

# Desafio do Módulo 2

**Entrega** 17 nov em 21:00

**Pontos** 40

**Perguntas** 15

**Disponível** até 17 nov em 21:00

**Limite de tempo** Nenhum

## Instruções

O Desafio do Módulo 2 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 e leia as orientações e enunciados com atenção. Em caso de dúvidas, utilize o "Fórum de dúvidas do Desafio do Módulo 2".

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

Lembre-se de conferir todas as questões antes de enviar a atividade, e clique em "Enviar teste" **somente** quando você tiver terminado.

Uma vez terminado o prazo de entrega, a plataforma enviará as tentativas não finalizadas automaticamente. Por isso, fique atento ao prazo final.

Novas tentativas só serão concedidas perante a apresentação de atestado médico.

O gabarito será disponibilizado a partir de sexta, **18/11/2022**, às 23h59.

Bons estudos!

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

[Enunciado Desafio - Módulo 2 - Bootcamp Cientista de Dados de Renda Fixa.pdf](#)

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<u>Tentativa 1</u>	11 minutos	40 de 40

⚠ As respostas corretas estarão disponíveis em 18 nov em 23:59.

Pontuação deste teste: **40** de 40

Enviado 17 nov em 18:29

Esta tentativa levou 11 minutos.

### Pergunta 1

2,67 / 2,67 pts

Uma corretora deseja iniciar um novo departamento que tenha a responsabilidade de, a partir de dados obtidos de diversas áreas da empresa, entender as variáveis que mais influenciam na captação de novos clientes e recursos, no *churn* (cancelamento do serviço), resultados de pesquisas de satisfação (NPS) e, com isso, gerar insights e relatórios internos para a tomada de decisões estratégicas junto aos clientes. Qual será o tipo de processamento de dados dominante nesse departamento?

- ☐ Online Transaction Processing (OLTP)
- ☐ Não relacional
- ☐ Nenhuma das anteriores
- ☒ Online Analytical Processing (OLAP)

### Pergunta 2

2,67 / 2,67 pts

Uma instituição financeira deseja avaliar o risco associado a um determinado título de dívida emitido por uma grande corporação do ramo de alimentos. Para tanto, opta por avaliar as relações existentes entre essa empresa e diversos outros *players* da indústria: quem são os sócios de cada empresa e qual a participação de cada um deles nelas; quais os principais fornecedores de matéria prima e o fluxo financeiro entre as empresas e seus fornecedores; de quem os principais distribuidores estão comprando e em qual quantidade; etc. Que modelo de dados mais se adequa à representação desse conjunto de relações entre diferentes entidades?

- ☐ Nenhuma das demais alternativas
- ☒ Modelo baseado em grafos

☐ Modelo baseado em documentos

☐ Modelo relacional

### Pergunta 3

2,67 / 2,67 pts

Em um projeto junto ao departamento de risco de uma seguradora, a equipe de cientistas de dados desenvolveu uma solução que, apesar de funcional, não resolvia exatamente a dor central que o departamento buscava solucionar com esse projeto. Dentro do framework do CRISP-DM, que etapa do projeto não foi satisfatoriamente desenvolvida para que a solução final não atendesse às necessidades da área?

☐ Modelagem

☐ Entendimento dos dados

☐ Implantação

☒ Entendimento do negócio

### Pergunta 4

2,67 / 2,67 pts

Qual o termo utilizado em Processamento de Linguagem Natural para um conjunto de palavras com pouco ou nenhum valor informativo dentro de determinado texto?

☐ Nenhuma das demais alternativas

☐ Bigramas

☒ Stopwords

☐ Tokens

### Pergunta 5

2,67 / 2,67 pts

Qual o significado do framework CRISP-DM?

- ☒ Cross Industry Standard Process for Data Mining
- ☐ Cross International Standard Process for Data Modeling
- ☐ Cross Industry Standard Process for Delivering Models
- ☐ Cross Industry Science and Process for Data Modeling

### Pergunta 6

2,67 / 2,67 pts

Qual dos modelos de dados estudado é caracterizado por maior versatilidade na organização dos dados (*schema on read*) e por localidade do dado.

- ☐ Nenhuma das demais alternativas
- ☒ Modelo baseado em documentos
- ☐ Modelo relacional
- ☐ Modelo baseado em grafos

### Pergunta 7

2,67 / 2,67 pts

Quais das características abaixo podem ser associadas aos processamentos OLTP e OLAP?



OLTP: leitura de pequena quantidade de registros / OLAP: agregação sobre grande quantidade de dados



OLTP: cargas de dados via ETL / OLAP: baixa latência como padrão de escrita OLTP: usado para análise sobre histórico de eventos ao longo do tempo / OLAP: usado na interação do usuário final com aplicações



Nenhuma das demais alternativas

### Pergunta 8

2,67 / 2,67 pts

Como podemos caracterizar o framework CRISP-DM?



Processo linear em que cada etapa só é executada uma única vez, de modo a evitar retrabalhos futuros



Processo cíclico e iterativo no qual constantemente se retorna a etapas anteriores do fluxo para aprofundamentos, correções, etc.



Processo no qual não se leva em consideração o entendimento do negócio, dado que o foco está totalmente no dado e na tecnologia



Framework que ignora a implantação da solução, dado que esta é uma atribuição relacionado aos times de tecnologia

**Pergunta 9****2,67 / 2,67 pts**

Após retiradas as pontuações e stop words (da mesma maneira que se fez nas aulas) e tomando-se a última ata do Copom de 2021 para análise, qual o número de tokens únicos e número total de tokens encontrado no texto?

- ☐ 405 e 988, respectivamente
- ☐ 721 e 1.145, respectivamente
- ☐ 392 e 1.042, respectivamente
- ☒ 510 e 839, respectivamente

**Pergunta 10****2,67 / 2,67 pts**

Qual pesquisa serve como um dos parâmetros para decisões do Copom e que, portanto, é recorrentemente citado nas atas, aparecendo frequentemente entre os principais bigramas nas atas do Copom?

- ☒ Pesquisa Focus
- ☐ Pesquisa Nacional por Amostra de Domicílios (PNAD)
- ☐ Pesquisa Mensal de Emprego (PME)
- ☐ Pesquisa Eleitoral

**Pergunta 11****2,67 / 2,67 pts**

Qual o tipo de dado envolvido no desafio prático?

- ☒ Dado não estruturado
- ☐ Nenhuma das demais alternativas
- ☐ Dado estruturado
- ☐ Dado semi-estruturado

**Pergunta 12****2,67 / 2,67 pts**

Quais dos serviços de armazenamento da AWS cobertos nas aulas pode ser utilizado para armazenar os dados do desafio?

- ☐ Amazon Athena
- ☐ Amazon SageMaker
- ☒ Amazon S3
- ☐ Amazon Glue

**Pergunta 13****2,67 / 2,67 pts**

Qual a média de palavras distintas por ata em cada nos anos de 2019, 2020 e 2021, respectivamente, após retirados todos os *stopwords* e pontuações? (números arredondados)

- ☐ 671, 612 e 674
- ☒ 567, 542 e 525
- ☐ 757, 722 e 798
- ☐ 474, 418 e 455

**Pergunta 14****2,67 / 2,67 pts**

Ao eliminarmos Stop Words da mesma maneira que fizemos nas aulas práticas de explicação do desafio, qual a palavra mais frequente nas atas do Copom entre 2019 e 2021?

☐ Monetária☐ Economia☐ Cenário☒ Inflação**Pergunta 15****2,62 / 2,62 pts**

Em que mês/ano a ata do Copom começa a citar o termo “pandemia” entre as 10 palavras mais frequentes e em que mês/ano o número de citações à pandemia atinge o seu pico?

☐ Janeiro/2020 e Junho/2020☐ Março/2020 e Junho/2021☒ Março/2020 e Junho/2020☐ Dezembro/2019 e Junho/2020**Pontuação do teste: 40 de 40**