**Projeto Pandas**

**Desenvolvido por: Danilo Morales Teixeira**

**Data de criação: 28/03/2019**

**Instruções**

Este repositório fornece exemplos utilizando a biblioteca PANDAS do Python, onde são fornecidas questões para o usuário desenvolver um programa em Python utilizando ou não o Jupiter Notebook. Cada exercício contém um arquivo em Python com a sua respectiva solução e um arquivo geral utilizando o Jupiter Notebook com a solução de todos os exercícios.

Dúvidas, comentários ou sugestões podem ser enviadas para o e-mail [danilomorales.astro@gmail.com](mailto:danilomorales.astro@gmail.com)

**Questões**

**Parte 1: Séries em Pandas**

1. Escreva um programa que utilize Pandas para criar e exibir um arranjo unidimensional contendo um arranjo de dados com pelo menos 5 valores.
2. Escreva um programa que converta uma série obtida através do módulo Pandas para uma lista. Por fim, exiba o tipo da série e da lista
3. Crie um programa que realize as operações de soma, subtração, multiplicação e divisão de duas séries Pandas onde cada série deverá ter pelo menos 5 elementos.
4. Escreva um programa que converta um dicionário simples para uma série em Pandas
5. Escreva um código que converta um arranjo Numpy para uma série em Pandas
6. Escreve um programa que converta uma série qualquer em Pandas para uma série do tipo numérico onde pelo menos um dos elementos da série original seja uma string.
7. Escreva um programa utilizando Pandas que converta a primeira coluna de um DataFrame para uma serie.
8. Escreva um programa que converta uma determinada série em Pandas para um arranjo.
9. Escreva um programa utilizando Pandas para converter séries de listas para uma única série.
10. Escreva um programa para ordenar uma série aleatória. Faça com que o mesmo programa ordene a serie de forma inversa.
11. Escreva um programa utilizando Pandas que adicione elementos numa série existente.
12. Crie um subconjunto a partir de uma série em Pandas utilizando a série e condições fornecidas.

Arranjo = [1,2,3,5,8,12,15,20,40,50,80,100,150,180,250,280,290,300]

1. Para valores menores do que 10
2. Para valores maiores do que 50
3. Para valores maiores do que 30 e menores do que 250
4. Para valores maiores igual a 8 e menor igual a 250
5. Escreva um programa que calcule a média e desvio padrão de uma série fornecida.

array = [1,2,4,5,6,7,8,9,12,14,15,18,19,25,29,35,38,40,45,48,49,50]

**Parte 2: Arranjos em Pandas**

1. Utilizando o dicionário e labels fornecidos abaixo, escreva um programa que utilize Pandas para realizar as seguintes tarefas:
2. Converta o dicionário para um arranjo utilizando índices como índice do arranjo
3. Exiba as informações básicas deste arranjo
4. Obtenha as três primeiras linhas do arranjo
5. Selecione as colunas nome e nota
6. Selecione as mesmas duas colunas, porém somente da terceira até sexta linha
7. Selecione as linhas onde o número de tentativas é maior do que 2
8. Conte o número de linhas e colunas
9. Selecione as linhas em que as notas estejam como NaN
10. Selecione as linhas em que as notas sejam maiores do que 5 e menores e iguais a 10
11. Selecione as linhas em que o número de tentativas seja menor do que 2 e a nota seja maior do que 6
12. Altera a nota da linha d para 7.0
13. Determine a soma do número de tentativas
14. Determine a média das notas
15. Adicione o seguinte aluno ao arranjo:

Nome:Aluno11, Nota = 6.7, Tentativas = 1, Aprovado = Sim

1. Remova o aluno 4
2. Organize o arranjo por nome em ordem decrescente e as notas por ordem crescente
3. Remova a coluna com o número de tentativas
4. Insira a seguinte coluna:

faltas = [0,3,1,2,2,0,0,4,1,0]

1. Crie um loop que itere sobre as linhas e colunas, exibindo apenas as colunas nome, nota e faltas.
2. Obtenha uma lista com os nomes das colunas

Dicionário:

Info\_alunos = {'nome' : ['Aluno1', 'Aluno2', 'Aluno3', 'Aluno4', 'Aluno5', 'Aluno6', 'Aluno7', 'Aluno8', 'Aluno9', 'Aluno10'],

'nota' : [7.5, 8.2, 3.4, np.nan, 8.8, 1.2, np.nan, 4.5, 6.7, 9.4],

'tentativas' : [2,1,3,0,2,1,0,1,3,2],

'aprovado' : ['Sim','Sim','Não','Não','Sim','Não','Não','Não','Sim','Sim']}

labels\_alunos = ['a', 'b','c','d','e','f','g','h','i','j']

1. Dado o arranjo abaixo, renomeie as colunas para:

Nome das colunas: [Coluna 1, Coluna 2, Coluna 3]

Arranjo = {‘A’ : [1,2,3,4], ‘B’ : [‘a’,’b’,’c’,’d’], ‘C’ : [‘aa’,’bb’,’cc’, ‘dd’]}

1. Escreva um programa que selecione linhas de um arranjo baseado no valor de alguma coluna.
2. Escreva um programa que mude a ordem das colunas de um arranjo.
3. Dado um arranjo com nome de pessoas e cidade onde residem, escreva um programa que conte o número de habitantes de cada cidade
4. Escreva um programa utilizando Pandas que remova linhas baseado num valor de uma determinada coluna
5. Escreva um programa utilizando apenas contenha com arranjo com três colunas e de linhas. Exiba somente 5 linhas deste arranjo e apenas duas colunas do mesmo.
6. Escreva um programa que exiba apenas uma linha do arranjo.
7. Escreva um programa utilizando Pandas que conte a quantidade de valores NaNs e substitua todos os valores eles por 0. Determine a média antes e depois da substituição.
8. Escreva um programa utilizando Pandas que combine duas séries
9. Converta um arranjo em Pandas no formato mês/dia/ano para o formato dia/hora.
10. Escreva um programa utilizando Pandas que crie um arranjo a partir de um arranjo NumPy e especifique os índices e nomes das colunas.
11. Escreva um programa que verifica se uma determinada coluna está presente no arranjo.
12. Utilizando o arranjo do exercício 14, escreva um programa utilizando Pandas que converta o tipo de dados da coluna das notas de float para inteiro. Converta os NaNs para 0.
13. Crie um programa utilizando Pandas que substitua valores infinitos por zero.

**Parte 3: Arquivos CSV**

1. Crie um programa utilizando Pandas que salve um determinado arranjo num arquivo do tipo CSV, utilizando vírgula como separador
2. Dados os arquivos Populacao\_Individual.csv e Populacao\_Total.csv localizados no diretório CSV, escreva um programa utilizando Pandas que realize as seguintes tarefas:
3. Leia e armazene os dados de ambos os arquivos num arranjo de dados
4. Exiba as informações de cada um deles
5. Exiba as cinco primeiras linhas
6. Exiba as cinco últimas linhas
7. Determine o número de elementos do tipo NaN
8. Exiba os nomes das colunas
9. Remova os países desnecessários (países presentes nas primeiras 33 linhas)
10. Exiba os 5 países mais populosos no ano de 2014
11. Crie um dicionário que armazene num arranjo os dados destes cinco países nos anos de 1960, 1970, 1980, 1990, 2000 e 2010
12. Na mesma figura faça um gráfico de barras horizontal e vertical da população destes países nos anos mencionados acima
13. Faça as mesmas operações com os 5 países menos populosos
14. Compute a população percentual os cinco países mais populosos em relação a população mundial
15. Compute os dados da população do Brasil
16. Faça um gráfico de espalhamento (scatter) da população do Brasil em função do do tempo. Faça o mesmo com o porcentual da população urbana

**Parte 4: Arquivos Excel**

1. Faça a leitura do arquivo Excel Funcionarios.xlsx localizado na pasta Excel e realize as seguintes tarefas utilizando Pandas
2. Armazene os dados dos funcionários junior e sênior em arranjos diferentes
3. Concatene ambos os tipos de funcionários num único arranjo
4. Determine os salários médios de cada tipo de funcionário e de todos juntos
5. Determine o menor salário exibindo o nome do funcionário
6. Determine o maior salário exibindo o nome do funcionário
7. Ordene os funcionários por salário
8. Determine o faturamento médio
9. Determine os faturamentos máximo e mínimo
10. Faça o gráfico do faturamento num gráfico de barras
11. Salve este gráfico numa imagem do tipo PNG

**Parte 5: Tabela Pivot**

1. Faça a leitura do arquivo Felicidade.csv no diretório CSV e realize as seguintes tarefas utilizando Pandas
2. Ordene os dados por ano e índice de felicidade
3. Exiba as dez primeiras linhas
4. Determine o número de linhas e colunas
5. Determine se existem valores do tipo NaN
6. Crie uma tabela Pivot categorizando por ano
7. Crie uma tabela Pivot categorizando por região
8. Crie uma tabela Pivot categorizando por ano e região
9. Utilize ano como coluna
10. Utilizando ano como coluna, faça um gráfico de barras do índice de felicidade
11. Utilize funções agregadas para determinar os valores da média, mediana, mínimo, máximo e desvio padrão
12. Utilize as mesmas funções agregadas e uma função lambda para determinar o número médio de países numa determinada região num dado ano
13. Remova outliers
14. Crie uma tabela Pivot para a região asiática
15. Faça o mesmo com a região europeia
16. Extraia os valores das colunas região e ano da região africana
17. Separe os dados em três quartilhes
18. Preencha os valores NaNs com zero