

Projeto Final

Objetivo do Projeto

O principal objetivo deste projeto é importar um conjunto de dados existente na web, de modo a limpar e organizar esses dados e realizar análises exploratórias dos mesmos.

Dados do projeto

Você é livre para escolher o conjunto de dados que irá trabalhar. Porém, esses dados precisam estar disponíveis gratuitamente online e precisam ser aprovados por mim. Considerando que esta cadeira foca em data science, espero que seu conjunto de dados tenha certas características que ilustre seu conhecimento sobre os assuntos abordados em sala de aula. A seguir são listadas algumas características dos dados que procurarei nos seus projetos (note que não precisa conter todas essas características, apenas alguns):

- Vários tipos de dados (numéricos, caracteres, datas, etc.)
- Características não normalizadas (podem conter separador decimal distinto do adotado, letras maiúsculas e minúsculas, etc).
- Variáveis que precisam ser separadas ou unidas.
- Dois ou mais conjuntos de dados que precisam ser mesclados.
- Formato largo ou longo onde você precisa reformular os dados.
- Variáveis que precisam ser criadas (ou seja, os dados podem conter variáveis de receita e despesa, mas você deseja analisar as economias de modo que seja necessário criar uma variável de economia a partir das variáveis de receita e despesa).
- Dados que precisam ser filtrados.

Relatório de Projeto

Seção	Características	Pontos
Introdução	1.1 Forneça uma introdução que explique a declaração do problema que você está endereçando. Por que você e eu deveríamos estar interessados nisso? 1.2 Forneça uma breve explicação de como você planeja abordar essa declaração do problema (os dados usados e a metodologia empregada) 1.3 Discuta sua abordagem proposta/técnica adotada que você acha que abordará (total ou parcialmente) esse problema. 1.4 Explique como sua análise ajudará os potenciais clientes de sua análise.	8
Pacotes Requeridos	2.1 Todos os pacotes utilizados são carregados antecipadamente para que o leitor saiba quais são os requisitos necessários para replicar sua análise. 2.2 Uma explicação deve ser fornecida sobre o propósito de cada pacote.	4
Preparação dos dados	3.1 A fonte original em que os dados foram obtidos deve ser citada e, se possível, com hiperlink.	20

	<p>3.2 Os dados da fonte devem ser minuciosamente explicados (qual era o propósito original dos dados? quando foi coletado? quantas variáveis o original tinha? explique quaisquer peculiaridades dos dados de origem como os valores ausentes estavam registrados ou como os dados foram incluídos, etc.).</p> <p>3.3 As etapas de importação e limpeza de dados devem ser explicadas no texto (diga-me por que você está realizando as atividades de limpeza de dados que você executa) e siga um processo lógico.</p> <p>3.4 Quando seus dados estiverem limpos, mostre como é o conjunto de dados final. No entanto, não imprima uma data frame com mais de 200 linhas; mostre-me os dados da forma mais condensada possível.</p> <p>3.5 Forneça informações resumidas sobre as variáveis de interesse em seu conjunto de dados limpo. Não apenas imprima um monte de trechos de código com <code>str()</code>, <code>summary()</code>, etc. Em vez disso, forneça uma explicação consolidada, seja com uma tabela que forneça informações resumidas para cada variável ou um parágrafo de resumo bem escrito com código embutido.</p>	
Análise exploratória dos dados	<p>4.1 Descobrir novas informações nos dados que não são auto evidentes (ou seja, não apenas plotar os dados como estão; em vez disso, dividir os dados de maneiras diferentes, criar novas variáveis ou unir data frames separados para criar novas informações de resumo).</p> <p>4.2 Fornecer descobertas na forma de gráficos e tabelas. Mostre-me que você pode exibir as descobertas de maneiras diferentes.</p> <p>4.3 O(s) gráfico(s) devem ser cuidadosamente ajustados para o propósito desejado. Um gráfico ilustra um ponto primário e é apropriadamente formatado (títulos de plotagem e eixos, legenda se necessário, escalas apropriadas, etc.).</p> <p>4.4 Tabela(s) cuidadosamente construídas podem facilitar a realização de comparações importantes. Formatos cuidadosos podem destacar características importantes. O tamanho da tabela deve ser apropriado.</p> <p>4.5 Os insights obtidos das análises devem ser minuciosamente explicados, porém de forma sucintamente. Precisam ser fáceis de ver e entender as descobertas interessantes que você descobriu.</p>	40
Conclusões	<p>5.1 Conclua a declaração do problema que você abordou.</p> <p>5.2 Conclua como você abordou essa declaração de problema (os dados usados e a metodologia empregada).</p> <p>5.3 Conclua os insights interessantes que sua análise forneceu.</p> <p>5.4 Conclua as implicações para os potenciais clientes de sua análise.</p> <p>5.5 Discuta as limitações de sua análise e como você, ou outra pessoa, poderia melhorar ou construir em cima da sua análise.</p>	20
Formatação e Outros Requisitos	<p>6.1 O estilo adequado de codificação deve ser seguido e o código deve ser bem comentado.</p> <p>6.2 A codificação deve ser sistemática - problemas complicados devem ser divididos em subproblemas que são individualmente muito mais simples de serem resolvidos. O código deve ser</p>	8

	<p>eficiente, correto e mínimo. Ele também precisa usar estruturas de dados apropriadas (lista, data frame, vetor, matriz e fatores).</p> <p>6.3 Realização, domínio, esperteza e criatividade: As ferramentas e técnicas do curso devem ser aplicadas de forma competente e, talvez, de alguma forma criativa.</p> <p>6.4 Ir além do que era esperado e requerido através do uso de pacotes, ferramentas e métodos não abordados pelo curso.</p>	
--	---	--

Total de pontos possíveis: 100

Espero que o seu relatório conte uma história com os dados. Eu **NÃO** quero que você apenas relate algumas estatísticas que você encontra, mas, ao invés disso, forneça uma narrativa coerente de suas descobertas. Abaixo estão alguns exemplos do tipo de relatório que gostaria de ver:

- Capital punishment in America (<http://rpubs.com/hailce17/301571>)
- Violência contra a mulher na cidade do Recife (<https://rpubs.com/naanadr/violencia-mulher-recife>)

Entrega do Projeto

Eu irei disponibilizar uma planilha onde vocês irão adicionar o link do relatório publicado no RPubs (vários formatos podem ser encontrados em: <https://github.com/juba/rmdformats>). A **entrega** e a **apresentação** deverão ser feitas no dia **04/12/2025**. Projetos atrasados **NÃO** serão aceitos.

Avaliação do Projeto

O projeto será avaliado por mim (ou monitor) e por mais 4 alunos da disciplina. Iremos verificar cada item da tabela acima, atribuindo pontos de acordo com a coluna “Pontos”. Ao final, uma nota única de 0 (zero) a 100 (cem) será atribuída.

Note que cada aluno será responsável pela avaliação de **quatro projetos** que serão escolhidos e indicados individualmente por mim. Esta é uma tarefa **obrigatória** e contará como parte da nota final de cada aluno. Cada projeto **NÃO** avaliado pelo aluno resultará num decréscimo de **10%** na nota final do projeto do mesmo.