

Desafio



Laércio Serra

Data Engineer

laercio.serra@gmail.com

Agenda

- Plano de Trabalho
- Os desafios e os resultados
- Pontos fortes e fracos
- Encerramento

Plano de Trabalho



Task Timer

Use this tool to keep track of time that you spend on work, studies, or any other type of task.

Double-click a task name, time, or goal to edit it.

Task	Time Spent	Goal	Progress	Options
Analising the questions	0:30:00	0:30:00	<div><div></div></div>	▶ i 🔄 -
Planning the tasks	0:15:00	0:15:00	<div><div></div></div>	▶ i 🔄 -
Ex4 - Creating the tables	0:01:08	0:02:00	<div><div></div></div>	▶ i 🔄 -
Ex4 - Populating the tables	0:02:00	0:02:00	<div><div></div></div>	▶ i 🔄 -
Ex4 - Solving the problem	0:04:47	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex3 - Creating the tables	0:01:55	0:02:00	<div><div></div></div>	▶ i 🔄 -
Ex3 - Populating the tables	0:03:00	0:02:00	<div><div></div></div>	▶ i 🔄 -
Ex3 - Solving the problem	0:04:18	0:02:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Creating the table	0:04:03	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Creating a code to do a webcrawler	0:07:04	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Testing the webcrawler code	0:01:51	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Executing the webcrawler code	0:03:53	0:15:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Creating a code to merge the files	0:05:24	0:10:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Testing the merging code	0:00:31	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Merging the files	0:00:33	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Creating a code to import the merged file to Postgres	0:06:18	0:10:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Importing the merged file to Postgres	0:02:37	0:10:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Creating analysis and graphs	0:30:00	0:30:00	<div><div></div></div>	▶ i 🔄 -
Ex1 - Exporting the results from analysis	0:05:00	0:05:00	<div><div></div></div>	▶ i 🔄 -
Ex2 - Defining the tables, fields and relationships	0:30:00	0:30:00	<div><div></div></div>	▶ i 🔄 -
Ex2 - Designing the logical model (transational)	0:15:01	0:15:00	<div><div></div></div>	▶ i 🔄 -
Ex2 - Defining the star-schema	0:30:00	0:30:00	<div><div></div></div>	▶ i 🔄 -
Ex2 - Designing the logical model (dimensional)	0:15:00	0:15:00	<div><div></div></div>	▶ i 🔄 -
Totals	3:39:23	4:15:00	<div><div></div></div>	▶ i 🔄 -

Os desafios e os resultados

```

-- Exercise 4
-- Before Last Transaction By Salesperson
select a.salesperson_id, max(a.date) as date
from
(
  select salesperson_id, max(date) as date
  from transactions
  group by salesperson_id
) as b
where
(
  a.salesperson_id = b.salesperson_id and
  a.date < b.date
)
group by a.salesperson_id
order by a.salesperson_id
  
```

Task Timer

Use this tool to keep track of time that you spend on work, studies, or any other type of task.

Double-click a task name, time, or goal to edit it.

Task	Time Spent	Goal	Progress
Analising the questions	0:30:00	0:30:00	<div></div>
Planning the tasks	0:15:00	0:15:00	<div></div>
Ex4 - Creating the tables	0:01:08	0:02:00	<div></div>
Ex4 - Populating the tables	0:02:00	0:02:00	<div></div>
Ex4 - Solving the problem	0:04:47	0:05:00	<div></div>

```

INSERT 0 1
postgres=# INSERT INTO emails (id, contact_id, email, email_type) Values('5', '3', 'mt.chael.jones@hotmail.com', 'office');
INSERT 0 1
postgres=# INSERT INTO emails (id, contact_id, email, email_type) Values('6', '4', 'ra.lf.schmid@amazon.com', 'office');
INSERT 0 1
postgres=# INSERT INTO emails (id, contact_id, email, email_type) Values('7', '4', 'ra.lf.schmid@rocket.com', 'office');
INSERT 0 1
postgres=# INSERT INTO emails (id, contact_id, email, email_type) Values('8', '4', 'ra.lf.schmid@zaland.com', 'office');
postgres=# INSERT INTO emails (id, contact_id, email, email_type) Values('9', '5', 'l.jackson@googlemail.com', 'private');
postgres=# select * from emails;
 id | contact_id | email | email_type
-----+-----+-----+-----
  5 |          3 | mt.chael.jones@hotmail.com | office
  6 |          4 | ra.lf.schmid@amazon.com | office
  7 |          4 | ra.lf.schmid@rocket.com | office
  8 |          4 | ra.lf.schmid@zaland.com | office
  9 |          5 | l.jackson@googlemail.com | private
  
```

Qty Accidents in Brazil - Top Cities

This exercise requires all necessary code to be written in any programming language (Preferably Python or Java). The goal is to retrieve the metric of municipal work related accidents provided by the data based on the Public Brazilian Social Security API (<http://api.dataprev.gov.br/doc/index.html>). As a first step, retrieve data for all years saving it in a .csv file and then transform and analyze the data to get the top 3 Brazilian cities on the dataset, which grew the metric the most between 2003 and 2007, by geopolitical region.

Importando as bibliotecas que serão usadas. Neste notebook estou usando o Python3 e as seguintes bibliotecas: pandas e matplotlib.

```

In [1]: import pandas as pd
import matplotlib
import matplotlib.pyplot as plt

%matplotlib inline
matplotlib.style.use('ggplot')
  
```

Definição das variáveis de localização do arquivo (caminho) e nome-do-arquivo (CSV) que contem os dados que iremos analisar.

```

In [2]: path_acc = '/home/df/ContaAzul/Ex1/data/accidents_brazil.csv'
path_reg = '/home/df/ContaAzul/Ex1/data/regions_brazil.csv'
  
```

Criação de um dataframe contendo os dados de acidentes de todo o Brasil entre o periodo de 2003 a 2007. Estes dados são carregados a partir do arquivo CSV.

```

In [3]: u_cols = ['ano', 'mun_cod_ibge', 'mun_nome', 'uf', 'qtd_acid_c_cat_tipicos', 'qtd_acid_c_cat_trajeto',
'qtd_acid_c_cat_doenca_profiss', 'qtd_obitos', 'qtd_acid_s_cat']
df_acc_brz = pd.read_csv(path_acc, names=u_cols, encoding='latin-1', skiprows=1)
  
```

```

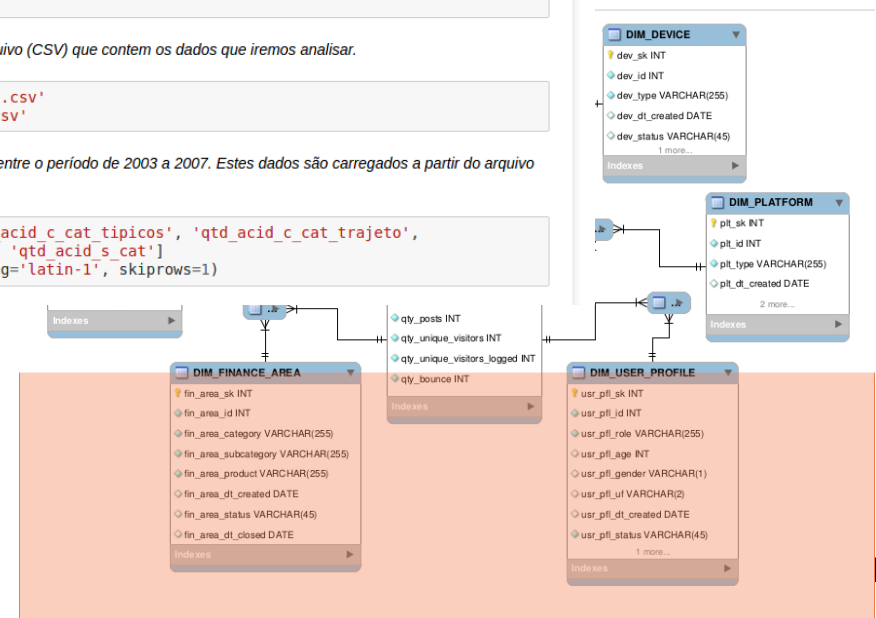
# Exercise 1
import pandas as pd
from sqlalchemy import create_engine

path = '/home/df/ContaAzul/Ex1/data/accidents_brazil.csv'
engine = create_engine('pd')
df = pd.read_csv(path + '.csv')
df.to_sql('conta_azul', engine, if_exists='replace')

print(">> Has been imported")
print(">> * 100")
print(str(df.count()))
print(">> * 100")
print(">> Importing concl")

>> Has been imported:
ano      23745
Municipio_COD-IBGE 23745
Municipio_Nome    23745
UF                23745
Quantidade_Acidentes_com_Cat_Tipicos 23745
Quantidade_Acidentes_com_Cat_Trajeto 23745
Quantidade_Acidentes_com_Cat_Doenca_Profissional 23745
Quantidade_Obitos 23745
Quantidade_Acidentes_Sem_CAT 23745
dtypes: int64 1
>> Importing concluded successfully!
(venv) dfdatafresh:~/ContaAzul/Ex1$
  
```

Table	Column	Type	Modifiers
conta_azul	ano	bigint	
conta_azul	Municipio_COD-IBGE	bigint	
conta_azul	Municipio_Nome	text	
conta_azul	UF	text	
conta_azul	Quantidade_Acidentes_com_Cat_Tipicos	bigint	
conta_azul	Quantidade_Acidentes_com_Cat_Trajeto	bigint	
conta_azul	Quantidade_Acidentes_com_Cat_Doenca_Profissional	bigint	
conta_azul	Quantidade_Obitos	bigint	
conta_azul	Quantidade_Acidentes_Sem_CAT	bigint	
conta_azul	dtypes	int64	



Pontos fortes e fracos



Encerramento

[Manifesto](#) [História](#) [Notícias](#) [Imprensa](#) [Investidores](#) [Nosso dia a dia](#) [Contato](#)

Manifesto

Gerenciar uma empresa é uma tarefa honrosa e de mérito supremo. Aqui na ContaAzul acreditamos que a tarefa de gerenciar o financeiro, estoque ou mesmo os contatos de uma empresa deve ser fácil, rápido e até mesmo automático.

Sabemos que o futuro está nas micro e pequenas empresas e faz parte de nossa luta viabilizar que toda a gestão seja simples e eficiente. Simples o suficiente para que o empreendedor utilize seu tempo em expandir o seu network, fazer bons negócios e crescer em progressões geométricas. Eficiente o suficiente para que as decisões de negócio sejam maximizadas pela informação organizada e de qualidade.

Pensamos também que toda empresa merece deter ferramentas para controlar o presente e prever o futuro com o mínimo esforço, mantendo um padrão organizacional digno de uma multi-nacional.

Este é o nosso sonho e podemos dizer com orgulho que cada funcionalidade em que trabalhamos foi, é e será pensada para que esta seja a realidade de toda micro e pequena empresa Brasileira.

