Análise de Dados para *(Trocatine)*

Nome da startup

Gabriela Machado Ribeiro, Sarah de Moraes Ferreira Batista

Sumário

[Introdução 2](#_Toc177728807)

[Objetivo do aplicativo 2](#_Toc177728808)

[Objetivo do trabalho de Análise de Dados 2](#_Toc177728809)

[Levantamento dos dados 2](#_Toc177728810)

[Busca dos dados 2](#_Toc177728811)

[Justificativa de uso 2](#_Toc177728812)

[Descrição da base de dados encontrada 2](#_Toc177728813)

[Limpeza dos dados 3](#_Toc177728814)

[Condicionamento para alimentar o modelo de ML 3](#_Toc177728815)

[Condicionamento inicial 3](#_Toc177728816)

[Condicionamento após rodar o modelo pela primeira vez 3](#_Toc177728817)

[Condicionamento após rodar o modelo pela segunda vez 3](#_Toc177728818)

[AQUI COMEÇA A ANÁLISE DE DADOS (MODOLO) 3](#_Toc177728819)

## Introdução

O Trocatine busca proporcionar um espaço seguro e de fácil acesso, para vendas e trocas de artigos infantis. O intuito é que o aplicativo seja utilizado por famílias com crianças pequenas que desejam se desfazer de objetos que seus filhos não utilizam mais ou, que como crescem de forma rápida, devem trocá-los frequentemente. Entretanto as mesmas que não sabem como descartar ou como proceder diante da situação, desta forma são levadas frequentemente a decisão de descartar no lixo comum os itens, até mesmo os que estão em bom estado, e comprar novos itens com preços absurdos.

## Objetivo do aplicativo

O objetivo do aplicativo é auxiliar a introduzir a sustentabilidade na vida de famílias e incentivar a troca de itens infantis entre as pessoas que realmente precisam, assim evitando o descarte indevido e incentivando a economia circular, além de disponibilizar os produtos por um preço mais acessível.

# Objetivo do trabalho de Análise de Dados

Buscar e limpar uma base de dados com a finalidade de alimentar uma IA que irá determinar se uma pessoa, com base no seu perfil, é um cliente em potencial para o aplicativo ou não. Além disso iremos recomendar produtos com base nos dados pessoais do cliente coletados por meio de sondagem própria das equipes envolvidas.

# --------------------------------------------------------------------------------------

# Levantamento dos dados

## Busca dos dados

Aqui, o grupo descreve os critérios e os métodos para buscar os dados. Por exemplo:

* Levantamento de dados por questionário criado pelo próprio grupo
* Dados em bases oficiais/do governo/da ONU
* Bases de veículos de comunicação – jornais, revistas especializadas (não serve a Caras)

Coloque aqui os questionários que a equipe de Dados fez e indique quantas pessoas responderam. Se os questionários foram feitos pelos outros integrantes da equipe do Projeto (Dev ou 1° ano), não é necessário incluir aqui.

## Justificativa de uso

Explique por que é importante esses dados que foram escolhidos, qual a importância deles para o trabalho. Essas justificativas podem ser colocadas em uma seção separada, ou então indicadas junto da descrição da busca de dados.

# Descrição da base de dados encontrada

Aqui, faça como nas atividades do primeiro semestre. Inclua coisas como:

* Tipos dos dados: inteiro, data/hora, string etc.
* Valores-limite: máximo e mínimo para dados numéricos, lista de valores para dados categóricos
* Plotagem dos gráficos de análise preliminar dos dados brutos usando Python.
* Parâmetros estatísticos que descrevem os dados: média, moda, desvio padrão, variância etc.

Sempre indique como fez a análise, seja com seu código Python próprio, ou dizendo qual a biblioteca de análise exploratória de dados (por exemplo, ydata-profiling).

No relatório, coloque os resultados da análise dos dados brutos (parâmetros estatísticos e gráficos). O código (notebook) deve ser fornecido em um anexo ou como um link para o GitHub.

# Limpeza dos dados

Descrição do procedimento de limpeza e preparação dos dados para poder prosseguir para a análise preditiva, ou seja:

* quais colunas/campos foram limpos e por quê (remoção de outliers, dados claramente incorretos etc)
* gráficos que ajudem na justificativa
* como transformou os dados (mudança de tipo, arredondamento etc)

No relatório, coloque os resultados da limpeza (parâmetros estatísticos e gráficos). O código (notebook) deve ser fornecido em um anexo ou como um link para o GitHub.

# Condicionamento para alimentar o modelo de ML

O condicionamento de dados será realizado antes de rodar o modelo de ML. Também será feito após rodar o modelo de ML e for constatado que é necessário algum ajuste para reprocessamento.

Este item deverá conter uma subseção para cada etapa de condicionamento dos dados.

## Condicionamento inicial

* Normalização, padronização
* Redução de dimensionalidade
* Escolha de uma variável dentro de um conjunto de variáveis com alta correlação
* Avaliação dos dados preparados: valores estatísticos, histograma, boxplot, média/desvio padrão etc. usando Python.

## Condicionamento após rodar o modelo pela primeira vez

Indicar aqui o que foi alterado

## Condicionamento após rodar o modelo pela segunda vez

Indicar aqui o que foi alterado

# AQUI COMEÇA A ANÁLISE DE DADOS (MODOLO)

Os tópicos serão definidos junto com o prof. Módolo e inseridos aqui na próxima versão do documento.