

# Desafio Final

**Entrega** 7 mai em 23:59**Pontos** 100**Perguntas** 15**Disponível** até 7 mai 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, **07/05/2021** às 23h59.

Bons estudos!

### 2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio Final - Bootcamp Engenheiro\(a\) de Dados.pdf](#)

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<a href="#">Tentativa 1</a>	6 minutos	100 de 100

⚠ As respostas corretas estarão disponíveis em 7 mai em 23:59.

Pontuação deste teste: **100** de 100

Enviado 1 mai em 13:01

Esta tentativa levou 6 minutos.

**Pergunta 1****6,6 / 6,6 pts**

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

☒ 1809,3.☐ 1733,6.☐ 1796,5.☐ 1932,4.**Pergunta 2****6,6 / 6,6 pts**

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

☐ 2899,5.☒ 2795,9.☐ 3107,4.☐ 2456,3.**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)?

☐ 1899,7.

☐ 1967,5.☐ 2165,4.☒ 2034,1.**Pergunta 4****6,6 / 6,6 pts**

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

☐ Piauí.☐ Maranhão.☐ Ceará.☒ Alagoas.**Pergunta 5****6,6 / 6,6 pts**

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

☐ Rio de Janeiro.☐ Paraná.☒ Distrito Federal.☐ São Paulo.**Pergunta 6****6,6 / 6,6 pts**

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

- ☐ 11,8.
- ☐ 12,0.
- ☒ Não é possível responder.
- ☐ 12,4.

### 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?

- ☐ Não é possível responder.
- ☒ 12,6.
- ☐ 13,2.
- ☐ 12,1.

### 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?

- ☐ 2012,3.
- ☐ 2104,6.
- ☒ 2231,9.

☐ 1984,6.

### 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.)

☐ 394763.

☒ 371952.

☐ 403746.

☐ 369846.

### 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?

☐ 1476,9.

☐ 1213,5.

☐ 1196,3.

☒ 1391,7.

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

☐

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

☐

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

☒

Todas estão corretas.

## Pergunta 12

6,6 / 6,6 pts

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

☐

Nenhuma das anteriores.

```
pnadDF
```

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

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

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

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

☒

```
.show()
```

pnadDF

.groupBy("uf")

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

.agg(avg('anosesco').alias("med\_esco"))

.orderBy('med\_esco.desc)

☐ .show()

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:

- ☐ Escalabilidade.
- ☐ Recuperação automatizada contra falha.
- ☐ Alta disponibilidade.
- ☒ 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?

- ☒ MongoDB.
- ☐ Google BigQuery.
- ☐ Azure SQL.
- ☐ Amazon RedShift.

### Pergunta 15

7,6 / 7,6 pts

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

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

Pontuação do teste: **100** de 100