

MÓDULO - ESTATÍSTICA I

Este projeto busca avaliar os conhecimentos sobre o módulo de probabilidade e estatística. Para isso, vocês devem utilizar um dataset com uma série temporal.

Obs.: Como caráter de desenvolvimento, disponibilizamos alguns outros datasets interessantes, caso você tenha interesse de praticar os conteúdos aprendidos no módulo de Estatística I:

- Covid-19: https://brasil.io/dataset/covid19/caso_full/
- NBA Players: <https://www.kaggle.com/datasets/justinas/nba-players-data>
- BTC Dataset: <https://www.cryptodatadownload.com/data/poloniex/>
- Ticket de Cinema: https://www.kaggle.com/datasets/arashnic/cinema-ticket?select=cinemaTicket_Ref.csv
- Machine Predictive Maintenance Classification: <https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance/discussion>

ESTATÍSTICA DESCRITIVA (DATA DE ENTREGA: 16/10/2024)

Cada grupo deverá fazer uma análise do tipo de cada variável e apresentá-las graficamente. Deve-se usar as medidas de resumo para explicar a concentração dos dados, quais são os valores com maiores frequências, sua distribuição e como cada variável correlaciona-se com a outra.

Além disso, o grupo deve fazer uma análise sobre dados anômalos (os *outliers*) e dizer, em cada contexto, qual seria a melhor opção a ser tomada (manter os dados ou descartá-los).

O Material deve ser enviado em um notebook, com as imagens e as análises feitas para cada grupo.

- *Para o trabalho foque em apresentar um storytelling com os dados. Utilize as perguntas abaixo para guiar o seu trabalho e não limite-se apenas a respondê-las (somente sugestão):*
 - *Quais são as variáveis (colunas) neste conjunto de dados e qual é o tipo de cada uma?*
 - *Há valores ausentes no conjunto de dados? Explique-os*

- *Qual é a estatística descritiva básica para as variáveis numéricas, como média, mediana, mínimo, máximo e desvio padrão?*
- *Como as variáveis numéricas estão distribuídas?*
- *Existe alguma correlação entre as variáveis numéricas?*
- *Como os dados categóricos estão distribuídos?*
- *Existe alguma relação entre as variáveis numéricas?*

SÉRIES TEMPORAIS (DATA DE ENTREGA: 21/10/2024)

Séries temporais geralmente contém informações muito importantes para a análise humana. Considere a série temporal selecionada para o grupo. Com base nas aulas, faça uma análise completa sobre o comportamento da série temporal.

Questionamentos e sugestões para motivação da análise:

- Mostre graficamente o comportamento da série temporal. Realce os valores extremos (mínimo e máximo). Existem registros atípicos?
- Quais são os componentes da série temporal? Faça a análise de tais componentes e plote os gráficos.
- Os resíduos da série seguem uma distribuição normal?