



Estatística Aplicada à Computação/Telemática

Projeto I

Acidentes de Trânsito em Rodovias Federais

Apresentação

Nesse projeto nós trabalharemos com os dados disponibilizados pela Polícia Rodoviária Federal (PRF), no seu [portal de dados abertos](#).

Na página é possível acessar dados bem detalhados sobre acidentes e infrações de trânsito registrados pela PRF em rodovias federais de todo o país.

Nosso foco nesse projeto será nos dados de **acidentes**, especificamente dos anos de 2019 e 2020.

Esse projeto será guiado, seguindo um roteiro, disposto a seguir, para ajudar vocês na elaboração das análises.

1ª Tarefa

Baixe e carregue em *dataframes* os conjuntos de dados referentes ao ano de 2020, disponível [aqui](#) e de 2019, disponível [aqui](#). Corrija quaisquer erros que apareçam no carregamento dos arquivos.

2ª Tarefa

Explore os conjuntos de dados, respondendo as questões:

- (a) quantos acidentes ocorreram em cada ano?
- (b) quantas variáveis cada conjunto de dados registra?
- (c) quais as cinco cidades brasileiras onde mais ocorreram acidentes em rodovias federais?
- (d) quantos acidentes com feridos graves aconteceram na Paraíba em 2019?

3ª Tarefa

Para cada indicação abaixo, construa um novo *dataframe*, salvando-o com a terminação *.csv*.

- (a) *ranking* de acidentes por estado para ambos os anos (um só arquivo combinando a informação dos dois anos);
- (b) acidentes por dia da semana para ambos os anos (um só arquivo combinando a informação dos dois anos).

4ª Tarefa

Com relação a ocorrência dos acidentes, responda:

- (a) qual a causa mais frequente e a mais rara de acidentes registrados nos dois anos?
- (b) qual a proporção de pessoas ilesas e de feridos graves por mês em cada um dos anos considerados? (proporção = número de ilesos ou feridos graves/ número total de pessoas envolvidas no acidente);
- (c) mostre, reorganizando o conjunto de dados, se a pandemia conseguiu diminuir ou não a incidência de acidentes nas rodovias federais.

OBSERVAÇÕES

- o trabalho é individual;
- somente serão considerados os resultados apresentados em notebooks jupyter;
- o projeto é o principal componente da nota (vale 60%), portanto, deixar de fazê-lo implica em prejuízo na avaliação;
- obviamente, é **mandatório** que todas as implementações feitas sejam discutidas, com explicações claras e diretas sobre como foi feito cada passo constante em seu trabalho. Projetos que contenham apenas o código serão considerados incompletos, assim como projetos com comentários iguais, **gramatica ou semanticamente**, serão considerados **PLÁGIO**. E plágio é crime!