

Apresentação

EMANUELA CAETANO
LEONARDO FABRÍCIO
LUCAS ALCÂNTARA
TALYTTA OLIVEIRA
VITOR BRANDÃO





CONSULTORIA ORIENTADA
POR DADOS

TEMA: MERCADO DE BEBIDAS

ASSUNTO: MERCADO DE VINHOS

MUNDIAL



Principais objetivos

IDEIA

Desenvolver uma empresa com base em nossos princípios para a prestação de consultoria para a transformação de dados em informações.

OBJETIVO

Cumprir os requisitos obrigatórios e desejáveis do trabalho com excelência.

JUSTIFICATIVA

Disponibilidade de dados gratuitos na internet e relevância do assunto para o tema.

Missão

Entregar aos nossos clientes análises objetivas e eficientes sobre a condição do respectivo segmento de mercado

Valores

Ser referência na área de dados como empresa que visa entregar o que é proposto com responsabilidade e qualidade

Propósito

Integridade, honestidade, responsabilidade, compromisso com os clientes, diversidade e inclusão, aprendizado, trabalho em equipe e qualidade

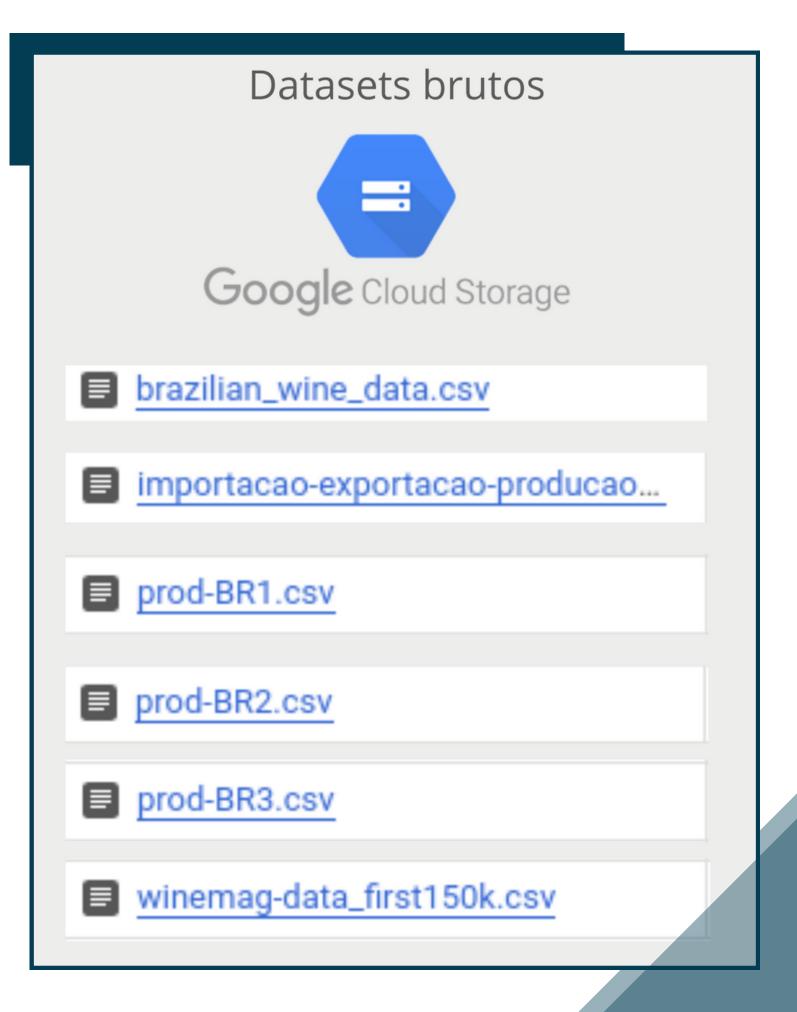
Datasets iniciais

Fontes:

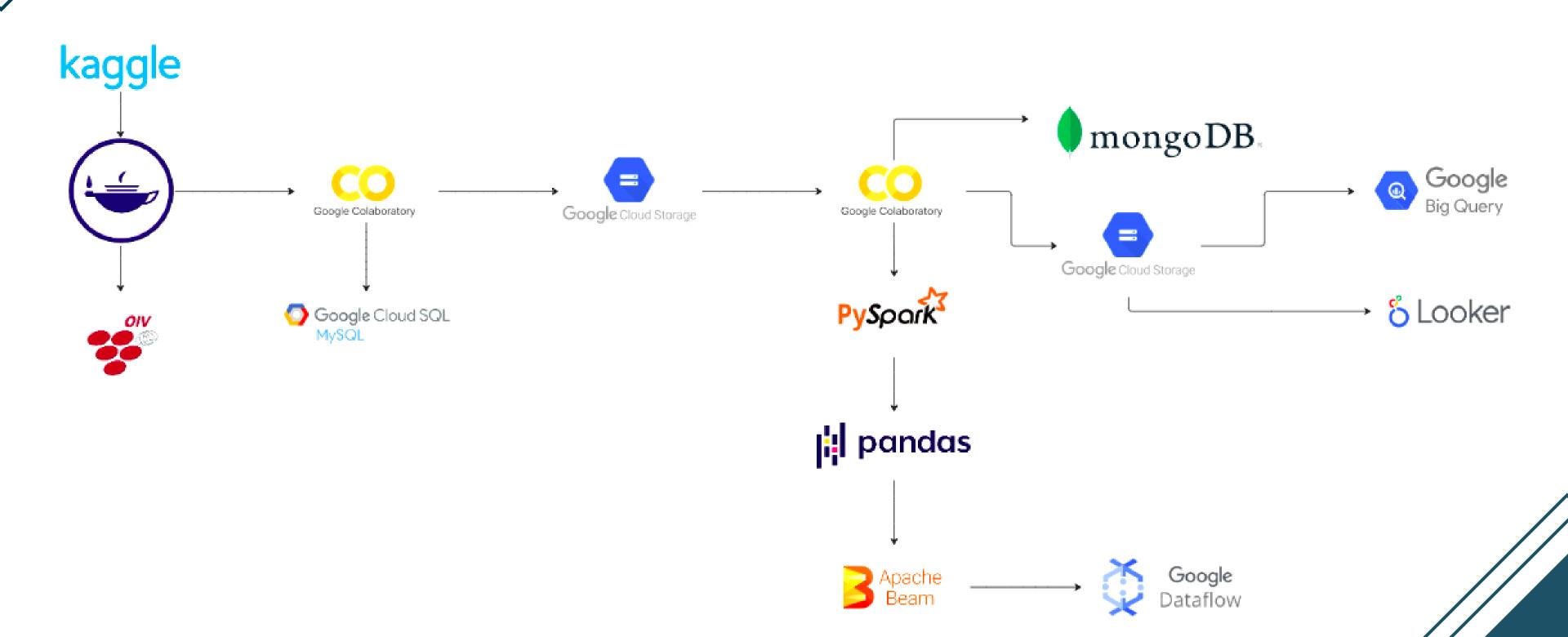
Organização Internacional da Vinha e do Vinho

Kaggle

Taylor & Francis Online



WorkFlow



Conexão MySQL

```
Conexão
     1 #FUNÇÃO PARA CONECTAR AO BANCO
     2 def conexao db(servidor, usuario, senha, nome do banco):
         conexao = None
         try:
           conexao = mysql.connector.connect(host=servidor, user=usuario, passwd=senha, database=nome_do_banco)
           print(f'Conexão ao banco {nome_do_banco} efetuada com sucesso')
         except:
           print(f'Erro ao conectar ao banco {nome_do_banco}')
         return conexao
     1 con = conexao_db('34.162.75.223', 'root', 'root', 'mercado_vinho')
     1 pd.read_sql('show databases',con)
```

Envio MySQL

Realizar conexão com o servidor | criar uma SQLAlch:emy engine para conectar com o MySQL | convertendo os dataframes em uma table do sql

Database MySQL

```
mysql> show databases;
  Database
  Z README TO RECOVER
  information schema
 mercado vinho
 mysql
 performance schema
  373
6 rows in set (0.02 sec)
```

Tabela MySQL

```
mysql> use mercado vinho;
Database changed
mysql> show tables;
 Tables_in_mercado_vinho
 avaliacoes-vinhos-global
 dados vinho brasil
 prod-vinho-BR1
| prod-vinho-BR2
prod-vinho-BR3
| producao_consumo_importacao_exportacao_uvas_vinhos_global
6 rows in set (0.03 sec)
```

Pesquisa MySQL

```
mvsgl> select *
    -> from producao consumo importacao exportacao uvas vinhos global
    -> where Product = 'wine' and Variable = 'Consumption' and Year='2020';
  index | Continent | Region/Country
                                                                  Product
                                                                             Variable
                                                                                           Year |
                                                                                                   Unit
                                                                                                               Quantity
  34451 |
          Asia
                       Afghanistan
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                       Albania
                                                                  Wine
                                                                             Consumption |
                                                                                                   1000 hl
                                                                                                                    205
  34557
          Europe
                                                                                            2020
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                    455
  34666
          Africa
                       Algeria
                                                                  Wine
  34724
                       American Samoa
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
          Oceania
                                                                                                                      0
  34803
          Africa
                       Angola
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                    317
                                                                  Wine
  34884
          America
                       Antigua and Barbuda
                                                                  Wine
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                     15
                                                                             Consumption |
  34990
                       Argentina
                                                                  Wine
                                                                                            2020
          America
                                                                             Consumption |
                                                                                                   1000 hl
                                                                                                                   9430
  35099
                                                                  Wine
                                                                                            2020
          Asia
                                                                             Consumption |
                                                                                                   1000 hl
                                                                                                                    105
                       Armenia
  35261
                       Australia
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                                   5950
          Oceania
                                                                                                   1000 hl
  35370
                       Austria
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                   2306
          Europe
  35479
          Asia
                       Azerbaijan
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                    108
  35562
                       Bahamas
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                     37
          America
  35643
          Asia
                       Bahrain
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                     15
  35699
          Asia
                       Bangladesh
                                                                  Wine
                                                                             Consumption |
                                                                                            2020
                                                                                                   1000 hl
                                                                                                                      1
                                                                                            2020
                                                                                                                     26
  35778
          America
                       Barbados
                                                                  Wine
                                                                             Consumption |
                                                                                                   1000 hl
                                                                                                                    982
  35884
                       Belarus
                                                                  Wine
                                                                                            2020
          Europe
                                                                             Consumption |
                                                                                                   1000 hl
```

Removendo colunas irrelevantes

```
[ ] 1 df1.drop(['id', 'type_id', 'vintage_type', 'region_id', 'region_country_name', 'region_country_most_used_grapes_0_has_detailed_info'\n
2 'winery_id', 'taste_flavor_0_stats_mentions_count', 'taste_flavor_0_primary_keywords_0_id', 'taste_flavor_0_primary_keywords_0_count'\n
3 'taste_flavor_0_primary_keywords_1_id', 'taste_flavor_0_primary_keywords_2_count', 'taste_flavor_0_stats_count'], axis=1, inplace=True)
```

Criando um dicionário para renomear as colunas e executando a função para renomear as colunas

Traduzindo os valores da coluna continente

```
[ ] 1 df2['continente'] = df2['continente'].replace({'Asia':'Asia', 'Europe':'Europa', 'Africa':'Africa', 'America':'Amer
```

Traduzindo os valores da coluna regiao_pais

```
1 df2['regiao_pais'] = df2['regiao_pais'].replace(
       {'Afghanistan': 'Afeganistão', 'Algeria': 'Argélia', 'Antigua and Barbuda': 'Antígua e Barbuda', 'Azerbaijan': 'Azerbaijão',
        'Belarus':'Bielorrússia', 'Belgium':'Bélgica', 'Belgium and Luxemburg':'Bélgica e Luxemburgo', 'Bhutan':'Butão',
        'Bosnia and Herzegovina':'Bósnia e Herzegovina', 'Botswana':'Botsuana', 'Brazil':'Brasil', 'Brunei Darussalam':'Brunei',
        'Cambodia':'Camboja', 'Cameroon':'Camarões', 'Cayman Islands':'Ilhas Cayman',
        'Central African Republic': 'República Centro-Africana', 'Chad': 'Chade', 'Comoros': 'Comores', 'Cook Islands': 'Ilhas Cook',
 6
        "Côte d'Ivoire": 'Costa do Marfim', 'Croatia': 'Croácia', 'Cyprus': 'Chipre', 'Czech Republic': 'Tchéquia',
        'Denmark':'Dinamarca', 'Dominican Republic':'República Dominicana', 'Ecuador':'Equador', 'Egypt':'Egito',
 8
        'Eswatini':'Suazilandia', 'Ethiopia':'Etiopia', 'Faroe Islands':'Ilhas Faroe', 'Finland':'Finlandia', 'France':'França',
 9
        'French Guiana': 'Guiana Francesa', 'French Polynesia': 'Polinesia Francesa', 'Gabon': 'Gabão', 'Germany': 'Alemanha',
10
        'Ghana':'Gana', 'Greece':'Grécia', 'Greenland':'Groelândia', 'Grenada':'Granada', 'Guadeloupe':'Guadalupe','Guinea':'Guiné',
11
        'Guinea-Bissau':'Guiné-Bissau', 'Guyana':'Guiana', 'Hungary':'Hungria', 'Iceland':'Islândia', 'Iran':'Irã', 'Iraq':'Iraque',
12
        'Ireland':'Irlanda', 'Italy':'Itália', 'Japan':'Japão', 'Jordan':'Jordânia', 'Kazakhstan':'Cazaquistão', 'Kenya':'Quênia',
13
        'Kiribati':'Quiribati', 'Korea (the Republic of)':'Coreia do Sul', 'Kyrgyzstan':'Quirguistão',
14
        "Lao People's Democratic Republic": Laos', 'Latvia': Letonia', 'Lebanon': Libano', 'Lesotho': Lesoto', 'Libya': Libia',
15
        'Lithuania':'Lituânia', 'Luxemburg':'Luxemburgo','Macao':'Macau', 'Malaysia':'Malásia', 'Maldives':'Maldivas',
16
        'Martinique':'Martinica', 'Mauritius':'Ilhas Maurício','Moldova':'Moldávia', 'Morocco':'Marrocos', 'Mozambique':'Moçambique',
17
        'Netherlands': 'Países Baixos', 'Netherlands Antilles': 'Antilhas Neerlandesas', 'New Caledonia': 'Nova Caledônia',
18
        'New Zealand':'Nova Zelândia', 'Niger':'Nigéria', 'Norway':'Noruega', 'Oman':'Omã', 'Pakistan':'Paquistão',
19
        'Papua New Guinea': 'Papua-Nova Guiné', 'Paraguay': 'Paraguai', 'Philippines': 'Filipinas', 'Poland': 'Polônia',
20
        'Qatar':'Catar', 'Republic of North Macedonia':'Macedônia do Norte','Republic of Türkiye':'Turquia',
21
```

Analisando o conteúdo da coluna produto

Traduzindo os valores da coluna produto

```
1 df2['produto'] = df2['produto'].replace({
2    'Dried Grapes':'Uvas Secas', 'Table Grapes':'Uvas de Mesa',
3    'Fresh Grapes':'Uvas Frescas', 'Vineyard':'Vinhas',
4    'Wine':'Vinho'
5 })
```

Visualização inicial do df4

U	1	ат4.пе	ead()					
		Year	Imports amounta	Exports amounta	Imports-exports amounta	Imports valueb	Exports valueb	Import-exports value
	0	1990	8.0	3.4	4.6	15.3	3.6	11.
	1	1991	8.0	4.3	3.7	14.1	4.2	9.

, .	. u. j	()				
	Year	RS total (in tons)	% of Brazilian production	PE total (in tons)	% of Brazilian production.1	Brazil total production
0	1990	538,705	66.94	14,483	1.80	804,774
1	1991	396,318	61.16	17,163	2.65	648,026

Fazendo o merge dos dfs 4 e 5

```
[ ] 1 merged_df = pd.merge(df4, df5, on='Year')
```

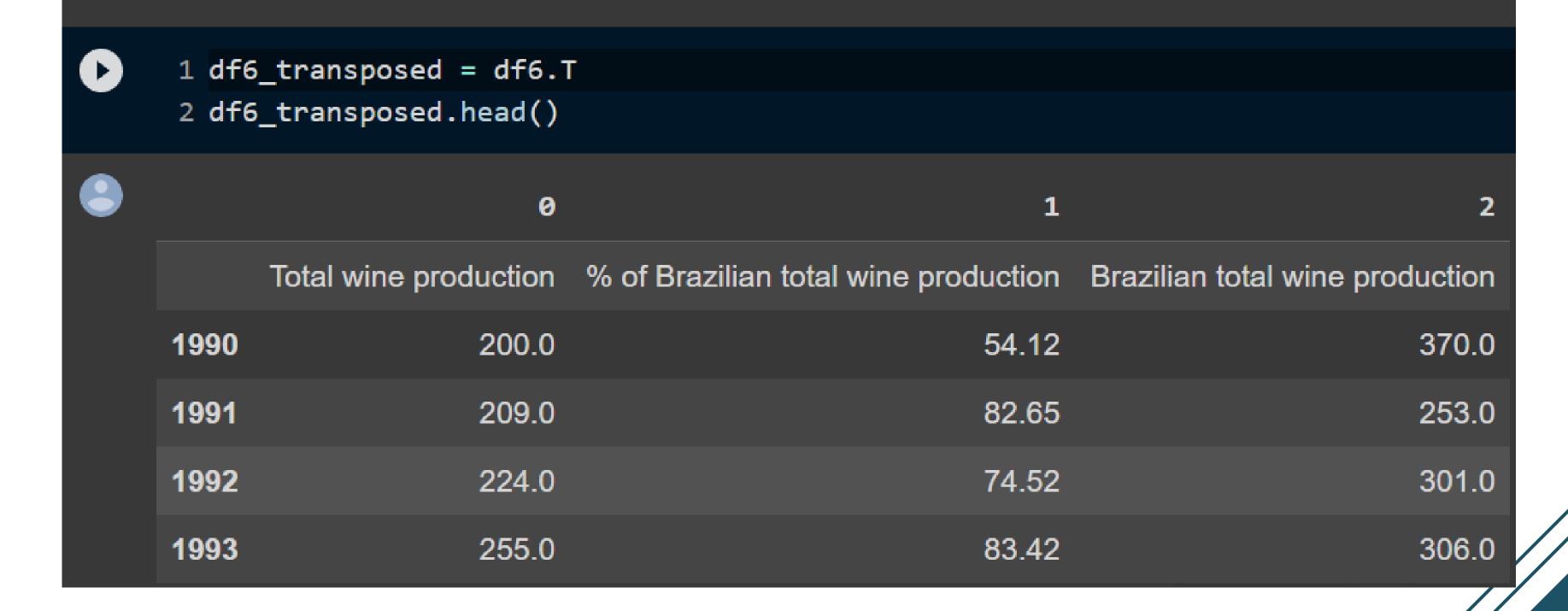
Analisando o novo dataframe

•	<pre>1 merged_df.head()</pre>			

•		Year	Imports amounta	Exports amounta	Imports- exports amounta	Imports valueb	Exports valueb	Import- exports valueb	RS total (in tons)	% of Brazilian production	PE total (in tons)	% of Brazilian production.1	Brazil total production
	0	1990	8.0	3.4	4.6	15.3	3.6	11.7	538,705	66.94	14,483	1.80	804,774
	1	1991	8.0	4.3	3.7	14.1	4.2	9.9	396,318	61.16	17,163	2.65	648,026
	2	1992	6.0	7.4	-1.4	10.2	7.5	2.7	505,462	63.17	18,510	2.31	800,112
	3	1993	12.0	20.2	-8.2	18.8	14.8	4.0	489,464	62.16	26,475	3.36	787,363
	4	1994	21.4	14.8	6.6	35.0	12.7	22.3	479,034	59.32	30,821	3.82	807,520



Invertendo as colunas pelas linhas no df6 e checagem da operação



Convertendo o tipo da coluna "Year" de string para int para que a junção com o df final possa ser bem sucedida

- 1 df6_transposed['index'] = df6_transposed['index'].astype(str).astype(int)

 1 df6_transposed.rename(columns={'index':'Year'}, inplace=True)

 1 df6_transposed.head()

 Year Total wine production % of Brazilian total wine production Brazilian total wine production
 - **0** 1990 200.0 54.12 370.0 1991 209.0 82.65 253.0 **2** 1992 224.0 74.52 301.0 1993 255.0 83.42 306.0 274.0 349.0 **4** 1994 78.41

Fazendo o merge do df6 com o df novo(junção dos dfs 4 e 5)

[] 1 merged_df = pd.merge(merged_df, df6_transposed, on='Year')

<u> </u>															
0	1 merged	_df.head	()												
•	Year	•	Exports amounta	Imports- exports	Imports valueb	Exports valueb	Import- exports	RS total (in	% of Brazilian	PE total (in	% of Brazilian	Brazil total	Total wine	% of Brazilian total wine	Brazilian total wine

ction total wine	total wine production
200.0 54.12	370.0
209.0 82.65	253.0
224.0 74.52	301.0
255.0 83.42	306.0
274.0 78.41	349.0
	224.0 74.52 255.0 83.42

Criando um dicionário para traduzir as colunas e executando a função para traduzir as colunas

1 merged_df.rename(columns=dict_rename_merged, inplace=True)

Veri	Verificando se as colunas foram traduzidas													
[21]	[21] 1 merged_df.head()													
		ano	qtd_litros_importacao(milhoes)	qtd_litros_exportacao(milhoes)	valor_em_dolar_importacao(milhoes)	valor_em_dolar_exportacao(milhoes)	Producao_uvas_RS(toneladas)							
	0	1990	8.0	3.4	15.3	3.6	538,705							
	1	1991	8.0	4.3	14.1	4.2	396,318							
	2	1992	6.0	7.4	10.2	7.5	505,462							
	3	1993	12.0	20.2	18.8	14.8	489,464							
	4	1994	21.4	14.8	35.0	12.7	479,034							

Traduzindo os valores de inglês para português da coluna grupo_de_sabores

```
1 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'non_oak', 'sem_barrica'))
2 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'citrus_fruit', 'frutas_citricas'))
3 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'earth', 'terra'))
4 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'tree_fruit', 'arvore_frutifera'))
5 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'oak', 'barricado'))
6 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'tropical_fruit', 'frutas_vermelhas'))
7 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'tropical_fruit', 'frutas_tropicais'))
8 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'microbio'))
9 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'black_fruit', 'frutas_negras'))
10 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'spices', 'especiarias'))
11 sparkDF1 = sparkDF1.withColumn('grupo_de_sabores', F.regexp_replace('grupo_de_sabores', 'dried_fruit', 'frutas_secas'))
```

Alteração do tipo da coluna "natural" de Boolean para String para realizarmos a renomeação dos valores

```
1 sparkDF1.withColumn("natural", col("natural").cast(StringType()))
Verificando se há valores únicos dentro da coluna natural
      1 sparkDF1.select('natural').distinct().show()
     natural
        true
       false
Alteração dos valores da coluna natural
```

1 sparkDF1 = sparkDF1.withColumn('natural', F.regexp_replace('natural', 'false', 'Não'))

2 sparkDF1 = sparkDF1.withColumn('natural', F.regexp_replace('natural', 'true', 'Sim'))

```
Verificando se há valores únicos dentro da coluna "grupo_de_sabores"
     1 sparkDF1.select('grupo_de_sabores').distinct().show(100, truncate=False)
     grupo_de_sabores
     non oak
     citrus_fruit
     earth
     oak
     tree fruit
     red_fruit
     NaN
     tropical fruit
     microbio
     black_fruit
     spices
     floral
     vegetal
     dried fruit
```

```
Removendo vírgulas do dataframe

[ ] 1 sparkDF_merged = sparkDF_merged.withColumn("Producao_uvas_RS(toneladas)", F.regexp_replace('Producao_uvas_RS(toneladas)', ',', ''))

[ ] 1 sparkDF_merged = sparkDF_merged.withColumn("Producao_uvas_PE(toneladas)", F.regexp_replace('Producao_uvas_PE(toneladas)', ',', ''))

[ ] 1 sparkDF_merged = sparkDF_merged.withColumn("Producao_de_uvas_brasil", F.regexp_replace('Producao_de_uvas_brasil', ',', ''))
```

Pipeline

Criação de um modelo para a GCP

```
1 pipeline_options = {
       'project': 'projeto-final-373619',
       'runner':'DataFlowRunner',
      'region':'us-east5',
       'staging_location':'gs://datasets-tvlle/staging/',
       'temp_location':'gs://datasets-tvlle/temp/',
       'template location': 'gs://datasets-tvlle/models/modelo batch'
 8 }
10 serviceAccount = '/content/projeto-final-373619-f6460a3b65bd.json'
11 os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = serviceAccount
12
13
14 pipeline options = PipelineOptions.from dictionary(pipeline options)
15 p1 = beam.Pipeline(options=pipeline options)
16
17 vinho = (
18
      p1
       |'Extrair do CSV'>> beam.io.ReadFromText('gs://datasets-tvlle/tratados/producao consumo importacao exportacao uvas vinhos trat.csv',skip header lines=1)
19
       'Separador de dados'>> beam.Map(lambda record: record.split(','))
20
       |'Filtro vinho'>> beam.Filter(lambda record: record[3] == 'Vinho')
21
       |'Load final'>> beam.io.WriteToText('gs://datasets-tvlle/tratados/vinhotr',file_name_suffix='.csv')
22
23)
24
25 p1.run()
```

Overview Real Time Metrics Collections Search Profiler Performance Advisor Online Archive Cmd Line Tools

DATABASES: 1 COLLECTIONS: 4

+ Create Database

Q Search Namespaces

db-MercadoBebidas

avaliacoes_vinhos_gl...

producao_consumo_i...
producao_vinhos_br_tr...
vinhos_brasil_tratado

db-MercadoBebidas.avaliacoes_vinhos_global_trat

STORAGE SIZE: 15.16MB LOGICAL DATA SIZE: 33.65MB TOTAL DOCUMENTS: 150930 INDEXES TOTAL SIZE: 5.09MB

Find Indexes Schema Anti-Patterns (1) Aggregation Search Indexes •

FILTER { field: 'value' }

QUERY RESULTS: 1-20 OF MANY

_id: ObjectId('63bd805d5ccf64cd4a836f69')
index: 0
pais: "Estados Unidos"
designacao: "Martha's Vineyard"
pontos: 96
preco: 235
provincia: "California"
regiao1: "Napa Valley"
regiao2: "Napa"
variedade: "Cabernet Sauvignon"
adega: "Heitz"

_id: ObjectId('63bd805d5ccf64cd4a836f6e')
index: 5

Datasets finais

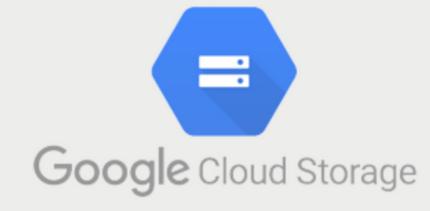
Avaliações de vinhos no mercado mundial;

Consumo, produção, importação e exportação de uvas e vinhos no mercado mundial;

Produção de vinhos no Brasil;

Avaliações de vinhos no mercado brasileiro;

Datasets tratados



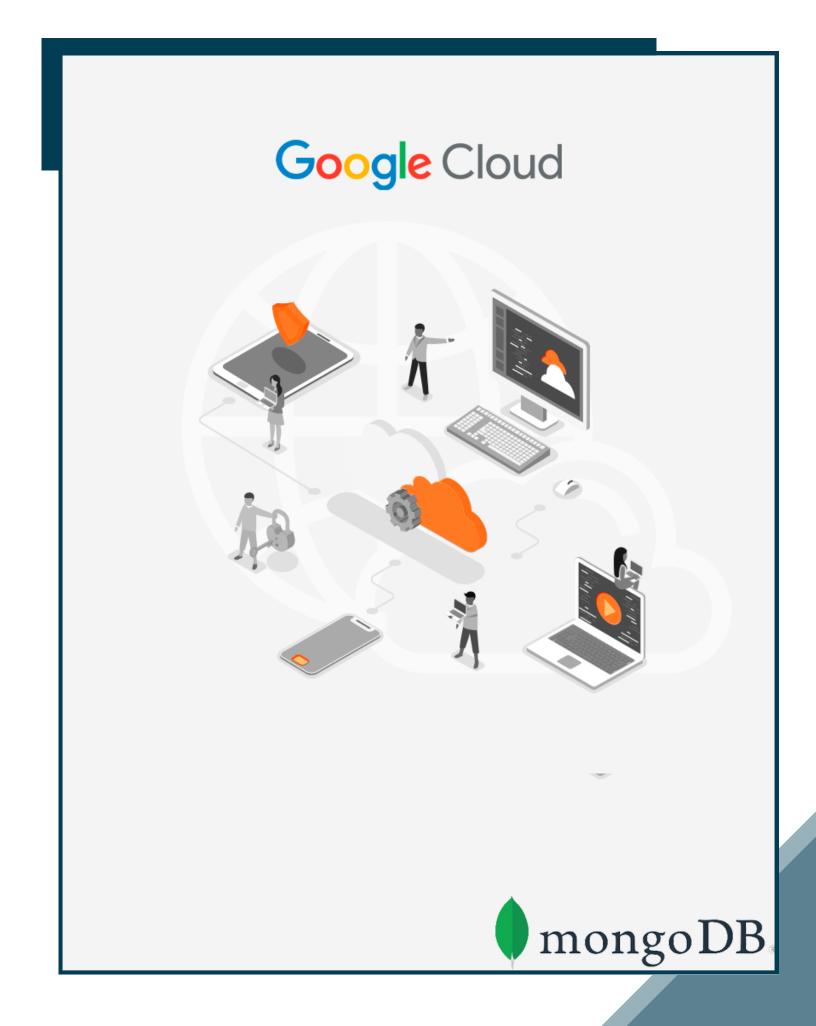
- avaliacoes_vinhos_global_trat.csv
- producao_consumo_importacao_expo
- producao_vinhos_brasil_trat.csv
- vinhos_brasil_tratado.csv

Armazenamento

Storage da GCP (Bucket)

MongoDB

BigQuery



BigQuery & LookerStudio

OBJETIVO

• Representação dos dados

OVITOM

• Otimização

Conclusão



Viabilidade, competitivo

Alta demanda do mercado



Grande mercado nacional e mundial

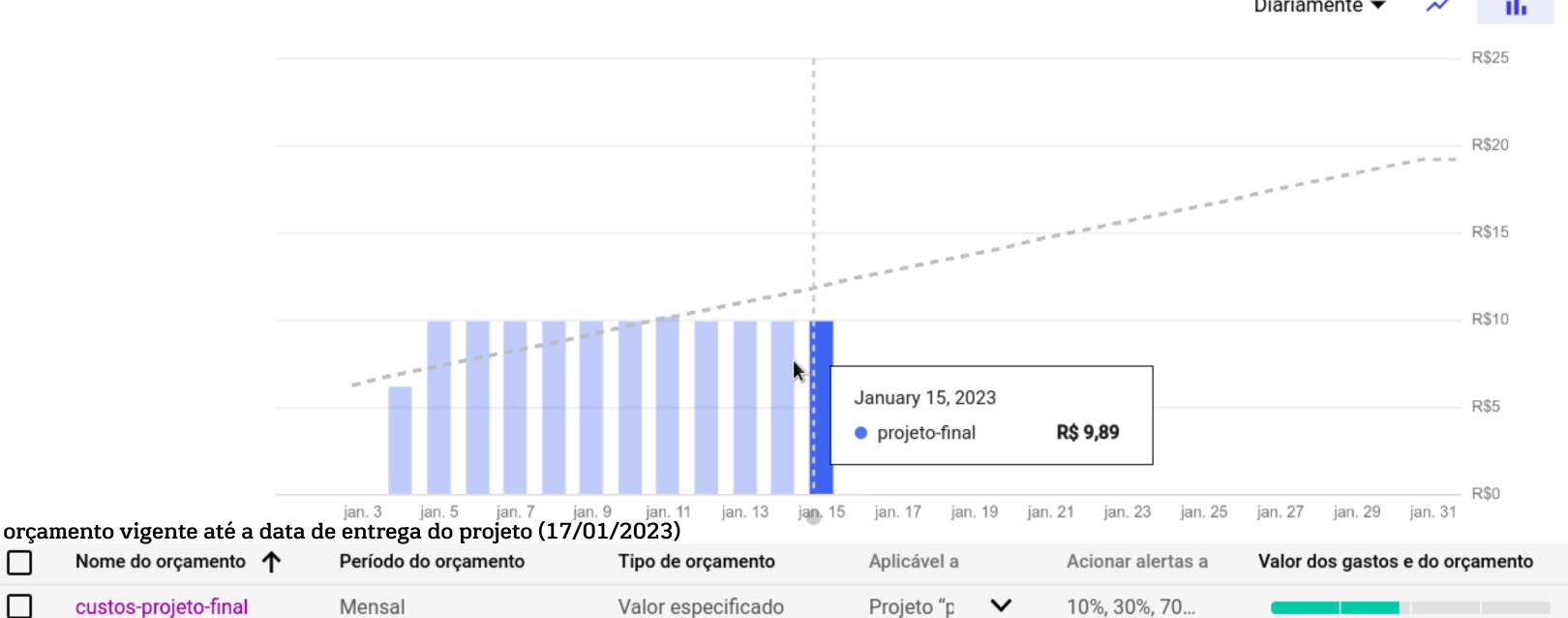


Crescimento

Investimento

Orçamento





R\$ 115,12/R\$ 250,00

Não inclui -R\$ 115,12 de crédito

