



PROJETO 2 - FINAL

GRUPO LAMEVV

Artur de Miranda Rodrigues - 224538

Eliel Oliveira da Silva - 221437

Luiz Guilherme Sousa Nascimento - 230667

Victor Wu - 231467

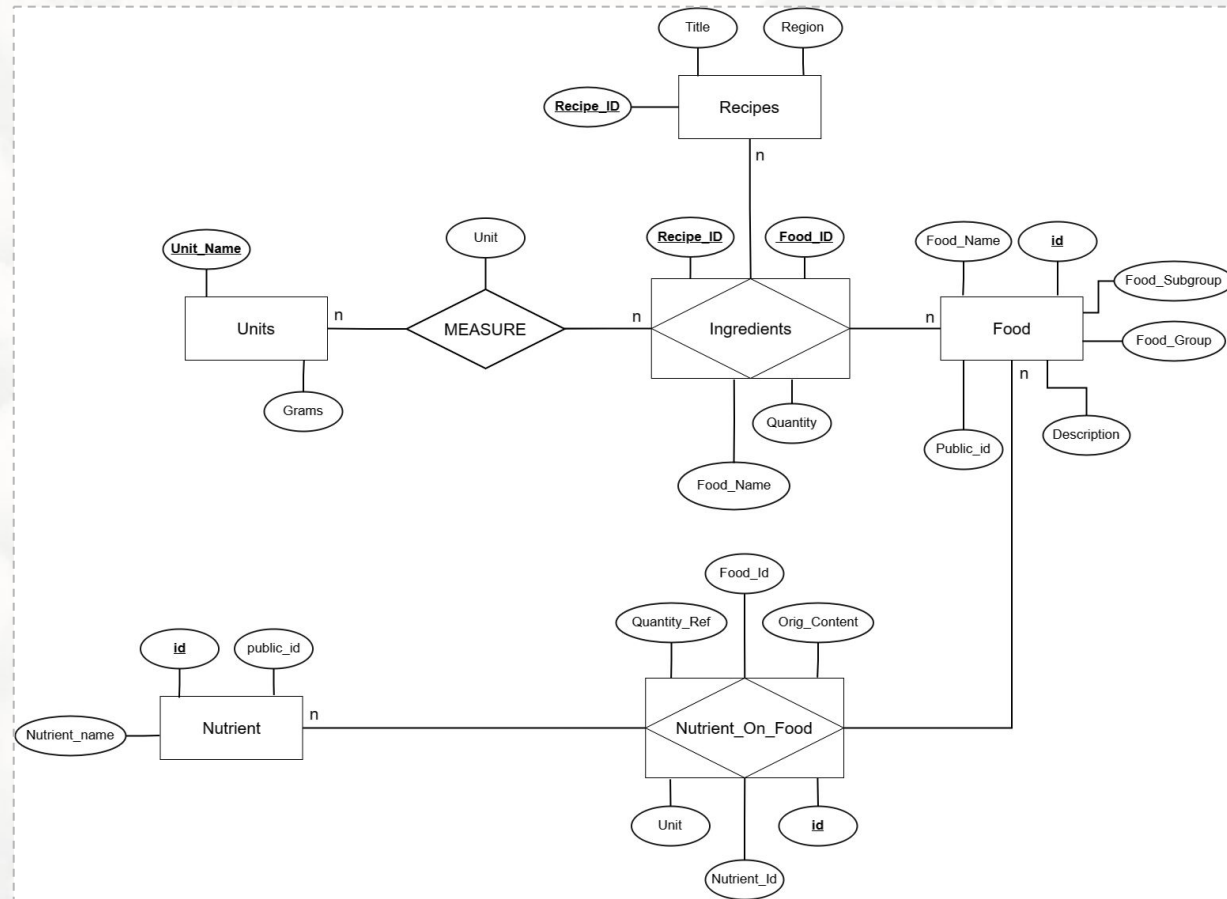
Vítor Paziam Magalhães - 238316

MOTIVAÇÃO E CONTEXTO

O projeto proposto tem como foco central a investigação da composição e análise das receitas culinárias de diferentes regiões do mundo, visando compreender e mapear os macronutrientes presentes em cada prato, e concentração de determinado um macronutriente se relaciona com os hábitos alimentares de cada região.

Motivação: - conscientização sobre a importância da alimentação para a saúde;
- compreensão da diversidade gastronômica global.

Modelo Conceitual - ER



Modelo Lógico Relacional

Nutrient(_ID_, Public_id, Nutrient_name)

Food(_ID_, Public_id, Food_name, description, food_group, food_subgroup)

Nutrient_On_Food(_ID_, nutrient_ID, food_ID, Orig_Content, unit, quantity_Ref)

 nutrient_ID: chave estrangeira para Nutrient

 food_ID: chave estrangeira para Food

Recipes(_Recipe_ID_, Title, Region)

Units(_Unit_Name_, grams)

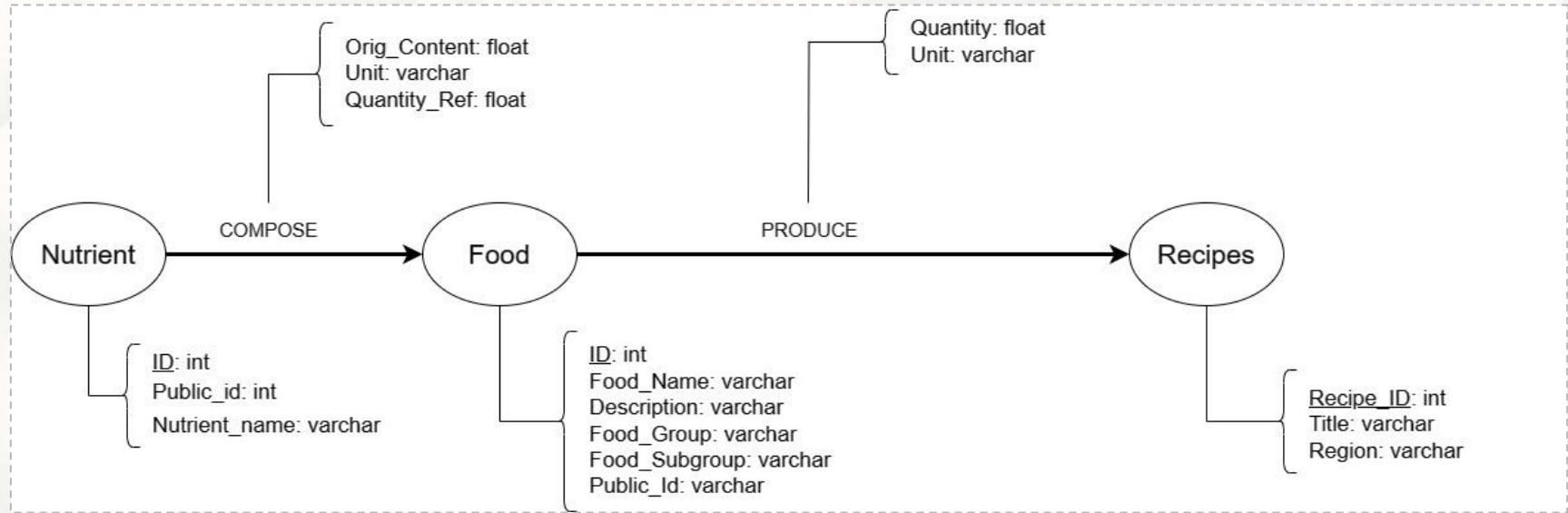
Ingredients(_Recipe_ID_, _Food_ID_, quantity, food_name, unit)

 Recipe_ID: chave estrangeira para Recipes

 Food_ID: chave estrangeira para Food

 Unit: chave estrangeira para Units

Modelo Lógico de Grafos



Transformação e adequação dos dados das receitas do *CulinaryDB*

- Tabela *Recipes.csv*
- Etapas de filtragem e renomeação das colunas
- Modificação de valores → normalização e padronização dos dados de algumas regiões

```
import pandas as pd

# Leitura da tabela de receitas do CulinaryDB através de um DataFrame
df_recipes = pd.read_csv('../data/raw/culinarydb/01_Recipe_Details.csv', sep=',', header=0, encoding='UTF-8')

# Filtragem e renomeação das colunas com informações relevantes
df_recipes = df_recipes.rename(columns={'Cuisine': 'Region'})[['Recipe ID', 'Title', 'Region']]

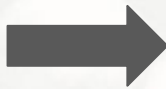
# Padronização/Normalização de regiões específicas
df_recipes['Region'] = df_recipes['Region'].apply(lambda x: x.replace('Misc.: ', ''))
df_recipes['Region'] = df_recipes['Region'].apply(lambda x: x.replace('DACH Countries', 'Germany, Austria, and Switzerland'))

# Exportação dos dados processados para o formato csv (Recipes.csv)
df_recipes.to_csv('../data/interim/Recipes.csv', index=False)
```

Tratamento dos dados da composição das receitas do *CulinaryDB*

- Tabela ***04_Recipe-Ingredients_Aliases*** do *CulinaryDB* não segue uma padronização específica para apresentar a participação de um ingrediente em uma dada receita
- Informação condensada em um texto no atributo ***Original Ingredient Name***

<i>Original Ingredient Name</i>
3/4 cup mayonnaise
1 tablespoon butter
2 (4 ounce) skinless, boneless chicken breast halves
1 1/2 pounds ground beef
1 clove garlic, crushed
...



<i>Quantity</i>	<i>Unit</i>
0.75	cup
1	tablespoon
4	ounce
1.5	pound
1	unit
...	...

Tratamento dos dados da composição das receitas do *CulinaryDB*

- Notebook [units-of-measurement-data](#) → série de extrações e normalizações (Regex)
- Tabela de unidades de medida → unidades selecionadas equivalência padrão em gramas
 - Simplificação e padronização
 - *Unit* → unidade genérica e arbitrária com equivalência de 100 gramas
 - Para unidades como *liter*, *milliliter* e *cup*, consideramos aproximações tomando a água como elemento padrão

Unit of Measurement	Approximation in grams
unit	100
tablespoon	15
pound	454
ounce	28
cup	250
liter	1000
milliliter	1
...	...

Integração dos dados de ingredientes do *FooDB* com o *CulinaryDB*

```
import pandas as pd

# Importação dados processados com quantidade e unidade de medida dos ingredientes nas receitas (CulinaryDB)
df_with_unit = pd.read_csv('../data/interim/Recipe-Ingredients-With-Quantity.csv', sep=',', header=0, encoding='UTF-8')

# Renomeação das colunas
df_with_unit = df_with_unit.rename(columns={'Recipe ID': 'Recipe_ID', 'Original Ingredient Name': 'Original_Ingredient_Name', 'Aliased Ingredient Name': 'Aliased_Ingredient_Name', 'Entity ID': 'Entity_ID'})

# Importação da tabela Food.csv com os dados dos ingredientes do FooDB
df_food = pd.read_csv('Food.csv', sep=',', header=0, encoding='UTF-8')

# Filtragem das colunas com informações relevantes para a integração
df_food = df_food[['id', 'name']]
```

Integração dos dados de ingredientes do *FoodDB* com o *CulinaryDB*

- Criação da coluna *Food_ID* no arquivo das receitas → verificar se o ingrediente tinha correspondência com algum alimento do *FoodDB*
 - exclusão dos registros sem equivalência

```
def get_food_id(food_name):  
    filtered = df_food[df_food['name'].str.contains(food_name, case=False)]  
    if not filtered.empty:  
        return filtered.iloc[0]['id']  
    return None  
  
df_with_unit['Food_ID'] = df_with_unit['Aliased_Ingredient_Name'].apply(get_food_id)  
  
# Filtragem das colunas com informações relevantes e exclusão dos registros com NaN  
df_with_unit = df_with_unit.dropna()[['Recipe_ID', 'Food_ID', 'Aliased_Ingredient_Name', 'Quantity', 'Unit']]  
  
# Exportação dos dados processados para o formato csv (IngredientOnFood.csv)  
df_with_unit.to_csv('../data/interim/IngredientOnFood.csv', index=False)
```

Quais os ingredientes mais utilizados em cada região?

Utilizando funções de agregação, foi possível fazer essa análise. Seleccionamos os ingredientes com uma frequência maior do que a média das frequências em cada região. Como esperado, obtivemos ingredientes comuns, tais como ovo, água e leite.

index	REGION	FOOD_ID	FOOD_NAME	FREQUENCIA
39	USA	92	potato	1188
36	USA	50	cinnamon	1293
32	USA	183	thyme	1403
31	USA	10	lemon	1417
29	USA	15	celery	1460
28	USA	27	mustard	1501
25	USA	131	parsley	1585
24	USA	645	vinegar	1607

Quais os alimentos e receitas que mais contribuem para a ingestão de açúcares, gorduras e proteínas para cada região?

Traçamos um perfil nutricional de cada receita, permitindo filtrar receitas com diferentes índices desses macronutrientes.

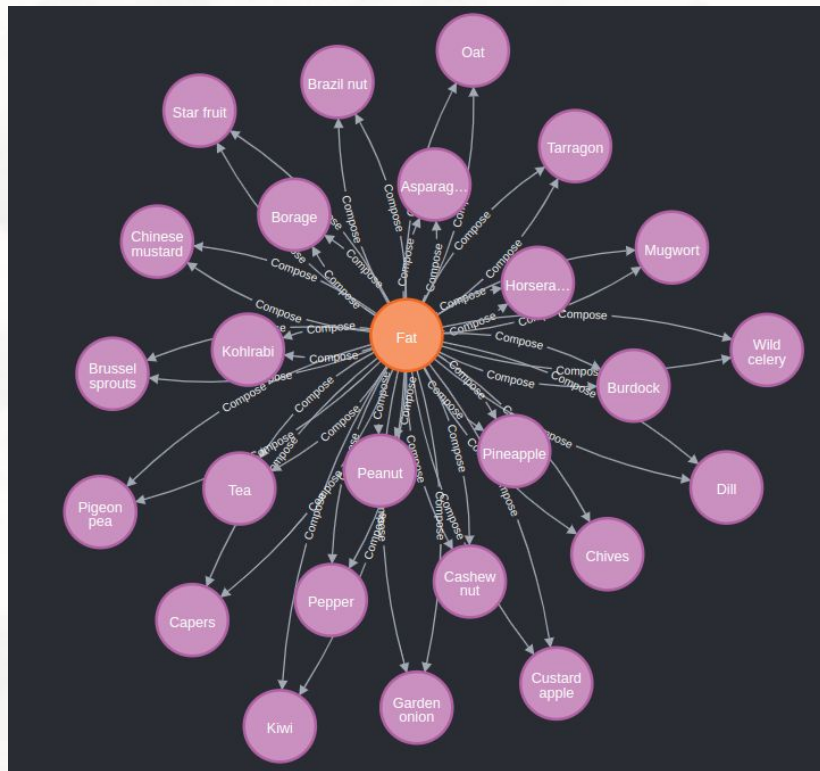
index	RECIPE_ID	REGION	TITLE	RECIPE_WEIGHT	PER_FAT	TOTAL_FAT
13016	34583	USA	Voodoo Empress	42	0.752	31.584
401	5552	Australia &	Fairy Bread	62.5	0.706	44.125
2874	8753	Greece	Greek Spaghetti I	125.35	0.7040426804946152	88.25175
14128	45659	USA	Spiced Pecans	401.25	0.7032461993769471	282.1775375
13412	43358	USA	Spicy Toasted Pecans	672.5	0.699277063197026	470.263825
13569	43856	USA	Candied Curried Pecans	835	0.6764180793413174	564.80909625
11392	29528	USA	White and Dark Chocolate Dogw	918	0.6692810457516339	614.4
10343	17362	USA	Quick Zesty Green Beans	97.5	0.6363076923076924	62.04
929	18710	Canada	Spicy Glazed Pecans	452.5	0.626271270718232	283.38775

Quais das receitas são equilibradas em termos de macronutrientes para cada região?

Através do perfil nutricional de cada receita, podemos definir se a receita é balanceada com base em parâmetros padrões de proteínas, gorduras e carboidratos. Na nossa análise, consideramos balanceado uma receita que seja composta com até 10% de gordura, 50% de carboidrato e pelo menos 10% de proteína.

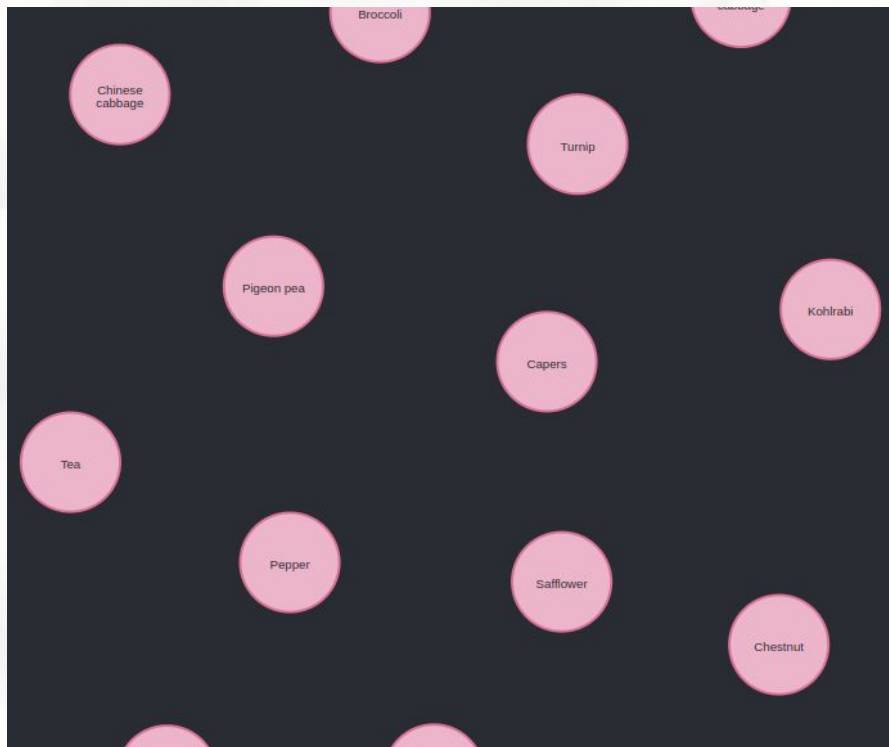
index	RECIPE_ID	TITLE	REGION	RECIPE_WEIGHT	PER_FAT	PER_CARBO	PER_PROTEIN	EQUILIBRADO
41	2651	Zimbabwean Chicken	Africa	1866.25	0.01961564	0.14226188881	0.0551417280643	false
336	18898	Za'atar Puff Pastry	Africa	850	0.07195882	0.28514705882	0.04125	false
291	18851	Yogurt Marinated Gr	Africa	602.5	0.04652782	0.325613278008	0.1157742738589	true
439	34729	Yellow Rice	Africa	1772	0.07142271	0.32783103837	0.0468022573363	false
417	34707	Ye'abesha Gomen (f	Africa	548.75	0.06986788	0.07702505694	0.0080956719817	false
458	34748	Yassa au Poulet II (f	Africa	1240	0.05962584	0.30322580645	0.0703169354838	false
328	18890	Yassa Chicken with	Africa	3580	0.02037002	0.239813547486	0.0707056564245	false
378	18942	Yassa Chicken	Africa	2615	0.02202861	0.276746271516	0.0825945506692	false

Perguntas Não Implementadas - Modelo Lógico de Grafos



Por meio dos nós apresentados no modelo lógico e seus respectivos atributos, é possível identificar relações que poderiam ser construídas para responder questões sobre o perfil nutricional das regiões, utilizando ferramentas de rede complexa

Perguntas Não Implementadas - Modelo Lógico de Grafos



1. Qual o perfil nutricional de uma região baseado nos alimentos mais populares?
2. É possível construir uma relação de perfil nutricional entre as regiões?

Criar uma relação entre nós do tipo Food, chamada Same_Region, em que dois alimentos são conectados, quando aparecem em receitas de uma mesma região.