



Trabalho - Programação Básica de Computadores

Análise de Dados

23 de outubro de 2023

1. Objetivo

O grupo deve criar um programa para análise de dados sobre um *dataset* público obtido na internet (sites como *Kaggle*, Instituto Brasileiro de Geografia e Estatística - IBGE, *Quadi*, Censo da Educação Superior (INEP), Microdados do Enem, Data.gov, etc.). É importante a seleção de uma base de dados que permita a análise de estatísticas, geração de gráficos e tabelas e aplicação de técnicas de ciência de dados. Assim, espera-se a aplicação de técnicas de análise e visualização de dados para extrair informações úteis e apresentá-las de forma clara, como médias, desvios padrões e comparações entre amostras/populações em diferentes períodos de consulta/coleta. **Seu programa deve apresentar, no mínimo, DEZ (10) análises diferentes para o dataset escolhido, incluindo a visualização destas análises por gráficos, tabelas e arquivos de acompanhamento do processamento (.log).**

As características do conjunto de dados utilizado na análise devem ser apresentadas de forma clara e organizada. Deve-se explicar e apresentar como os dados foram coletados, qual é o tamanho da amostra analisadas e quais variáveis estão presentes. A utilização de gráficos ou tabelas para ilustrar os dados e torná-los mais compreensíveis é necessária. Os algoritmos, modelos ou métodos estatísticos empregados devem ser detalhados. Por fim, para cada análise realizada sobre os subconjuntos ou conjunto total de dados, arquivos de *log* (.log) devem ser gerados com informações referentes aos dados (quantidade de dados utilizados, nome dos arquivos, situação com relação à base de dados completa, etc.) e as conclusões obtidas.

*Dica: Utilize as bibliotecas **Pandas** (para manipulação dos dados) e a **Matplotlib** (para visualização dos dados e dos resultados).

2. Pré-projeto

O pré-projeto será a primeira etapa do trabalho. Ele deverá ser entregue via Google Classroom até **23/11/2023**. O grupo só deverá iniciar o trabalho após avaliação do pré-projeto. Caso seja necessário, ele será devolvido para o grupo com algumas sugestões de melhoria, e deverá ser enviado novamente até ser aprovado para execução

O pré-projeto **DEVE** conter:

- a. Nome e descrição do banco de dados a ser utilizado, além de onde ele foi obtido (seja uma página da web, um link, um e-mail, etc).

- b. Detalhamento dos dados presentes no dataset escolhido. Número de atributos, número de amostras, quais os principais atributos lá presentes, etc.
- c. Quais características da base de dados o grupo irá focar na análise.
- d. Descrição de quais análises serão efetuadas, como, por exemplo: médias, medianas, correlações, tendências, evolução de uma característica, etc.

3. Trabalho

No trabalho, vocês deverão criar um código, em python, que lê (carrega) e analisa/processa uma base de dados. Dessa análise, serão gerados resultados que deverão estar em um relatório a ser entregue por vocês.

O relatório do trabalho deverá conter, no mínimo:

- a. O conteúdo do pré-projeto.
- b. Resultados obtidos (estatísticas, gráficos, etc)
- c. O código elaborado, que deve estar bem comentado
- d. Interpretação e discussão sobre os dados analisados, resultados e conclusões obtidas.

4. Entrevista

Após a entrega do relatório do trabalho, serão agendadas entrevistas com cada grupo. Nestas, serão feitas perguntas sobre o trabalho para cada membro do grupo, que deverão ser respondidas individualmente. Todos do grupo deverão saber o que foi feito e por qual motivo foi feito, além de apresentarem a participação individual dos integrantes.

5. Avaliação

A nota final do trabalho (T) será composta pela nota da execução (Ex) e da entrevista (En) do pré-projeto (Pp), da seguinte forma:

$$T = (Pp + Ex) * En$$

O pré-projeto vale no máximo 1 ponto, enquanto a nota da execução será de no máximo 3 pontos. A entrevista será um fator multiplicador para a nota, podendo ter valores entre ½ e 1. Desta forma, a nota máxima do trabalho será de 4 pontos (ou 40% da nota da disciplina).

Todos os trabalhos irão passar por uma verificação automatizada de plágio. Trabalhos identificados como iguais, em termos de programação (por exemplo, mudar nomes de variáveis e funções entre outros não faz dois trabalhos serem diferentes), serão **penalizados com a nota zero**. Isso também inclui a pessoa que forneceu o trabalho, sendo portanto, de sua obrigação a proteção de seu trabalho contra cópias ilícitas. Proteja seu trabalho e não esqueça cópias do seu código nas máquinas de uso comum. É importante lembrar que **plágio é crime**, e será tratado como tal. Além da nota zero, a situação será relatada para instâncias superiores da UFES e os alunos envolvidos estarão sujeitos a penalidades adicionais após abertura de processo disciplinar.

6. Entrega

As datas da entrega de cada etapa serão publicadas no Google Classroom. As orientações aqui dadas estão sujeitas a modificação para melhor adequação à proposta do trabalho. Caso existam mudanças, estas serão anunciadas via Google Classroom.