

Relatório da análise de dados

Programa Lighthouse -Indicium

Guilherme Frutuoso de Almeida

Após o recebimento dos dados e contextualização dos mesmos, o fluxo de análise de dados foi posto em prática.

Primeiramente, o processo de ELT(Extraction, Load and Transform), extração, carregamento e transformação foi iniciado para termos os dados “em mãos” em primeiro lugar. Os dados em formato CSV foram incorporados em Dataframes, utilizando a biblioteca Pandas[Extração e Carregamento]. Logo em seguida, vem o processo de limpeza e manipulação dos dados, manipulando colunas do Dataframe e fazendo operações de limpeza básica, como remoção de NaN values e exclusão de outliers.

Após o processo de ELT finalizado, é hora demonstrar informações estatísticas dos dados, tanto textualmente como visualmente.

A escolha por medidas de tendências centrais e medidas de dispersão é a chave para saber como os dados estão distribuídos.

-Média, Variância e Desvio Padrão

Os valores de correlação e covariância também são importantíssimos para saber como os dados se relacionam. Os mesmo foram, também, escolhidos na análise.