

MÓDULO - ESTATÍSTICA I

Este projeto busca avaliar os conhecimentos sobre o módulo de probabilidade e estatística.

Obs: Como caráter de desenvolvimento, deixo algumas sugestões de dataset.

- NBAPlayers: <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

ESTATÍSTICA DESCRITIVA (DATA DE ENTREGA: 21/10/2024)

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?