



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE EDUCACAO ABERTA E A DISTANCIA
CURSO DE SISTEMAS DE INFORMAÇÃO
ANÁLISE DE DADOS**

TRABALHO II

O trabalho consiste em fazer a análise de dados de um *dataset*. No trabalho deve conter os seguintes itens:

1. *Dataset* no formato CSV. De preferência, o *dataset* deve ser obtido de algum *site* ou local público.
2. Descrição do *dataset*. Se necessário, descrever os campos do conjunto de dados.
3. Apresentar como o *dataset* foi obtido.
4. Apresentar informações sobre o conjunto de dados, como:
 - a) estatísticas: máximo, mínimo, média, ...
 - b) especificidades sobre um dado (ou conjuntos de dados) em particular.
 - c) gráficos.
5. Operações desejadas para obter as informações: *groupby* e *pivoteamento*.
6. Espera-se que seja encontrado algum padrão nos dados, como, por exemplo, alguma regressão.
7. *Link* da apresentação.

O trabalho deve contemplar todos os itens acima, onde devem haver pelo menos 3 tipos agregação diferentes e 3 tipos de gráficos diferentes. Todas as informações obtidas devem prover algum valor semântico relevante e não óbvio em relação aos dados.

O trabalho pode ser feito em grupo, com no máximo 03 integrantes e deve ser entregue em um arquivo Jupyter Notebook. Cada um dos comandos usado deve ser explicado semanticamente no notebook. Adicionalmente, o trabalho também deve ser apresentado em forma de seminários. O próprio arquivo do Jupyter pode ser usado como apresentação. O arquivo do trabalho deve ser enviado na tarefa do Sigaa. O *link* da apresentação deve ser postado no fórum de *Seminários do Trabalho II* no Sigaa. Todos os integrantes do grupo devem apresentar o trabalho e devem aparecer na apresentação.

Observação adicional:

1. Pontuações não previstas na tabela acima pode ser acrescentadas ou decrementadas na nota de trabalho de acordo com correção do professor.