

Faça o download dos arquivos TRAB2.zip

Neste arquivo temos dados de promoções e vendas realizadas em um rede de lojas

Este é um trabalho guiado opcional. Para cada um dos itens abaixo solicitados você deve gerar o código spark (SQL, PySpark, Scala) utilizado.

Colunas contidas no arquivo:

1. Faça a importação do arquivo TXT
2. Apresente o esquema da estrutura importada
3. Verifique a quantidade de linhas da estrutura importada

Resultado(s): 550068

4. Faça uma análise estatística descritiva básica(média, desvio padrão, quantidade, min, max) para a coluna Purchase (transforme a coluna purchase para o tipo de dados inteiro)

Resultado(s):

summary		PurchaseI
count		550068
mean	9263.96	
stddev	5023.06	
min		12
max		23961

5. Verifique o número de produtos distintos contidos nesta base de dados

Resultado(s): 3631

6. Verifique o número de pessoas que compraram os produtos por idade cruzados por gênero

Possíveis Resultado(s):

Age_Gender	F	M
0-17	5083	10019
46-50	13199	32502
18-25	24628	75032
36-45	27170	82843
55+	5083	16421
51-55	9894	28607
26-35	50752	168835

7. Crie outro data frame em que o valor da compras(purchase) foi maior que 20.000 e depois exiba o número de linhas:

Resultado(s): 12691

8. Crie outro data frame em que o valor da compras(purchase) foi maior que 20.000 e menor que 45000 somente para as mulheres('F'). Depois exiba o número de linhas:

Resultado(s): 2822

9. Encontre a média de compras (purchase) de cada uma das faixas de idade:

Resultado(s):

Age	AVG(PurchaseI)
51-55	9534.80
46-50	9208.62
0-17	8933.46
36-45	9331.35
26-35	9252.69
55+	9336.28
18-25	9169.66

10. Encontre a média de compras(purchase) de cada um por genero:

Resultado(s):

Gender	AVG(PurchaseI)
F	8734.56
M	9437.52

11. Faça uma simulação gerando um novo dataframe com a média de compras por idade aumentada em 10%:

Resultado(s):

Age	AVG(PurchaseI)	Pur_10_percent
51-55	9534.80	10488.28
46-50	9208.62	10129.48
0-17	8933.46	9826.81
36-45	9331.35	10264.48
26-35	9252.69	10177.95
55+	9336.28	10269.90
18-25	9169.66	10086.62