# 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 e responder a primeira pergunta.

### **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:

Como é possível observar nos gráficos acima, os aplicativos que tem maior frequência de acesso nas semanas são principalmente o Instagram, Netflix e Whatsapp respectivamente. Além disso, é preciso descobrir o tempo de uso de tela total durante o mês, para descobrir quais são os aplicativos mais usados nesse período.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.Como o dataset apresenta um único mês, que no caso é o mês 8, em agosto, a operação para descobrir esse resultado se torna mais simples, e nesse caso, foi preciso criar um subset que some todo o tempo de uso durante o mês para cada aplicativo, através de um agrupamento:

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.E como é possível verificar no resultado acima, os aplicativos com maior tempo de uso são Instagram, Netflix e Whatsapp, assim como foi verificado antes, e para uma melhor visualização desses dados, foi criado o seguinte gráfico:

**Respondendo a primeira pergunta:**

A partir dos dados análisados e dos gráficos, os aplicativos que mais tiveram tempo de uso na semana no mês, foram Instagram, Netflix e Whatsapp na maior parte das vezes respectivamente, sendo que o Instagram apresenta aproximadamente 25% do tempo de uso total do mês, Netflix 24% e WhatsApp 16%. Além disso, a média de tempo de uso em horas dos aplicativos no mês é de aproximadamente 15.5 horas, e o Instagram está aproximadamente 16 horas acima da média, a Netflix a aproximadamente 14.5% horas acima da média e o Whatsapp 4.3% horas acima.

E para complementar a resposta dessa pergunta, foi feita outra pergunta que foi analisada na linguagem SQL, para complementar o resultado da análise feita em pandas, e a pergunta é a seguinte:

### **Quais são os 3 aplicativos com maior variação no tempo de uso ao longo dos dias?**

Interface gráfica do usuário, Aplicativo

O conteúdo gerado por IA pode estar incorreto.Porém, antes de realizar a análise em SQL, é necessário fazer a importação de dados para o PostgreSQL, e para que essa base de dados possa ser importada, foi criado um banco de dados chamado “screen\_time” para que possa armazenar os dados que serão importados, e após isso, foi criado a tabela que irá receber os dados da base de dados, chamada de “screentime\_analysis”, e após isso é feita a cópia de todos os dados do arquivo para essa tabela:

Após executar esses comandos, o resultado foi o seguinte:

Uma imagem contendo Tabela

O conteúdo gerado por IA pode estar incorreto.

Tabela

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

O conteúdo gerado por IA pode estar incorreto.Como se pode visualizar, o resultado foi bem-sucedido, o que indica que as operações já podem ser iniciadas. Dessa forma, para responder a pergunta proposta, foi feita uma operação que agrupe os 3 aplicativos que tiveram maior dispersão de tempo de uso:

**Respondendo a segunda pergunta:**

Como é possível visualizar no resultado, os três aplicativos com maior variação ao longo do tempo são o Instagram, Netflix e o Whatsapp, os três aplicativos que também possuem o maior tempo de uso durante todo o período da base de dados, com um desvio padrão que é calculado e mostrado na execução do código.

### **Existe correlação entre notificações e o tempo de uso?**

Para que essa pergunta possa ser respondida, é necessário realizar um cálculo para descobrir a correlação entre essas duas variáveis, ou seja, se a variável de notificações influencia positivamente ou negativamente na variável de tempo de uso, e vice e versa, e o cálculo para descobrir esse resultado é através da correlação de Pearson, em que se o resultado for um número positivo, então a correlação é positiva, caso contrário, a correlação é negativa, e se o número for 0, a correlação é neutra.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.Para realizar o cálculo da correlação de Pearson, foi usada uma função da biblioteca pandas, que é o **corr()**, que calcula a correlação entre as colunas do dataframe ou tabela, mas antes de usar essa função, é necessário separar as colunas específicas em um subset:

Padrão do plano de fundo

O conteúdo gerado por IA pode estar incorreto.Agora que as colunas já estão separadas, a função **corr()** já pode ser utilizada, e o resultado foi o seguinte:

Quadrado

O conteúdo gerado por IA pode estar incorreto.Como é possível ver nesse resultado, a correlação entre a variável de notificações e de tempo de uso é positiva, porém é uma correlação fraca, pois o número está mais próximo de 0 do que de 1, o que pode ser melhor visualizado no gráfico a seguir:

Outra demonstração de que há uma correlação entre as duas variáveis em questão, é utilizando como exemplo o aplicativo Instagram, que foi o aplicativo com maior tempo de uso da tabela, e para descobrir a relação entre as notificações e o tempo de uso do instagram, foi feito a seguinte operação:

Tela de computador com jogo

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.Através dessa operação, foi criado um subset que mostra o tempo de uso e a quantidade de notificações do Instagram por semana, e no gráfico criado a seguir, mostra as variações de valores entre essas variáveis de forma melhor:

**Respondendo a terceira pergunta:**

Através dos resultados obtidos durante essa análise, é notável que existem sim uma correlação positiva entre as notificações e o tempo de uso, apesar de ser fraca. Portanto, um exemplo que mostra essa correlação é o Instagram, que da primeira pra segunda semana, teve um aumento de 4 horas de tempo de uso e um aumento de 87 notificações, e da quarta semana para a quinta semana, houve uma queda de 7 horas de tempo de uso, e também houve uma queda de 285 notificações nesse mesmo período, o que confirma que as notificações contribuem para o aumento de tempo de tela.

### **Qual a relação entre quantidade de aberturas e tempo de uso nos aplicativos mais populares?**

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.Para descobrir a correlação entre essas duas variáveis, foi feita uma operação em SQL, para calcular o coeficiente de correlação envolvendo a função CORR() do PostgreSQL, e como foi descoberto anteriormente, os aplicativos mais populares, ou seja, que tem o maior tempo de uso, são Instagram, Netflix e Whatsapp, que foram os aplicativos usados na operação a seguir:

Aplicativo, Tabela

O conteúdo gerado por IA pode estar incorreto.

**Respondendo a quarta pergunta:**

Através do número da correlação, existe uma relação entre o tempo de uso e a quantidade de aberturas, os aplicativos mais populares como Instagram e Netflix tiveram uma correlação positiva, porém o Whatsapp teve uma correlação negativa, o que significa que o aumento da quantidade de aberturas influenciam no aumento do tempo de uso, com exceção do Whatsapp, que tem uma quantidade de vezes aberto maior que a quantidade de tempo de uso.

### **Quais dias da semana apresentam maior uso de aplicativos específicos?**

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.Para responder essa pergunta, é preciso criar uma coluna separada que indique o dia da semana de cada data, como foi feito na operação a seguir, com a criação de um subset e uma nova coluna chamada “Days of the week”:

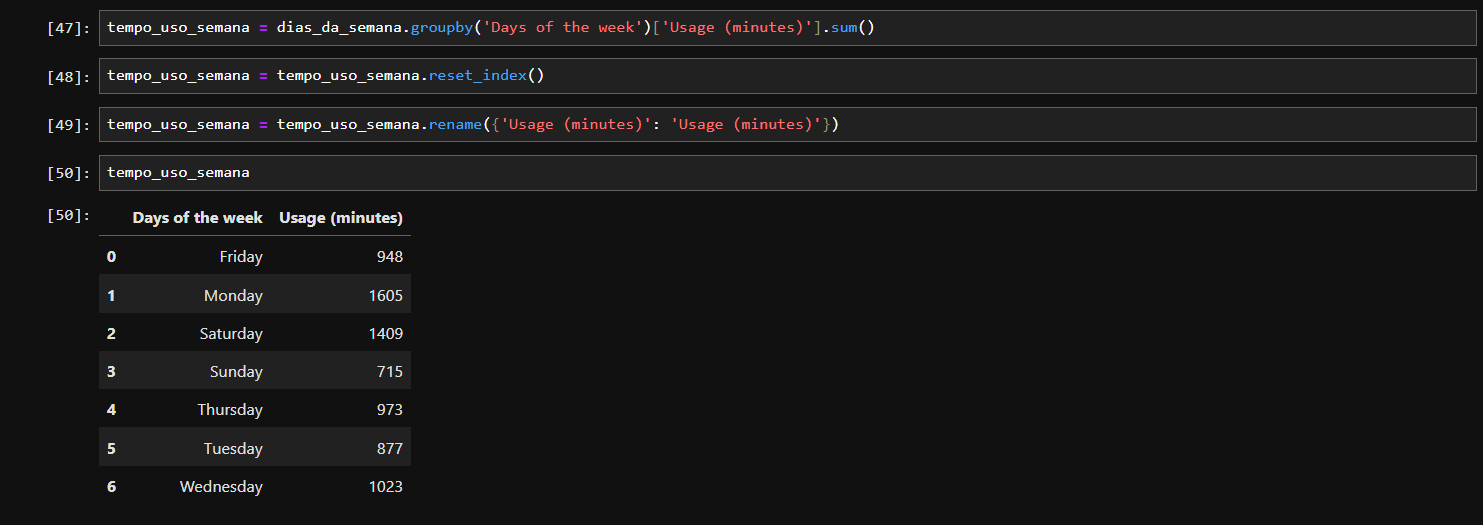
Tela de computador com fundo preto

O conteúdo gerado por IA pode estar incorreto.Após isso, foram adicionadas duas novas colunas a esse subset, a coluna “App” e a coluna “Usage (minutes)” da base de dados principal:

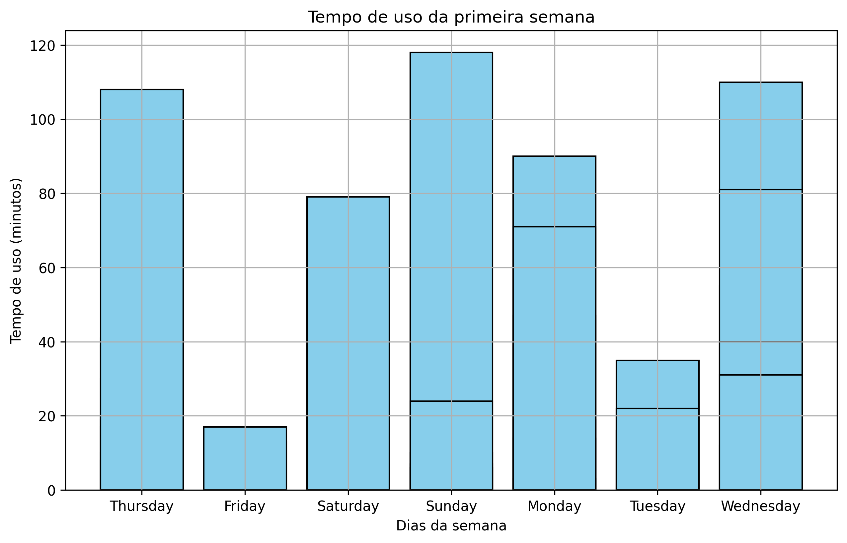
Tela de computador com fundo preto

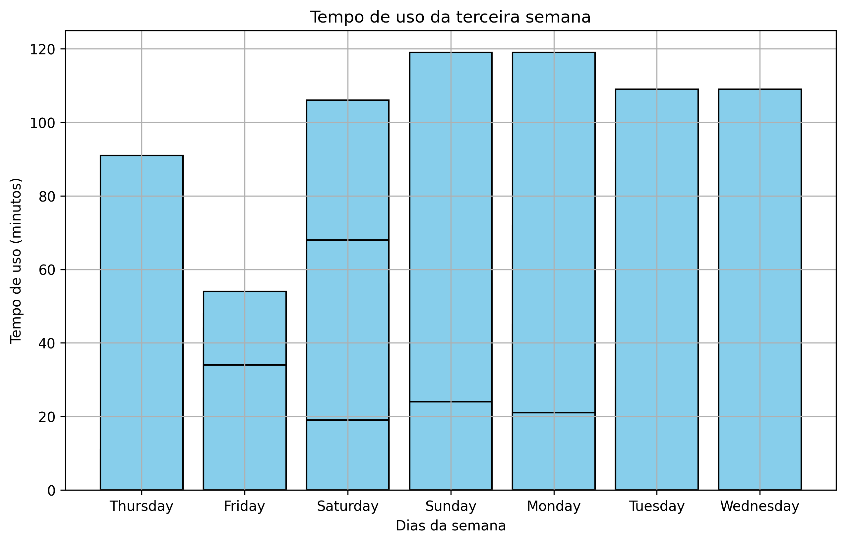
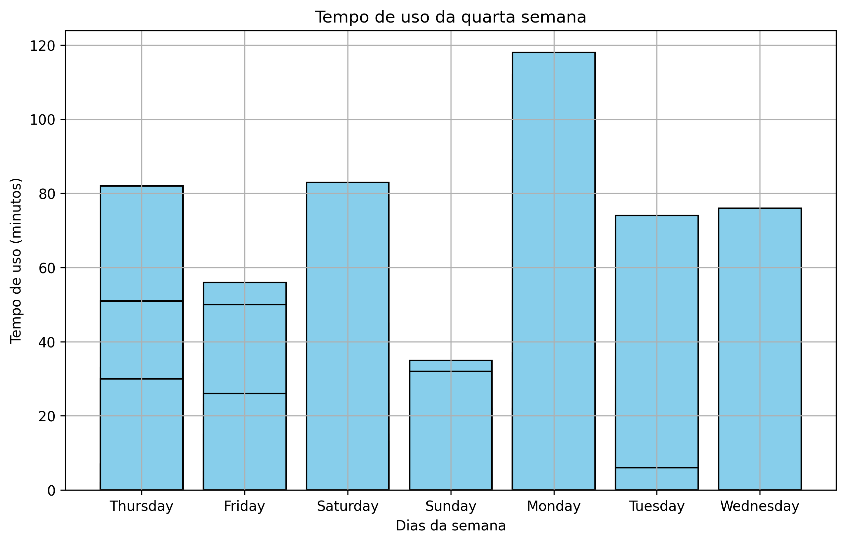
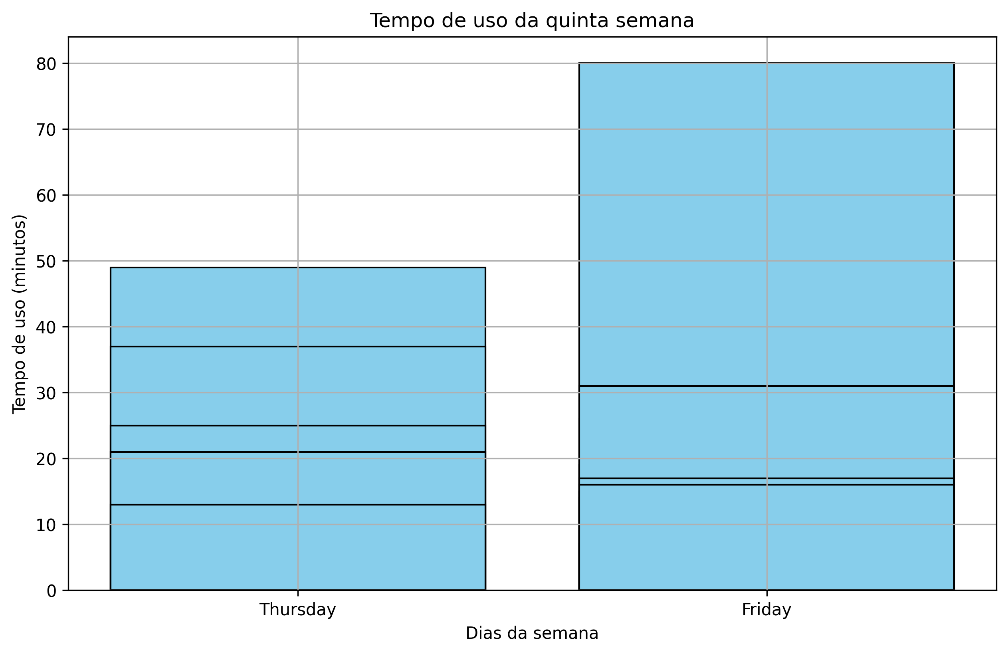
O conteúdo gerado por IA pode estar incorreto.Agora para descobrir quais aplicativos tiveram um maior tempo de uso em cada semana, foi feita a seguinte operação:

Como é possível visualizar, o Instagram foi o aplicativo mais usado durante as semanas, tendo um maior tempo de uso principalmente no dia de domingo e segunda-feira.

Além disso, outra informação importante é sobre quais dias da semana possuem um maior tempo de uso durante todo o mês, e o resultado foi descoberto com a operação abaixo:

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.Os dias da semana que possuem um maior tempo de uso durante o mês, são os dias de segunda feira, sábado e quarta feira. Ademais, para uma melhor visualização desses dados, foram feitos gráficos que detalham quais dias tiveram maior tempo de uso durante as semanas:



**Respondendo a quinta pergunta:**

Os dias da semanas com maior tempo de uso para aplicativos específicos é variado, sendo os dias da semana mais frequentes no domingo, sábado e segunda-feira. Porém, ao somar o tempo de uso de todos os aplicativos nas semanas, o dia de segunda-feira teve o maior tempo de uso, com aproximadamente 27 horas de uso de aplicativos, sábado com 24.5 horas e quarta-feira com 17 horas, sendo que domingo apenas teve 12 horas de tempo de uso, mas se for considerar os aplicativos mais usados, que no caso é Instagram, Netflix e Whatsapp, o maior tempo de uso do Instagram foi no domingo e com um tempo de uso de aproximadamente 2 horas no dia, o dia com maior tempo de uso da Netflix foi na segunda-feira, com um tempo de uso de aproximadamente 2 horas também, e o maior tempo de uso do Whatsapp foi na terça-feira, com um tempo de uso de apenas aproximadamente 1 hora e meia.

### **Quais aplicativos são abertos mais vezes, mas usados por menos tempo?**

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.Para responder essa pergunta, é necessário fazer um agrupamento de dados, tendo como referência a coluna de aplicativos, em que para cada aplicativo, serão mostrados o tempo de uso total em minutos e a quantidade de vezes aberto no mês, através da seguinte operação:

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.Para uma melhor visualização desse resultado, foi criado um gráfico de barras empilhadas que entrega a relação entre a quantidade de vezes abertos e a quantidade de uso em minuto em todo o mês, para cada aplicativo:

**Respondendo a sexta pergunta:**

A partir dos dados relacionados e o gráfico acima, é possível observar que o único aplicativo que tem a quantidade de vezes aberto maior que o tempo de uso é o Whatsapp, onde o tempo médio de uso em horas é 0.8 horas, enquanto a média da quantidade de vezes que esse aplicativo foi aberto, foi de 68 vezes aproximadamente. Além disso, a partir de um cálculo entre essas duas variáveis, para descobrir a quantidade de tempo de uso por vezes aberto, o resultado foi aproximadamente 0,70 minutos, ou seja, a cada vez que o Whatsapp é aberto, ele é usado por aproximadamente 42 segundos apenas.

### **Os aplicativos mais notificados são também os mais abertos?**

Uma imagem contendo Texto

O conteúdo gerado por IA pode estar incorreto.Para descobrir a relação entre as notificações e a quantidade de vezes em que um aplicativo é aberto, é preciso calcular a correlação entre essas duas variáveis e agrupar o resultado para cada aplicativo, através da seguinte operação feita em SQL:

Tabela

O conteúdo gerado por IA pode estar incorreto.

Como é possível verificar no resultado da operação acima, os aplicativos que possuem uma correlação positiva entre as notificações e a quantidade de vezes aberto, são o Safari, Linkedin, Instagram e Whatsapp, já os aplicativos como Facebook, Netflix, X e 8 Ball Pool tiveram uma correlação negativa.

Tabela

O conteúdo gerado por IA pode estar incorreto.Texto

O conteúdo gerado por IA pode estar incorreto.E para ter uma noção da quantidade de notificações e da quantidade vezes em que os aplicativos foram abertos no total, foi feita a seguinte operação:

**Respondendo a sétima pergunta:**

Aplicativos como Safari, Linkedin e Instagram tiveram uma correlação positiva, sendo Safari com uma maior correlação, de 0.22, ou seja, a maioria dos aplicativos tem uma correlação positiva fraca e até negativa, sendo o 8 BALL Pool com a menor correlação negativa, de -0.30, seguindo de X com -0.28 e Netflix com -0.25. Além disso, a soma do total de notificações e do tempo de uso dos aplicativos, indicam que os aplicativos com mais notificações não são os mais usados, por exemplos, o Whatsapp tem uma quantidade de notificações de 2498, mas um tempo de uso de 1204 minutos, já o Instagram teve uma quantidade de 1245 notificações e uma quantidade de tempo de uso de 1898 minutos. Então tudo indica que as notificações dos aplicativos mais notificados não são os mais abertos.

### **Quais mudanças no uso diário podem otimizar o tempo gasto no celular?**

**Respondendo a quinta pergunta:**

Depois de responder todas as perguntas anteriores, foi possível descobrir que os aplicativos mais usados foram Instagram, Netflix e Whatsapp, então deve-se reduzir o tempo de uso principalmente desses três aplicativos, isso é possível por meio do estabelecimento de limites de tempo de uso e substituindo esse tempo por atividades mais produtivas, como leitura, prática de atividades físicas, assistir documentários e entre outros, mas não tirando totalmente o uso desses aplicativos, outra alternativa útil também seria o uso de aplicativos de monitoramento de tempo de tela. Além disso, as notificações possuem uma correlação positiva com o tempo de uso, ou seja, as notificações corroboram para o aumento do tempo de uso, apesar de ser uma correlação fraca, então para esse problema, é necessário desativar as notificações desses aplicativos em momentos de produtividade, não é preciso desativar totalmente, pois as notificações também são importantes para a comunicação das pessoas.