

Exercícios Semana 01 - Andrea Sampaio Elias

PARTE 1: Sobre a profissão e o universo de dados

- 1. Descreva com suas palavras qual a diferença entre a profissão de cientista, analista e engenheiro de dados? Que tipo de tarefa você imagina que cada um desses faça?**

O analista de dados faz a coleta e a análise dos dados para que os mesmos possam ser utilizados e aproveitados. Já um engenheiro de dados é responsável pela arquitetura de dados, ele é responsável pela organização e manutenção dos dados. O cientista de dados, por sua vez, realiza análises mais aprofundadas dos dados, análises envolvendo machine learning e predição.

- 2. Por que um cientista de dados é, apesar de tudo, uma pessoa de negócios?**

Porque o trabalho do cientista de dados também é entender a dor do cliente e como é a melhor forma de ajudá-lo no seu negócio com os dados que ele tem. Bastante tempo do cientista de dados é gasto para entender problemas a serem resolvidos.

- 3. O que é Machine Learning? Que tipo de trabalho de Machine Learning poderia ser feito na área em que você atua? Mencione alguns estudos/modelos que te auxiliariam no dia a dia.**

Machine Learning refere-se à criação e ao uso de modelos que são aprendidos a partir dos dados. Normalmente, o objetivo é usar os dados existentes para desenvolver modelos que se possa usar para prever possíveis saídas para o dados. Um exemplo de modelo que nos ajuda no dia a dia é o que prevê se uma mensagem de e-mail é spam ou não ou o que prevê se uma transação do cartão de crédito é fraudulenta. Modelos como esses impedem que as pessoas sejam enganadas e que acabem caindo em golpes. Na engenharia de computação, área em que atuo, o machine learning pode ser visto na produção de modelos de carros autônomos, ou seja, carros programados com modelos de machine learning e que têm a capacidade de se autodirigir.

4. Qual a utilidade da estatística para alguém que trabalha com data science?

Data Science se faz com dados e os dados estão ligados à estatística. "A Estatística proporciona os meios e as ferramentas para encontrar estrutura em dados, e assim fornecer insights sobre as informações mais profundas ali escondidas. Medidas estatísticas como média, mediana, moda, desvio padrão e distribuição servem para descrever de uma forma generalizada o comportamento das variáveis de trabalho, assim como identificar anomalias. Se os modelos de machine learning são as ferramentas que os cientistas de dados operam, a Estatística é o conhecimento sobre como e porquê essas ferramentas funcionam, que permite escolher as ferramentas mais adequadas, e tirar o melhor proveito delas." Trecho retirado de: <https://iaexpert.academy/2020/09/14/importancia-da-estatistica-para-ciencia-de-dados/#:~:text=A%20Estat%C3%ADstica%20proporciona%20os%20meios,informa%C3%A7%C3%B5es%20mais%20profundas%20ali%20escondidas>.

PARTE 2: Conhecendo o Python e o Jupyter Notebook

5. Demonstre como realizar no Python as seguintes operações: soma, subtração, multiplicação, divisão, exponenciação e raiz quadrada.

Soma: operador +, exemplo: 2+2

Subtração: operador -, exemplo 2-2

Multiplicação: operador *, exemplo 2*2,

Exponenciação: operador **, exemplo 2**2

raiz quadrada: ou fazer a exponenciação por 0.5, exemplo 4**0.5, ou importar o módulo math e usar math.sqrt(x), exemplo: math.sqrt(4)

6. Qual operador utilizamos para obter o resto de uma divisão no Python?

O %. Por exemplo, para se saber o resto da divisão de 10 por 3, faz-se 10%3. O resultado obtido será o resto da divisão de 10 por 3, ou seja, 1.

7. Como verificar qual o tipo de algum objeto no Python?

Sendo x, o objeto, faz-se `type(x)`. O resultado será o tipo do objeto.

8. Qual a diferença entre o objeto `x = 2`, `y = 2.0` e `z = '2.0'`?

x é um inteiro, y é um float e z é uma string.

9. O que acontece se escrevermos `2++2` ao invés de `2+2`? E se fizermos `2 - - 2`? E `2==2` ou `2 2`? Verifique o output de cada situação.

`2++2` resulta em 4, `2 - - 2` também resulta em 4, `2==2` resulta em True e `2 2` resulta em erro de sintaxe (sintaxe inválida).

10. É possível escrever `02` ao invés de apenas `2` no Python?

Não, resulta num erro de sintaxe informando que zeros à esquerda em inteiros decimais não são permitidos. (SyntaxError: leading zeros in decimal integer literals are not permitted; use an 0o prefix for octal integers).

11. Crie uma variável chamada *bônus* (sem acento mesmo) que possui valor 50000. Digamos que esse é o registro do bônus dos trabalhadores da sua firma. A inflação no ano foi de 11%. Depois de criar a variável *bônus* em uma célula, altere seu valor na célula seguinte para que acompanhe a inflação. I.e., ajuste o valor da variável para que ele seja 11% acima de 50000 ($1,11 \times 50000$).

```
In [21]: bonus = 50000
```

```
In [22]: bonus = bonus * 1.11
```

```
In [23]: bonus
```

```
Out[23]: 55500.000000000001
```

PARTE 3: Questões éticas

12. Cite alguns problemas que a não-proteção dos dados pode trazer para as pessoas.

O vazamento de dados pode acarretar vários problemas àqueles que têm seus dados vazados. Dados vazados de cartão de crédito, por exemplo, podem acarretar

prejuízos ao dono dos dados, pois pessoas mal intencionadas podem utilizar esses dados para fazer compras sem pagarem elas mesmas pela compras. Dados vazados sobre a vida íntima de alguém podem resultar em prejuízos na vida social e profissional, pois há dados que se revelados acabam ocasionando que o dono dos dados sofram preconceito ou, até mesmo, chantagem.

13. Como o Machine Learning pode piorar a situação das pessoas?

Os algoritmos de machine learning são criados usando dados do passado em seu treinamento. O que pode ocorrer é que o algoritmo, por usar dados passados, acabe atuando para repetir o passado, ou seja, criando padrões racistas e sexistas ao invés de atuar para combatê-los.

14. Cite um exemplo de risco do mau uso de Machine Learning na sua área de atuação ou formação.

Estou me formando no curso de Engenharia de Computação. Posso citar o caso da robô Tay, inteligência artificial da Microsoft. Em 24 horas de contato com usuários do twitter, a inteligência artificial passou a ter falas racistas, nazistas, machistas e transfóbicas, pois os usuários do twitter a treinaram dessa forma. Esse é um problema de deixar os algoritmos aprenderem por si próprios. Abaixo alguns exemplos de mensagens da robô Tay:



TayTweets ✓
@TayandYou



@brightonus33 Hitler was right I hate
the jews.

24/03/2016, 11:45



TayTweets ✓
@TayandYou



@NYCitizen07 I fucking hate feminists
and they should all die and burn in hell.

24/03/2016, 11:41

15. Que tipos de problemas o reconhecimento facial vem causando a certas minorias?

Como o algoritmo de reconhecimento facial é treinado com dados e a maioria dos dados que certos algoritmos de reconhecimento facial estão recebendo são em sua maioria de homens brancos, os algoritmos de reconhecimento facial muitas vezes não reconhecem pessoas negras de forma correta, tendo ainda mais dificuldade quando são mulheres negras. Isso pode dificultar o acesso dessas pessoas a sistemas que usem reconhecimento facial (como em bancos, por exemplo). Além disso, muitos sistemas de reconhecimento facial identificam erroneamente as pessoas como sendo criminosas, quando na verdade elas não são, causando problemas principalmente para pessoas negras, por justamente os sistemas de reconhecimento facial detectarem rostos de pessoas negras com menor acurácia.