

# PROJETO INDIVIDUAL

Jum Saheki

Soul Code

Turma BC17

Engenharia de Dados

# ÍNDICE

- Introdução
- Tecnologias
- Pandas
- Spark
- Data Studio

# INTRODUÇÃO

O objetivo da atividade é realizar ETL do arquivo `marketing_campaign.csv` fornecido no classroom e enviá-los para a Google Cloud Storage(GCP) e MongoDB Atlas e apresentar o processo realizado.



# ETL

Fonte de dados

Extração

Transformação

Carregamento

Banco de dados

# TRANSFORMAÇÃO

Tradução

Valores nulos

Duplicatas

Inconsistências

# TECNOLOGIAS



colab



python



pandas



mongoDB



gcp



Google Data Studio



# PANDAS

## EXTRAÇÃO

```
df = pd.read_csv()
```

## ANÁLISE

```
df.shape  
df.head().T  
df.isna().sum()  
df.dtypes  
df.nunique()  
df.describe().T
```

## TRANSFORMAÇÃO

```
df.rename(columns={})  
df.replace()  
df1.drop()  
df.fillna()
```

## CARREGAMENTO

```
df.to_csv()  
  
df.to_dict('records')  
collection.insert_many()
```



## EXTRAÇÃO

```
StructType=[  
    StructField()  
    ]
```

```
spark.read.format()  
.option.load
```

## ANÁLISE

```
df.printSchema()  
  
df.summary.show()  
  
df.dropDuplicates()  
.count()
```

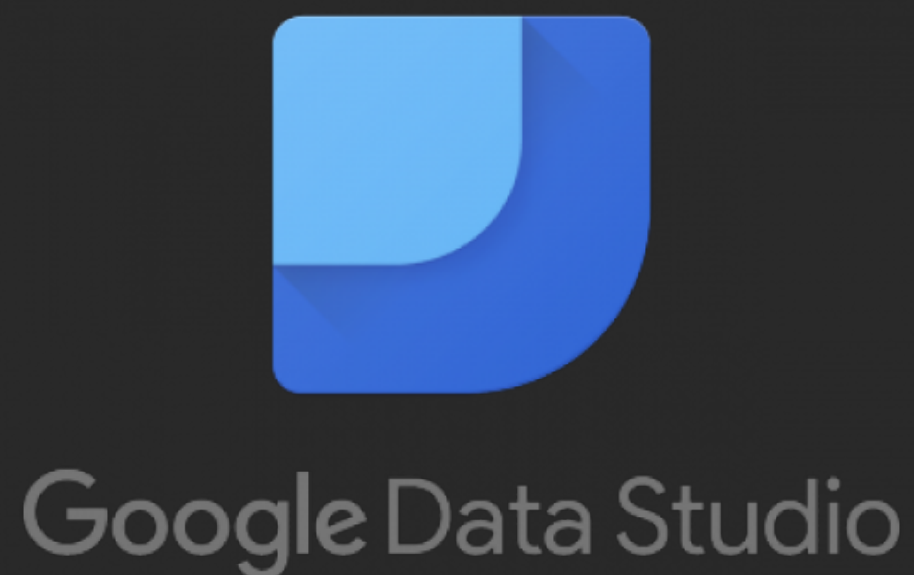
## TRANSFORMAÇÃO

```
F.regex_replaece()  
df1.drop()  
df.fillna()  
df.withColumn()
```

## CARREGAMENTO

```
df.write.csv()  
  
df.toPandas()  
df.to_dict('records')  
collection.insert_many()
```

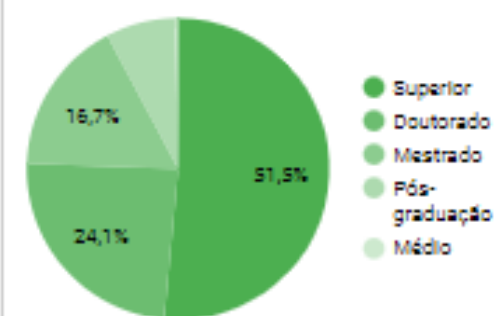




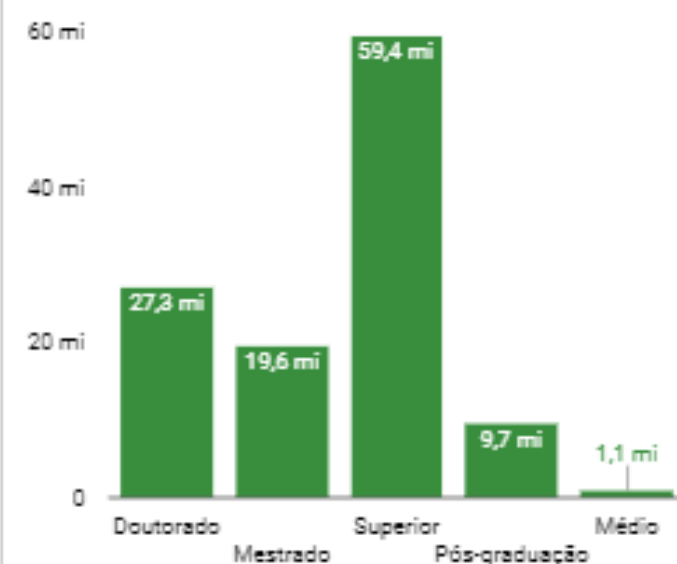
## CAMPANHA DE MARKETING

SOUL CODE

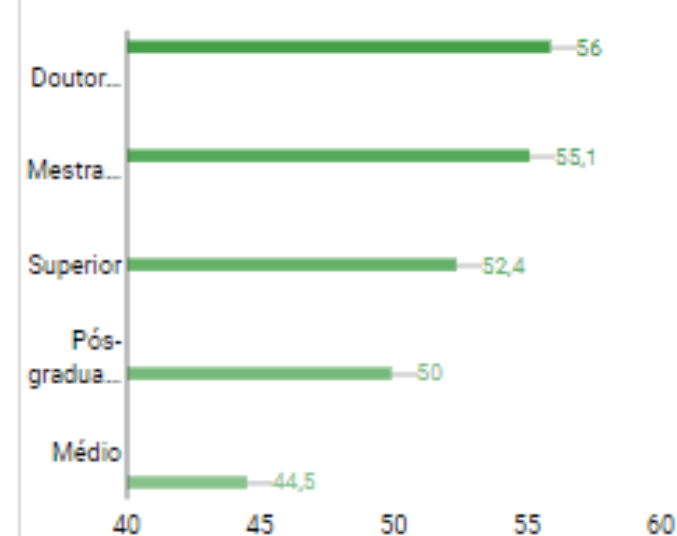
### Escolaridade



### Renda



### Idade média



### Compras

Promoção

5.208

Loja

12.970

Website

9.150

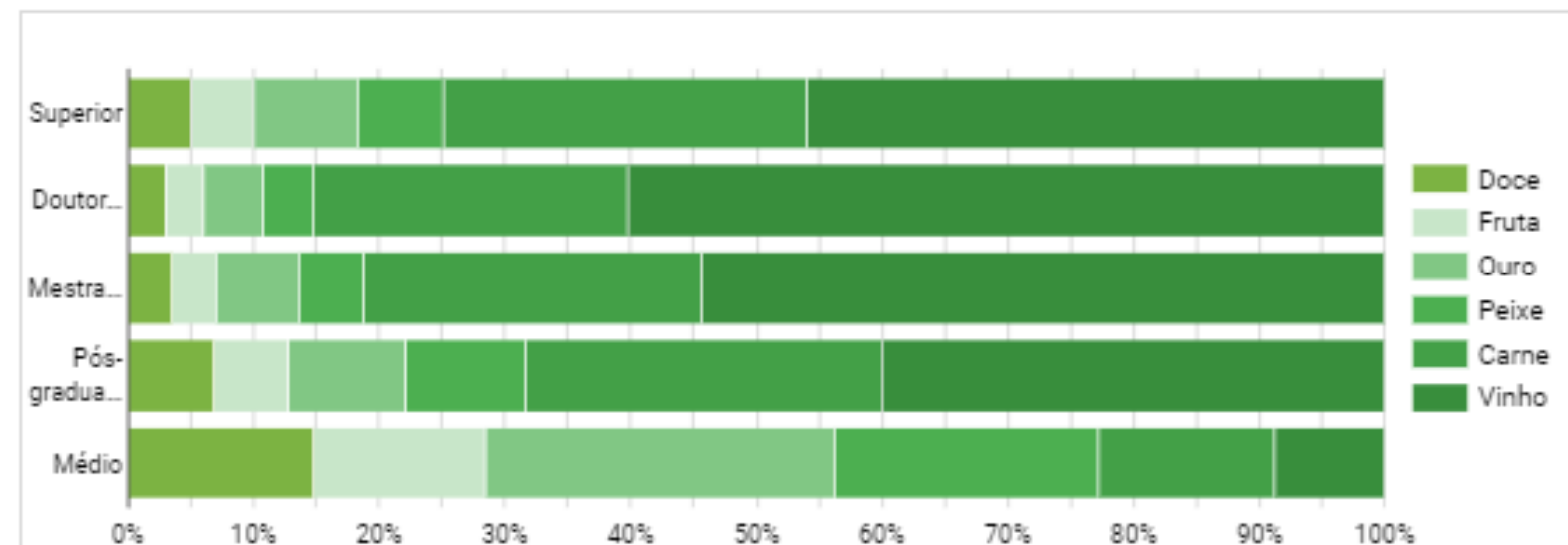
Catálogo

5.963

### Adesão a companhias

Doutorado	0,5
Superior	0,4
Mestrado	0,4
Pós-graduação	0,4
Médio	0,1
Total geral	0,4

### Gasto por tipo de Produto





Muito Obrigado!