

SCC0252 - Visualização Computacional (2024) - 2º Semestre 2004

Prof. Dra. Maria Cristina Ferreira de Oliveira

PAE: Priscylla Silva

Projeto de Exploração e Análise Visual de Dados

Objetivo do Trabalho

Demonstrar capacidade de aplicar os conceitos e técnicas de visualização aprendidas na disciplina para explorar, analisar e comunicar insights a partir de um conjunto de dados.

Equipes

O projeto deve ser realizado em **duplas**.

Bases de dados

Escolher uma das bases de dados disponíveis no link: [📄 Datasets](#)

Avaliação

O projeto será dividido em duas entregas. A nota final será composta por 30% da primeira entrega e 70% da segunda entrega.

Primeira Entrega (Data: 30/10/2024)

- **Objetivo**

Exploração inicial dos dados, visualizações clássicas e multidimensionais, pré-processamento e redução de dimensionalidade (opcional).

- **Entrega**

Apresentação de slides com a seguinte estrutura:

- **Introdução**

- **Data profiling:** Faça uma contextualização da base de dados escolhida (conteúdo, tipos de variáveis, tamanho, ...);

- Com base em uma análise preliminar da base de dados, crie algumas questões que podem ser respondidas ou exploradas a partir dos dados;
- Pré-processamento
 - Descrever a etapa de limpeza e transformação dos dados (tratamento de valores faltantes, normalização, criação de novas features, redução de dimensionalidade, ...);
- Visualizações
 - Identificar visualmente tendências, outliers e distribuições relevantes;
 - Aplicar técnicas clássicas e multidimensionais;
- Discussões preliminares
 - Descrever os *insights* iniciais obtidos a partir da exploração e visualização de dados
- Próximos passos
 - Descrever as direções para a segunda etapa do projeto (novas questões para serem exploradas, ou o aprofundamento de alguma questão específica)

Segunda Entrega (Data: 26/11/2024)

- **Objetivo**

Continuar a análise feita na primeira parte do projeto, podendo ser inseridas novas visualizações e/ou novas questões, ou a dupla pode aprofundar as respostas para alguma questão já apresentada anteriormente. O principal objetivo é realizar análises mais aprofundadas, criar visualizações avançadas e interativas, apresentar um *storytelling* e comunicar melhor os aspectos interessantes dos dados e os resultados das análises. As visualizações interativas serão criadas utilizando a biblioteca [Vega-Altair](#).

- **Entrega**

Relatório com vídeo de apresentação.

O relatório deve seguir a seguinte estrutura:

- Capa
 - Título, nome dos integrantes do grupo e código USP;
- Introdução
 - Utilizar o conteúdo da introdução dos slides da entrega 01 (contextualização da base de dado e questões a serem respondidas);
 - Deve conter as questões apresentadas nos slides da entrega 01 com o complemento do que foi proposto em “próximos passos”;
- Análise


- Utilizar o conteúdo dos slides da entrega 01 para identificar visualmente tendências, outliers e distribuições relevantes;
- Responder às questões levantadas na entrega 01 do projeto utilizando visualizações e técnicas de análise exploratória de dados;
- Responder novas questões ou aprofundar as existentes;
- Visualizações Interativas
 - Utilizar o Vega-Altair para o desenvolvimento de visualizações interativas, permitindo a exploração dinâmica dos dados;
 - Justificar a escolha da visualização e o que pode ser explorado usando as interações;
- Conclusão
 - Interpretação dos resultados obtidos para as questões levantadas;
 - Reflexão sobre o processo de aprendizagem e lições aprendidas durante o projeto.
- Anexo
 - Inserir link para o google colab ou download do Jupyter Notebook;

Sobre a gravação do vídeo:

- O vídeo deve ser curto (máximo de 6 minutos);
- Deve seguir a estrutura do relatório;
- Deve mostrar como as visualizações interativas estão conectadas;

Critérios de Avaliação

- **A avaliação da entrega 01 (slides) vai considerar:**
 - Introdução (10%)
 - Base de dados descrita de forma clara e concisa;
 - Apresentação de pelo menos 3 questões potenciais que podem ser respondidas ou exploradas a partir dos dados;
 - Pré-processamento (20%)
 - Escolha e aplicação adequada das técnicas de pré-processamento;
 - Visualizações (50%)
 - Utilizar ao menos 3 tipos de visualizações diferentes;
 - Utilizar visualizações claras, informativas e que revelam *insights* sobre os dados com base nas questões levantadas;
 - Coerência entre a escolha das visualizações com as questões levantadas;
 - Discussões preliminares (10%)
 - Demonstrar capacidade de revelar *insights* relevantes para a base de dados;

- 
- Responder as questões com clareza;
 - Qualidade visual da apresentação (10%)
 - Slides visualmente bem organizados e com *design* limpo;
 - Visualizações legíveis;
 - **A avaliação da entrega 02 vai considerar:**
 - Relatório (50%)
 - Vídeo (50%)