

 unesp <small>Centro Universitário</small>	UNIESP – CENTRO UNIVERSITÁRIO		
	CURSO: Bacharelado em Sistemas de Informação		TURNO: Tarde
	UNIDADE CURRICULAR: Projeto de Extensão		
	PERÍODO: 2º	C.H.: 60 h/aula	SEMESTRE: 2023.2
	PROFESSOR: Nisston Moraes Tavares de Melo		

LISTA DE EXERCÍCIO – 01 –Python (Pandas)

Construa um programa para atender as questões apresentadas abaixo, deve ser feito em Python e disponibilizado no GitHub.

- Um gerente de vendas deseja analisar os dados de carros de sua empresa. Ele possui um arquivo CSV chamado "carro.csv" com as seguintes colunas: 'Veiculo', 'UF', 'Combustivel', 'Marca', 'Modelo', 'AnoFab', 'AnoMod', 'Categoria', 'Cor', 'DataCad', 'idinstrutor', 'Situacao'. Utilize a biblioteca pandas para responder às seguintes questões, coloque o comando necessário:
 - Quantos carros estão cadastrados?
 - Quais os combustíveis cadastrados?
 - Quais as cores cadastradas?
 - Qual o ano de fabricação mais antigo?
 - Qual o ano de fabricação mais novo?
 - Qual a quantidade de carro por ano de fabricação?
 - Qual a quantidade de carro por cor?
 - Quais os “idinstrutor” que estão cadastrados com os carros da Marca Honda?
 - Qual a marca que apresenta mais carros cadastrados?
 - Qual o combustível que apresenta mais carros cadastrados?
 - Quais os carros que estão cadastrados com combustível Diesel?
 - Quantos carros estão em situação “0”? O campo situação só contém valores 1 ou 0.
 - Quantos carros estão cadastrados com o ano de fabricação igual a 2005?
 - Quais os carros com o ano de fabricação igual a 2004?
 - Qual o veículo com maior ocorrência nos dados?
 - Quais os tipos de categorias cadastrados?
 - Qual o código do instrutor com maior ocorrência?
 - Quais os carros da cor Verde?
 - Quais as cores dos carros do modelo “CELTA”?
 - Quais os anos de fabricação dos carros de modelo “CELTA”?

2. Um analista de dados está trabalhando com um conjunto de dados de oficina em um arquivo CSV chamado "conserto.csv". O conjunto de dados possui as colunas 'IDCarro', 'Oficina', 'Servico', 'Mecanico', 'DataEntrada', 'DataSaida', 'Responsavel', 'Obs', 'ValorServico'. Utilize a biblioteca pandas para responder às seguintes questões:

- a) Quantas linhas tem o arquivo de conserto.csv?
- b) Encontre os 10 mecânicos com maior ocorrência?
- c) Encontre o mecânico com maior ocorrência no ano de 2006?
- d) Quais os 10 serviços mais realizados?
- e) Quais os 5 serviços mais realizados no ano de 2006 com base na Data de Saída?
- f) Qual foi o mês com maior ocorrência de registros no ano de 2006 com base na Data de Saída?
- g) Qual foi a média por mês dos valores pagos no ano de 2005 com base na Data de Saída?
- h) Qual foi a oficina com maior ocorrência de carro?
- i) Quanto foi faturado pela oficina 'MECANICA UCHOA'?
- j) Quais os anos que estão presentes no arquivo com base na data de saída?
- k) Qual foi o mês com maior ocorrência de registros com base na data de saída no ano de 2004?
- l) Qual foi o faturamento por mês com base na data de saída do ano de 2004?
- m) Quanto foi faturado por mês pela oficina 'OFICINA NOSSA SENHORA DA PENHA' no ano de 2004?
- n) Quem é o responsável com maior ocorrência em 2006 com base na data de saída?
- o) Quem são os mecânicos da oficina 'OFICINA NOSSA SENHORA DA PENHA'?
- p) Qual foi os valores pagos pelos serviços realizados pelo Mecânico Toinho e Piloto?
- q) Qual foi os valores pagos pelos serviços realizados pelo Mecânico Toinho e Piloto no ano de 2006 com base na data de saída?
- r) Quais os mecânicos que realizaram o serviço de "Troca de Pneus"?
- s) Quais os 10 serviços realizados pelo mecânico Piloto?
- t) Quais os mecânicos que realizaram serviço pela oficina "WCAR PEÇAS E SERVIÇOS"?

****Gabarito:****

****Questão 1:****

- a) `df = pd.read_csv("sales.csv")``
- b) `df['Total_Vendas'] = df['Quantidade'] * df['Valor']``
- c) `produto_mais_vendido = df.groupby('Produto')['Quantidade'].sum().idxmax()``
- d) `df['Data'] = pd.to_datetime(df['Data'])`` e então `df.groupby(df['Data'].dt.to_period('M'))['Total_Vendas'].sum()``
- e) `dia_maior_receita = df[df['Total_Vendas'] == df['Total_Vendas'].max()]['Data'].iloc[0]``

****Questão 2:****

- a) `df = pd.read_csv("movies.csv")``
- b) `filme_maior_avaliacao = df[df['Avaliação'] == df['Avaliação'].max()]['Título'].iloc[0]``
- c) `media_avaliacoes_por_genero = df.groupby('Gênero')['Avaliação'].mean()``
- d) `df['Ano'] = df['Ano'].astype(str)`` e então `df['Ano'].value_counts().sort_index().plot(kind='bar')``
- e) `filmes_recentes_aval_acima_8 = df[(df['Ano'] > '2010') & (df['Avaliação'] > 8)]``

****Questão 3:****

- a) `df = pd.read_csv("sales_by_region.csv")``
- b) `regiao_maior_lucro = df.groupby('Região')['Lucro'].sum().idxmax()``
- c) `total_vendas_por_produto_regiao = df.groupby(['Região', 'Produto'])['Vendas'].sum()``
- d) `df.groupby('Região')['Vendas'].sum().plot(kind='pie', autopct='%1.1f%%')``
- e) `produto_mais_vendido_por_regiao = df.groupby(['Região'])['Produto'].apply(lambda x: x.value_counts().idxmax())``