

Desafio do Módulo 1

Entrega 18 mar em 21:00	Pontos 40	Perguntas 15
Disponível até 18 mar em 21:00	Limite de tempo Nenhum	

Instruções

O Desafio do Módulo 1 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 1".

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 partir de sexta, **19/03/2021**, às 23h59.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio – Módulo 1 – Bootcamp Cientista de Dados.pdf](#)

[Consumo.csv](#)

[Apoio_Desafio.ipynb](#)

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	71 minutos	40 de 40

⚠ As respostas corretas estarão disponíveis em 19 mar em 23:59.

Pontuação deste teste: **40** de 40

Enviado 18 mar em 10:52

Esta tentativa levou 71 minutos.

Pergunta 1

2,67 / 2,67 pts

A princípio, quais eram os três V's do Big Data?

- ☐ Velocidade, Visibilidade, Validade.
- ☒ Volume, Velocidade, Variedade.
- ☐ Volume, Velocidade, Variabilidade.
- ☐ Volume, Valor, Volatilidade.

Pergunta 2

2,67 / 2,67 pts

Quais são os tipos de aprendizagens existentes em Machine Learning?

- ☒ Supervisionada, não supervisionada, por reforço.
- ☐ Underfitting, Overfitting, Árvore de decisão.
- ☐ Underfitting, supervisionada, não supervisionada.
- ☐ Overfitting, não supervisionada, por reforço.

Pergunta 3

2,67 / 2,67 pts

Em linhas gerais, quais habilidades constroem um bom cientista de dados?



Programação, matemática e estatística, além de conhecimento de negócio. Esses três pilares, juntos, diferenciam o cientista de dados de um desenvolvedor de software, machine learning ou pesquisador tradicional.



Conhecimentos em matemática, estatística e ciência da computação, com programação.



Basta ter conhecimento de negócio.



Programação, tecnologia e facilidade com problemas de negócio.

Pergunta 4

2,67 / 2,67 pts

Qual das bibliotecas abaixo é a mais recomendada para gerar o gráfico de dispersão de Salários x Gastos do exemplo?



Matplotlib.



Sklearn.



Pandas.



AptSources.

Pergunta 5

2,67 / 2,67 pts

Qual é a ideia principal do Self Service Analytics?

☐

Somente a área de tecnologia, elaborar projetos de Data Science num núcleo central focado no atendimento das demandas de negócio, sem a participação dos profissionais de negócio.

☒

Todas as informações necessárias estarem disponíveis para as áreas de negócio, sem o contato direto com a TI, possibilitando uma tomada de decisão muito mais ágil.

☐

Nenhuma das alternativas.

☐

Acesso irrestrito de todas as áreas de negócio em todos os dados.

Pergunta 6

2,67 / 2,67 pts

Para discutir a implantação de um Self Service Analytics é necessário que sejam discutidos os seguintes pontos, EXCETO:

☐

Treinamento e capacitação para os usuários realizarem o que for necessário.

☐

Melhor organização dos dados e disponibilidade.

☐

Privacidade dos dados.

☒

Disponibilizar apenas ferramentas de visualização.

Pergunta 7

2,67 / 2,67 pts

Qual é o objetivo do algoritmo K-Means?

- ☐ Classificação.
- ☐ Correlação.
- ☒ Clusterização.
- ☐ Regressão.

Pergunta 8

2,67 / 2,67 pts

O algoritmo K-Means é baseado em qual tipo de aprendizagem?

- ☒ Relacionamento entre variáveis.
- ☐ Segmentação de mercado.
- ☐ Agrupamento de notícias e documentos.
- ☐ Agrupamento de clientes/usuários similares.

Pergunta 9

2,67 / 2,67 pts

É um dos principais usos do algoritmo K-Means, EXCETO:

- ☒ Não supervisionada.

☐ Nenhuma das anteriores.

☐ Por reforço.

☐ Supervisionada.

Pergunta 10

2,67 / 2,67 pts

Qual é o primeiro passo do K-Means?



Inicializar os centroides aleatoriamente (necessário saber o valor de k antes de começar).



Calcular a média de todos os pontos relacionados a um centroide e definir um novo centroide.



Para cada ponto, calcular a distância para cada centroide e associar ao que está mais próximo.



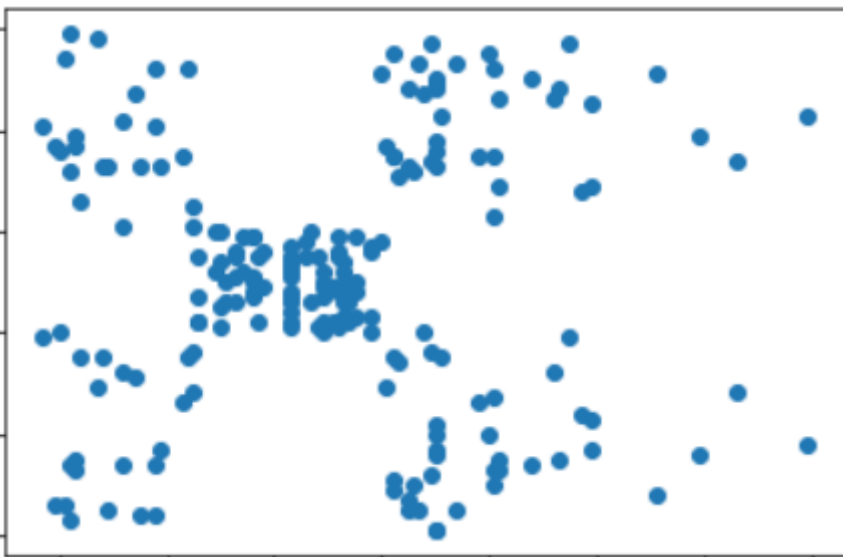
Recalcular as instâncias, uma vez que pode ser que o centroide mude de cluster.

Pergunta 11

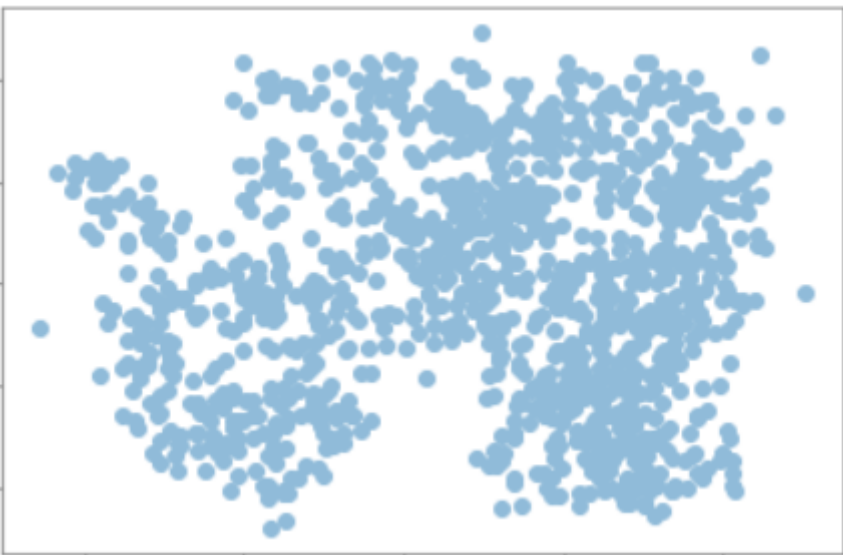
2,67 / 2,67 pts

Considerando o dataset gerado a partir das instruções do enunciado, qual das imagens representa o gráfico CORRETO?

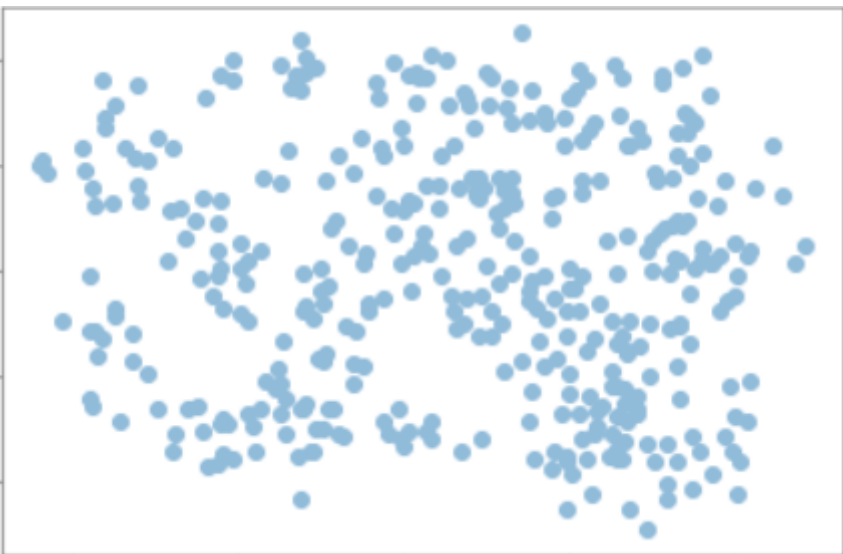
○

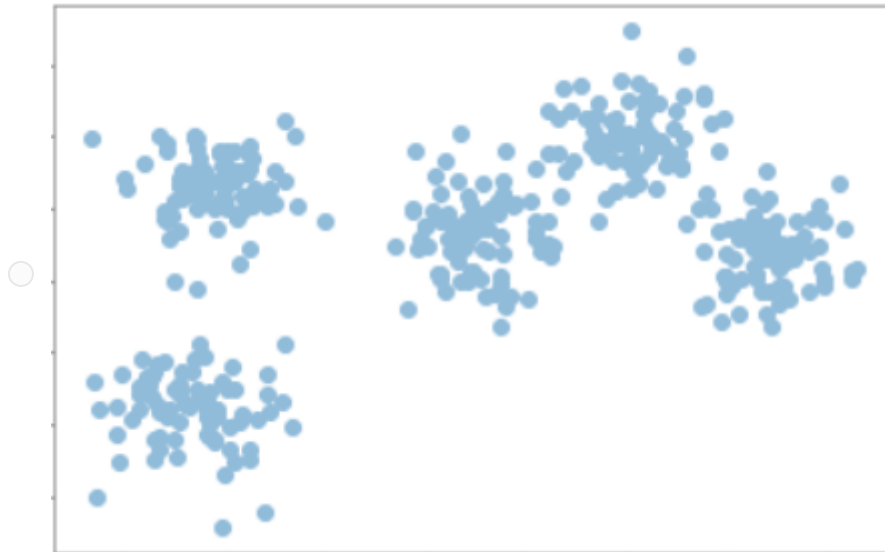


○



○





Pergunta 12

2,67 / 2,67 pts

Considere o dataset gerado a partir das instruções do enunciado. Caso o algoritmo K-Means seja executado buscando apenas 1 cluster, qual será a coordenada do centroide ao final da execução?

☐ [32.50 86.1]

☐ [10.16 90.6]

☒ [70.14 50.2]

☐ [60.12 80.7]

Pergunta 13

2,67 / 2,67 pts

Considere o dataset gerado a partir das instruções do enunciado. Caso o algoritmo seja executado buscando 6 clusters, qual será a soma dos quadrados intra-clusters (wcss) ao final da execução?

- ☐ 7934.8795858501
- ☒ 44532.0999700246
- ☐ 4218.057891958863
- ☐ 27110.400194118167

Pergunta 14

2,67 / 2,67 pts

De acordo com a curva do cotovelo e com as configurações das iterações do kmeans apontadas no enunciado, qual foi o número ideal de clusters indicado pelo método para essa base Consumo?

- ☒ 5.
- ☐ 1.
- ☐ 2.
- ☐ 3.

Pergunta 15

2,62 / 2,62 pts

Qual biblioteca do python já possui o Kmeans implementado?

- ☐ Matplotlib.
- ☐ Numpy.
- ☐ Pandas.

☒ Sklearn.

Pontuação do teste: **40** de 40