e Fagioli	Italy	8.536000000000000
2	26227	"Ayuh" Seafood Paella	Spain	11.234000000000000
3	24292	"Blanched" Basil Pesto	Italy	5.505000000000000
4	39340	"Blanketed" Eggplant	Italy	7.510000000000000
5	40184	"Bow-Tie" Pasta with Zucchini	Italy	3.252
6	36506	"California Roll" Salad	Japan	19.867
7	30042	"Canned" Quiches	United States	10.427
8	38275	"Cocotte" of Vegetables	France	7.462
9	25973	"Cowboy" Stew	Ireland	5.642999999999999
10	25973	"Cowboy" Stew	United Kingdom	4.275
11	24104	"Da Beef" Italian Pot Roast Sandwich	Italy	8.536000000000000
12	28749	"Dixie Polenta" (I.e., Creamy Roasted Garlic Grits)	United States	8.258
13	33422	"Dressed" American Wagyu Flat Iron Steak	United States	20.892000000000000
14	26650	"Eat Your Vegetables!": Green Bean Salad with Red Onion and Tomato	United States	5.874999999999999
15	33863	"Egg"Stratordinary Scramble	United States	11.809000000000000



# Resultados finais

Valor da receita em cada país

```
CREATE TABLE Recipes_Prices_per_Country AS
SELECT IR.recipe_id, R.Title, P.country, SUM(COALESCE(CAST(price AS FLOAT), 0)) as price
FROM Prices P, Ingredients_Series_per_Recipes IR, Countries C, Recipes R
WHERE P.category = IR.series_name AND R.Cuisine = C.region AND C.country = P.country AND R.recipe_id = IR.recipe_id
GROUP BY IR.recipe_id, P.country
ORDER BY R.Title, P.country
```



# Resultados finais

Quantas receitas podem ser feitas baseado no salário mínimo local

index	MIN_WAGE_FRACTION	COUNTRY
0	37.556007450361896	Burundi
1	43.23835677536106	Central African Republic
2	55.70078213030439	Mozambique
3	56.9332099538404	Guinea
4	63.294196319584444	Sierra Leone
5	63.61929046993425	Guinea-Bissau
6	74.16481498733577	Liberia
7	75.86223443755922	Malawi
8	78.7444148490596	Burkina Faso
9	86.94476550792045	Niger
10	91.2158549677491	Chad
11	102.50499274832644	Madagascar
12	110.67382238412362	Benin
13	113.10939098592337	Mali
14	118.55469885715496	Ethiopia
15	123.63309205131542	Lesotho
16	126.91301875242848	Uganda



# Resultados finais

Quantas receitas podem ser feitas baseado no salário mínimo local

```
CREATE TABLE GDP_per_Average_Recipe_Price AS
SELECT (G.gdp)/AVG(COALESCE(CAST(price AS FLOAT), 0)) AS min_wage_fraction, R.country
FROM Gdp_per_Country G, Recipes_Prices_Per_Country R
WHERE G.country = R.country
GROUP BY R.country
ORDER BY min_wage_fraction
```

# Análises - Implementadas

- 1) Como avaliar diversidade alimentar e nutritiva perante dietas de diferentes culturas?

Realizou-se uma projeção sobre os nós do tipo "Receita", de modo a interligá-los caso possuam no mínimo 3 ingredientes em comum com outra receita. Assim, é possível comparar o número de receitas semelhantes entre si através de todas as subregiões. Consequentemente, tende-se a concluir que subregiões com uma grande quantidade de receitas semelhantes possuem menor diversidade alimentar. É possível, portanto, avaliar qual parte do globo possui maior variação dentro da sua dieta.

```
MATCH (r1:Receita)-[:Contem]->(i:Ingrediente)<-[:Contem]-(r2:Receita)
WITH r1, r2, i, COUNT(*) AS arestasComum
WHERE arestasComum >= 3
MERGE (r1)-[s:Semelhante]->(r2)
ON CREATE
    SET s.weight = 1
ON MATCH
    SET s.weight = s.weight + 1

MATCH (r1:Receita)-[:Semelhante]-(r2:Receita)
WHERE r1.subregiao = r2.subregiao
WITH r1.subregiao AS subregiao, COUNT(*) AS num_semelhantes
RETURN subregiao, num_semelhantes
ORDER BY num_semelhantes DESC
```



# Análises - Implementadas

## 2) Como avaliar o poder de compra em cada país a partir do custo das receitas?

A partir das nossas bases de dados Countries e Prices, criamos a tabela Recipes\_Prices\_per\_Country, que mostra os preços de cada receita presente no CulinaryDB em todos os países. Para nossa análise, acrescentamos ainda a base de dados World Development Indicators (DataBank), a qual mostra o PIB per capita de cada país, o que nos permite analisar o poder de compra. Por poder de compra, entende-se a razão entre o custo médio das receitas e o PIB per capita (PIB per capita / custo médio).

index	MIN_WAGE_FRACTION	COUNTRY
0	37.556007450361896	Burundi
1	43.23835677536106	Central African Republic
2	55.70078213030439	Mozambique
3	56.9332099538404	Guinea
4	63.294196319584444	Sierra Leone
5	63.61929046993425	Guinea-Bissau
6	74.16481498733577	Liberia
7	75.86223443755922	Malawi
8	78.7444148490596	Burkina Faso
9	86.94476550792045	Niger
10	91.2158549677491	Chad
11	102.50499274832644	Madagascar

```
CREATE TABLE GDP_per_Average_Recipe_Price AS
SELECT (G.gdp)/AVG(COALESCE(CAST(price AS FLOAT), 0)) AS min_wage_fraction, R.country
FROM Gdp_per_Country G, Recipes_Prices_Per_Country R
WHERE G.country = R.country
GROUP BY R.country
ORDER BY min_wage_fraction
```

# Análises - Implementadas

## 3) Como analisar o perfil nutricional das regiões baseado no seu PIB e uso de ingredientes?

Com a base World Development Indicators (DataBank), somos capazes de avaliar o PIB de diferentes regiões. Considerando as regiões Africa e France presentes na nossa base CulinaryDB, obtemos que a região Africa possui um PIB menor que a região France. Isso se torna um fato relevante ao analisarmos a ocorrência de categorias de ingredientes nas receitas de cada região. Na Africa, por exemplo, a taxa de ingredientes vegetais por receita é 3, enquanto na France é 7. Portanto, o perfil nutricional da France pode ser classificado como mais próximo daquele que é ideal para uma vida saudável. Dessa forma, somos capazes de avaliar o perfil nutricional de uma região baseado no seu PIB e uso de ingredientes, sendo, inclusive, possível detectar uma possível insegurança alimentar na região Africa.

```
CREATE TABLE Ingredients_Ratio_France AS
SELECT ISR.series_name, COUNT(ISR.series_name)/COALESCE(CAST(16218 AS FLOAT), 0) AS ratio
FROM Ingredients_Series_per_Recipes ISR, Recipes_per_Region RR
WHERE ISR.recipe_id=RR.recipe_id AND RR.region='France'
GROUP BY ISR.series_name
ORDER BY ratio
```

index	SERIES_NAME	RATIO
0	Cost of oils and fats	0.017758046614872364
1	Cost of legumes, nuts and seeds	0.1564927857935627
2	Cost of fruits	0.7173510913799482
3	Cost of starchy staples	1.6629670736219015
4	Cost of animal-source foods	2.358490566037736
5	Cost of vegetables	3.832408435072142

index	SERIES_NAME	RATIO
0	Cost of oils and fats	0.03533026113671275
1	Cost of legumes, nuts and seeds	0.1858678955453149
2	Cost of fruits	0.8279569892473119
3	Cost of animal-source foods	1.2104454685099846
4	Cost of starchy staples	1.3840245775729647
5	Cost of vegetables	7.307219662058372



# Análises - Não Implementadas

## 1) Qual o impacto do preço dos ingredientes para cada região no seu uso em receitas?

Com `Category_Prices_per_Region` iremos escolher duas regiões e comparar a correlação entre o preço de cada categoria de ingredientes e o seu uso nas receitas típicas da região. Assim, para cada categoria será possível analisar o quão impactante o preço é na presença dos ingredientes na alimentação de uma região. O resultado esperado é quanto maior o preço, menor a presença, porém o impacto do preço não é o mesmo para todas regiões, o que será observado com a análise.

Para isso usamos `Category_Prices_per_Region` para escolher duas regiões para comparação. Usaremos `Foodb_to_Food_Prices` e `Ingredients_to_Foodb_Groups` para obter os ingredientes que pertencem às categorias que estamos analisando. Agora, com `04_Recipe-Ingredients_Aliases` e `01_Recipe_Details` podemos contar quantas vezes uma categoria de alimentos foi utilizada em uma receita, e então contar quantas dessas receitas aparecem em cada uma das regiões analisadas. Por fim, com o preço médio de uma categoria, e o número de usos dessa categoria nas receitas da região, poderemos calcular um coeficiente de correlação entre elas e analisar o quão impactante o preço foi no uso dos ingredientes nas receitas.



# Análises - Não Implementadas

2) Quais são os grupos alimentares mais populares? Quais países possuem os menores custos para tais grupos?

A partir da base CulinaryDB, poderíamos projetar um grafo onde um ingrediente recebe arestas das receitas da qual ele faz parte. Assim, selecionaríamos os ingredientes que recebessem ao menos N receitas, assim podendo selecionar os ingredientes mais populares entre as receitas (chamados de ingredientes populares). Com isso, seria possível visualizar quais grupos alimentares contêm mais desses ingredientes, e poderíamos calcular o valor médio de um determinado grupo a partir de seus ingredientes populares. Selecionando os 4 grupos mais populares, conectaríamos eles aos países do banco Countries, onde a aresta entre eles seria o preço médio daquele grupo naquele local. Concluídas essas etapas, filtraríamos as ligações para que fosse projetado apenas as que possuísem valores menores que X. Dessa maneira, seria possível visualizar quais países têm os grupos alimentares mais populares no menor custo.