

TRABALHO INDIVIDUAL

Análise Exploratória de Dados

MAP2112 – Introdução a Lógica de Programação e Modelagem Computacional

Prof.Dr. Luis Carlos de Castro Santos

1º semestre 2020

lsantos@ime.usp.br

ENTREGA até 12/07/2020 23:59

A Análise Exploratória de Dados segue tipicamente um roteiro composto das seguintes fases:

1. **Formulação das questões relevantes e hipóteses**
2. **Teste dessas questões com visualizações**
3. **Identificação de tendências nos dados**
4. **Busca de relações entre as variáveis**
5. **Comunicação dos resultados da análise (conclusões)**

Esse processo não é linear, algumas tentativas são realizadas, os dados visualizados e analisados, e filtrados de acordo com hipóteses e observações. Finalmente as conclusões são apresentadas e sustentadas por visualizações gráficas adequadas.

Usem como o exemplo o exemplo do post:

<https://medium.com/@williamkoehrsen/exploratory-data-analysis-with-r-f9d3a4eb6b16>

O autor analisa dados de absenteísmo do sistema de saúde brasileiro. Não é obrigatório utilizar os pacotes apresentados na análise, mas sintam-se para incorporar o que acharem relevante. O aprendizado de R e da Ciência de Dados se dá pela incorporação das melhores práticas de diversos autores. É o uso dessas práticas para extrair conteúdo de informação dos dados que importa.

ROTEIRO DO TRABALHO

Existe um repositório bem amplo que atualmente conta com mais de 40.000 datasets

<https://www.kaggle.com/datasets>

Há a necessidade de se registrar para baixar os dados. Usem a palavra 'Brazil' no campo de buscas e haverá uma longa lista de datasets sobre os mais diversos aspectos da sociedade brasileira. Escolha pelo menos um desses datasets e realize sua análise.

A entrega consiste numa pasta compactada (.zip ou .rar) contendo o relatório, no formato pdf, e todos os scripts R que produzem os resultados enviada por e-mail para: lsantos@ime.usp.br até a data e horário limite.

A avaliação do trabalho irá considerar os seguintes aspectos:

- a) Qualidade do texto (40%) – Clareza na descrição das atividades e análise dos resultados.
- b) Qualidade dos gráficos (40%) – Legibilidade das informações relevantes
- c) Aspecto de originalidade (20%) – Inclusão de algum elemento que não tenha sido apresentado durante o curso e demonstre que houve pesquisa e aprendizado de uma habilidade nova. (ex. novos pacotes, novas formas de visualização). Seja explícito no texto indicando a sua escolha para esse quesito.