

1 Projeto 1

Daniel Ermelino Carvalho - RA 11092613

Elton Mauricio Da Silva - RA

link de arquivos: <https://github.com/dcarve/data_visulization_ufabc>

2 Questao 1

Código

```
import pandas as pd
import numpy as np
import matplotlib

# pr convertido para utf-8
df = pd.read_csv('VIS_Pr_01_Vendas.csv')

df = df.groupby('Region').agg(
    {
        'Sales': ['sum'],
        'Profit': ['sum'],
        'Discount': ['mean']
    }
).reset_index()

df.columns = ["_".join(t).strip('_') for t in df.columns]

#exportando csv configura es do MS excel
df.to_csv(
    'result_q1.csv',
    sep=';',
    index=False,
    encoding='utf-8-sig'
)
```

Output file

Quadro 1 – File - result_q1.csv

| Region | Sales_sum | Profit_sum | Discount_mean |
|---------|-------------|-------------|---------------------|
| Central | 501239.8908 | 39706.3625 | 0.24035299182092124 |
| East | 678781.24 | 91522.78 | 0.14536516853932585 |
| South | 391721.905 | 46749.4303 | 0.1472530864197531 |
| West | 725457.8245 | 108418.4489 | 0.10933499843896347 |

Comentários

O arquivo veio com o encoder ANSI, que é proprietário da Microsoft, o que atrapalha a leitura no linux, logo o arquivo tem que ser converdido para um outro encoder,

no caso, o arquivo foi convertido para utf-8. No windows, o arquivo pode ser lido usando

`pd.read_csv('VISPr01Vendas.csv', encoding='ansi')`, mas não funciona no linux. Já a exportação do

8 – sig.

3 Questao 2

Código

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# pr convertido para utf-8
df = pd.read_csv('VIS_Pr_01_Vendas.csv')

df['order_year'] = pd.to_datetime(df['Order Date']).dt.strftime('%Y')

df = df.groupby(['Sub-Category', 'order_year']).agg(
    {'Profit': ['sum']}
).reset_index()

df.columns = [
    "_".join(t).strip('_').replace('-', '_').lower()
    for t in df.columns
]

df = df.pivot(
    index='sub_category',
    columns='order_year',
    values=['profit_sum']
).reset_index()

df.columns = [
    (t[1] if ('profit_sum' in t) else t[0])
    for t in df.columns
]

sub_category = df['sub_category'].to_list()
values = df.drop('sub_category', axis=1).to_dict('list')

x = np.arange(len(sub_category))
width = 0.15
multiplier = 0

fig, ax = plt.subplots(
    constrained_layout=True,
    figsize=(15, 7)
)
```

```

for attribute, measurement in values.items():

    rgb = (
        (12*(5**(-multiplier))*(10**multiplier))/255,
        (34+(45*multiplier))/255,
        (150+(30*multiplier))/255,
    )

    offset = width * multiplier
    rects = ax.bar(
        x + offset,
        measurement,
        width,
        label=attribute,
        color=rgb)

    multiplier += 1

ax.set_ylabel('profit')
ax.set_xticks(x + width, sub_category)
ax.legend(loc='upper left')
ax.grid(axis = 'y')

plt.show()

```

plot

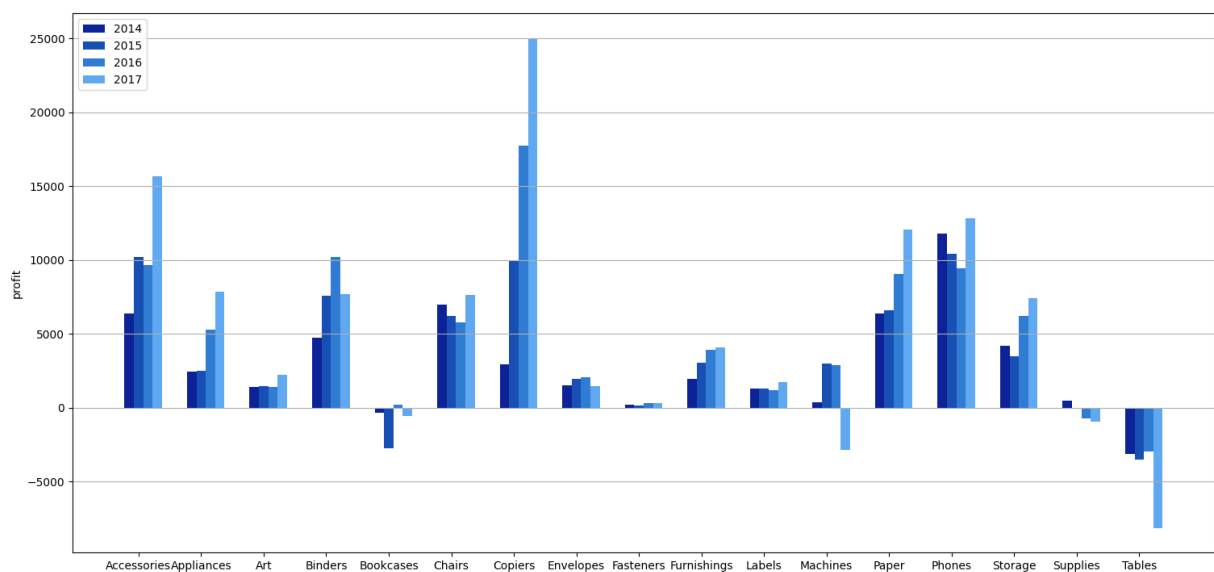


Figura 1 – diagrama do modelo - dedução da matriz

Comentários

Os gráficos do matplotlib embora podem ser facilmente personalizados, exigem um código mais extenso para o caso de houver mais dados e quiser manter a personalização de esquemas de cores do gráfico.

4 Questao 3

Código

```
import pandas as pd

# pr convertido para utf-8
df = pd.read_csv('VIS_Pr_01_Vendas.csv')

df = df.copy()

df['perform'] = df['Profit']/(df['Sales']-df['Discount'])

df = df.groupby(['Customer Name', 'Segment']).agg(
    {'perform': ['mean']}
).reset_index()

df.columns = [t[0] for t in df.columns]

def f(x):

    if x<0.1:
        return "E"
    elif (x>-0.1) and (x<0.15):
        return "D"
    elif (x>-0.15) and (x<0.2):
        return "C"
    elif (x>-0.2) and (x<0.25):
        return "B"
    elif (x>-0.25):
        return "A"

df['perform'] = df.apply(
    lambda row: f(row['perform']),
    axis=1
)

#exportando csv configura es do MS excel
df.to_csv(
    'result_q3.csv',
    sep=';',
    index=False,
```

```
encoding='utf-8-sig',  
)
```

Output file

Quadro 2 – File - result_q23.csv

| Customer Name | Segment | perform |
|--------------------|-------------|---------|
| Aaron Bergman | Consumer | D |
| Aaron Hawkins | Corporate | A |
| Aaron Smayling | Corporate | E |
| Adam Bellavance | Home Office | A |
| Adam Hart | Corporate | C |
| Adam Shillingsburg | Consumer | E |
| Adrian Barton | Consumer | E |
| ... | ... | ... |
| Zuschuss Carroll | Consumer | E |
| Zuschuss Donatelli | Consumer | A |

Comentários

Esse é caso em que falta é necessário supor algumas informação. após fazer a conta de performance, para os caso de houver mais de uma categoria vendida por consumidor, o resultado da performance é uma soma ou média? Nesse caso foi suposto uma média, pois faria mais sentido com as condições estabelecidas depois, em que converte o valor da performance em letras de "A" a "E"