

Avaliação A2

Este documento inclui as instruções para a realização da avaliação A2 da disciplina de Introdução à Computação para Ciência de Dados.

- **Arquivo de dados:** Um arquivo `bilheteria.db` contendo os dados a ser utilizados no trabalho, está disponível para download no Eclass. Este banco de dados contém os dados de bilheteria dos cinemas coletados pela Ancine no ano de 2023.
- **Notebook inicial:** Um notebook Jupyter inicial `a2.ipynb` está disponível no Eclass. Este notebook contém o código inicial para a realização do trabalho, incluindo a importação dos dados e a criação de um dataframe com os dados de bilheteria.

Forma da entrega

O trabalho deve ser feito em duplas. A entrega deve ser feita através do Eclass, na forma de um arquivo compactado (`.zip`) contendo os 3 seguintes arquivos:

- **Relatório em formato PDF**, juntamente com o seu código fonte em LaTeX com os resultados gerados; O relatório deve também conter uma breve explicação sobre o que foi feito em cada função e como foi feito;
- **Arquivo `relatorio.tex`** com o código fonte em LaTeX do relatório; As figuras necessárias para compilação do relatório em LaTeX devem estar no mesmo diretório do arquivo `relatorio.tex`
- **Arquivo fonte `a2.py`** com o código fonte em Python; para a geração dos resultados. Este arquivo deve conter uma variável global denominada `AUTORES = ['Nome1', 'Nome2']`, onde `Nome1` e `Nome2` devem ser substituídos pelos nomes dos integrantes da dupla. Caso o grupo prefira poderá também desenvolver as soluções diretamente a partir do notebook Jupyter `a2.ipynb` fornecido. Neste caso, terá que entregar o notebook modificado.

Parte 1: Análise exploratória dos dados (5 pontos)

Para cada questão listada abaixo, uma função deve existir no arquivo fonte `a2.py`, ou `a2.ipynb` que recebe como argumento o nome do database. O nome da função deve ser o mesmo nome da questão, por exemplo:

```
def questao_1(database):  
    # código com a implementação da solução da questão 1  
    return resultado
```

1. Qual o total de bilheteria de todos os filmes, ou seja, o público que foi aos filmes listados?
2. Qual o filme de maior bilheteria em 2023, por país de origem?
3. Crie um dataframe com as 100 cidades com maior bilheteria em 2023, ordenadas de forma decrescente de bilheteria.
4. Qual o filme com maior bilheteria em cada cidade? Retorne um dataframe com as colunas `CIDADE`, `FILME` e `BILHETERIA`.
5. Quais as cidades com as maiores bilheterias para filmes brasileiros? Retorne um dataframe com as colunas `CIDADE`, `BILHETERIA_BR`, `BILHETERIA_ESTRANGEIRA`.

Parte 2: Análise estatística dos dados (5 pontos)

Nesta segunda parte, vocês terão que construir um relatório analítico sobre os dados de bilheteria, utilizando o LaTeX. Os aspectos dos dados que serão analisados e metodologia de análise empregada fica a critério do grupo. O relatório deve conter as seguintes seções: 1. **Introdução:** Uma breve introdução sobre o que é o relatório e o que ele pretende mostrar. 2. **Análise exploratória dos dados:** Uma breve descrição dos dados utilizados, incluindo o número de registros, colunas e tipos de dados. 3. **Análise estatística descritiva dos dados:** Uma análise estatística descritiva dos dados, incluindo gráficos e tabelas. Vocês devem incluir *pelo menos* 3 gráficos e 2 tabelas. 4. **Conclusão:** Uma breve conclusão sobre os dados analisados e o que foi aprendido com a análise.