



# ATIVIDADE INDIVIDUAL

Carlos Bahia

# PROJETO INDIVIDUAL

1. Nível Infra
2. Nível Pandas
3. Nível PySpark
4. Nível Postgres
5. Nível DataStudio



# PROJETO INDIVIDUAL

## 1.Nível Infra

### 1.1 O Dataset deve ser salvo em ambiente cloud (Cloud Storage)

BC17

Search Products, resources, doc

←

Bucket details

atividade-individual

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Buckets > atividade-individual

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

Filter by name prefix only

Filter

Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created	Storage
<input type="checkbox"/>	original/	—	Folder	—	—
<input type="checkbox"/>	tratado/	—	Folder	—	—

atividade-individual

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Buckets > atividade-individual > original

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

DELETE

Filter by name prefix only

Filter

Filter objects and folders

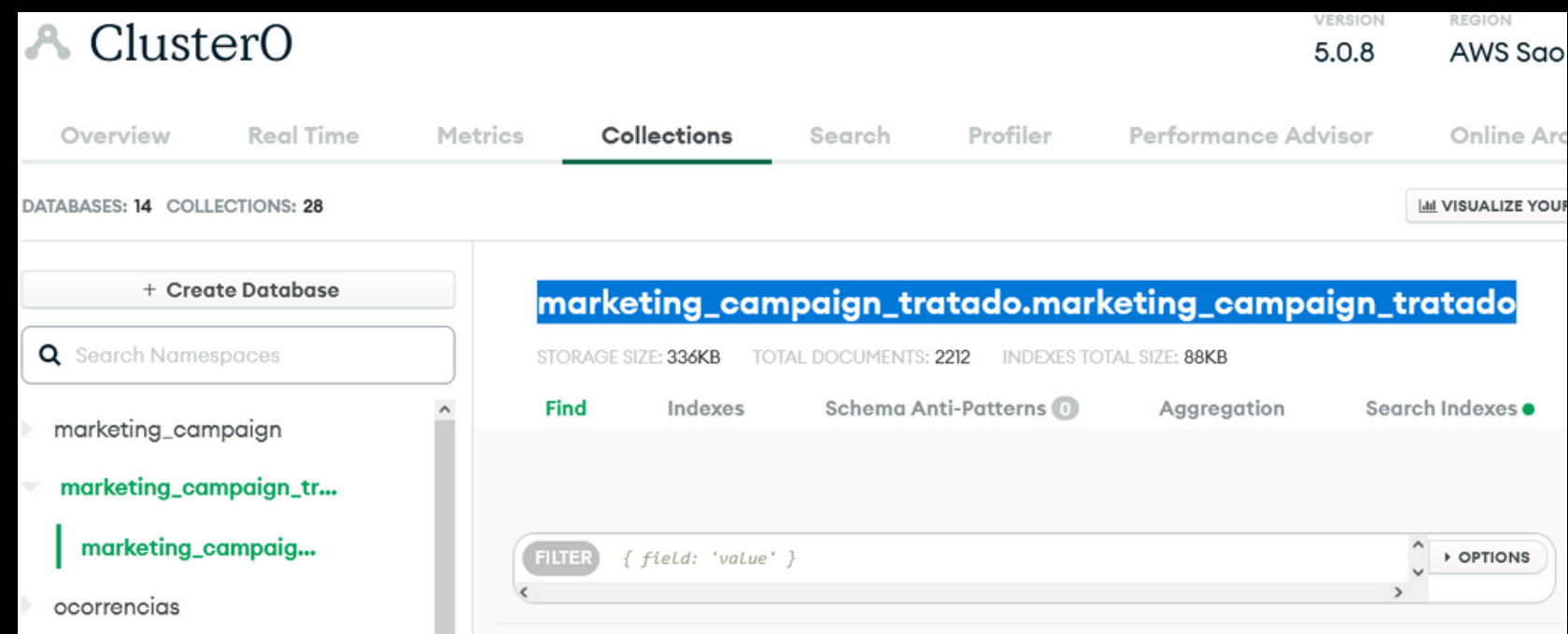
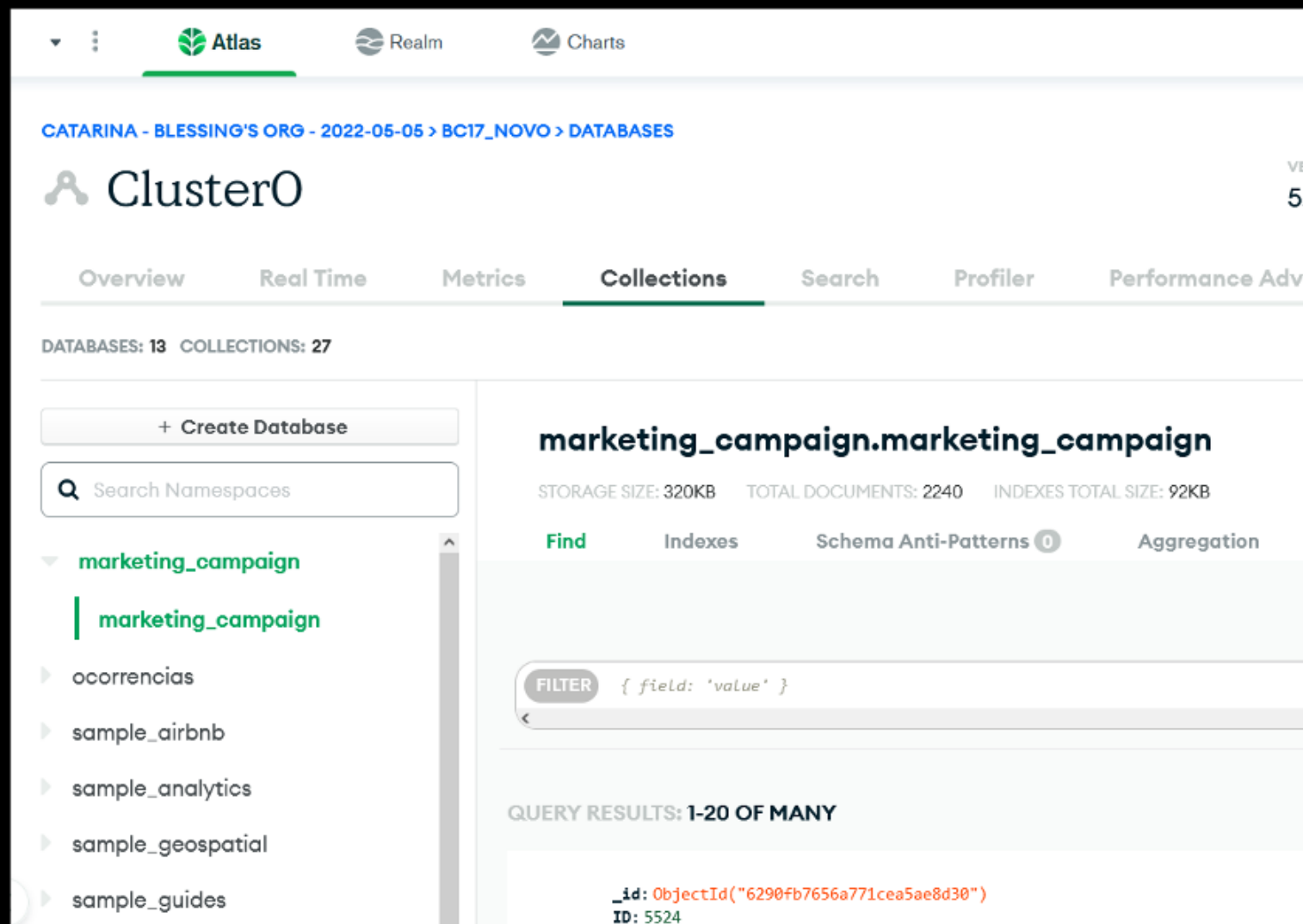
<input type="checkbox"/>	Name	Size	Type	Created	Storage class
<input type="checkbox"/>	marketing_campaign.csv	215 KB	text/csv	May 28, 20...	Standard



# PROJETO INDIVIDUAL

## 1. Nível Infra

- 1.2 O arquivo original e tratado deve ser salvo em MongoDB Atlas em coleções diferentes





# PROJETO INDIVIDUAL

## 1.Nível Infra

### 1.3 O Dataset deve ser salvo em ambiente cloud (Cloud Storage)

BC17

Search Products, resources, doc

← Bucket details

atividade-individual

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Buckets > atividade-individual

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

Filter by name prefix only

Filter

Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created	Storage
<input type="checkbox"/>	original/	—	Folder	—	—
<input type="checkbox"/>	tratado/	—	Folder	—	—

atividade-individual

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

Buckets > atividade-individual > tratado

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

Filter by name prefix only

Filter

Filter objects and folders

<input type="checkbox"/>	Name	Size	Type
<input type="checkbox"/>	marketing_campaign_tratado.csv	234.2 KB	text/csv



# PROJETO INDIVIDUAL

## 2. Nível Pandas

- 2.1 O arquivo está em outra linguagem e deve ter seus dados traduzidos para Português-BR  
antes disso, no Colab, tenho que preparar o ambiente e carregar o arquivo original
- 2.2 Verificar a existência de dados inconsistentes e realizar a limpeza para NaN ou NA
- 2.3 Realizar o drop (se necessário) de colunas do dataframe realizando o comentário do porque da exclusão



# PROJETO INDIVIDUAL

## 3. Nível PySpark

- 3.1 Carregando o dataframe no PySpark

antes disso, no Colab, tenho que preparar o ambiente para trabalhar com PySpark

- 3.2 Ajustando o StructType do dataframe
- 3.3 Verificar a necessidade de drop em colunas ou linhas. Caso seja necessário, fazer comentário do porque não houve drop, pois a atividade foi realizada no pandas anteriormente
- 3.4 Realizar a mudança de nome de pelo menos 2 colunas
- 3.5 Deverá criar pelo menos duas novas colunas contendo alguma informação relevante sobre as outras colunas
- 3.6 Deverá utilizar filtros, ordenação e agrupamento, trazendo dados relevantes para o negócio em questão
- 3.7 Utilizar Window Functions



## 4. Nível PySparkSQL

- 4.1 - Analisando a renda familiar mínima por grau de instrução do individuo
- 4.2 - Analisando a renda familiar máxima por grau de instrução do individuo
- 4.3 - Analisando a renda familiar média por grau de instrução do individuo
- 4.4 - Analisando a renda familiar mínima pelo estado civil do individuo
- 4.5 - Analisando a renda familiar máxima pelo estado civil do individuo
- 4.6 - Analisando a renda familiar média pelo estado civil do individuo



5. Salvando o dataframe tratado no Google Cloud Storage (ver Colab)

6 - inserindo arquivo tratado no MongoDB (ver Colab)

7 - Análise do dataframe no DataStudio



# PROJETO INDIVIDUAL

Muito obrigado

Carlos Bahia  
carlos.bahia.br@gmail.com