

MAE399 - Análise de Dados e Simulação

Prof. Fábio Machado

Lista 2 - Entregar dia 10/04/2022

1. O objetivo é analisar os dados de um mês do Citi Bike em NYC. Para isto é necessário acessar ao endereço

<https://www.citibikenyc.com/system-data>

e fazer o download do arquivo **202102-citibike-tripdata.csv**. Este arquivo pode ser acessado diretamente em

<https://s3.amazonaws.com/tripdata/202102-citibike-tripdata.csv.zip>

Neste arquivo há informações relevantes sobre o uso do sistema de bicicletas. Até 2019 cada arquivo continha:

- Tempo de duração da viagem (segundos)
- Data e tempo da partida
- Data e tempo da chegada
- Nome da estação da partida
- Nome da estação da chegada
- Identificação da estação
- Lat/Long da estação
- Identificação da bicicleta
- Tipo de usuário (*24-hour pass*, *3-day pass user* ou *Annual Member*)
- Sexo (Zero=não informado; 1=masculino; 2=feminino)
- Ano do nascimento

Depois houve algumas modificações. Agora encontramos em cada arquivo:

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID
- End station name
- End station ID
- Start latitude

- Member or casual ride

Roteiro para uma primeira análise estatística dos dados do Citi Bike:

- Escolha um mês: **fevereiro 2019** (formato antigo) ou **fevereiro 2021** (formato novo);
- A todo o momento questione-se se há ou não *outliers* nas variáveis quantitativas. Retire estes dados caso você entenda que há motivos para tal;
- Faça estatísticas básicas sobre as variáveis, respeitando os seus tipos (quantitativa ou qualitativa);
- Faça gráficos para as variáveis, respeitando os seus tipos (quantitativa ou qualitativa);
- Faça análises bi-variadas de um ou mais pares de variáveis que você imagine estarem relacionadas. Exemplo: boxplots da variável *Tempo de duração da viagem* por cada valor da variável *Sexo*. O mesmo para *Tipo de usuário*
- Apresente alguma observação (mesmo que ainda inconclusiva) que você acha relevante sobre esta base de dados e que possa ser suportada pelos dados que você está analisando.
- Apresente tudo isto em um **mini-relatório de 3 a 6 páginas** que pode ou não conter as linhas de comando usadas, mas deve conter as estatísticas, tabelas e gráficos além de suas observações pessoais baseadas nos dados. Os dados em si não devem aparecer no relatório.