

# Relatório de Pré processamento - inspeção de datasets

Cláudia Patrícia Silva Pimentel

E-mails: [cacaupimentel@hotmail.com](mailto:cacaupimentel@hotmail.com) (<mailto:cacaupimentel@hotmail.com>), [cacaupimentel@gmail.com](mailto:cacaupimentel@gmail.com) (<mailto:cacaupimentel@gmail.com>), [claudiapimentel@aluno.uema.br](mailto:claudiapimentel@aluno.uema.br) (<mailto:claudiapimentel@aluno.uema.br>)

## 3. Resumo e inspeções dos atributos de [Wholesale customers Data Set](https://archive.ics.uci.edu/ml/datasets/Wholesale+customers) (<https://archive.ics.uci.edu/ml/datasets/Wholesale+customers>)

### 3.1 Informações sobre a Fonte

1. Autores: Margarida GMS Cardoso e Nuno Gonçalo Costa Fernandes Marques de Abreu;
2. Artigo: Análise do perfil do cliente Recheio e desenvolvimento de um sistema promocional;
3. Publicação: ISCTE-IUL, Lisbon;
4. Ano: 2011;
5. Objetivo da Pesquisa: foi o desenvolvimento de um sistema promocional com base na análise do perfil do cliente Recheio com o intuito do incremento de vendas das principais categorias (Abreu, 2011);
6. Resultados da Pesquisa: desenvolvimento de um sistema de promocional (Abreu, 2011).

### 3.2 Importando o [Wholesale customers Data Set](https://archive.ics.uci.edu/ml/datasets/Wholesale+customers) (<https://archive.ics.uci.edu/ml/datasets/Wholesale+customers>)

Foram instalados os pacotes do pandas, numpy e pandas\_profiling, para exploração dos dados.

In [ ]:

```
# Instalando o pacote pandas
%pip install pandas
```

In [ ]:

```
# Instalando o pacote numpy
%pip install numpy
```

In [ ]:

```
# Instalando o pacote pandas_profiling
%pip install pandas_profiling
```

In [1]:

```
# Importando as bibliotecas necessárias
import pandas as pd
import numpy as np
from pandas_profiling import ProfileReport
```

### 3.3 Procedimentos de limpeza do arquivo

Na importação dos dados não houve nenhuma alteração ou ajuste de imediato para ser feito.

In [2]:

```
#importando o data set sem tratamento inicial apenas
clientes = pd.read_csv('DataSets/Wholesale customers data.csv', sep = ',')
clientes
```

Out[2]:

	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	2	3	12669	9656	7561	214	2674	1338
1	2	3	7057	9810	9568	1762	3293	1776
2	2	3	6353	8808	7684	2405	3516	7844
3	1	3	13265	1196	4221	6404	507	1788
4	2	3	22615	5410	7198	3915	1777	5185
...	...	...	...	...	...	...	...	...
435	1	3	29703	12051	16027	13135	182	2204
436	1	3	39228	1431	764	4510	93	2346
437	2	3	14531	15488	30243	437	14841	1867
438	1	3	10290	1981	2232	1038	168	2125
439	1	3	2787	1698	2510	65	477	52

440 rows × 8 columns

### 3.4 Inspeção e descrição dos Atributos

Abreu (2011) informa que o trabalho foi precedido de uma revisão de literatura para ulterior recolhimento dos dados, o qual ocorreu por intermédio de um "(...) de forma a segmentá-los e proceder ao desenvolvimento de um sistema promocional adequado." Assim, as informações dos atributos disponibiliza de forma simplificada como a seguinte descrição:

- 1) CHANNEL: Canal do cliente:  
Horeca (Hotel / Restaurante / Café),  
canal Retalho (Nominal);
- 2) REGION: Região do cliente:  
Lisbon,  
Porto,  
Outra Região;
- 3) FRESH: gasto anual (mu) com produtos frescos (Contínuo);
- 4) MILK: gasto anual (mu) com produtos lácteos (Contínuo);
- 5) GROCERY: gasto anual (mu) com produtos de mercearia (Contínuo);
- 6) FROZEN: gasto anual (mu) com produtos congelados (Contínuo)
- 5) DETERGENTS\_PAPER: gasto anual (mu) com detergentes e produtos de papel (Contínuo)
- 6) DELICATESSEN: gasto anual (mu) com produtos e charcutaria (Contínuo);

Segundo os autores, o resultado foi apresentado em Estatística Descritiva, apresentando os Mínimo, Máximo, Média, Desvio Padrão de FRESH, MILK, GROCERY, FROZEN, DETERGENTS\_PAPER, DELICATESSEN. Quantificando as frequências por região. É importante registrar que o conjunto de dados é originado de um

banco de dados maior referenciado na dissertação de Mestrado em Marketing de Abreu (2011).

In [3]:

```
# Inspeccionando os atributos do dataset  
clientes.describe()
```

Out[3]:

	Channel	Region	Fresh	Milk	Grocery	Frozen	Deter
count	440.000000	440.000000	440.000000	440.000000	440.000000	440.000000	
mean	1.322727	2.543182	12000.297727	5796.265909	7951.277273	3071.931818	
std	0.468052	0.774272	12647.328865	7380.377175	9503.162829	4854.673333	
min	1.000000	1.000000	3.000000	55.000000	3.000000	25.000000	
25%	1.000000	2.000000	3127.750000	1533.000000	2153.000000	742.250000	
50%	1.000000	3.000000	8504.000000	3627.000000	4755.500000	1526.000000	
75%	2.000000	3.000000	16933.750000	7190.250000	10655.750000	3554.250000	
max	2.000000	3.000000	112151.000000	73498.000000	92780.000000	60869.000000	4

## 2.5 Analisando resultados

A seguir tens a análise completa pelo "ProfileReport".

In [5]:

```
profile = ProfileReport(clientes, title='Pandas Profiling Report', explorative=True)
profile.to_notebook_iframe()
```

```
HBox(children=(HTML(value='Summarize dataset'), FloatProgress(value=0.0, max=21.0), HTML(value='')))
```

```
HBox(children=(HTML(value='Generate report structure'), FloatProgress(value=0.0, max=1.0), HTML(value='')))
```

```
HBox(children=(HTML(value='Render HTML'), FloatProgress(value=0.0, max=1.0), HTML(value='')))
```

## Referências

Abreu, N. (2011). Análise do perfil do cliente Recheio e desenvolvimento de um sistema promocional. Mestrado em Marketing, ISCTE-IUL, Lisbon.

ABREU, Nuno Gonçalo Costa Fernandes Marques de - Análise do perfil do cliente Recheio e desenvolvimento de um sistema promocional [Em linha]. Lisboa: ISCTE-IUL, 2011. Dissertação de mestrado. [Consult. Dia Mês Ano] Disponível em www:<http://hdl.handle.net/10071/4097> (<http://hdl.handle.net/10071/4097>).

