



Bruno Yaporandy

# Projeto Individual

**Marketing\_Campaign.csv**

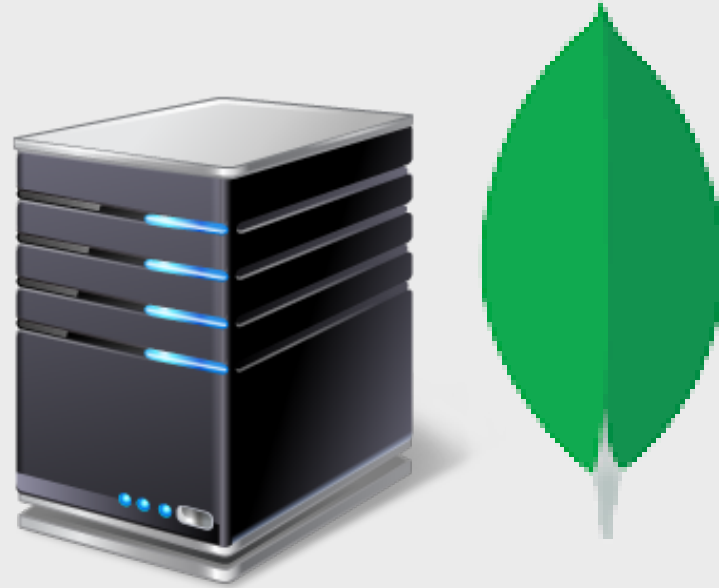
@yaporandy

# Ciclo ETL

**Extract, Transform and Load**



# Extraíndo o DataSet



**Banco de Dados  
Servidor do Cliente**



**Google Storage**

- Conexão com MongoDB Atlas,
- Transformação da coleção em DataFrame,
- Envio desse DataFrame para um Bucket (Google Storage),
- Início da etapa de Transformação.

# Extraíndo o DataSet

## Principais comandos:

### Definindo parâmetros de conexão

```
uri      = 'mongodb+srv://<user>:<password>@  
cluster0.3jmog.mongodb.net...'  
client = MongoClient( uri )
```

### Transformação da coleção em DataFrame

```
df = pd.DataFrame(collection.find())
```

### Envio do DataFrame para o Bucket

```
df.to_csv('gs://path/df-original-proj-individual.csv')
```

# Tratamento de Dados



**Google Storage**

**Pandas (Python)**

**Google Storage**

- Tradução para PT-BR,
- Pré-Analise,
- Correção de Inconstâncias,
- Validação de Dados com PANDERA,
- Início da 2ª etapa de Transformação.

# Tratamento de Dados

## Principais comandos:

### Pré-Análise

```
df.isna().sum()
```

```
df.duplicated().sum()
```

```
df.drop(['Z_CostContact'], axis=1, inplace=True)
```

```
df.drop(['Z_Revenue'], axis=1, inplace=True)
```

### Tradução para PT-BR

```
df.rename(columns={'Year_Birth':'Ano_Nascimento'},  
inplace=True)
```

### Validação de dados com PANDERA

```
schema.validate(df)
```

# Tratamento de Dados



**Google Storage**



**PySpark**



**Google Storage**

- Estruturado através do StructType,
- Pré-Analise,
- Criação de 2 colunas (pedido do cliente),
- Filtros, Ordenações e Agrupamentos de dados relevantes para o negócio,
- Window Functions,
- Início da etapa de Carregamento.

# Tratamento de Dados

## Principais comandos:

### StructType

```
esquema = ( StructType([  
    StructField('Ano_Nascimento', IntegerType(), True),  
    StructField('Estado_Civil', StringType(), True),  
    StructField('Data_Inscrição_Cliente', DateType(), True)...
```

### Filtros, Ordenação e Agrupamentos

```
df_spark.groupBy('Escolaridade').sum('Gasto_Total').show()  
df_spark.select(F.avg('Renda_Anual')).show()
```

### Window Functions

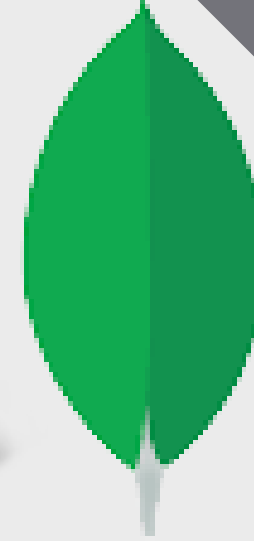
```
w0 = Window.partitionBy(F.col('Escolaridade')).orderBy('Renda')  
df_spark.withColumn('rank', F.rank().over(w0)).show(10)
```



# Carregamento(Load)



**Google Storage**



**Banco de Dados  
Servidor do Cliente**

- Inserção do DataFrame no Banco de Dados do cliente (Banco noSQL).

# Carregamento (Load)

## Principais comandos:

### Transformando em Dicionário

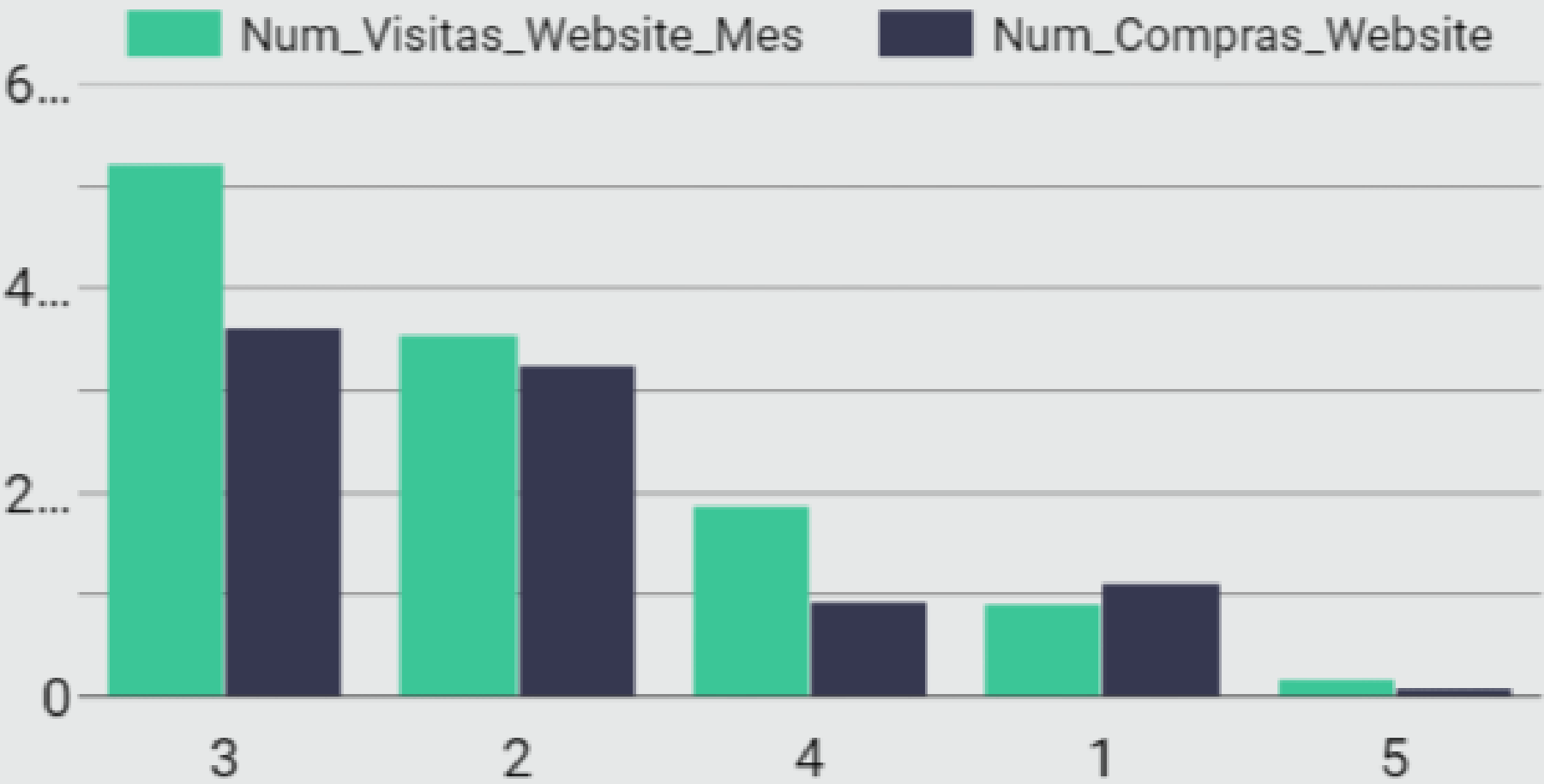
```
df_dicio = df.to_dict("records")
```

### Inserindo Coleção

```
collection.insert_many(df_dicio)
```

# Insights

Relação de Compras x Visitas pela Web das Famílias



	Escolaridade	Gasto_Total ▾
1.	Nível Superior	693.740
2.	PhD	323.656
3.	Mestrado	321.464
4.	2o Grau	4.417

Quantidade de Reclamações

20

Resultado da Ultima Campanha



Marketing\_Campaign.csv

@yaporandy

# Obrigado,

## Bruno Yaporandy

### Imagens retiradas de:

- [https://miro.medium.com/max/780/1\\*Rmc568knYGLn7kJ3B97WUQ.png](https://miro.medium.com/max/780/1*Rmc568knYGLn7kJ3B97WUQ.png)
- <https://logodownload.org/wp-content/uploads/2021/06/google-cloud-logo-1.png>
- <http://assets.stickpng.com/images/5848152fcef1014c0b5e4967.png>
- [https://upload.wikimedia.org/wikipedia/commons/thumb/f/f3/Apache\\_Spark\\_logo.svg/2560px-Apache\\_Spark\\_logo.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/f/f3/Apache_Spark_logo.svg/2560px-Apache_Spark_logo.svg.png)
- <https://www.freepnglogos.com/uploads/server-png/home-server-icon-icons-and-png-backgrounds-30.png>

### Dashboard Disponível em:

<https://datastudio.google.com/reporting/939b5227-a097-41fe-ac85-73b9c014c49b>

### Agradecimento aos colegas:

- Lucas David, . . . . .
- Danilo Ferrari, . . . . .
- Raquel Nunes, . . . . .
- Stéphanie Pirajá