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

Atividade A9 – 06/agosto/2021 – Distribuição de Frequência de Dados

INSTRUÇÕES:

a) Leia as instruções atentamente e consulte o professor em caso de dúvida;

b) Entregar o documento com os resultados e a análise, juntamente com o código (arquivo compactado);

c) Enviar (INDIVIDUALMENTE) o documento no link indicado no ***Blackboard*** até as 23h00 do dia 09/agosto/2021.

d) 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: Gabriel Scholze Rosa

Equipe: Gabriel Scholze Rosa, Manoel Felipe Bina, Matheus Leindorf Muller e Nicholas Davi da Cruz

**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://www.kaggle.com/datasets>

**2 –** Para a base escolhida apresente as seguintes informações:

* Indicar o nome da base;

Developers and programming languages

<https://www.kaggle.com/jaimevalero/developers-and-programming-languages>

* 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;

Habilidade que cada usuário do GitHub possui em determinada linguagem de programação

* 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);

Jaime Valero disponibilizou essa base no site kaggle. A data de publicação da base foi no dia 29 de novembro de 2017, a última atualização do documento foi realizada no dia 03 de dezembro de 2017 e não temos a informação se a coleta foi interrompida durante um período de tempo.

* Quantos atributos (variáveis) a base possui e para cada atributo indicar o seu: nominal, ordinal, discreto, contínuo;

Assembly(contínuo), C(contínuo), C#(contínuo), C++(contínuo), CSS(contínuo), go(contínuo), Haskell(contínuo), HTML(contínuo), Java(contínuo), Javascript(contínuo), Objective-C(contínuo), PHP(contínuo), Powershell(contínuo), Python(contínuo), Ruby(contínuo), SQL(contínuo), Swift(contínuo), User\_id(nominal).

* Indicar se existem informações já publicadas sobre seu uso e, em caso positivo, onde foi publicado, por quem, o quê.

Não foi possível encontrar nenhum trabalho com a mesma proposta que a nossa.

**3 –** 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> .

**4 –** Implemente um programa em Python que leia a base de dados e plote a Tabela de Distribuição de Frequência da variável escolhida pelo usuário.