

Desafio do Módulo 2

Entrega 9 ago em 23:59 **Pontos** 40 **Perguntas** 15 **Disponível** até 9 ago em 23:59
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 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".

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.

O gabarito será disponibilizado partir de domingo, **09/08/2020**, às 23h59.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio - Módulo 2 - Cientista de Dados.pdf](#) 

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	37 minutos	40 de 40

⚠ As respostas corretas estarão disponíveis em 9 ago em 23:59.

Pontuação deste teste: **40** de 40

Enviado 4 ago em 2:40

Esta tentativa levou 37 minutos.

Pergunta 1

2,67 / 2,67 pts

Em Python, a função `sentiment.polarity` do pacote `TextBlob` foi usada para calcular a polaridade do tweets. A função retorna um valor entre -1 e 1 sobre um determinado texto. Sobre a polaridade calculada por esta função, podemos afirmar que:

☐ Uma polaridade for maior que 0(zero), indica que o texto tem uma tendência indicar um sentimento negativo.

☐ Uma polaridade for menor que 0(zero), indica que o texto tem uma tendência indicar um sentimento positivo.

☐ As polaridades -1 e 1 podem ser consideradas polaridades neutras.

☒ Quando uma polaridade for maior que 0(zero), indica que o texto tem uma tendência indicar um sentimento positivo.

Pergunta 2

2,67 / 2,67 pts

2. Considere a função `search` do pacote `tweepy` para coletar tweets que atendam a condição de possuir uma determinada palavra chave conforme exemplo abaixo. Qual das assertivas abaixo representam um valor inválido que parâmetro `result_type` da função pode assumir?

`search(q=keyword, result_type=?)`

- ☒ trend
- ☐ popular
- ☐ recent
- ☐ mixed

Pergunta 3

2,67 / 2,67 pts

Considere a função `search` do pacote `tweepy` para coletar tweets que atendam a condição de possuir uma determinada palavra chave conforme exemplo abaixo. Qual das assertivas abaixo está CORRETA?

- ☐ Quando usado o parâmetro `tweet_mode='extended'` será recuperado, além dos tweets, as respostas ou comentários a cada um dos tweets.
- ☒ O parâmetro `tweet_mode='extended'` indica que deve ser recuperado o texto total do tweet e não o texto truncado.
- ☐ Para recuperar os trend tweets deve-se usar o parâmetro `tweet_mode='extended'` na função `search`.
- ☐ O parâmetro `tweet_mode='extended'` indica que deve ser recuperado o texto truncado do tweet.

Pergunta 4

2,67 / 2,67 pts

Sobre a API do Twitter para coletar de dados, qual das assertivas abaixo está INCORRETA?

- ☐ São necessárias 4 chaves de acesso para usar a API, são elas: API key, API secret key, Access token e Access token secret.
- ☐ Para solicitar uma conta de desenvolvedor é necessário ser um usuário cadastrado, ou seja, ter uma conta no Twitter.
- ☐ Após receber o acesso de desenvolvedor é necessário criar uma aplicação para receber as chaves de acesso.
- ☒ Não é necessário ter uma conta no Twitter, basta cadastrar um e-mail no Twitter e solicitar uma conta de desenvolvedor.

Pergunta 5

2,67 / 2,67 pts

Considere o trecho de código abaixo no Python referente à análise de sentimento dos *tweets*. Tal exemplo foi apresentado na videoaula e na Segunda Aula Interativa. Marque a alternativa INCORRETA.

```
1 polarities = []
2
3 for tweet in tweets:
4     analysis = tb(tweet.text)
5
6     if analysis.detect_language() != 'en':
7
8         traducao = tb(str(analysis.translate(to='en')))
9
10        polarity = traducao.sentiment.polarity
11    else:
12        polarity = analysis.sentiment.polarity
13
14    polarities.append(polarity)
```

- ☒ A linha 4 realiza a análise de polaridade do texto do tweet e passa o valor encontrado para a variável analysis.
- ☐ A linha 6 representa uma estrutura condicional e análise o idioma no qual o tweet foi escrito.
- ☐ Na linha 10, a variável polarity recebe o valor da polaridade do texto do tweet já traduzido, no qual é obtido pela função sentiment.polarity.
- ☐ A linha 14 refere-se a uma estrutura que insere valores em uma lista existente.

Pergunta 6

2,67 / 2,67 pts

Considere o trecho de código abaixo e marque a alternativa CORRETA. Tal exemplo foi apresentado na videoaula e na Segunda Aula Interativa.

```
1 #Definição do dataframe com informação básicas dos tweets
2 tweets_df = pd.DataFrame(tweets, columns=['Tweets'])
3
4 tweets_df['len'] = np.array([len(tweet) for tweet in tweets])
5 tweets_df['ID'] = np.array([tweet.id for tweet in info])
6 tweets_df['USER'] = np.array([tweet.user.screen_name for tweet in info])
7 tweets_df['userName'] = np.array([tweet.user.name for tweet in info])
8 tweets_df['User Location'] = np.array([tweet.user.location for tweet in info])
9 tweets_df['Language'] = np.array([tweet.user.lang for tweet in info])
10 tweets_df['Date'] = np.array([tweet.created_at for tweet in info])
11 tweets_df['Source'] = np.array([tweet.source for tweet in info])
12 tweets_df['Likes'] = np.array([tweet.favorite_count for tweet in info])
13 tweets_df['Retweets'] = np.array([tweet.retweet_count for tweet in info])
14 tweets_df['Geo'] = np.array([tweet.geo for tweet in info])
15 tweets_df['Coordinates'] = np.array([tweet.coordinates for tweet in info])
16 tweets_df['Place'] = np.array([tweet.place for tweet in info])
17
18 likes_max = np.max(tweets_df['Likes'])
19 retweet_max = np.max(tweets_df['Retweets'])
20
```



Nas linhas 18 e 19 a função max do numpy identifica respectivamente o valor mínimo das colunas Likes e Retweets do dataframe.



Nas linhas 18 e 19 a função max do numpy identifica respectivamente o valor máximo das colunas Likes e Retweets do dataframe.



Nas linhas 18 e 19 a função max do numpy identifica respectivamente o valor total de dados existentes nas colunas Likes e Retweets do dataframe, ignorando aquelas no qual o valor é 0(zero).



Nas linhas 18 e 19 a função max do numpy identifica respectivamente a média de valores das colunas Likes e Retweets do dataframe.

Pergunta 7

2,67 / 2,67 pts

Considere o trecho de código abaixo, apresentado na videoaula e na Segunda Aula Interativa, e marque a alternativa CORRETA:

```

In [11]: 1 sources = [] #Lista para armazenar a fontes
          2 for source in tweets_df['Source']:
          3     if source not in sources:
          4         sources.append(source)
          5
          6 percent = np.zeros(len(sources))
          7
          8 for source in tweets_df['Source']:
          9     for index in range(len(sources)):
10         if source == sources[index]:
11             percent[index] += 1
12         pass

In [12]: 1 sourceDF = pd.DataFrame({
          2     'source':percent,
          3 }, index=sources)
          4
          5 sources_sorted = sourceDF.sort_values('source',ascending=False)
          6 ax = sources_sorted.source.plot(kind='barh',color='#B8860B')
          7 ax.get_xaxis().set_major_formatter(plt.FuncFormatter(lambda x, loc: "{:,}".format(int(x)

```

☐

Este trecho de código pode ser usado para identificar a fonte(origem) do tweet e gerar um gráfico apontando se o tweet da respectiva fonte é positivo ou negativo.

☒

Este trecho de código pode ser usado para identificar a fonte(origem) do tweet e gerar um gráfico apontando a quantidades de tweets por fonte.

☐

Este trecho de código pode ser usado para identificar a localização geográfica de onde o tweet foi postado.

☐

Este trecho de código pode ser usado para identificar o sentimento expresso pelo tweet.

Pergunta 8

2,67 / 2,67 pts

Considere o trecho de código abaixo e marque a alternativa CORRETA:

```

1
2 for tweet in tweets_df['Tweets']:
3     analysis = tb(tweet)
4
5     polarity = analysis.sentiment.polarity
6
7     polarities.append(polarity)
8

```

☒

Este trecho de código pode ser usado para identificar o sentimento expresso pelo tweet.

☐

Este trecho de código pode ser usado para identificar a localização geográfica de onde o tweet foi postado.

☐

Este trecho de código pode ser usado para identificar a fonte(origem) do tweet e gerar um gráfico apontando a quantidades de tweets por fonte.



Este trecho de código pode ser usado para identificar a fonte(origem) do tweet e gerar um gráfico apontando se o tweet da respectiva fonte é positivo ou negativo.

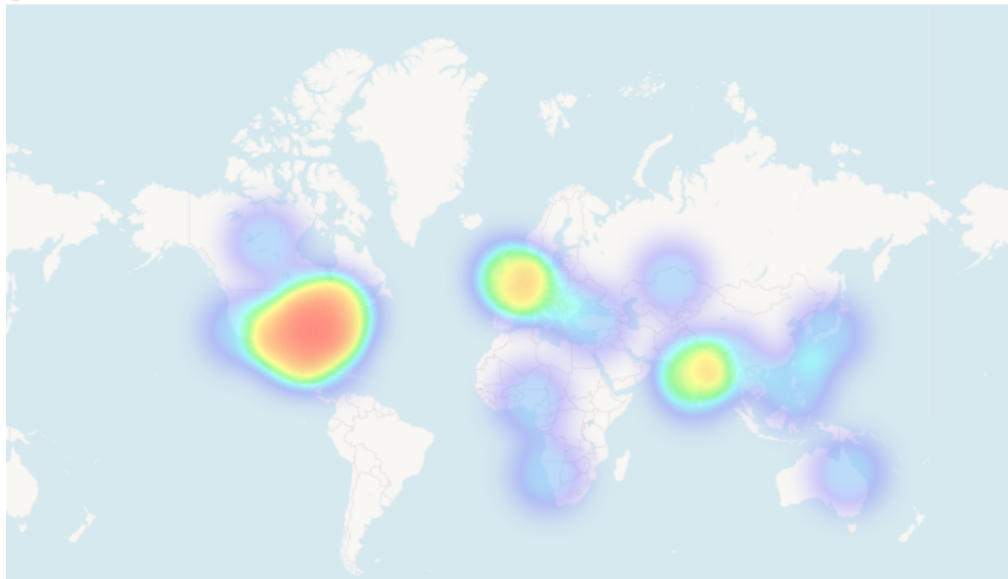
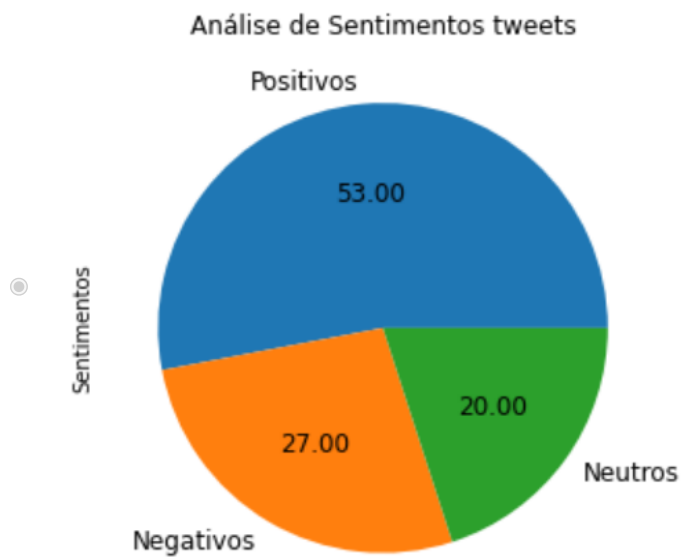
Pergunta 9

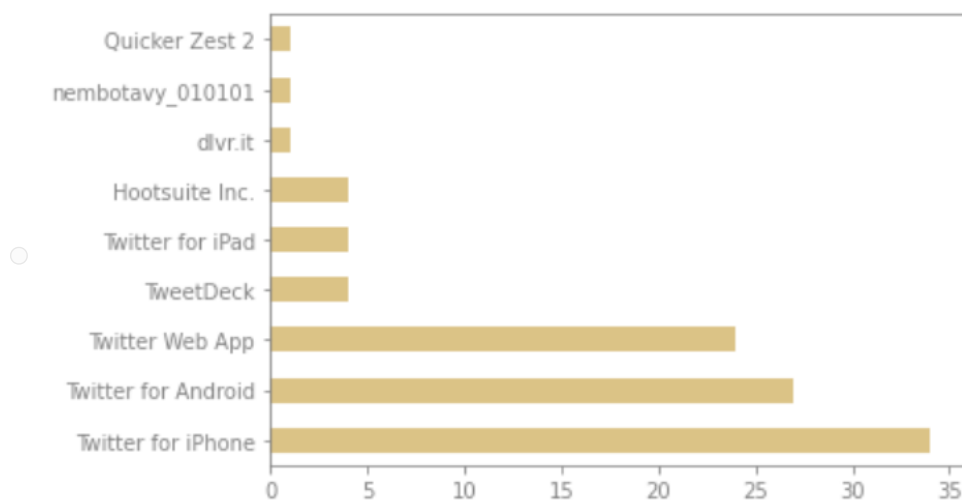
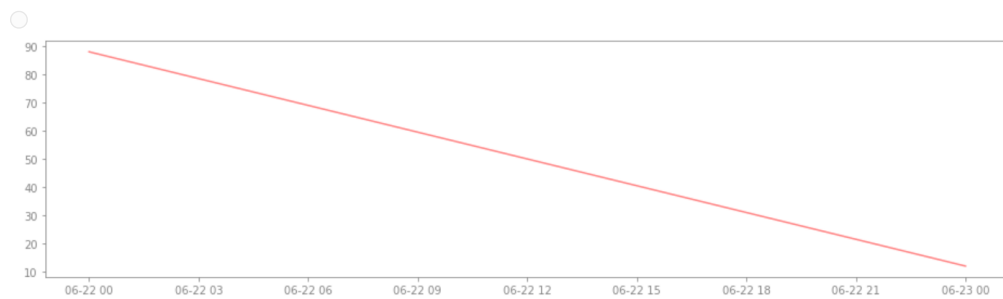
2,67 / 2,67 pts

Considere o trecho de código abaixo e marque a alternativa que indica o gráfico gerado pelo código:

```
In [32]: ▶ 1 sentiments = ['Positivos', 'Negativos', 'Neutros']  
          2 percents = [pos_pct, neg_pct, neu_pct]  
  
In [33]: ▶ 1 pie_chart = pd.Series(percents, index=sentiments, name='Sentimentos')  
          2 pie_chart.plot.pie(fontsize=12, autopct='%.2f', figsize=(5, 5), title="Análise de Sentimentos tweets")
```

Observação: os valores do gráfico podem ser diferentes, pois cada coletor de *tweet* pode coletar dados diferentes. O objetivo é identificar qual tipo de gráfico gerado



**Pergunta 10**

2,67 / 2,67 pts

Considere o trecho de código abaixo e marque a alternativa INCORRETA:

```

3 class Item(scrapy.Item):
4     titulo_pagina = scrapy.Field()
5     url_pagina = scrapy.Field()
6     categoria = scrapy.Field()
7
8 class blogSpider(scrapy.Spider):
9     name = 'IGTIBlog'
10    start_urls = ['http://www.igti.com.br/blog/']
11
12    def parse(self, response):
13        categories = response.xpath("//nav[@class='gridlove-main-navigation']//li//a")
14        item = Item()
15        for category in categories:
16            url = category.xpath('@href').extract_first()
17            self.log('Categoria %s' % category.xpath('text()').extract_first())
18            yield response.follow(url, self.parse_category)
19        item['titulo_pagina'] = response.css("title ::text").extract_first()
20        item['url_pagina'] = response.url
21        item['categoria'] = 'HOME'
22        yield item
23
24    def parse_category(self, response):
25        item = Item()
26        item['titulo_pagina'] = response.css("title ::text").extract_first()
27        item['url_pagina'] = response.url
28        item['categoria'] = response.xpath("//h1[@class='h2']/text()").extract_first()
29        yield item

```



Os métodos parse(self, response) e parse_category(self, response) podem ser considerados métodos de call-back do spider.



A classe Item() é a classe de representação da informação do seu crawling ou scraping.



A instrução yield item das linhas 22 e 29 correspondem a chamada de rastreamento que o spider assina para navegar entre as páginas encontradas.



A instrução response.url das linhas 20 e 27 são usadas para coletar a url da página corrente de seu spider.

Pergunta 11

2,67 / 2,67 pts

Considere as colunas 1 e 2 a seguir e marque a assertiva que melhor representa a correspondência entre as colunas.

Coluna 1

- 1- Web Crawling
- 2- Web Scraping
- 3- Web content mining
- 4- Web usage mining

Coluna 2

- a) É o processo de recuperar automaticamente informações específicas de um documento da web e coletar informações dele.
- b) É o processo de obtenção de conhecimento a partir da organização da web e da referência cruzada de ligações entre os recursos da web.

5- Web structure mining

c) É o processo de localização de informações na web, indexando todas as palavras de um documento web, salvando estas palavras, e seguindo todos os hiperlinks e índices.

d) É o processo de descobrir conhecimento a partir do conteúdo das páginas disponíveis na web.

e) É o processo de descoberta de conhecimento a partir de padrões de uso e dos registros de acesso da web.

☐ 1-a, 2-e, 3-d,4-b, 5-c

☐ 1-b, 2-c, 3-e,4-d, 5-b

☐ 1-d, 2-a, 3-b,4-e, 5-c

☒ 1-c, 2-a, 3-d,4-e, 5-b

Pergunta 12

2,67 / 2,67 pts

Considere as colunas 1 e 2 a seguir e marque a assertiva que melhor representa a correspondência entre as colunas.

Coluna 1

- 1- Nltk
- 2- Scrapy
- 3- Tweepy
- 4- Textblob

Coluna 2

- a) Pacote usado no Python para acessar a API do Twitter.
- b) Pacote que pode ser usado para criar um projeto de raspagem de páginas web.
- c) Pacote usado para realizar atividade de mineração de texto na linguagem Python.
- d) Este pacote possui uma função que calcula a polaridade de um texto.

☒ 1-d, 2-b, 3-a, 4-d

☐ 1-a, 2-b, 3-d, 4-c

☐ 1-b, 2-c, 3-d, 4-a

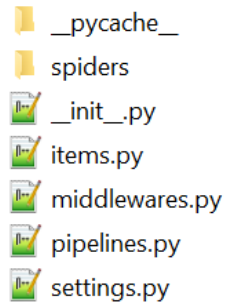
☐ 1-b, 2-a, 3-c, 4-d

Pergunta 13

2,67 / 2,67 pts

Em um projeto de web crawling criado pelo pacote scrapy da linguagem Python é gerada a seguinte estrutura de pastas e arquivos:

📁 > This PC > OS (C:) > Bootcamp > crawlerIGTI > crawlerIGTI



Considere as assertivas abaixo e marque a assertiva INCORRETA.

- ☐ No arquivo items.py são definidas as classes de representação da informação do seu crawling ou scraping.
- ☒ O arquivo pipelines.py é o arquivo destinado a implementação do spider do projeto.
- ☐ No arquivo middlewares.py é definido como a resposta http será processada e enviada para o spider.
- ☐ As configurações do seu projeto Scrapy devem ser definidas no arquivo settings.py.

Pergunta 14

2,67 / 2,67 pts

Sobre o pacote scrapy da linguagem Python, considere as assertivas abaixo e marque a assertiva CORRETA.

- ☒ A implementação do método start_requests() do Spider é opcional.
- ☐ O callback padrão do spider é definido pelo método request.follow() do objeto.
- ☐ Pode criar um Spider no Scrapy utilizando o comando scrapy runspider nome-do-spider.
- ☐ O comando scrapy runspider nome-do-spider serve para criar um projeto de scraping/crawling a partir do template padrão do pacote.

Pergunta 15

2,62 / 2,62 pts

O Scrapy usa objetos de Request e Response para rastrear sites. Considere as assertivas abaixo e marque a assertiva INCORRETA.

- ☐ Para recuperar o elemento title de uma página usando CSS podemos utilizar o seguinte comando:
response.css('title')
- ☐ O elemento response.url pode ser usado para recuperar a url de uma página.



Para recuperar o elemento title de uma página usando Xpath podemos utilizar o seguinte comando:
`response.xpath('//title')`



Utilizamos o elemento `response.follow(url,self.parse)` para recuperar a url de uma página.

Pontuação do teste: **40** de 40