# ANÁLISE DE TEMPO DE TELA

Projeto de análise de tempo de tela usando python e suas bibliotecas (Pandas e Matplotlib) e SQL

## INTRODUÇÃO

Como meu primeiro projeto de análise de dados, busco explorar e responder certas perguntas que ajudam na construção do relatório, apresentando resultados através dos dados analisados com Pandas e SQL, as ferramentas de análise de dados em que eu tenho mais experiência atualmente. Portanto, com base nos dados disponibilizados no site Kaggle, foi feito a extração do arquivo em CSV que contém a tabela que com os dados necessários para a análise, de acordo com o contexto dessa tabela.

*Fonte dos dados:*

<https://www.kaggle.com/datasets/anandshaw2001/mobile-apps-screentime-analysis>

## OBJETIVO

A partir do auxílio da ferramenta Chat GPT, foi possível a criação de algumas perguntas para que a análise possa ser feita e disponibilizar uma visão sobre as informações que os dados entregam, e para que isso possa ser possível, foram utilizados nesse projeto ferramentas como o Jupyter Notebook para que os códigos em python possam ser executados, assim como suas bibliotecas (Pandas, Matplotlib e Seaborn), e o PostgreSQL, como sistema de gerenciamento de banco de dados usado para extrair os dados do arquivo CSV e executar as consultas em SQL. Desse modo, o objetivo desse projeto é responder as perguntas e gerar informações para a análise a partir dos dados fornecidos e também para procurar estratégias para reduzir o tempo de tela excessivo das pessoas.

## ESTRUTURA DOS DADOS

A tabela utilizada para a análise se chama “screentime\_analysis”, e contém as seguintes colunas:

. **Date**:Data da coleta dos dados.

. **App**:Nomes dos aplicativos.

. **Usage\_minutes**: Tempo de uso do aplicativo em minutos

. **Times\_opened**: Quantidade de vezes que o aplicativo foi aberto

. **Notifications**:Quantidade de notificações recebidas no aplicativo no dia.

## METODOLOGIA

A análise realizada foi feita com base nas 3 etapas que serão mostradas a seguir:

### Limpeza e verificação dos dados

A limpeza de dados é um processo de identificar, corrigir e remover erros e dados incoerentes que estão na base de dados, como dados nulos, valores ausentes ou dados duplicados, para que a análise possa ser feita com dados relevantes e de forma precisa. Além disso, é preciso converter certos tipos de dados para que a análise possa ser feita do jeito correto.

### Análise exploratória de dados (EDA)

A análise exploratória de dados é feita para entender as características dos dados de forma estatística, o que permite descobrir padrões, tendências, relações entre variáveis e possíveis anomalias, tudo isso para que possa ser feito a extração de informações de maneira eficiente para responder as perguntas que interessam, formular hipóteses e suposições na análise. Além disso, exemplos de operações que podem ser feitas para esse tipo de análise, são cálculos estatísticos (como média, desvio padrão e distribuições), análise de correlação entre variáveis, filtragem de dados e visualização de dados para facilitar a interpretação

### Extração de Insights e conclusões

Após a análise exploratória de dados, em que são extraídas as informações necessárias para responder as perguntas, é preciso analisar todos esses resultados e transformar em um conhecimento útil, identificando padrões, tirando as conclusões para resolver o problema proposto e auxiliar na tomada de decisões.

## PERGUNTAS DA ANÁLISE

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.Essa é a etapa onde se inicia a análise dos dados para responder as perguntas propostas, porém, antes de começar essa etapa é preciso fazer a importação da biblioteca pandas para fazer as operações e ler o arquivo CSV:

Tela de computador com fundo preto

O conteúdo gerado por IA pode estar incorreto.Texto

O conteúdo gerado por IA pode estar incorreto.Logo após a importação e a leitura da base de dados, serão feitas as operações para verificar as 5 primeiras linhas e as 5 últimas linhas, para uma verificação rápida e inicial do dataframe, além das suas informações:

Após essa etapa, todas as operações foram feitas com sucesso, sem nenhum erro no dataframe, e a partir das informações, é possível concluir que esse dataframe possui duas colunas do tipo object, que é um conjunto de caracteres, que no caso são as colunas “Date” e “App”, e também possui três colunas do tipo inteiro, que são as colunas “Usage\_minutes”, “Notifications” e “Times\_Opened”. Além disso, serão feitas as operações para verificar se existem Tela de computador com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.dados incoerentes dentro do Dataframe:

Como é possível ver, foram feitas operações para verificar se há valores NaN ou indefinido, dados nulos e dados duplicados, e em todos os resultados, não existem erros ou inconsistências no dataframe, o que quer dizer que não é necessário que haja limpeza de dados, logo já é possível começar a fazer a análise.

### **Quais aplicativos consomem mais tempo semanalmente/mensalmente?**

Antes de começar a análise, é preciso fazer uma operação para converter o tipo de dado da coluna “Date” para o tipo “DateTime”, para que as operações envolvendo essa coluna sejam feitas da melhor forma e que não gere erros:

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.Nessa operação, foi feito a conversão da coluna “Date” e a ordenação do Data frame tendo a data como referencial. Além disso, para que possa ser feita uma análise de forma semanal para os aplicativos que mais consumiram, foi preciso criar uma nova coluna chamada “Week”, que divide todas as datas em semanas, a partir de um cálculo envolvendo os dias:

Após isso, foi criado um subset para separar a nova coluna criada, a coluna “App” e a coluna “Usage (minutes)”, para que facilite a visualização dos aplicativos com mais tempo de tela na semana.

Texto

O conteúdo gerado por IA pode estar incorreto.Para organizar todos esses dados, foi feito um agrupamento com a somatória do tempo de uso de cada aplicativo por semana:

Nesse resultado, para cada semana, todos os aplicativos tiveram seu tempo de uso somados, para que possa ser mais fácil criar uma visualização desses dados nos gráficos. Além disso, para facilitar mais a visualização desses gráficos, foram criadas variáveis que armazenam o tempo de uso total de cada aplicativo por semana específica, e após isso, foi feita a importação da biblioteca Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.Matplolib para a criação dos gráficos através das variáveis criadas.

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.A partir dessa separação de dados organizados por semanas, foi possível produzir os gráficos em linhas com o matplotlib para visualizar quais aplicativos tiveram o maior tempo de uso durante as semanas:samurai