



## Descrição do Projeto Final

### Instruções

Este trabalho pode ser realizado em grupo de no **máximo** 5 pessoas. Cada grupo deve entregar todo o conteúdo relativo ao trabalho em um arquivo .jar (que deve ser **executável**). Os fontes do programa desenvolvido devem ser disponibilizados via repositório **GIT**. Serão avaliados:

- Organização do Código
- Padrões de projeto utilizados
- Requisitos do projeto (apresentados posteriormente neste documento)
- Usabilidade da aplicação
- Apresentação do projeto

Caso o grupo não respeite as regras apresentadas ou falte com os critérios de avaliação, sofrerá redução na nota do trabalho.

### Entregas

O projeto final será avaliado em 4 etapas conforme as datas que serão divulgadas. A descrição de cada etapa e o que deve ser entregue é apresentado abaixo:

- Checkpoint **#1** – Devem ser entregues o projeto de interface gráfica do sistema;
- Checkpoint **#2** – Deve ser apresentado o diagrama de classes;
- Checkpoint **#3** – Nesta etapa o grupo deverá apresentar descrição dos padrões de projeto implementados;
- Checkpoint **#4** - Nesta entrega final, o grupo deverá realizar uma apresentação oral do sistema **funcionando completamente**. O grupo deve entregar os fontes e um arquivo .jar executável (disponibilizar também via GIT).

# Analizador Descritivo de Dados

O objetivo deste projeto é desenvolver uma aplicação que permita o usuário carregar um arquivo de dados **csv** (*comma-separated values*) e, sobre ele, calcular métricas estatísticas descritivas. Tem-se de implementar, pelo menos, as seguintes métricas:

- Univariadas:
  - Média
  - Moda
  - Mediana
  - Mínimo
  - Máximo
  - Variância
  - Desvio Padrão
  - *Skewness*
  - *Kurtosis*
- Multivariadas (apresentar em tabela):
  - Covariância
  - Coeficientes de Correlação (Pelo menos o de Pearson)

A aplicação deve permitir também a apresentação dos seguintes gráficos:

- Dados quantitativos (numéricos):
  - Histograma
  - Tabela de Frequências
  - Boxplot (pode ser apenas um atributo ou todos ao mesmo tempo)
  - *Scatterplot* entre 2 (dois) atributos **quaisquer** dos dados
- Dados qualitativos (labels/textos):
  - Tabela de contingência
  - Tabela de Frequências (absoluta e relativa)
  - Gráfico de barras com contagem de ocorrência dos valores (absolutos e em porcentagem)

Observações:

- Informações sobre estatística descritiva pode ser encontrada em [https://en.wikipedia.org/wiki/Descriptive\\_statistics](https://en.wikipedia.org/wiki/Descriptive_statistics) .
- Os alunos são encorajados a pesquisar outras métricas e gráficos para analisar os dados.
- O sistema deverá ser desenvolvido utilizando as linguagens Java e/ou Scala, **apenas**. Pode-se utilizar bibliotecas externas para apresentação dos gráficos.
- Durante a apresentação, os alunos deverão mostrar a aplicação funcionando e deverão fazer uma análise de um conjunto de dados de sua escolha. Deverão também explicar as métricas estatísticas utilizadas.
- O sistema poderá, caso necessário, descartar dados que estejam incompletos (algumas informações faltando).
- Os alunos poderão utilizar uma IDE para a **organização do código e geração do .jar**. Caso seja utilizada uma IDE para automatizar a geração da interface gráfica, haverá **penalidade** na nota.
- As apresentações terão duração de até 40 minutos + 20 minutos para responder dúvidas.