|  |  |
| --- | --- |
|  | PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  Escola Politécnica  Curso: Ciência da Computação Disciplina: Métodos Quantitativos |

Atividade A9 – 23/fevereiro/2024 – Distribuição de Frequência de Dados INSTRUÇÕES:

1. Leia as instruções atentamente e consulte o professor em caso de dúvida;
2. Entregar o documento com os resultados e a análise, juntamente com o código (arquivo compactado);
3. Enviar (INDIVIDUALMENTE) o documento no link indicado no ***Canvas*** até as 23h00 do dia 01/março/2024.
4. O código-fonte deverá ser entregue em formato IPYNB, juntamente com o documento e a base de dados original escolhida (bem como a versão transformada, se for o caso).

Nome: Rafael Vargas e Lucas Braga

Equipe:

1. **–** Verifique os quatro repositórios de dados abaixo e escolha uma base de dados para os trabalhos, atentando para as limitações de tamanho da mesma, conforme indicado pelo professor.

<https://data.census.gov/cedsci/> [https://dados.gov.br](https://dados.gov.br/) <https://archive.ics.uci.edu/ml/datasets.php> <https://www.kaggle.com/datasets>

1. **–** Para a base escolhida apresente as seguintes informações:
   * Indicar o nome da base;
   * Que tipo de dado (ou problema) caracteriza esta base de dados, ou seja, o que se pode extrair de conhecimento a partir dos dados lá disponíveis;
   * Indicar quem forneceu esta base, a data da coleta dos dados (início e final), além de particularidades (por exemplo, a coleta foi interrompida durante um período de tempo e depois retomada);
   * Quantos atributos (variáveis) a base possui e para cada atributo indicar o tipo: nominal, ordinal, discreto, contínuo;
   * Indicar se existem informações já publicadas sobre seu uso e, em caso positivo, onde foi publicado, por quem, o quê.
2. **–** Construa a Tabela de Distribuição de Frequência de cada uma das variáveis existentes na base, de acordo com a sequência apresentada no vídeo <https://www.youtube.com/watch?v=qzao79-zEZE>.
3. **–** Implemente um programa em Python que leia a base de dados e plote a Tabela de Distribuição de Frequência da variável escolhida pelo usuário