

Desafio Final

Entrega 12 fev em 23:59 **Pontos** 100 **Perguntas** 15
Disponível até 12 fev em 23:59 **Limite de tempo** Nenhum

Instruções

O Desafio Final está disponível!

1. Instruções para realizar o desafio

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Desafio Final".

Para iniciá-lo clique em "Fazer teste". Você tem somente **uma** tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você concluí-lo. Antes de enviar confira todas as questões.

Caso o teste seja iniciado e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independente do progresso no teste. Fique atento ao seu teste e ao prazo final, pois novas tentativas só serão concedidas em casos de questões médicas.

O gabarito será disponibilizado partir de sexta-feira, **12/02/2021**, às 23h59.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio Final - Bootcamp Engenharia de Dados.pdf](#)

Histórico de tentativas

| | Tentativa | Tempo | Pontuação |
|--------------|-----------------------------|---------------|------------|
| MAIS RECENTE | Tentativa 1 | 1.848 minutos | 100 de 100 |

⚠ As respostas corretas estarão disponíveis em 12 fev em 23:59.

Pontuação deste teste: **100** de 100

Enviado 6 fev em 23:30

Esta tentativa levou 1.848 minutos.

Pergunta 1

6,6 / 6,6 pts

Qual é a média de renda das pessoas presentes no seu banco de dados?

☒ 1809,3.

☐ 1796,5.

☐ 1733,6.

☐ 1932,4.

Pergunta 2

6,6 / 6,6 pts

Qual é a média de renda das pessoas que residem no Distrito Federal?

☐ 2456,3.

☐ 2899,5.

☐ 3107,4.

☒ 2795,9.

Pergunta 3

6,6 / 6,6 pts

Qual é a média de idade das pessoas que moram na região sudeste do país (dos que estão na sua amostra, é claro)?

☒ 2034,1.

☐ 1967,5.

☐ 1899,7.

☐ 2165,4.

Pergunta 4

6,6 / 6,6 pts

Qual é o estado brasileiro que possui a menor média de renda?

☐ Ceará.

☐ Maranhão.

☒ Alagoas.

☐ Piauí.

Pergunta 5

6,6 / 6,6 pts

Qual é o estado brasileiro que possui a maior média de escolaridade?

☐ São Paulo.

☒ Distrito Federal.

☐ Rio de Janeiro.

☐ Paraná.

Pergunta 6

6,6 / 6,6 pts

Qual é a média de escolaridade entre os homens do Mato Grosso?

☐ 12,0.

☐ 12,4.

☐ 11,8.

☒ Não é possível responder.

Pergunta 7

6,6 / 6,6 pts

Qual é a média de escolaridade entre as mulheres que moram no Paraná e estão entre 25 e 30 anos?

☐ 13,2.

☐ Não é possível responder.

☐ 12,1.

☒ 12,6.

Pergunta 8

6,6 / 6,6 pts

Qual é a média de renda para as pessoas na região Sul do país, que estão na força de trabalho e possuem entre 25 e 35 anos?

☐ 2104,6.

☐ 2012,3.

☐ 1984,6.

☒ 2231,9.

Pergunta 9**6,6 / 6,6 pts**

Qual é a renda média por mesorregião produzida no estado de MG?
(Para responder, você deve somar todas as rendas do estado e dividir pelo número de mesorregiões da UF.)

☐ 369846.☒ 371952.☐ 394763.☐ 403746.**Pergunta 10****6,6 / 6,6 pts**

Qual é a renda média das mulheres que residem na região Norte do Brasil, possuem graduação, tem idade entre 25 e 35 anos e são pretas ou pardas?

☐ 1196,3.☒ 1391,7.☐ 1213,5.☐ 1476,9.**Pergunta 11****6,6 / 6,6 pts**

Se você fosse realizar a limpeza inicial do seu dataset numa arquitetura distribuída utilizando Spark e Scala, qual seria a sintaxe CORRETA para selecionar mulheres de 20 a 40 anos?



```
pnadDF.filter(pnadDF("sexo") ===  
"Mulher").filter(pnadDF("idade").geq(25)).filter(pnadDF("idade").leq(35))
```



```
pnadDF.where(col("sexo") === "Mulher" && col("idade") >= 25 &&  
col("idade") <= 35)
```



Todas estão corretas.



```
pnadDF.where("sexo = 'Mulher' and idade >= 25 and idade <= 35")
```

Pergunta 12

6,6 / 6,6 pts

Para responder à pergunta 5, qual seria a sintaxe Scala que você implementaria?

```
pnadDF
```

```
.groupBy("uf")
```

```
.filter('sexo === "Mulher" && 'idade >= 20 && 'idade <= 40)
```

```
.agg(avg('anosesco).alias("med_esco"))
```

```
.orderBy('med_esco.desc)
```



```
.show()
```

```
pnadDF
```

```
.filter('sexo === "Mulher" && 'idade >= 20 && 'idade <= 40)
```

```
.groupBy("uf")
```

```
.agg(avg('anosesco).alias("med_esco"))
```

```
.orderBy('med_esco.desc)
```



```
.show()
```



Nenhuma das anteriores.

```
pnadDF
```

```
.groupBy("uf")
```

```
.agg(avg('anosesco').alias("med_esco"))
```

```
.filter('sexo === "Mulher" && 'idade > 20 && 'idade <= 40)
```

```
.orderBy('med_esco.desc)
```

☐ .show()

Pergunta 13

6,6 / 6,6 pts

Fazer a implantação do Airflow em vários containers que se comunicam ao invés de uma grande aplicação monolítica sendo executada em um mesmo local é uma sofisticação que está relacionada a estes princípios de desenho arquitetural, EXCETO:

☐ Recuperação automatizada contra falha.

☐ Alta disponibilidade.

☐ Escalabilidade.

☒ Otimização de custos.

Pergunta 14

6,6 / 6,6 pts

Qual das opções abaixo NÃO é uma alternativa viável para a construção de um Data Warehouse?

☐ Amazon RedShift.

☐ Google BigQuery.

☒ MongoDB.

☐ Azure SQL.

Pergunta 15

7,6 / 7,6 pts

Qual é a máxima prioridade do trabalho do Engenheiro de Dados no seu dia a dia?

- ☐ Garantir escalabilidade das soluções.
- ☐ Garantir a melhor performance dos pipelines.
- ☒ Garantir a qualidade e consistência dos dados da organização.
- ☐ Garantir otimização de custos com infraestrutura e nuvem.

Pontuação do teste: **100** de 100