

Fatos Mundiais

Giovanna Damian, Laura Corssac, Lucas Hagen

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)

{gvdamian, lbrcorssac, lucas.hagen}@inf.ufrgs.br

Resumo. *Este projeto tem como objetivo produzir análises sobre os países do mundo, além de possibilitar com que outros façam o mesmo. Os dados escolhidos possuem informações variadas sobre a maioria das nações, como consumo de álcool, produção de chocolate, quantidade populacional e Produto Interno Bruto (PIB). Através de um website interativo, esperamos que nós, assim como nossos usuários, possamos expandir nossos conhecimentos sobre as culturas e nações ao redor do mundo.*

1. Introdução

Com o mundo cada vez mais globalizado, a interconexão entre pessoas de diversos países cresce. Entretanto, conflitos entre nações e problemas humanitários tornam a desigualdade entre essas discrepante. Por isso, é cada vez mais importante usarmos a enorme fonte de dados na internet disponível para tomarmos conhecimento do mundo todo e não só de nossos arredores. Movimentos sociais como Global Citizen [glo 2021] buscam motivar as pessoas a engajarem-se com causas que afetam o mundo inteiro, e não apenas localidades próximas onde vivem. Esses movimentos, assim como os autores desse artigo, acreditam que informações sobre o mundo são cruciais para que todos tomem conhecimento de problemas que normalmente possuem localidades fisicamente distantes, mas que podem ser combatidos, ou, pelo menos, amenizados, com ações locais.

Um exemplo desses problemas humanitários é o trabalho escravo na produção de cacau. A Costa do Marfim produz 45% do cacau consumido no mundo, sendo grande parte exportado e destinado à produção de chocolate por grandes empresas. A produção é, todavia, feita por crianças em idade escolar em regime de trabalho forçado. [sla 2021] Dentre os conjuntos de dados que escolhemos para analisar, consta um que relaciona cada país a informações tais como consumo e produção de chocolate. Acreditamos que analisando-as, podemos produzir outras conclusões e prover às pessoas uma visualização mais detalhada a respeito do assunto. Assim, podemos alertar nossos usuários, fomentando neles um espírito de cidadão global.

Para deixar o trabalho mais interessante, selecionamos outros conjuntos de dados que também relacionam cada país a algum conjunto de informações, como consumo de álcool e informações demográficas. Essas fontes de dados possuem formato CSV e serão processado por scripts na linguagem de programação Python para por fim serem salvos em um banco de dados Postgres (SQL). É a partir das tabelas salvas nessa plataforma que realizaremos as consultas, assim, as habilidades de agregação, integração de dados vistas nas aulas teóricas poderão ser postas em prática.

Para que possamos visualizar nossas consultas e tirar conclusões sobre elas, assim como permitir que mais pessoas façam o mesmo, este projeto é composto por um Website

1	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
2	Afghanistan	0	0	0	0.0
3	Albania	89	132	54	4.9
4	Algeria	25	0	14	0.7
5	Andorra	245	138	312	12.4
6	Angola	217	57	45	5.9
7	Antigua & Barbuda	102	128	45	4.9
8	Argentina	193	25	221	8.3
9	Armenia	21	179	11	3.8

Figura 1. Parte da Tabela Drinks

interativo. Seu código fonte está disponibilizado publicamente no GitHub. [git 2021] Dessa forma, os interessados devem baixar o repositório em suas máquinas e seguir a sequência de instruções para execução que lá consta.

Este artigo está organizado em mais 6 seções. A próxima descreve as bases de dados que escolhemos e como serão organizadas no banco. Na seção 3, comentamos sobre a parte técnica do processo de desenvolvimento, como sobre as plataformas utilizadas. A seção 4 apresenta as etapas do processo de desenvolvimento, como o processamento e a análise dos dados. Na seção 5, apresentamos as funcionalidades que a implementação do projeto contém e detalhes sobre como foram construídas. Na seção 6, expomos algumas análises interessantes que fizemos na plataforma. Por fim, na seção 7, relatamos nossas conclusões tiradas ao longo do desenvolvimento do trabalho.

2. Descrição da Base de Dados

Para fazer esse trabalho, inicialmente coletamos dados de três bases de dados. Todas estão no foram obtidas de fontes online e públicas, de forma que nenhum *Web Scrapping* foi necessário. A seguir descrevemos cada uma delas e posteriormente como iremos integrá-las.

2.1. Drinks

Esta base de dados conta com 193 linhas, cada possui um nome de país e quantas taças de vinho uma pessoa em média consome por ano, quantas latas de cerveja, e *shots* de destilados uma pessoa desse país consome em média por ano. Os dados foram coletados a partir de uma pesquisa da Organização Mundial da Saúde (OMS) em 2010. O autor da base de dados também adicionou uma nova coluna, que representa o total de litros de álcool puro consumido por pessoa. Esse foi estimado multiplicando o teor médio de álcool, o tamanho médio da dose para cada bebida e a quantidade de doses. Por fim, soma-se os três valores e se tem o total de álcool por pessoa por país. Uma amostragem de 8 linhas do total do arquivo de 4.28 KB pode ser visualizada na Fig. 1 e mais informações sobre ele podem ser encontradas em [dri 2014] e [dri 2010].

2.2. Chocolate

Cada linha desta base de dados representa uma avaliação de um crítico gastronômico a uma barra de chocolate. Dentre outras informações estão a nota atribuída, a nação produtora do chocolate e de qual país o grão de cacau foi extraído. A nota atribuída varia entre 1.0 a 5.0, sendo 1.0 a nota mais baixa e 5.0 a mais alta. O tamanho total do

id	reference	company	company_location	review_date	country_of_bean_origin	specific_bean_origin_or_bar_name	cocoa_percent	rating
0	2454	5150	U.S.A	2019	Madagascar	Bejofo Estate, batch 1	76.0	3.75
1	2458	5150	U.S.A	2019	Dominican republic	Zorzal, batch 1	76.0	3.5
2	2454	5150	U.S.A	2019	Tanzania	Kokoa Kamili, batch 1	76.0	3.25
3	797	A. Morin	France	2012	Peru	Peru	63.0	3.75
4	797	A. Morin	France	2012	Bolivia	Bolivia	70.0	3.5
5	1015	A. Morin	France	2013	Venezuela	Chuao	70.0	4.0
6	1019	A. Morin	France	2013	Peru	Chanchamayo Province	63.0	4.0
7	1011	A. Morin	France	2013	Ecuador	Equateur	70.0	3.75

Figura 2. Parte da Tabela *Chocolate*

Country	Region	Population	Area (sq. mi.)	Pop. Density (per sq. mi.)
Afghanistan	ASIA (EX. NEAR EAST)	31056997	647500	48,0
Albania	EASTERN EUROPE	3581655	28748	124,6
Algeria	NORTHERN AFRICA	32930091	2381740	13,8
American Samoa	OCEANIA	57794	199	290,4
Andorra	WESTERN EUROPE	71201	468	152,1
Angola	SUB-SAHARAN AFRICA	12127071	1246700	9,7
Anguilla	LATIN AMER. & CARIB	13477	102	132,1
Antigua & Barbuda	LATIN AMER. & CARIB	69108	443	156,0

Figura 3. Parte da Tabela *Countries of the World*

arquivo é igual a 470 KB, sendo uma pequena parte desse representada pela figura 2, e mais informações encontradas em [cho 2020].

2.3. Countries of the World

Esta base de dados possui 227 linhas, cada uma contendo um nome de país e algumas informações sobre esse, tais quais sua área (em metros quadrados), sua população, seu Produto Interno Bruto (PIB) per capita. Esses dados foram coletados a partir de diversas instituições do governo estadunidense, como a CIA. A base possui tamanho igual a 37.41 KB e seus dados são datados no período de 1970 - 2017. Uma parte da tabela pode ser vista na Fig. 3 e mais detalhes podem ser obtidos em [cou 2017].

Cada uma dessas bases de dados passará por um processo de limpeza de dados e posteriormente será representado como uma tabela no banco de dados. Assim, nossas análises irão consultar cada uma dessas tabelas resultantes e produzir mais informações.

3. Ambiente de Desenvolvimento

Considerando a modelagem dos dados coletados, escolhemos utilizar um banco de dados relacional. Esse modelo também nos trouxe a vantagem de podermos conectar representações de um mesmo país em tabelas diferentes usando chaves estrangeiras. Além disso, é fácil de ser expandido com a adição de novas tabelas e registros. Para então a implementação deste banco de dados, usamos o SGBD PostgreSQL [pos 2021], tendo em vista que essa plataforma é uma das mais utilizadas no mundo (Fig. 4), é open-source e gratuita. Para auxiliar o desenvolvimento, usamos também a plataforma Postbird, que é um cliente para o Postgres, no qual tabelas e registros podem ser criados, visualizados, e manipulados.

A respeito do website, usamos o framework Flask, da linguagem Python, que facilitou a integração do front-end (parte de interação com o usuário) e do back-end (ar-

378 systems in ranking, September 2021

Rank			DBMS	Database Model	Score		
Sep 2021	Aug 2021	Sep 2020			Sep 2021	Aug 2021	Sep 2020
1.	1.	1.	Oracle +	Relational, Multi-model	1271.55	+2.29	-97.82
2.	2.	2.	MySQL +	Relational, Multi-model	1212.52	-25.69	-51.72
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	970.85	-2.50	-91.91
4.	4.	4.	PostgreSQL +	Relational, Multi-model	577.50	+0.45	+35.22
5.	5.	5.	MongoDB +	Document, Multi-model	496.50	-0.04	+50.02
6.	6.	7.	Redis +	Key-value, Multi-model	171.94	+2.05	+20.08
7.	7.	6.	IBM Db2	Relational, Multi-model	166.56	+1.09	+5.32
8.	8.	8.	Elasticsearch	Search engine, Multi-model	160.24	+3.16	+9.74
9.	9.	9.	SQLite +	Relational	128.65	-1.16	+1.98
10.	11.	10.	Cassandra +	Wide column	118.99	+5.33	-0.18

Figura 4. Ranking de SGBD mais utilizados. Fonte: [ran 2021]

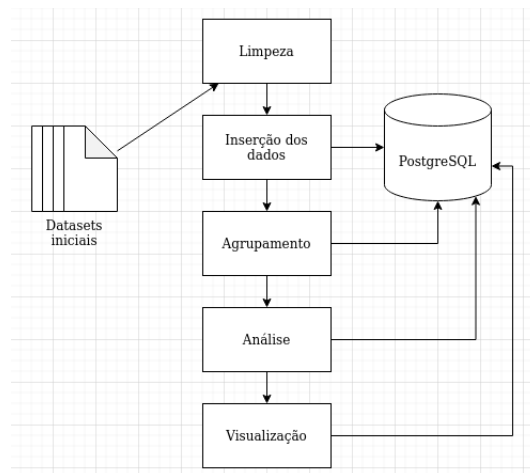


Figura 5. Etapas do Desenvolvimento

mazenamento de dados e consultas a eles). Para a visualização das consultas, além de HTML e CSS, utilizamos a biblioteca "matplotlib", também do Python, para a construção de gráficos. Escolhemos como plataforma de interface com o usuário um Website, e não, por exemplo, um aplicativo mobile, pelo fato dessa plataforma permitir uma visualização em telas maiores, o que é preferível para este projeto, que inclui gráficos e listagens. Também pensamos que um website pode ser acessado por uma gama maior de usuários, pois não é limitado a um sistema operacional ou um tipo de equipamento.

4. Metodologia

Nesta seção apresentamos as o caminho que um dado percorre desde os arquivos CSV iniciais até os gráficos e visualizações finais. Uma breve ilustração desse processo pode ser vista na Fig. 5. Também falaremos sobre as etapas de desenvolvimento feitas após a inserção dos dados no banco.

4.1. Coleta

A coleta de dados, como já mencionado, consistiu na obtenção dos data sets, todos em formato CSV. Como esse é um formato bem estruturado, sua conversão/exportação para o banco de dados relacional (Postgres) não necessitou de transformações de modelagem complexas.

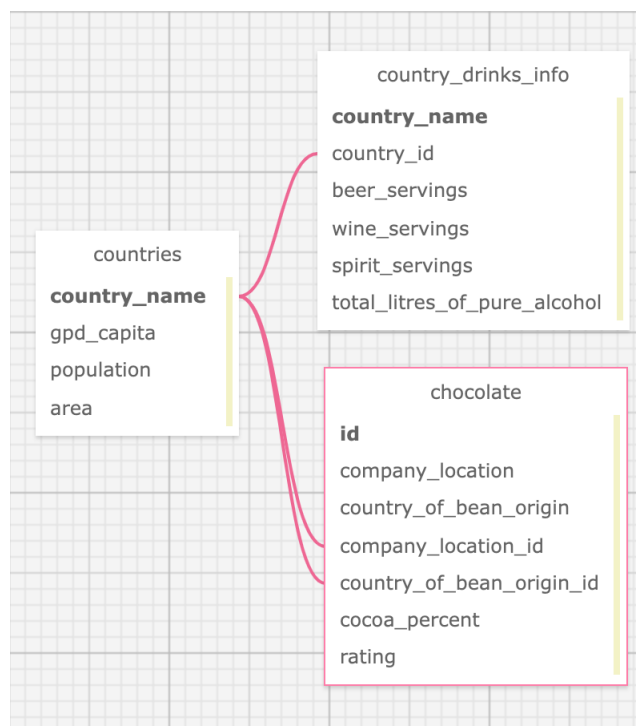


Figura 6. Tabelas inseridas no banco de dados

4.2. Limpeza e Agregação de Dados

Possuímos dados de diversas fontes, por isso, realizamos algumas operações de limpeza para agregá-los. Um dos desafios enfrentados foi o tratamento de nomes dos países. Como todos os conjuntos de dados foram escritos na língua inglesa, os países com nomes simples como *Brazil* são escritos da mesma forma em todos eles. Entretanto, nomes de países compostos possuem inúmeras variações. Por exemplo, os Estados Unidos da América é representado como *U.S.A* em uma das fontes e como *United States* em outra. A Coreia do Norte em alguns é grafada *Korea*, *North* e *North Korea* em outros. A Bósnia e Herzegovina é grafada como *Bosnia-Herzegovina* e *Bosnia & Herzegovina*. A República Tcheca às vezes é escrita como *Czech Republic* e em outras como *Czech Rep.*. E os exemplos são inúmeros. Para garantir uma devida agregação de dados, criamos uma rotina, em Python, que toma uso de expressões regulares para transformar os nomes dos países em um certo padrão. Essa rotina é invocada por outra, também escrita em Python, que lê cada entrada de um CSV e insere no banco de dados. Infelizmente, alguns poucos nomes tiveram que ser manualmente ajustados no arquivo fonte, pois continham erros de digitação.

Para associar dados de um mesmo país em diferentes tabelas, além de termos, pela padronização dos nomes, a garantia que cada registro referente a um mesmo país terá a grafia do nome desse igual, decidimos criar chaves estrangeiras. Assim, o atributo "country_name" da tabela "countries" é referenciado como chave estrangeira pelas duas outras tabelas. Caso tentemos adicionar uma país que não está presente nessa tabela de referência, o registro ainda sim será inserido ao banco, com chave estrangeira nula. Optamos por fazer dessa maneira, ao invés de descartarmos o registro, pois assim ainda podemos trabalhar com os dados que temos desse país. Por exemplo, o país Montenegro

está presente no arquivo 'drinks', mas não em 'countries'. Não é possível, por exemplo, obter o PIB desse país, mas ainda conseguimos visualizar dados do consumo alcoólico no país. Os itens da tabela "chocolate" possuem duas chaves estrangeiras. Uma para o país que originou o cacau do chocolate avaliado e uma para o país de origem da empresa que vendeu o cacau. Da mesma forma, ambos podem ser nulos. As chaves estrangeiras também permitem uma visualização dos dados mais fácil nos clientes de bancos de dados. Por exemplo, usando o Postbird, dado um registro da tabela drinks cuja chave estrangeira não seja nula, conseguimos facilmente encontrar mesmo país na tabela countries. Isso trouxe um ganho ao desenvolvimento.

4.3. Inserção no Banco de Dados

Após uma linha de um CSV de entrada ser limpa, ela é inserida no banco de dados, usando métodos em Python, que tomam usam de funções da biblioteca Flask. Nessa etapa, também selecionamos os itens que queremos de cada arquivo. Por exemplo, não inserimos, por enquanto, os dados de ingredientes de um chocolate no banco.

Outro processo de filtragem que fizemos foi descartar os itens do CSV "chocolate" que possuíam "blend" como país de origem do grão de cacau, que indica várias origens indefinidas. Como o escopo desse trabalho é fazer análises sobre os diferentes países, um dado não associado a um país definido não é interessante para gente no momento. Os modelos de tabelas resultantes podem ser vistos em 6. As tabelas "chocolate" foi originada a partir do CSV de mesmo nome. Já "countries" teve origem de "countries of the world" e o CSV "drinks" foi mapeado para a tabela "country_drinks_info". A partir de então, usaremos o nome "countries" e "chocolate" para nos referirmos às respectivas tabelas. Usaremos também os nomes "drinks" e "country_drinks_info" intercambiavelmente para nos referir à terceira tabela.

4.4. Implementação e análise de consultas

A implementação das consultas e a da interface gráfica do website tratou-se de um processo iterativo. Intercalamos o desenvolvimento das consultas com o a visualização dessas no website. Assim, a cada nova consulta, criávamos uma página simples em nosso website para que pudéssemos acessar e visualizar os resultados. Por exemplo, fizemos consultas de ordenamento, que inicialmente foram mostradas em uma listagem simples em uma parte do website. Depois de termos analisado os resultados e estarmos satisfeitos com esse tipo de consulta, passamos para próxima etapa.

4.5. Ajustes no website

Esta etapa consistiu na escolha de componentes gráficos que facilitem a visualização e tornem a experiência do usuário dentro do website mais agradável. Como esta etapa, apesar de ter enorme importância, não era o escopo principal do trabalho, procuramos sempre analisar a complexidade de implementação de cada escolha, preferindo sempre os elementos mais simples possíveis, mas que não deixavam de ser elegantes. Concluída esta etapa, a medida que tínhamos mais ideias de consultas, voltávamos a etapa anterior e, em seguida, a esta.

5. Funcionalidades

Nesta seção, detalhamos as funcionalidades disponíveis ao usuário no website e as consultas feitas para elas. Organizamos essa seção conforme o website é dividido. Esse

Countries of the World

[Home](#) [Correlation](#)

Criteria	<input type="text" value="Wine Servings"/>	Order	<input type="text" value="Ascending"/>	Filter	<input type="text" value="Show All"/>	Show Only:	<input type="text"/>	<input type="button" value="Enviar"/>
----------	--	-------	--	--------	---------------------------------------	------------	----------------------	---------------------------------------

1.	Afghanistan
2.	Albania
3.	Algeria
4.	American Samoa
5.	Andorra
6.	Angola
7.	Anguilla
8.	Antigua And Barbuda

Figura 7. Página inicial do website

possui três páginas, as quais chamamos de: Lista, Detalhes e Correlação. As próximas sub-seções expõem então detalhes de cada uma delas.

5.1. Lista

A interface gráfica resultante é mostrada na Fig. 7. Ela inicialmente apresenta a listagem de todos os países em ordem alfabética. Esses países são retirados das tabelas "drinks" e "countries". Assim, um país que está em pelo menos alguma dessas tabelas é listado. Escolhemos apenas essas tabelas por simplicidade. O caminho da URL para acessar essa página pode ser algum dos quais: '/', '/index', '/list' e a consulta feita é a seguinte:

```
SELECT ROW_NUMBER() OVER(ORDER BY country_name), \
country_name FROM
(
    SELECT country_name FROM countries
    UNION
    SELECT country_name FROM country_drinks_info
) AS subquery
ORDER BY country_name
```

É possível, nessa página, que o usuário monte suas consultas customizadas de ordenamento dos países baseando-se em algum critério, que é escolhido na barra superior da página, colorida de forma diferente. Da mesma forma, o usuário pode escolher uma ordem, um filtro e um limite para os resultados. Todos são detalhados abaixo.

- **Critério**

- **Wine Servings:** A quantidade de taças de vinho consumidas por pessoa durante o ano de 2010.
- **Beer Servings** A quantidade de garrafas de cerveja consumidas por pessoa durante o ano de 2010.
- **Spirit Servings:** A quantidade de doses de destilados de consumidas por pessoa durante o ano de 2010.

- **Total Alcohol:** A pela quantidade de total de álcool, em litros, consumida por pessoa durante o ano de 2010.
- **GDP:** O PIB per capita.
- **Population:** O tamanho populacional.
- **Area:** A extensão territorial, em metros quadrados.
- **Average Rating Chocolate Produced:** A média das avaliações dos chocolates cujos grãos de cacau são originados em cada país.
- **Average Rating Chocolate Sold:** A média das avaliações dos chocolates produzidos por empresas localizadas em cada país.

- **Ordem**

- **Ascending:** Os items são mostrados em ordem crescente com base no critério avaliado.
- **Descending:** Os items são mostrados em ordem decrescente com base no critério avaliado.

- **Filtro**

- **Show all:** O mesmo que não aplicar nenhum filtro. Todos os resultados são mostrados, obedecendo o limite estabelecido no próximo item.
- **Only Greater than Average:** Apresenta somente resultados que possuem o valor do critério selecionado maior que a média de todos os países.
- **Greater than or equal Average:** Apresenta resultados que possuem o valor do critério selecionado maior ou igual à média de todos os países.
- **Only Less than Average:** Apresenta somente resultados que possuem o valor do critério selecionado menor que a média de todos os países.
- **Less than or equal Average:** Apresenta resultados que possuem o valor do critério selecionado menor ou igual à média de todos os países.
- **Equal Average:** Apresenta somente resultados que possuem o valor do critério selecionado igual à média de todos os países.

- **Limite**

Esse campo é um limitador da quantidade de resultados a serem mostrados. Caso nulo, todos são mostrados, obedecendo o filtro estabelecido.

Para exemplificar, a consulta realizada para mostrarmos os 10 países que mais consomem cerveja, é a seguinte:

```
SELECT ROW_NUMBER() OVER(ORDER BY beer_servings DESC), \
country_name, \
beer_servings \
FROM country_drinks_info
ORDER BY beer_servings DESC
LIMIT 10
```

O usuário, para obter essa consulta basta selecionar "Beer Servings" no campo "Criteria", "Descending" no campo "Order" e digitar "10" no campo "Show Only". A forma como os items são listados pode ser vista em Fig. 8, que omite o cabeçalho da página para melhor visualização dos items. Essa página também pode ser obtida através do seguinte caminho da URL pode ser digitado:

"/ranking/?criteria=beer_servings&order=DESC&filter=all&limit=10"

1.	<u>Namibia</u>	376.00
2.	<u>Czech Republic</u>	361.00
3.	<u>Gabon</u>	347.00
4.	<u>Germany</u>	346.00
5.	<u>Poland</u>	343.00
6.	<u>Lithuania</u>	343.00
7.	<u>Venezuela</u>	333.00
8.	<u>Ireland</u>	313.00

Figura 8. Top 10 Consumo de Cerveja

Outro exemplo interessante, que exigiu uma consulta diferente em comparação às outras é o ordenamento dos países pelas médias dos chocolates cujos grãos foram produzidos neles. Essa consulta precisou do comando "GROUP BY", já que, ao contrário das outras tabelas, não há apenas um registro por país. Também tomamos uso de uma "VIEW" para facilitar a consulta.

```
DROP VIEW IF EXISTS sub_view;
CREATE VIEW sub_view AS
SELECT country_of_bean_origin, AVG(rating) AS avg_score
FROM chocolate
GROUP BY country_of_bean_origin
ORDER BY AVG(rating) DESC;

SELECT ROW_NUMBER() OVER(), \
country_of_bean_origin, avg_score \
FROM sub_view;
```

Da mesma forma que a consulta mostrada anteriormente, essa pode ser obtida selecionando "Average Rating Chocolate Produced" e "DESC" ou então digitando o seguinte caminho:

"/ranking/?criteria=country_of_bean_origin&order=DESC&filter=all&limit="

5.2. Detalhes

Nesta seção apresentamos detalhes de um país específico, dentre eles PIB per capita, e quantidade total de cerveja consumida. As informações são retiradas das tabelas "countries" e "drinks". Os países que estão presentes em apenas uma das tabelas mostram alguns dados como indefinidos. A consulta realizada para, por exemplo, a obtenção dos dados do Brasil é esta:

Brazil

Glasses of wine consumed per person: 16.0

Bottles of beer consumed per person: 245.0

Shots of spirits consumed per person: 145.0

Total alcohol (litres) consumed per person: 7.2

Population: 188078227

GPB (per capita): 7600.0

Area (square meters): 8511965.0

Figura 9. Página de detalhes sobre o Brasil

```
SELECT countries.country_name \
    beer_servings, \
    spirit_servings, \
    wine_servings, \
    total_litres_of_pure_alcohol, \
    population, \
    area, \
    gpd_capita \
FROM country_drinks_info
FULL OUTER JOIN countries
ON country_drinks_info.country_name \
= countries.country_name
WHERE countries.country_name = 'brazil' \
OR country_drinks_info.country_name = 'brazil'
```

Essa consulta pode ser obtida clicando no nome de nosso país na listagem da tela inicial e também com o seguinte caminho: `"/info/brazil"`. Os resultados aparecem conforme Fig. 9. Países com campos indefinidos, como Montenegro, aparecem conforme Fig. 10.

5.3. Correlação

Esta aba possibilita ao usuário fazer correlações dos mesmos critérios mostrados na seção 5.1. Quando os itens selecionados vêm ambos da tabela `"countries"` ou ambos da tabela `"drinks"`, a operação é a mais simples, sendo necessário o acesso a apenas a tabela em questão. Por exemplo, para relacionar doses de vinho com o total de álcool consumido, temos a seguinte consulta.

Montenegro

Glasses of wine consumed per person: 128.0

Bottles of beer consumed per person: 31.0

Shots of spirits consumed per person: 114.0

Total alcohol (litres) consumed per person: 4.9

Population: Undefined

GPB (per capita): Undefined

Area (square meters): Undefined

Figura 10. Página de detalhes sobre Montenegro

```
SELECT country_drinks_info.country_name, \
       country_drinks_info.wine_servings, \
       country_drinks_info.total_litres_of_pure_alcohol \
FROM country_drinks_info
```

Essa consulta pode ser obtida no selecionano "Wine Servings" para "X Criteria" e "Total Alcohol" para "Y Criteria", ou então através do seguinte caminho:

`"/correlation/?criteria1=wine_servings&criteria2=total_litres_of_pure_alcohol"`

Quando trata-se de registros de diferentes tabelas, é necessário fazer a operação de "INNER JOIN". No caso de um elemento da tabela "drinks" e outro da tabela "countries", como, por exemplo, a correlação entre o consumo de cerveja e o PIB per capita, a consulta resultante é a seguinte.

```
SELECT country_drinks_info.country_name, \
       country_drinks_info.beer_servings, \
       countries.gpd_capita \
FROM country_drinks_info
INNER JOIN countries
ON country_drinks_info.country_name \
= countries.country_name
```

Quando temos pelo menos um dos itens a ser retirado da tabela "chocolate", a correlação se torna um pouco mais complexa, pois criamos uma "VIEW" para cada um dos itens. Por exemplo, se desejarmos correlacionar a média das avaliações dos chocolates cujos grãos possuem origem em um país com a dos que foram vendidos por empresas estabelecidas no mesmo temos consulta mostrada abaixo.

```

DROP VIEW IF EXISTS view1;
CREATE VIEW view1 AS
SELECT country_of_bean_origin AS country_name, \
       AVG(rating) AS avg_rating
FROM chocolate
GROUP BY country_of_bean_origin;

DROP VIEW IF EXISTS view2;
CREATE VIEW view2 AS
SELECT company_location AS country_name, \
       AVG(rating) AS avg_rating
FROM chocolate
GROUP BY company_location;

SELECT view1.country_name, \
       view1.avg_rating, \
       view2.avg_rating
FROM view1
FULL OUTER JOIN view2
ON view1.country_name = view2.country_name

```

Note que é usado um "FULL OUTER JOIN" ao invés do "INNER JOIN". Resolvemos fazer uma exceção para essa consulta, em específico, pois queremos mostrar no gráfico aqueles países que apenas vendem chocolates e também aqueles que apenas produzem grãos. Se usássemos o "INNER JOIN", esses seriam descartados. Essa troca fez com que a análise da seção 6.6 ficasse mais interessante.

6. Análises

Nesta seção, vamos apresentar as análises que achamos mais interessantes.

6.1. Top 10 consumo de vinho

Motivados por curiosidades pessoais, buscamos pelos 10 maiores consumidores de vinho e obtivemos a seguinte lista.

1. França
2. Portugal
3. Andorra
4. Suíça
5. Dinamarca
6. Eslovênia
7. Luxemburgo
8. Croácia
9. Itália
10. Guiné Equatorial

Com a exceção do último país, todos eles se situam na Europa, o que é interessante. Não surpreendentemente, França, Portugal e Itália, famosos pela sua produção de

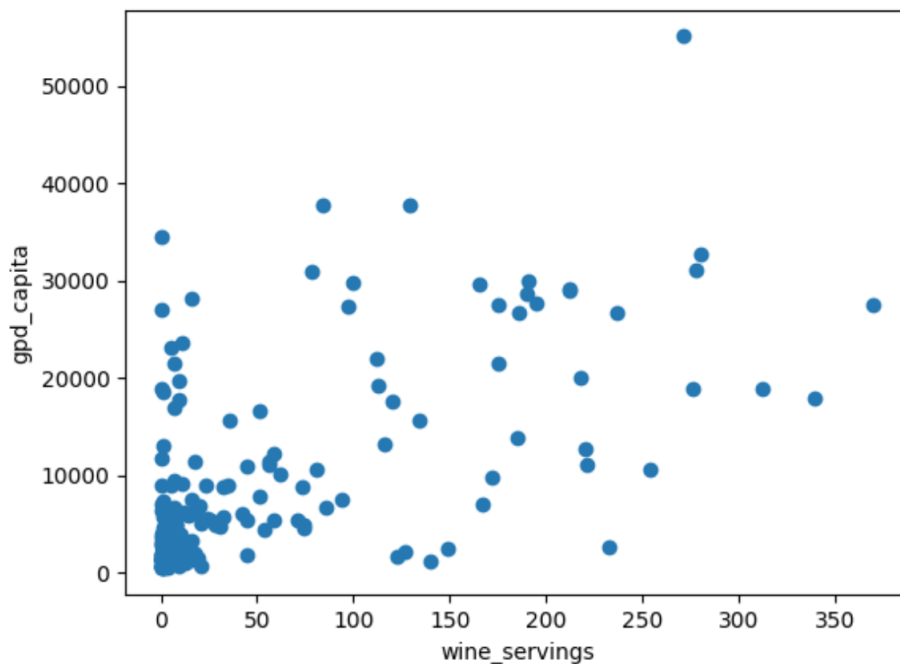


Figura 11. Correlação entre consumo de vinho e PIB per capita

vinho, estão presentes nessa listagem. Ao contrário do esperado, no entanto, o Chile, que também possui uma grande produção de vinhos de qualidade está em 23º lugar. Pensamos que como o vinho é uma bebida normalmente mais cara, os países mais desenvolvidos têm o maior costume de consumi-los.

6.2. Correlação entre consumo de vinho e PIB per capita

Baseando-nos na suposição criada no item anterior, de que países onde as pessoas tenham um maior poder aquisitivo são os mesmos que consomem mais vinho, criamos uma correlação entre PIB per capita e consumo de vinho, mostrada na Fig. 11. Ao contrário do esperado, o gráfico não mostra uma correlação forte entre os itens. Acreditamos que a falta de correlação tenha relação a divergências culturais entre os países. Alguns, como os Estados Unidos, que, apesar de ter o segundo maior PIB dentre os listados, possui consumo de destilado e cerveja cerca de duas e três vezes maior do que o de vinho, respectivamente. Há também, vários países com alto PIB, mas sem dados sobre seu consumo de bebidas alcoólicas, como Bermuda e Ilhas Cayman.

Esse gráfico, apesar de não apresentar uma correlação entre os critérios, nos faz refletir sobre a alta concentração de pontos no canto inferior esquerdo. Há muitos países nos quais o consumo de álcool é zero, pois é tido como ilegal. Aparentemente, a maioria desses países, como o Afeganistão, possui baixo valor PIB. Isso nos faz refletir sobre a relação entre a desenvolvimento econômico e a liberdade individual.

6.3. Top 10 consumo de cerveja

O objetivo dessa consulta foi de fato observar se, dentre os países que apresentam maiores índices de consumo de cerveja, há países com renda mais baixa, por se tratar de uma

bebida de custo mais baixo. A listagem completa da consulta é mostrada abaixo e, de fato, possui uma maior diversidade de países, espalhados por diversos continentes. O que não esperávamos era encontrar, dentre eles, países que desconhecíamos. Um desses, Gabão, possui uma posição maior do que a Alemanha, nação a qual pressupunhamos que estaria em primeiro lugar.

1. Namíbia
2. República Tcheca
3. Gabão
4. Alemanha
5. Polônia
6. Lituânia
7. Venezuela
8. Irlanda
9. Palau
10. Romênia

6.4. Ordenação decrescente de consumo de álcool com filtro "maior que a média"

Como percebemos que muitos dos países possuíam o consumo de álcool próximo de zero, buscamos saber a quantidade de nações que possuem o consumo de álcool total, em litros, acima da média e quais são eles. Obtemos então 55 entradas, sendo que o total de países analisados foi 230. O primeiro item da lista, ou seja, o que possui menor consumo de álcool maior que a média, foi as Bahamas com o total de 51 litros. O Brasil surpreendentemente não está nessa lista, pois possui o total de 7.20 litros de álcool por pessoa.

6.5. Ordenação crescente de consumo de álcool

Para analisarmos os menores índices de consumo de álcool, ordenamos os países em ordem crescente de consumo de álcool. Encontramos 40 que possuem essa métrica menor que um, dentre os quais Afeganistão, Irã, Bangladesh, Maldivas e Kwait, onde o consumo de álcool é proibido por lei [law 2017]. Quando listamos os países que possuem PIB maior que a média, nenhum desses exemplos é apresentado, o que é condizente com o que analisamos no item 6.2.

6.6. Correlação entre qualidade dos chocolates vendidos e dos originados a partir do cacau cultivado

A figura 12 tenta estabelecer uma correlação entre a qualidade dos chocolates compostos por grãos de cacau originados em um dado país com a qualidade do chocolate vendido pelo mesmo. Pensamos que um país que possua grande produção de cacau, como o Brasil, dê origem a chocolates de boa qualidade tanto nacionais como internacionais. Assim, esperávamos que haveria uma concentração de pontos no canto superior direito, o que é verdade. Pensávamos também que países pequenos, mas famosos pelas marcas de chocolate, como França e Suíça, ocupariam o canto superior esquerdo, pois não possuem grandes áreas de cultivo de cacau. Essa suposição também foi tida como certa, conforme mostra a figura. Podemos no entanto, ver uma concentração de pontos no canto inferior direito, indicando países que possuem produção de grãos que levam a bons chocolates, mas que não vendem esse tipo de alimento. Notamos que o gráfico também apresenta menos pontos que os demais, pois nem todos os países são ou produtores de grão de cacau ou possuem empresas produtoras de chocolate.

6.7. Ordenação decrescente por qualidade do chocolate vendido

Para investigar mais detalhes do item anterior, decidimos obter quais são os países que vendem os chocolates de melhor qualidade. Esperávamos que no resultado estivessem presentes países famosos por seus chocolates, como Suíça, Bélgica e Estados Unidos, além de países que sabemos que possuem grande cultivo de grãos de cacau, como Brasil e Costa do Marfim. No entanto, os 10 primeiros listados foram os seguintes:

1. Chile
2. Argentina
3. Polônia
4. Austrália
5. Emirados árabes
6. Suíça
7. Guatemala
8. Dinamarca
9. Vietnã
10. Noruega

O Brasil ocupa o 17º lugar, seguido pela Bélgica. A Costa do Marfim não é nem listada, pois não possui nenhum chocolate nacional avaliado. Os EUA ocupa o 34º lugar. Inicialmente pensamos que esses 10 países listados produziam chocolates de excelência e que não sabíamos disso. Posteriormente vimos, no entanto, que o Chile, primeiro colocado, só possui dois chocolates avaliados, ambos de mesma marca com uma pequena diferença em seus ingredientes. Ambas as barras foram avaliadas com 3.75, o que fez com o que ficasse com a média maior que, por exemplo os EUA, que possui 200 avaliações para os chocolates produzidos em seu país. A Bélgica e a Suíça possuem 50 e 40 avaliações, respectivamente. Conclui-se então que esse ordenamento é um tanto injusto, pois de algumas nações possuem uma gama maior de chocolates avaliados, de diferentes qualidades. Países como o Chile, que só tiveram boas marcas avaliadas, obtiveram vantagem no ranking.

6.8. Ordenação decrescente por qualidade do chocolate originado a partir do grão de cacau de cada país

Para seguirmos identificando mais informações sobre a relação entre a produção de cacau e a venda de chocolate, procuramos saber quais são as nações que produzem os grãos que originam os melhores chocolates. Partimos do pressuposto que países como Brasil, Peru e Costa do Marfim estariam no topo da lista, pois sabíamos que eram grandes cultivadores de cacau. Esses ocuparam os 10º, 24º e 54º (antepenúltimo) lugares, respectivamente. Novamente, a diversidade de entradas é algo que contribuiu para esses resultados. Foram avaliados 200 barras de chocolate cujos grãos foram produzidos no Peru, as quais foram produzidas em diferentes países. No caso do Brasil, 74 barras e da Costa do Marfim apenas 4.

6.9. Correlações entre PIB e Chocolate

Por último, decidimos obter a relação entre o PIB e a qualidade do chocolate vendido e a entre PIB e a qualidade dos chocolates originados dos grãos. Os gráficos são mostrados nas figuras 13 e 14, respectivamente. Embora o primeiro não mostre nenhuma correlação,

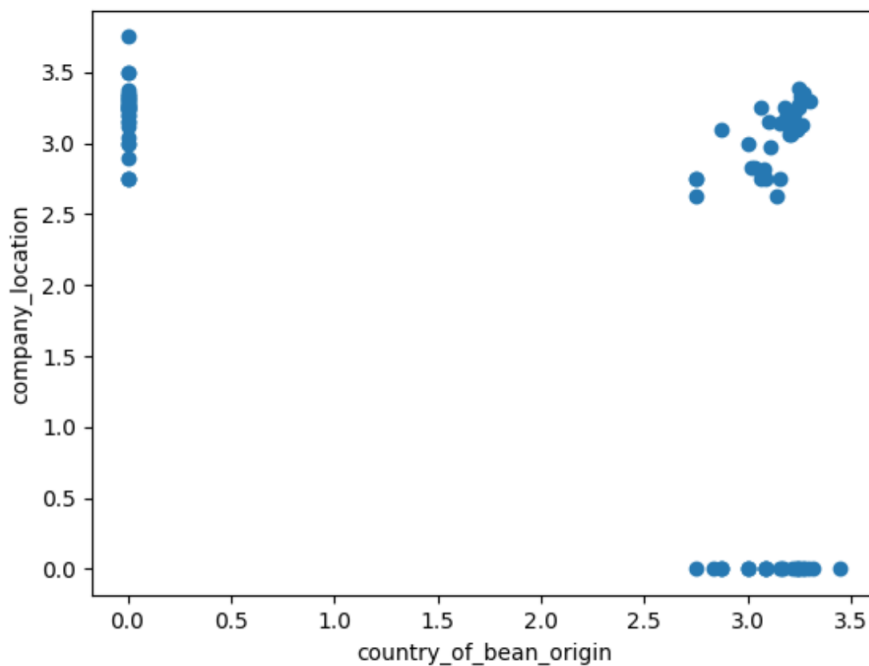


Figura 12. Correlação entre chocolate vendido e grão originado

pois, como já vimos, vários países com baixo PIB, como o Chile, produzem os melhores chocolates, segundo os dados utilizados. O segundo gráfico já é mais interessante, pois mostra que a concentração da produção de grãos de cacau está nos países de menor PIB. Esses grãos dão origem a chocolates de diferentes qualidades. Isso faz sentido para, pois acreditamos que o escore final de uma barra desse alimento depende de outros fatores na produção, além do grão de cacau.

7. Considerações Finais

Ao agregar informações sobre diferentes países e possibilitar a construção de análises de uma forma simples, visamos a estimular um espírito de cidadão global em nossos usuários, assim como em nós mesmos. Acreditamos que quando as pessoas se informam

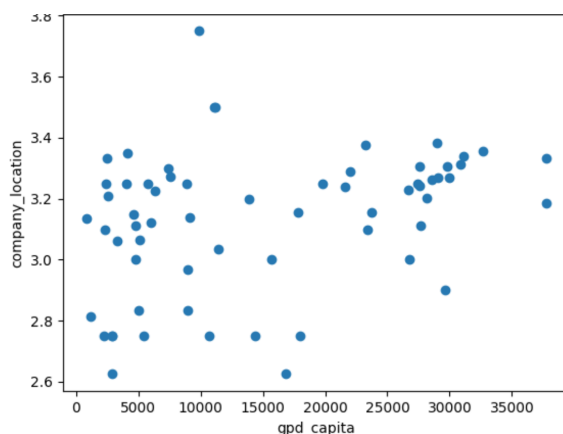


Figura 13. Chocolate vendido e PIB

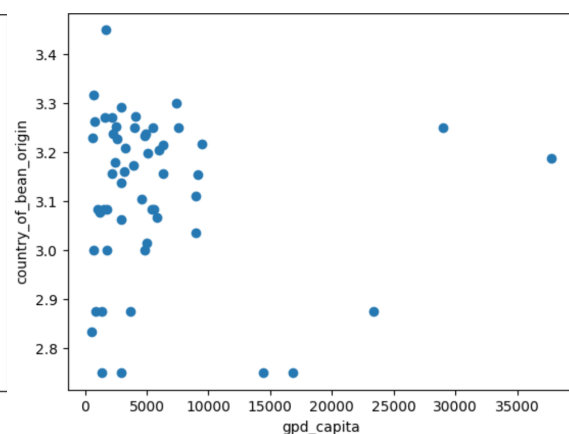


Figura 14. Produção de cacau e PIB

sobre outros países, suas visões de mundo são expandidas, assim como a solidariedade e respeito aos outros, por mais diferentes que sejam. Juntamos algumas informações de temas que pensamos que são mais interessantes ao público, como bebidas alcoólicas e chocolate, assim o trabalho, visa a gerar momentos não só de reflexão, como também de descontração. Tivemos a ideia de construir um website para possibilitar um acesso prático e interativo ao nosso projeto e esperamos que seus usuários tenham um bom momento usando-o, ao mesmo tempo que aprendem e despertam interesse pela cultura global. Esperamos também que esse interesse não acabe no encerramento da aplicação por parte do usuário, mas que a pessoa destine um tempo para refletir sobre as análises, exercite o pensamento interpretativo e se motive a pesquisar mais sobre o ambiente em que vivemos. É analisando as dados que as pessoas desconstroem os estereótipos que têm e constroem um pensamento verdadeiro.

Construindo esse projeto, além de todo o conhecimento técnico que aprendemos, sobre consultas SQL e desenvolvimento de websites, conseguimos atingir nosso objetivo expandir nossos conhecimentos sobre o mundo. As análises, principalmente aquelas que trouxeram resultados muito diferentes do esperado, trouxeram um ganho de conhecimento muito proveitoso, que nos motivou a pesquisar ainda mais conteúdo na internet em busca de explicações para as informações obtidas.

Apesar de satisfeitos com o desenvolvimento do projeto, concluímos que, para análises mais consistentes, devemos acrescentar ao sistema um maior volume de dados. Se conseguirmos, por exemplo, mais avaliações, de diferentes jurados e de diferentes barras de chocolate, poderemos no futuro tirar maiores conclusões sobre o que foi abordado. Sobre o consumo de álcool, seria também interessante dados de outras fontes confiáveis. Um ponto positivo do trabalho como um todo é a facilidade de expansão. Podemos acrescentar novos arquivos fonte e fazer diferentes operações sobre eles sem ter que adaptar a interface do usuário.

Por fim, portanto deixamos o código aberto no Github, para que as outras pessoas possam também executar o programa, conforme as instruções no repositório, e que outras pessoas desenvolvedoras possam aprender com nossa implementação. Além disso, outros desenvolvedores mais experientes poderão trazer sugestões que também nos agregue.

Referências

- (2010). Alcohol consumption around the world: <https://www.kaggle.com/codebreaker619/alcohol-comsumption-around-the-world>.
- (2014). Dear mona followup: Where do people drink the most beer, wine and spirits?: <https://fivethirtyeight.com/features/dear-mona-followup-where-do-people-drink-the-most-beer/-wine-and-spirits/>.
- (2017). 14 countries where alcohol is illegal: <https://www.worldatlas.com/articles/14-countries-where-drinking-alcohol-is-illegal.html>.
- (2017). Countries of the world: <https://www.kaggle.com/fernandol/countries-of-the-world>.

(2020). Chocolate bar 2020: <https://www.kaggle.com/soroushghaderi/chocolate-bar-2020>.

(2021). Db engines: <https://db-engines.com/en/ranking>.

(2021). Github countries of the world: <https://github.com/lauracorssac/countriesOfTheWorld>.

(2021). Mars, nestlé and hershey to face child slavery lawsuit in us: <https://www.theguardian.com/global-development/2021/feb/12/mars-nestle-and-hershey-to-face-landmark-child-slavery-lawsuit-in-us>.

(2021). Postgres sql web page: <https://www.postgresql.org/>.

(2021). Website oficial global citizen.

Todos os links foram visualizados por último em 29.11.2021.