

 <b>CIÊNCIA DA COMPUTAÇÃO</b>	<b>AVALIAÇÃO:</b> ( ) PARCIAL ( X ) BIMESTRAL ( ) 2º chamada ( ) Exame Final		
	<b>DISCIPLINA:</b> TÓPICOS ESPECIAIS EM CIÊNCIAS DA COMPUTAÇÃO		
	<b>ALUNO(A):</b>		
	<b>PROFESSOR(A):</b> Prof. Julio Cesar Pastoril		
<b>TURMA:</b> 1º Semestre		<b>DATA:</b>	<b>NOTA:</b>
<b>PROVAS</b>	<b>NÚMERO DAS QUESTÕES</b>	<b>VALOR DE CADA QUESTÃO</b>	
<b>OBJETIVAS</b>	<b>00</b>	<b>00</b>	
<b>DISCURSIVAS</b>	<b>06</b>	<b>1,0</b>	
<b>TOTAL DA PROVA</b>	<b>-</b>	<b>6,0</b>	

Obs: Utilize o conjunto de dados **casa\_para\_venda.csv** para resolver o que se pede nos exercícios 1 e 2 utilizando o google collab. Resolva os exercícios de forma organizada. Escreva o enunciado da alternativa no formato texto e abaixo desenvolva o código necessário.

1. Utilize o conjunto de dados casa\_para\_venda.csv e faça as análises iniciais que se pede:
  - i. Importe a biblioteca pandas para o projeto. Importe o conjunto de dados casa\_para\_venda no arquivo CSV em anexo em armazene em um pandas DataFrame.
  - ii. Inspecione as 10 primeiras linhas do DataFrame. Inspecione as 5 últimas linhas do DataFrame.
  - iii. Obtenha algumas meta informações do DataFrame. Dica: utilize os métodos *info* e *describe*. No formato texto, escreva os insights.
  - iv. Faça algumas seleções: Selecione a coluna "city" usando a notação de atributo (ponto); Qual o nome da cidade no index 35?; Selecione a coluna "rent amount"; Selecione a cidade no índice 100 usando o operador *iloc* (todas as colunas)! Qual o valor do aluguel?; Selecione a cidade no índice 123 usando o operador *iloc* (todas as colunas)! Quantos quartos tem?; Selecione as últimas 8 linhas e as colunas area, rooms, bathroom, floor usando o operador *iloc*. Qual o menor valor do aluguel?
2. Utilize o conjunto de dados casa\_para\_venda.csv e faça as análises iniciais que se pede (parte 2 da questão 1):
  - i. Importe o conjunto de dados casa\_para\_venda no arquivo CSV em anexo em armazene em um pandas DataFrame e defina a coluna city como index.
  - ii. Renomeie a coluna city para cidade.
  - iii. Verifique se todas as observações na coluna "cidade" são iguais a "São Paulo". Salve a resposta em uma variável.

Obs: Utilize o conjunto de dados **campeonato-brasileiro-full.csv** para resolver o que se pede nos exercícios 3, 4 e 5 utilizando o google collab. Resolva os exercícios de forma organizada. Escreva o enunciado da alternativa no formato texto e abaixo desenvolva o código necessário.

3. Utilize os dados campeonato-brasileiro-full.csv e faça as análises e alterações que se pede abaixo:
  - i. Importe a biblioteca pandas para o projeto. Importe o conjunto de dados casa\_para\_venda no arquivo CSV em anexo em armazene em um pandas DataFrame.
  - ii. Crie um multindex para o DataFrame com as colunas Dia e Estado Mandante
  - iii. Classifique os dois níveis de índice em ordem crescente e salve no próprio DataFrame

- iv. Inspeção o DataFrame
4. Agora que você já importou os dados e salvou em ordem crescente os níveis pedido, selecione o que se pede:
- i. Selecione todos os jogos com “Estado Mandante” sendo o RJ.
  - ii. Selecione todos os jogos com “Estado Mandante” sendo o RJ que ocorreram no Sábado.
  - iii. Selecione todos os jogos com “Estado Mandante” sendo “RS” e “SP” que ocorreram no Sábado e apenas as colunas “Arena” e “Rodada”.
  - iv. Verifique quantos dias na coluna “Dia” contém “feira”.
5. Utilize novamente o conjunto de dados do campeonato brasileiro e para selecionar um conjunto de amostra a partir do todo (população).
- i. Faça uma amostragem randomica simples contendo 100 jogos. Salve em um novo DataFrame e inspecione os dados e tire algumas informações. Use o formato texto do Collab para escrever as informações.
  - ii. Faça uma amostragem sistemática contendo 100 jogos. Salve em um novo DataFrame e inspecione os dados e tire algumas informações. Use o formato texto do Collab para escrever as informações.
6. Importe o conjunto de dados “cars\_unclean.csv” e inspecione. Em seguida, limpe o conjunto de dados fazendo o que se pede abaixo:
- i. Identificar os dados inconsistentes
  - ii. Cada coluna/feature deve ter o tipo de dado apropriado
  - iii. Identificar os dados faltantes (missing values)
  - iv. Limpe as colunas, deixando os elementos no formato correto
  - v. Identifique os outlier

### Observações finais.

Realize tudo em um unico arquivo de forma organizada e salve da seguinte forma:

TOPICOSESPECIAIS\_nome\_completo

EX:

TOPICOSESPECIAIS\_Julio\_cesar\_pastoril