



## Python para Ciências de Dados

**PROF.:** Leandro Lessa

### Objetivo:

O objetivo desta tarefa é incentivar a exploração criativa dos dados fornecidos, buscando gerar insights significativos por meio de gráficos, análises de agrupamentos e outras técnicas de análise de dados.

### Enunciado

Um dos principais trabalhos do cientista de dados é analisar dados para extrair insights valiosos e orientar a tomada de decisões informadas dentro das organizações. Neste contexto, apresentamos um conjunto de dados simulando informações demográficas e características de clientes. Sua missão é conduzir uma análise detalhada desses dados, identificando tendências, padrões e informações relevantes que possam beneficiar a organização.

Utilize técnicas estatísticas e de visualização de dados para explorar o conjunto de dados, respondendo às perguntas apresentadas anteriormente e descobrindo outros insights por conta própria. O conhecimento adquirido por meio dessa análise pode influenciar estratégias de negócios, marketing, atendimento ao cliente e muito mais. A análise de dados é uma ferramenta poderosa para auxiliar as organizações a tomar decisões embasadas em dados, resultando em melhorias significativas e vantagens competitivas.

Portanto, aproveite ao máximo essa oportunidade para explorar e analisar os dados fornecidos, contribuindo para o crescimento e sucesso da organização.

## Atividades

Para essa atividade, os alunos deverão realizar análise de dados utilizando as bibliotecas abordadas no curso.

1. Criar uma virtualenv e instalar as bibliotecas necessárias.
2. Coletar os dados dos arquivos:
  - a. dados\_caracteristicas\_fisicas.csv
  - b. dados\_pessoais.csv
  - c. Lista dos Estados Brasileiros

As bases de dados se encontram em:

[https://leandrolessa.com.br/datasets/?preview=true#demograficos\\_fisicos](https://leandrolessa.com.br/datasets/?preview=true#demograficos_fisicos)

<https://leandrolessa.com.br/lista-estados-brasileiros/>

3. Analisar os dados coletados.
4. Realizem a integração dos dados.
5. Avaliar a necessidade de tratamentos de dados ausentes/duplicados.
6. Manipulação e visualização dos resultados.
7. Criar análises descritivas, exploratórias e gráficas dos dados

## Tratamento de dados

1. Moda para as variáveis categóricas
2. Media para as variáveis numéricas

## Apoio para criação de gráficos

1. <https://seaborn.pydata.org/>
2. <https://matplotlib.org/stable/>

## Dicas do professor

### **Familiarize-se com os Dados**

Antes de iniciar qualquer análise, dedique um tempo para entender os dados fornecidos. Examine as diferentes variáveis, seus tipos e possíveis relações entre elas.

### **Pergunte-se**

Formule perguntas que possam ser respondidas com base nos dados. Questões abertas incentivam a exploração e podem levar a descobertas inesperadas.

### **Explore Graficamente**

Utilize gráficos adequados para visualizar diferentes aspectos dos dados. Experimente diferentes tipos de gráficos, como barras, linhas, dispersões e áreas, para representar as informações de maneira clara e eficaz.

### **Identifique Padrões e Tendências**

Procure por padrões ou tendências nos dados. Isso pode incluir identificar picos sazonais, correlações entre variáveis ou mudanças ao longo do tempo.

### **Realize Análises de Agrupamentos**

Aplique técnicas de análise de agrupamentos para identificar grupos ou segmentos naturais nos dados. Isso pode ajudar a entender melhor o comportamento dos clientes, padrões de compra ou características demográficas.

### **Extraia Insights**

Com base nas análises realizadas, extraia insights relevantes que possam beneficiar a organização. Estes podem incluir recomendações para estratégias de marketing, segmentação de clientes ou otimização de processos.

### **Comunique suas Descobertas**

Apresente suas descobertas de forma clara e concisa. Utilize visualizações de dados eficazes e forneça interpretações significativas para suas análises.

## Entrega

Prepare um relatório sucinto que documente suas análises, gráficos e insights. Este relatório deve ser apresentado de forma organizada e compreensível, destacando as descobertas mais relevantes.

**Atenção.** Por favor, cada aluno deve enviar seu relatório completo através do Canvas, contendo o documento no formato doc/pdf e o arquivo Python com os códigos desenvolvidos. Para facilitar o envio, utilize a função de compactação para criar um arquivo zip contendo todos os arquivos necessários e envie apenas esse arquivo.

Para trabalhos realizados em grupo, é importante que todos os membros do grupo sejam devidamente informados no cabeçalho do documento escrito..

## Observação

Esta tarefa destina-se a estimular sua criatividade e capacidade analítica. Sinta-se à vontade para explorar diferentes abordagens e experimentar novas técnicas de análise de dados. Lembre-se de que a análise criativa dos dados pode levar a descobertas inovadoras e insights valiosos para a organização.

Aqui estão algumas perguntas para orientar suas análises.

1. Qual é a cor de cabelo mais comum entre os clientes?
2. Existe alguma correlação entre a altura e o peso dos clientes?
3. Qual é a distribuição de tatuagens entre os clientes?
4. Quantos clientes têm o tipo sanguíneo AB-?
5. Qual é o tipo de pele mais prevalente entre os clientes?
6. Qual é o nível de escolaridade predominante entre os clientes?
7. Qual é a porcentagem de clientes que têm filhos?
8. Qual é o estado com a maior média de salário entre os clientes?
9. Existe uma relação entre a idade e o número de cartões de crédito dos clientes?
10. Quais são as profissões mais comuns entre os clientes?
11. Qual é a média de salário por estado/região?

Estas são algumas das perguntas iniciais. No entanto, o objetivo do nosso trabalho é buscar insights significativos a partir dos dados fornecidos. Desta forma, analisem os dados, sejam críticos e explorem além do óbvio. Estejam abertos a descobrir padrões incomuns e novas oportunidades. O objetivo é realizar uma exploração profunda dos dados para impulsionar o sucesso da nossa análise.