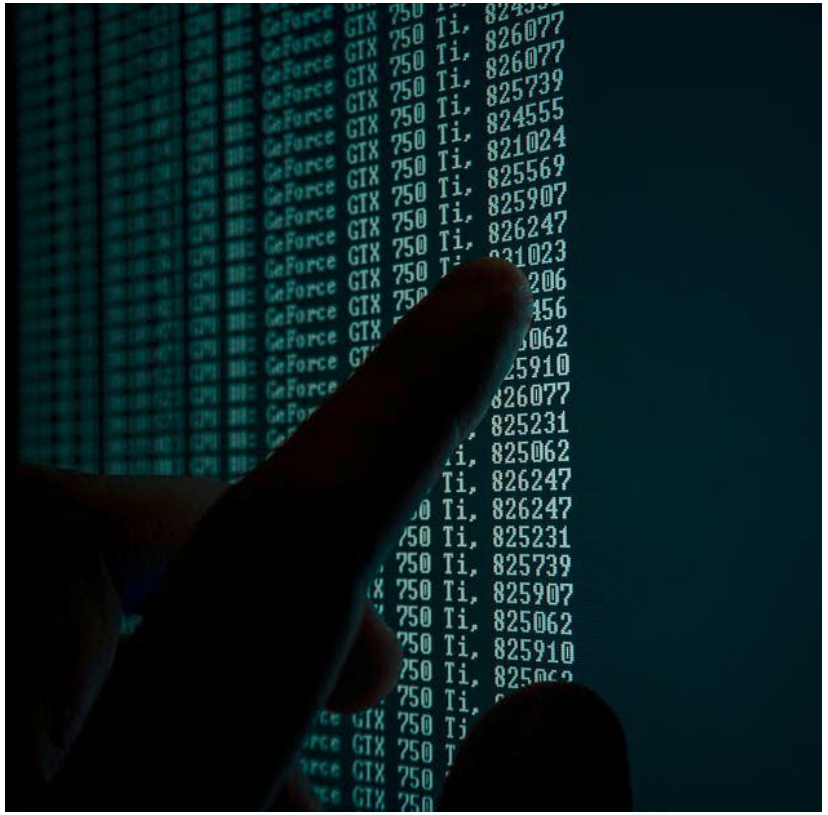
Aula de revisão para a prova

Dados

O que são dados?

Fatos e números coletados, analisados e sintetizados para a apresentação e interpretação.



Tipos de variáveis

Para cada tipo de variável, as funções tem comportamento diferente.

- Texto (str)
- Número inteiro (int)
- Número real (float)
- Booleano (bool), True ou False

Listas

Declarando listas

Uma lista pode ter vários argumentos de diferentes tipos.

```
minhaLista = [‘eu’, ‘amo’, ‘Python’]
minhaLista[0]
```

Declarando listas

Modificando elementos de uma lista

```
minhaLista = [‘eu’, ‘amo’, ‘Python’]
minhaLista[2] = ‘software livre’
```

idp

Dicionários

Dicionários

Declarando um dicionário com duas entradas e acessando cada uma delas:

```
meuDict = {‘linguagem’: ’Python’, ’idade’: ‘28 anos’}
meuDict[‘linguagem’]
meuDict[‘idade’]
```

Controle de fluxo

Laços

Em português, chamamos “loop” de “laço”

- Palavra-chave **for**

```
for cliente in clientes:
    print(cliente)
    if cliente == “Álvaro”:
        print(“Te peguei!”)
```

Laços for ou “para cada”

Com **for**, o bloco irá executar para cada elemento do objeto.

```
for cliente in clientes:
    print(cliente)
```

idp

Laços while ou “enquanto”

O **while** executará o bloco “enquanto” a condição a direita dele resultar em verdadeiro, ou booleano True.

```
while nome == “Álvaro”:
    print(“Olá, professor!”)
```

idp

Blocos if... else ou “se... senão”

O **if** ou “se”, executa o bloco se a condição for satisfeita.

```
if nome == “Álvaro”:
    print(“Bom dia, professor”)
else:
    print(“Bom dia, querido colega”)
```

idp

Operadores de comparação

==	Igual	x == y
!=	Diferente	x != y
>	Maior que	x > y
<	Menor que	x < y
>=	Maior ou igual	x >= y
<=	Menor ou igual	x <= y

Operadores de identidade e pertencimento

in Verdadeiro se o valor está presente no objeto

x in y

not in Verdadeiro se o valor não está presente no objeto

x not in y

Pandas e DataFrames

Pandas e Dataframes

Vamos usar o Pandas para acessar os nossos dados sobre Pokemons.

```
import pandas as pd
file = “pokemon_data.csv”
df = pd.read_csv(file)
print(df)
```

idp

Indexação de DataFrames

O DataFrames é organizado em colunas que podem ser acessadas a partir de seu nome.

Para determinar os nomes das colunas, pode-se escrever na tela os primeiros valores com a função columns

```
df.columns
```

Indexação de DataFrames

Para selecionar uma coluna de um DataFrame, a sintaxe é a mesma de um dicionário, com o nome da coluna no lugar da chave:

```
df[“Name”]
```

Para selecionar elementos, use a indexação de listas:

```
df[“Name”][0]
```

idp

Indexação de DataFrames

Slices ou fatias são seleções de mais de um elemento de um objeto ao mesmo tempo.

A faixa de índices é indicada entre o sinal “:”.

Ou seja, para selecionar os dez primeiros elementos, usa-se:

```
df[“Name”][0:10]
```

Indexação de DataFrames

Para modificar o valor de um DataFrame, utilize o indicador .loc:

df.loc[0,“Name”] = “Álvaro”

Para selecionar elementos sem fazer referência ao nome da coluna, use a indexação de listas com .iloc:

df.iloc[0,0] = “Xerxes”

idp

Indexação de DataFrames

Para acessar um elemento ou mais a partir de uma operação, pode-se utilizar a seguinte sintaxe:

```
df[df[“Nome”] == “Xerxes”]
```

Isso vai retornar todas as linhas em que o valor da coluna “Nome” seja “Xerxes”.

idp

Funções de DataFrames

DataFrames possuem diversas funções para facilitar a análise exploratória e de estatística descritiva. Algumas das funções que vamos utilizar são:

- describe()
- plot()

Funções de DataFrames

Funções de estatística descritiva:

- describe()
- count()
- sum()
- mean()
- median()
- mode()
- std()
- min()
- max()
- abs()

Funções de DataFrames

Algumas funções importantes para DataFrames são:

- groupby()
- sort_values()
- filter()
- value_counts()
- columns
- head()
- tail()
- values

Tratamento e limpeza de dados

Qualidade de dados

Dimensões da qualidade de dados:

- Completude ou integridade
- Conformidade
- Validade
- Acurácia e precisão

Qualidade de dados

Soluções para quando já não se pode corrigir a coleta:

- Descartar o registro ruim
- Atribuir um valor de sentinela
- Atribuir o valor médio ou mais frequente
- Calcular um valor substituto
- Atribuir um valor baseado em valores vizinhos

idp

INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA