

Projeto Aplicado II



COLLAM FILMS

Sistema de Recomendação de Filmes



AGENDA

01

Informações

02

Objetivo

03

Dataset

04

Apresentação
da empresa

05

Análise
Exploratória

06

Limpeza e
preparação dos
Dados

07

Aprendizado
de Máquina

08

Testes e
acurácia



Curso: Tecnologia em Ciências de Dados

Semestre: 3º

Componente curricular: Projeto Aplicado II

Professor: Anderson Adaime de Borba

Integrantes e TIA:

- **Adrieli Machado Zaluski - 22503668**
- **Caroline Ribeiro Ferreira - 22514635**
- **Lais César Fonseca - 22500790**
- **Liliane Gonçalves de Brito Ferraz - 22501142**
- **Múcio Emanuel Feitosa Ferraz Filho - 22515925**
- **Otavio Bernardo Scandiuzzi - 22511921**

Objetivos do Projeto



1

Melhorar a experiência do usuário ao fornecer recomendações personalizadas com base em seu histórico de filmes assistidos.

2

Criar um modelo eficaz de recomendação de filmes utilizando técnicas de aprendizado de máquina e análise estatística preditiva

3

















Reduzir o tempo de escolha do próximo filme a assistir, com as recomendações geradas pelo sistema, com base no filme que o usuário apresentar como favorito(s).



Como Base de dados para desenvolvimento do projeto e treinamento dados Públicos, da plataforma Kaggle.



Conjunto de Dados Full MovieLens - Metadados de Filmes é uma coleção de informações abrangentes sobre filmes lançados até julho de 2017. Ele inclui metadados detalhados sobre aproximadamente 45.000 filmes, oferecendo uma riqueza de informações relacionadas à indústria cinematográfica.

	Filme 1	Filme 2	Filme 3	Filme 4
Usuário 1				
Usuário 2				
Usuário 3				
Usuário 4				
Usuário 5				

A Base contem também 26 milhões de avaliações de 270 mil usuários para todos os 45 mil filmes. As Classificações estão em uma escala de 1 a 5, obtidas no site oficial do GroupLens.



APRESENTAÇÃO DA EMPRESA



A origem do nome da empresa “COLLAM FILMS”, nasceu da paixão por filmes e séries e da necessidade de tornar a experiência de assistir filmes ainda mais cativantes. Seu nome é uma fusão das iniciais dos nomes dos integrantes do grupo que deram vida a essa iniciativa, representando nossa colaboração e dedicação.

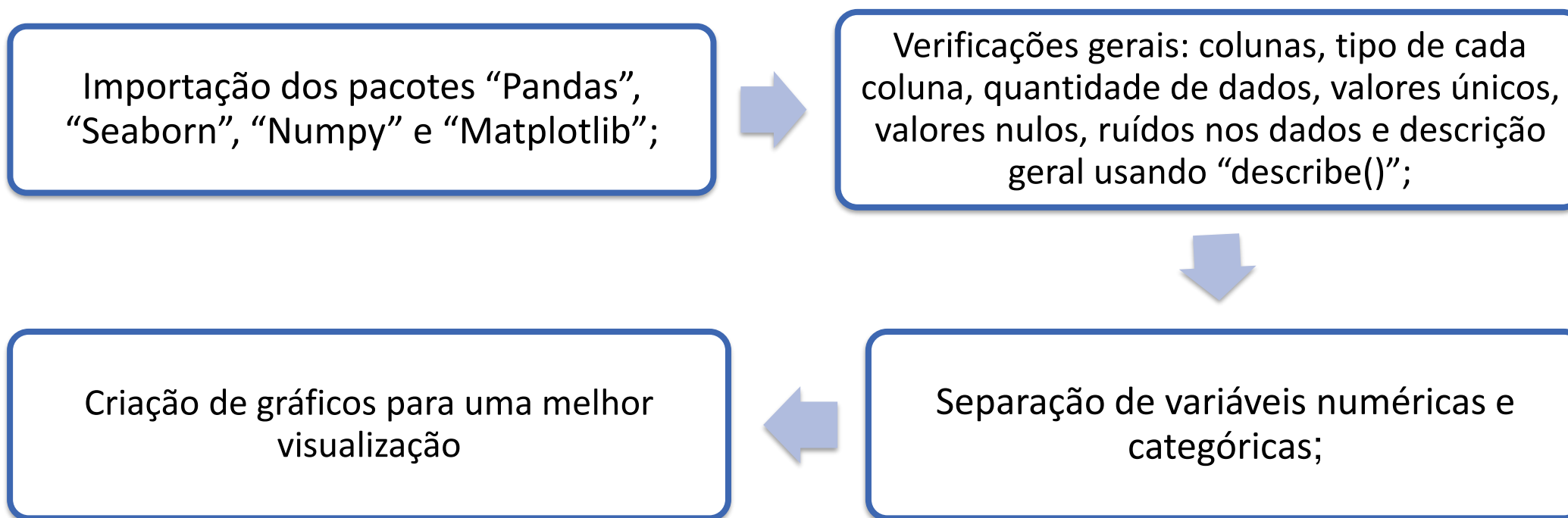
O nome “Collam” é uma celebração da união e a diversidade de habilidades que empregamos nesse projeto.



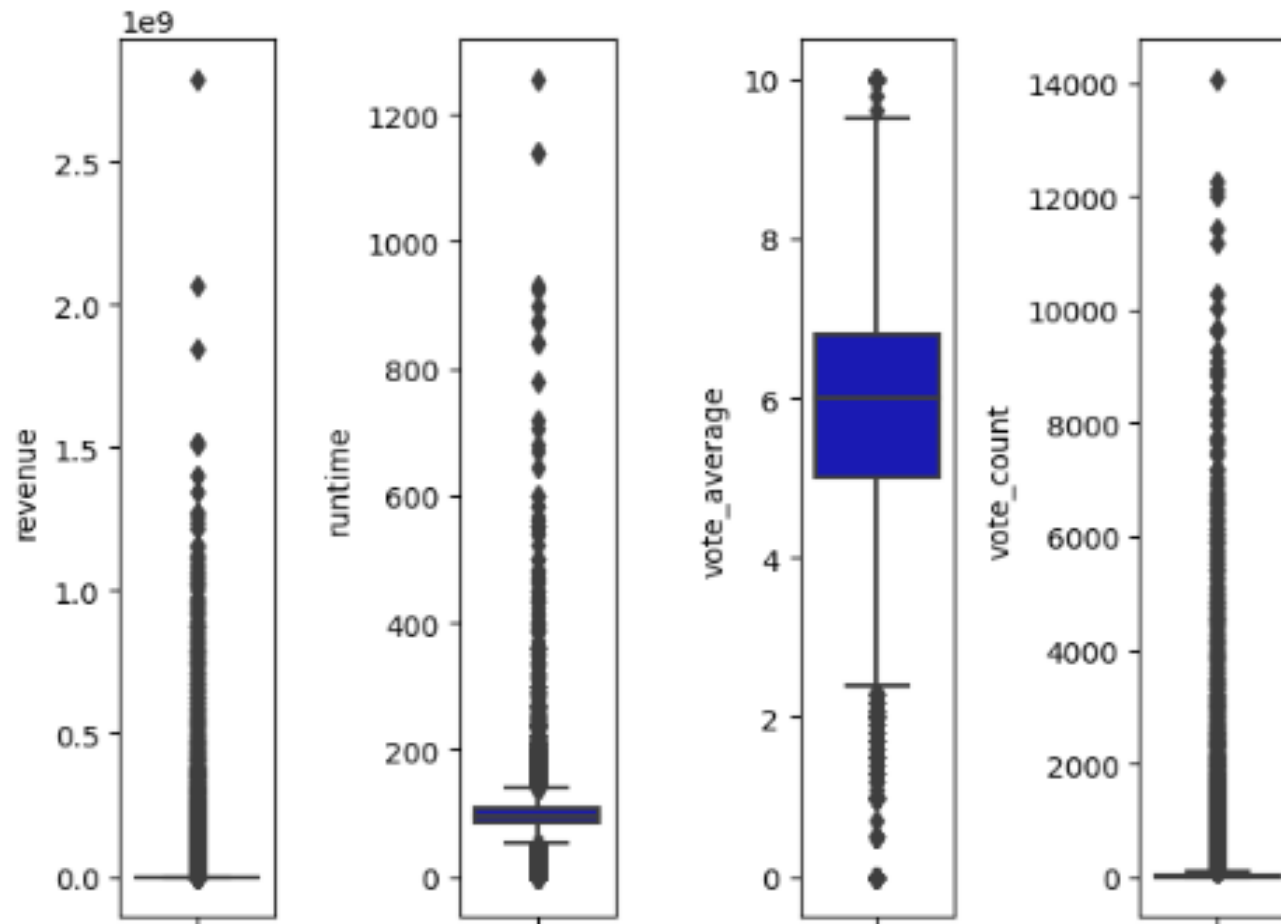
Análise Exploratória



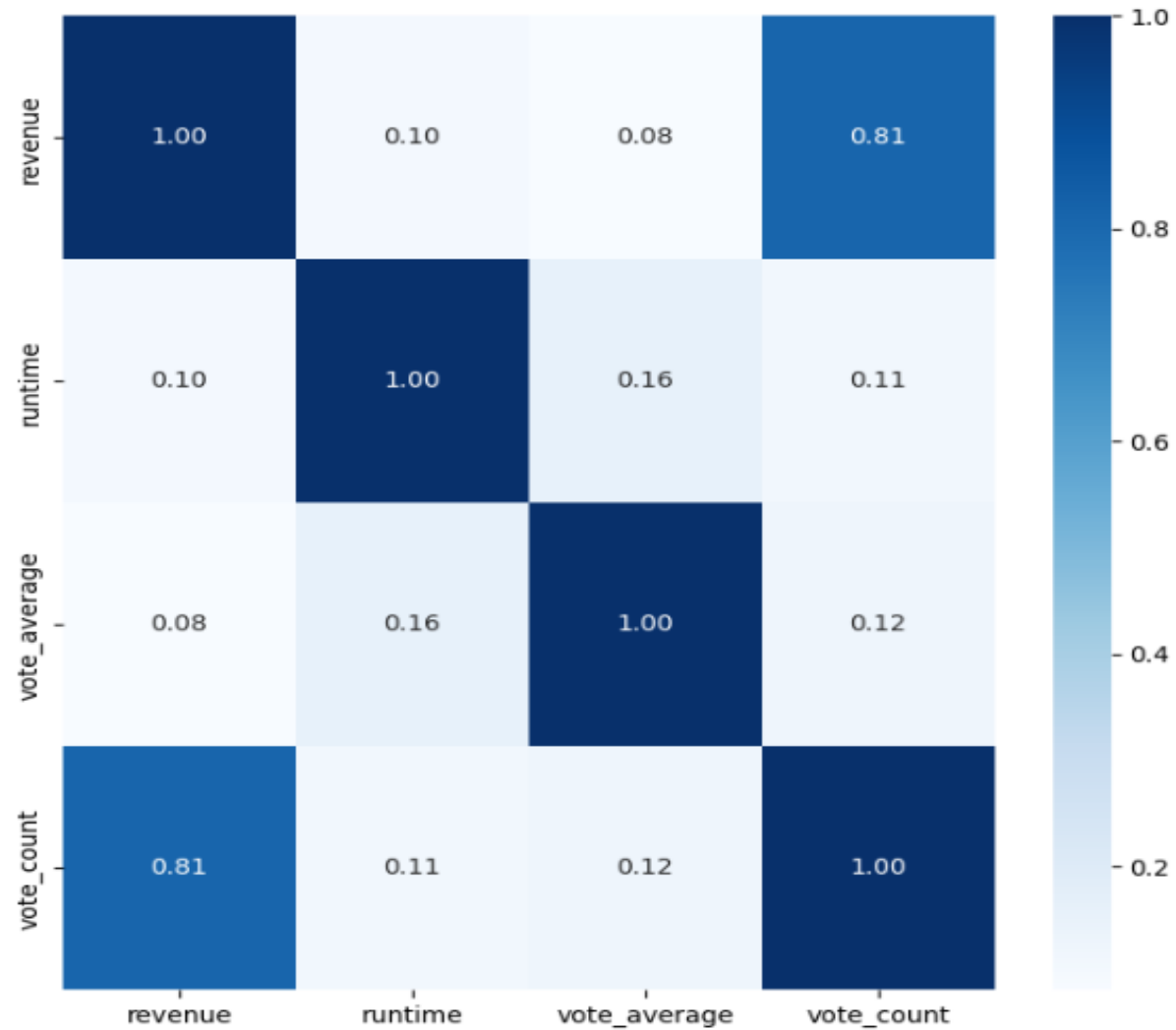
Durante o processo de exploração foi possível conhecer melhor os nossos dados através do uso de métodos estatísticos, para isso foram seguidos os seguintes passos:



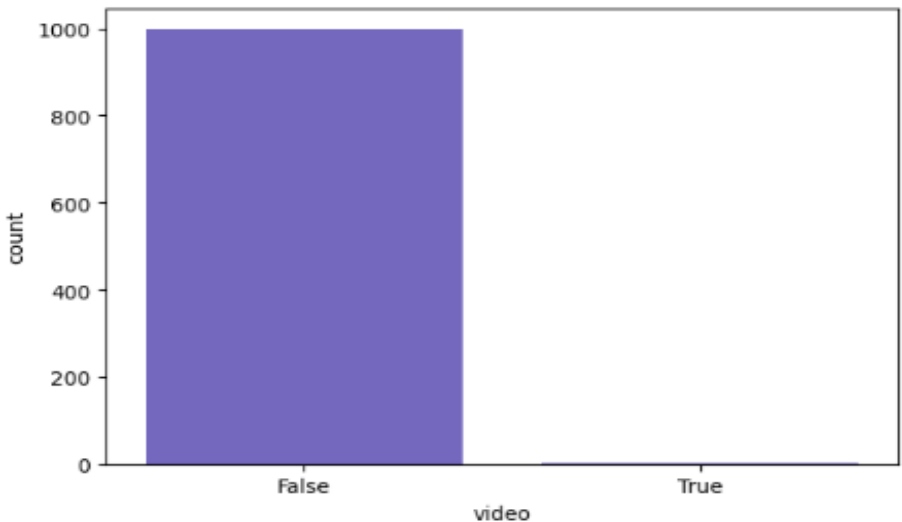
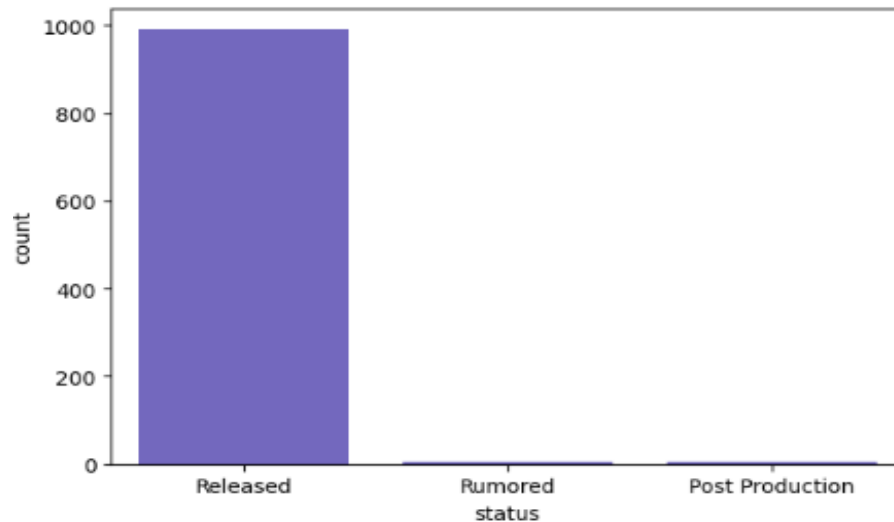
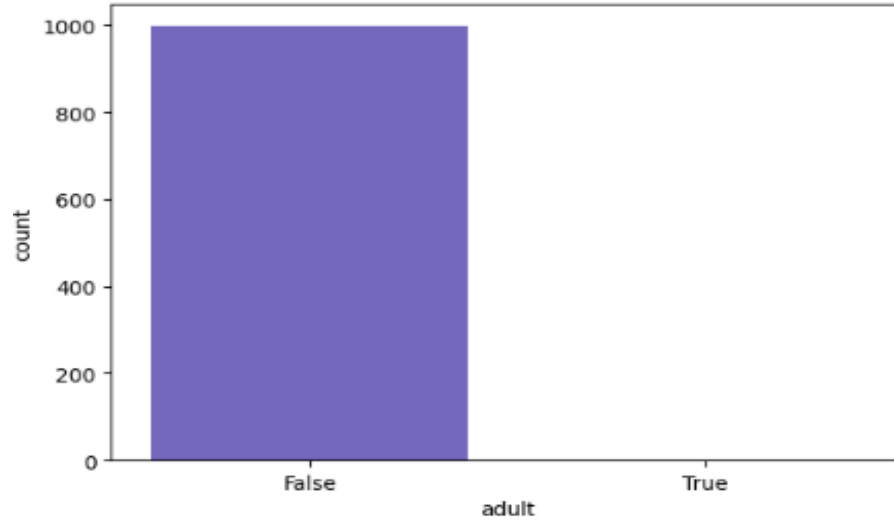
Alguns dos gráficos criados para a exploração:
Boxplots das variáveis numéricas:



Correlação entre as variáveis numéricas:



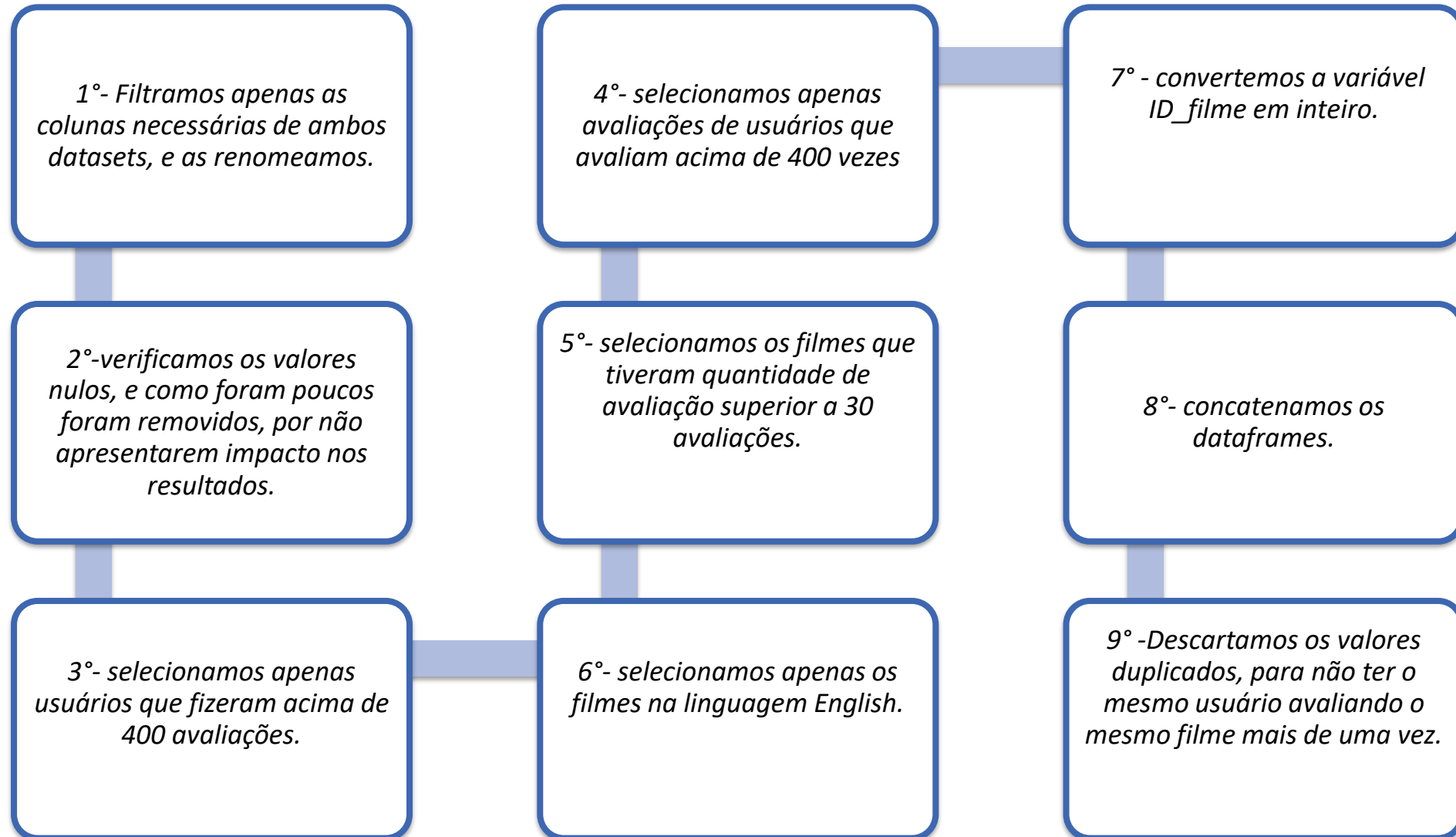
Distribuição das variáveis categóricas:



Limpeza e preparação dos Dados



Limpeza e preparação de dados



Aprendizado de Máquina



Utilizamos o sistema de recomendação para alcançar nossos objetivos com este projeto. **A distância euclidiana** foi a escolhida para analisar a similaridade entre os filmes e sua aplicação passou pelas seguintes etapas:

- 1) Criação de uma matriz que represente as avaliações dos usuários para os diferentes filmes;
- 2) Extração dos títulos únicos presentes na matriz;
- 3) Criação de um dataframe com estes títulos e exportação para um arquivo Excel;
- 4) Chamar a função definida para fazer a recomendação dos filmes pelo seu título usando a similaridade com base euclidiana;
- 5) Impressão das recomendações.

Para consolidar o aprendizado de máquina foi feito o pivoteamento da matriz, com o intuito de organizar as avaliações dos usuários, e chamamos a função definida “recomendar_filmes_por_titulos” para obter as recomendações com base em um filme de referência.

Ao executar essa função é feita a impressão das recomendações.

```
# Chamar a função `recomendar_filmes_por_titulo()`  
recomendacoes = recomendar_filmes_por_titulo('10,000 BC', matriz_usuario_filme, avaliacoes_e_filmes, num_recomendacoes=3)  
  
# Imprimir as recomendações  
print(recomendacoes)
```

```
('The King of Comedy', 2.345207879911715), ('The Quiet Man', 1.5), ('The Gate', 1.0)]
```

Teste e Acurácia



O teste é útil em um aprendizado de máquina para colocar em prático o modelo criado e checar o seu funcionamento. Já a acurácia serve para medir a capacidade deste funcionamento, classificando o seu desempenho.

Uma das medidas de acurácia mais utilizadas em sistemas de recomendações com classificações de usuários é o “Erro Médio Absoluto” (Mean Absolute Error - MAE), Que mede a diferença absoluta entre as classificações previstas e as reais.



Obrigado(a)!

