

# Challenge 2020

Processo de Seleção para DATA ENGINEER na BOA

# O desafio

Empresa	Contexto	Descrição do problema
<p>A TUBES &amp; CO é uma indústria familiar que produz tubos flexíveis. Estes tubos são muito utilizados em instalações residenciais e comerciais.</p>	<p>A TUBES &amp; CO está passando por um processo de transformação digital.</p>	<p>Levar os dados da TUBES &amp; CO para a nuvem. Usando os produtos e serviços do Google.</p>

# Sobre os dados do desafio

## Price Quote

Esse arquivo contém informações sobre price quotes dos fornecedores. Estes preços podem ser:

- Bracket pricing
- No bracket

## Bill of Materials

Esse arquivo contém a lista de componentes e as quantidades que foram utilizadas em cada conjunto de tubos.

## Comp Boss

Esse arquivo contém as informações sobre cada componente.

# Solução

## Usando o Google Cloud

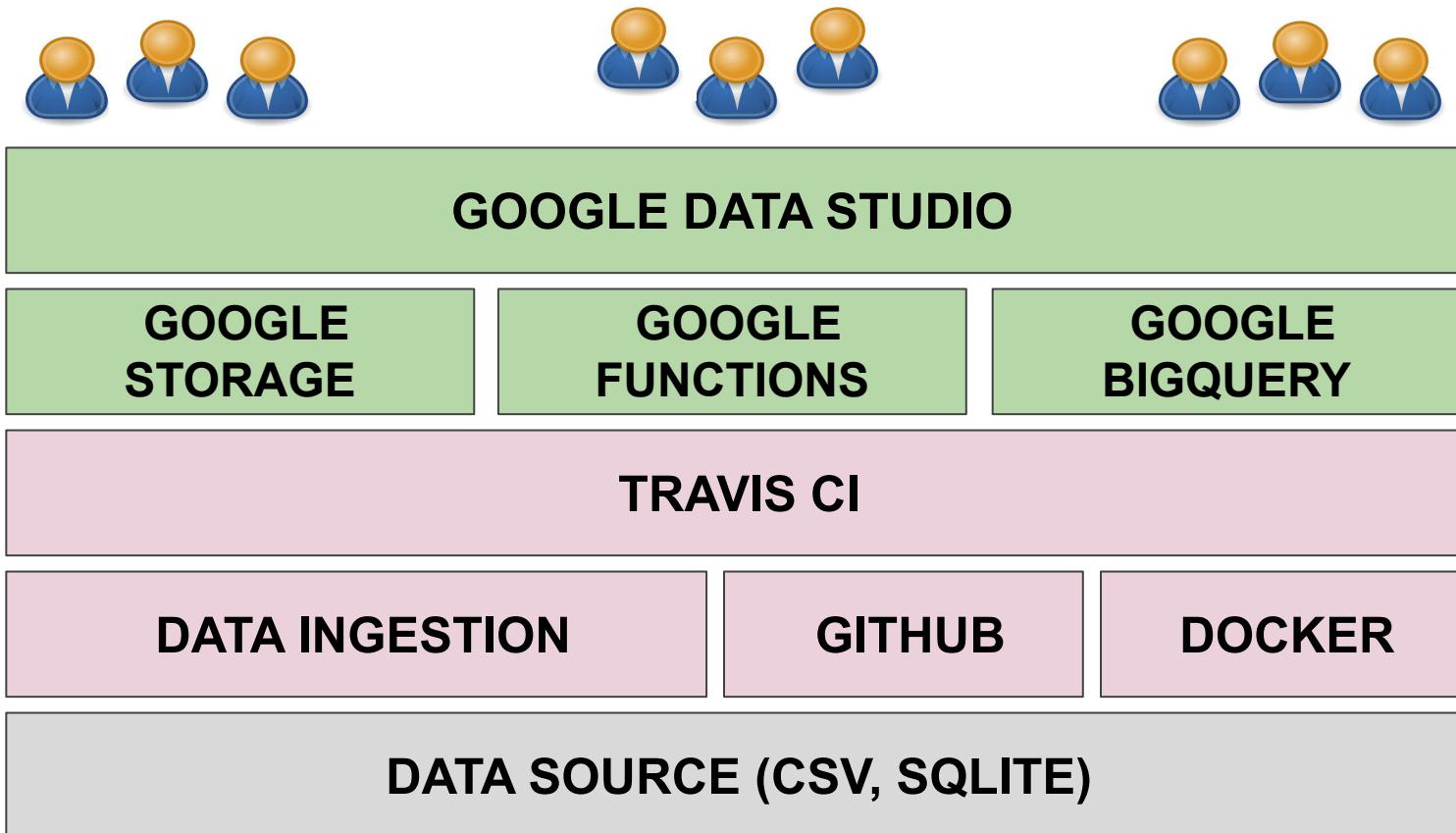
GCS - Cloud Storage

GCF - Cloud Functions (Python)

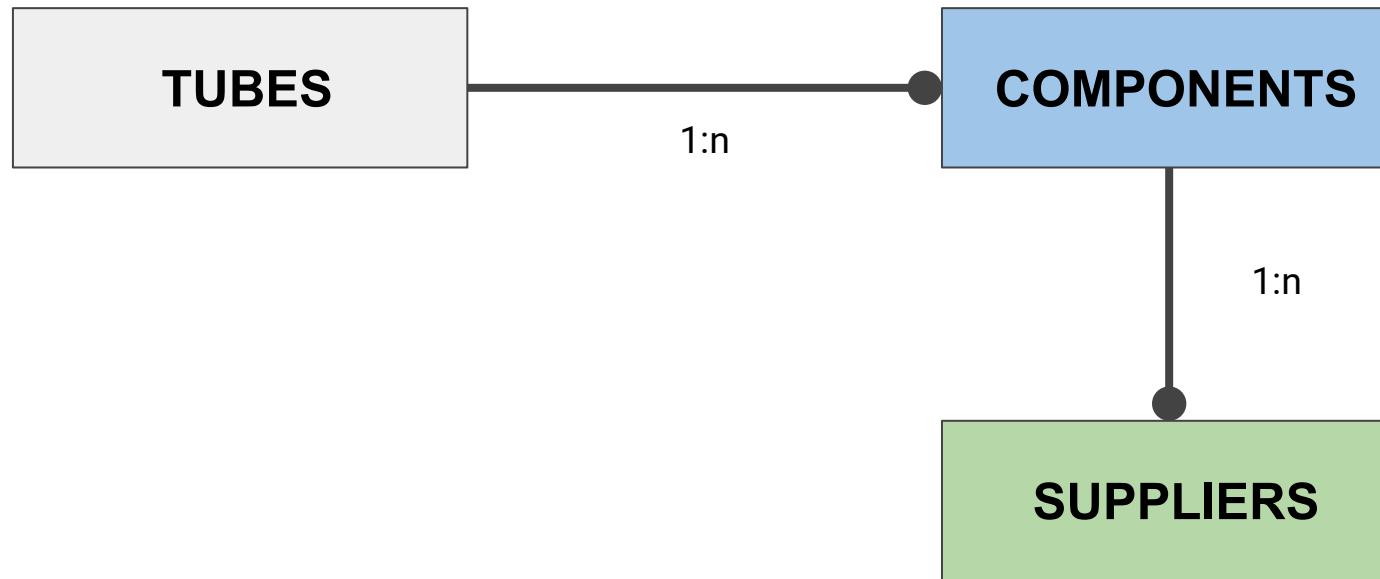
GBQ - Cloud BigQuery

---

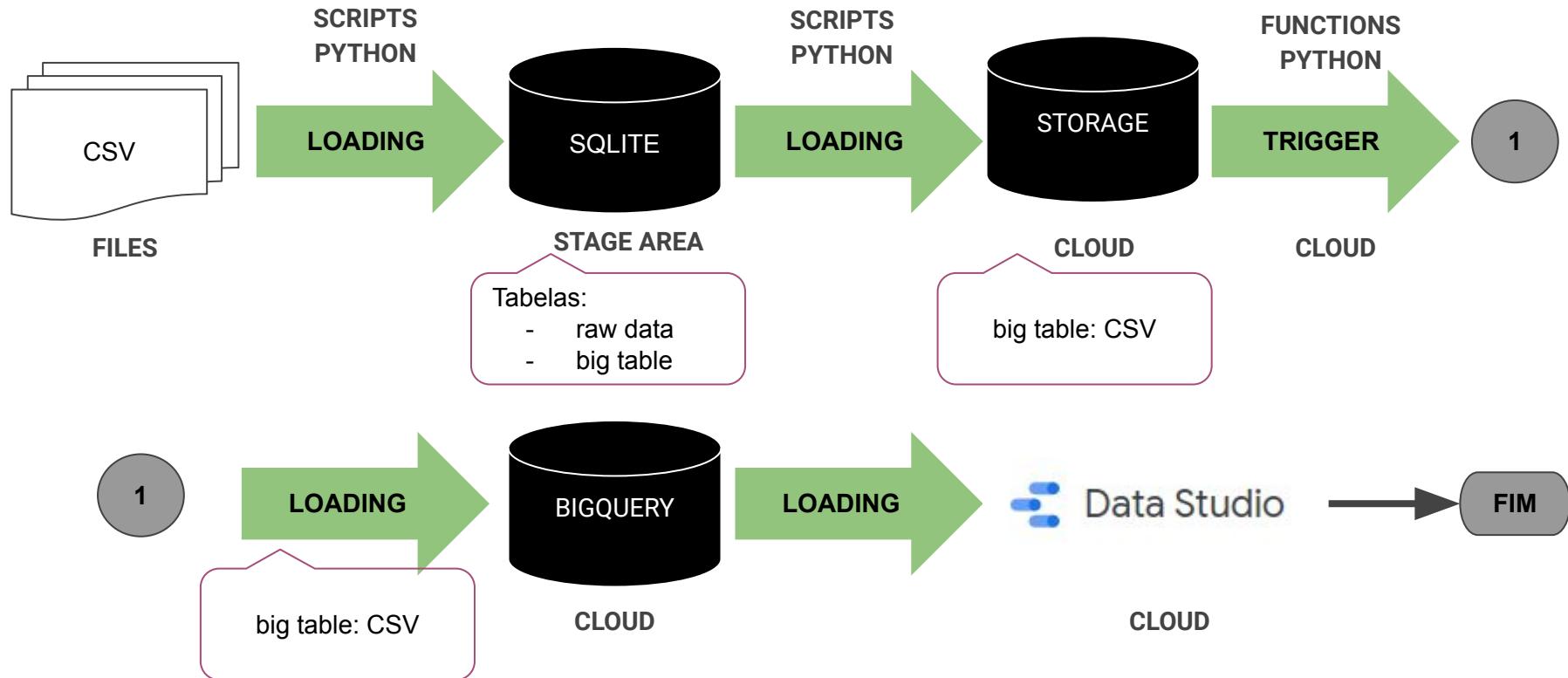
# Arquitetura



# Modelo de dados



# Fluxo de dados



# Timeline

Entrevista com o RH e  
recebimento do  
material do teste após  
a entrevista com o RH

Entrevista técnica e  
alinhamento de  
expectativas

Entrega do teste ao  
RH para análise  
técnica



Análise do material do  
teste, planejamento  
das atividades e  
feedback ao RH

Retomada das  
atividades do teste  
depois dos problemas  
técnicos

# Considerações Finais

## O processo

Estou gostando muito de conhecer a empresa, os desafios e as oportunidades que existem.

Também estou gostando bastante da forma como o processo está sendo conduzido.

## O desafio

Era pra ser um desafio DE BOA. Mas, acabei encarando alguns imprevistos no meio do caminho que me tiraram do foco. Mas, com perseverança e resiliência segui em frente!

## O futuro

**O futuro é incerto.**

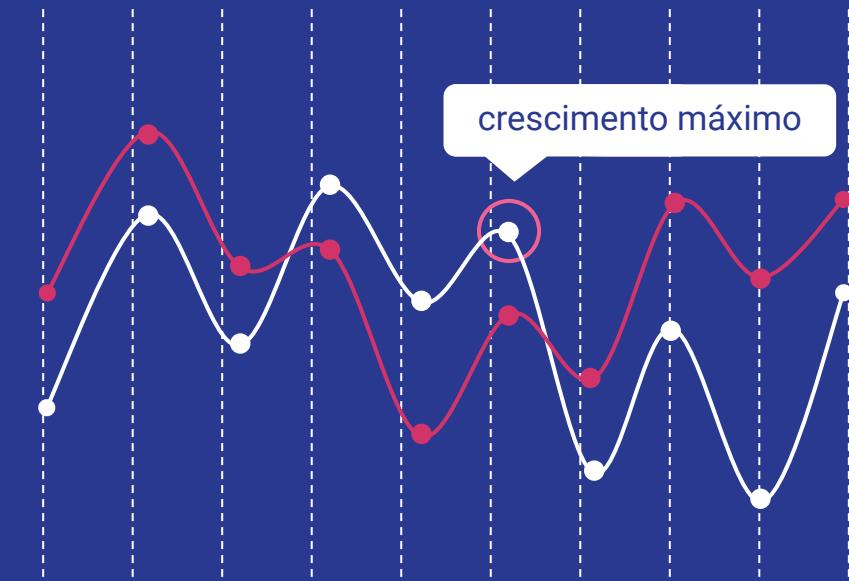
**Mas hoje a única certeza que eu tenho, é o meu desejo de fazer parte desta equipe e me tornar um DE BOA !**

**(Data Engineer na BOA).**

# Thank you!

**Laercio Serra**

laercio.serra@gmail.com



# Gerando Relatórios EDA

```
WELCOME TO BOA CHALLENGE FOR DATA ENGINEER!
NAME: LAERCIO SERRA
-----
=> [OK] - EXPLORATORY DATA ANALYSIS
-----
>> Generating Profile Report: comp_boss.csv
>> Exporting Profile Report to HTML: comp_boss.csv
-----
Summarize dataset: 100%|██████████| 29/29 [00:07<00:00,  3.76it/s, Completed]
Generate report structure: 100%|██████████| 1/1 [00:06<00:00,  6.28s/it]
Render HTML: 100%|██████████| 1/1 [00:00<00:00,  1.00it/s]
Export report to file: 100%|██████████| 1/1 [00:00<00:00, 259.47it/s]
-----
>> Generating Profile Report: bill_of_materials.csv
>> Exporting Profile Report to HTML: bill_of_materials.csv
-----
Summarize dataset: 100%|██████████| 31/31 [00:03<00:00,  7.94it/s, Completed]
Generate report structure: 100%|██████████| 1/1 [00:06<00:00,  6.88s/it]
Render HTML: 100%|██████████| 1/1 [00:00<00:00,  1.59it/s]
Export report to file: 100%|██████████| 1/1 [00:00<00:00, 311.57it/s]
-----
>> Generating Profile Report: price_quote.csv
>> Exporting Profile Report to HTML: price_quote.csv
```

# Google Cloud Storage

The screenshot shows the Google Cloud Platform interface for managing a bucket named 'my-bigdata-projects'. The top navigation bar includes the 'Google Cloud Platform' logo, the project name 'MyFirstProject', a search bar, and various navigation icons.

The main content area displays the 'Bucket details' page for 'my-bigdata-projects'. On the left, there's a sidebar with icons for Buckets, Objects, Configuration, Permissions, Retention, and Lifecycle. The 'OBJECTS' tab is selected, showing a list of objects under the path 'Buckets > my-bigdata-projects > data > csv'. The list includes one object: 'bt\_challenge\_box' (45.1 KB, text/csv, created Nov 2, 2020, Standard storage class, Not public access, Google-managed key encryption, no retention policy, and no holds).

Name	Size	Type	Created time	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
bt_challenge_box	45.1 KB	text/csv	Nov 2, 2020, 8:...	Standard	Nov 2, 202...	Not public	Google-managed key	—	None

# Google Cloud Functions

The screenshot shows the Google Cloud Platform Functions interface. At the top, there's a navigation bar with the project name "MyFirstProject". Below the navigation bar is a search bar labeled "Search products and resources". The main area has tabs for "Cloud Functions" and "Functions", with "Cloud Functions" currently selected. There are buttons for "CREATE FUNCTION", "REFRESH", "DELETE", and "COPY". On the right, there's a "SHOW INFO PANEL" button and a user profile icon. A "Filter functions" dropdown is open. The main table lists one function:

Name	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed	Authentication	Actions
CSV_LOADER_GCS_GBQ	us-central1	Bucket: my-bigdata-projects	Python 3.7	256 MiB	csv_loader	Nov 1, 2020, 9:51:39 PM		⋮

# Google Cloud Functions

Google Cloud Platform MyFirstProject Search products and resources 7:55:51 AM - 9:55:51 AM PAGE LAYOUT

Logs Explorer OPTIONS REFINE SCOPE Project Dismiss Learn more

New features are available in the Logs Explorer.

Query preview  
resource.type="cloud\_function" resource.labels.function\_name="CSV\_LOADER\_GCS\_GBQ" resource.labels.region="us-central1"

Save Run Query

Log fields Search fields and values

LOG NAME  
cloudfunctions.googleapis.com/cloud-functio... 7

FUNCTION\_NAME CSV\_LOADER\_GCS\_GBQ Clear X

PROJECT\_ID sturdy-conduit-260900 7

REGION us-central1 Clear X

RESOURCE TYPE Cloud Function Clear X

SEVERITY Info

Histogram

Nov 2, 7:54 AM 8:30 AM 9:00 AM 9:30 AM Nov 2, 9:56 AM

Query results Jump to Now Actions Configure

SEVERITY	TIMESTAMP	BRT	SUMMARY
> i	2020-11-02 08:54:00.439	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 Function execution started
> i	2020-11-02 08:54:01.324	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 >> Function: CSV LOADER GCS TO GBQ
> i	2020-11-02 08:54:01.325	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 >> Storage: gs://my-bigdata-projects/data/csv/bt_chall...
> i	2020-11-02 08:54:01.325	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 >> Starting Job: ef6e2ab0-c515-4415-bd0d-a520face7382
> i	2020-11-02 08:54:03.005	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 >> Total Rows: 342
> i	2020-11-02 08:54:03.005	BRT	CSV_LOADER_GCS_GBQ qq31491jbv92 >> Job finished successfully!

# Google BigQuery

Query editor

bt\_challenge\_boa

QUERY TABLE SHARE TABLE

Schema Details Preview

Run Save query Save view Schedule

bt\_challenge\_boa

Description Labels

Big Table Challenge BOA challenge : boa

Schema Details Preview

Field name	Type	Mode	Policy tags
tube_assembly_id	STRING	NULLABLE	
quantity_component	FLOAT	NULLABLE	
component_id	STRING	NULLABLE	
component_type_id	STRING	NULLABLE	
component_type	STRING	NULLABLE	
connection_type_id	STRING	NULLABLE	
outside_shape	STRING	NULLABLE	
base_type	STRING	NULLABLE	

Table info

Table ID	sturdy-conduit-260900:de_boa.bt_challenge_boa
Table size	46.2 KB
Number of rows	342
Created	Nov 1, 2020, 6:19:54 PM
Table expiration	Nov 11, 2020, 6:19:54 PM
Last modified	Nov 2, 2020, 11:25:34 AM
Data location	US

# Google BigQuery

Google Cloud Platform MyFirstProject Search products and resources

BigQuery FEATURES & INFO SHORTCUT

Query history Saved queries Job history Transfers Scheduled queries Reservations BI Engine Resources + ADD DATA

Search for your tables and datasets

sturdy-conduit-260900 listenbrainz

Query editor

+ COMPOSE NEW QUERY HIDE EDITOR FULL SCREEN

```
1 SELECT * FROM `sturdy-conduit-260900.de_boa.bt_challenge_boa` LIMIT 1000
```

Run Save query Save view Schedule query More This query will process 46.2 KB when run.

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.4 sec elapsed, 46.2 KB processed)

Job information Results JSON Execution details

Row	tube_assembly_id	quantity_component	component_id	component_type_id	component_type	connection_type_id	outside_shape	base_type	height_over_tube	bolt_pattern
1	TA-01947	1.0	C-0161	CP-020	null	9999	null	null	19.2	
2	TA-11878	1.0	C-0161	CP-020	null	9999	null	null	19.2	
3	TA-13310	4.0	C-0243	CP-019	null	B-012	null	null	36.0	
4	TA-14192	4.0	C-0243	CP-019	null	B-012	null	null	36.0	
5	TA-13359	1.0	C-0770	CP-020	null	9999	null	null	21.5	
6	TA-10009	1.0	C-0780	CP-020	null	9999	null	null	21.5	
7	TA-20386	2.0	C-1919	CP-019	null	B-012	null	null	35.0	

Rows per page: 100 1 - 100 of 342 First page < > >> Last page

# Google Data Studio

Data Studio

Pesquisar no Data Studio

?

⚙️

GRID

👤

 Criar

Recentes

Relatórios

Origens de dados

Explorador

Recentes

Começar com um modelo

Galeria de modelos

Compartilhados comigo

De minha propriedade

Lixeira

 Relatório em branco Data Studio

 Relatório de tutorial Data Studio

 Marketing da Acme Google Analytics

 Search Console Report Search Console

Nome

Pertencem a qualquer pessoa

Última vez que abri

 Executive Dashboard

Laércio Serra

10:16

# Google Data Studio

The screenshot shows the Google Data Studio interface. At the top, there is a navigation bar with the Data Studio logo, a search bar containing 'Pesquisar no Data Studio', and various settings icons. Below the navigation bar, there are four tabs: 'Recentes' (selected), 'Relatórios', 'Origens de dados' (highlighted in blue), and 'Explorador'. On the left side, there is a sidebar with categories: 'Recentes' (selected), 'Compartilhados comigo', 'De minha propriedade', and 'Lixeira'. A red box highlights the 'Origens de dados' tab and the list of data sources below it. The list includes a single item: 'BigQuery - 01/11/2020 23:36' by Laércio Serra, which was last opened at 00:49.

Nome	Pertencem a qualquer pessoa	Última vez que abri
BigQuery - 01/11/2020 23:36	Laércio Serra	00:49

# Google Data Studio



Pesquisar no Data Studio



Criar

Recentes

Relatórios

Origens de dados

Explorador

Recentes

Compartilhados  
comigo

De minha propriedade

Lixeira

Nome

Última modificação

10:16

Data Explorer

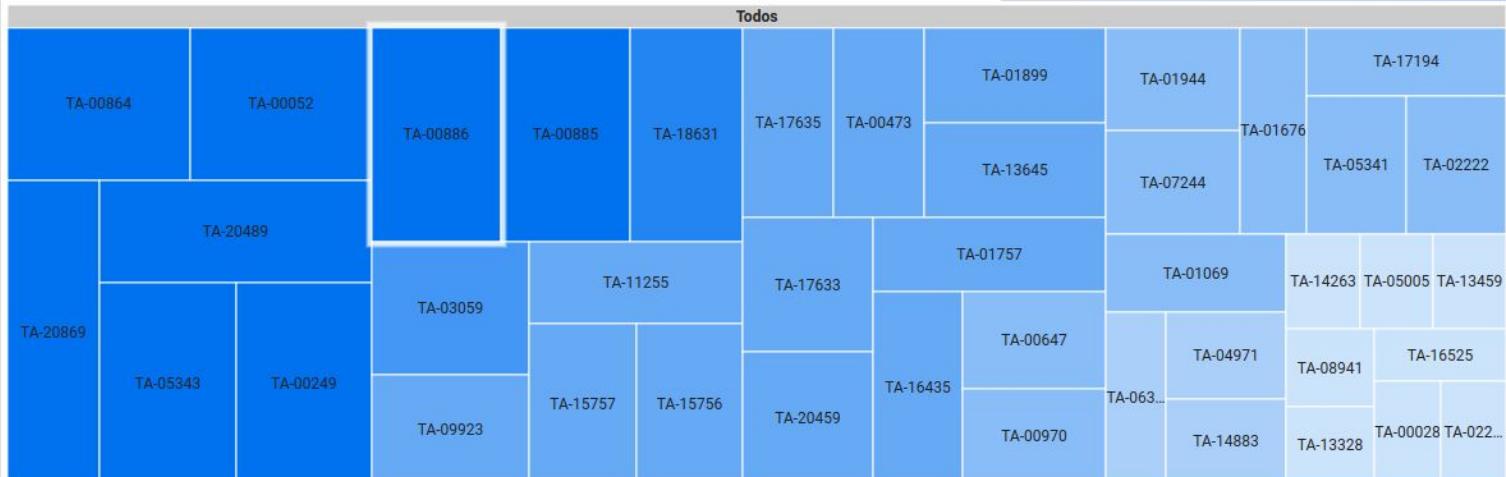
Nome	Última modificação
Data Explorer	10:16

# Google Data Studio

Executive Dashboard



## Produtos Mais Fabricados



## Custo do Produto por Componente

Comp ID / Custo

Tube ID	C-0699	C-0199	C-0817	C-1536	C-1231	C-1725
---------	--------	--------	--------	--------	--------	--------

## Top 10 Suppliers