

SENAI – GASPAR RICARDO JUNIOR
CIÊNCIA DE DADOS

ANÁLISE DE DADOS DE LOCADORA DE FILMES

Giovana Rafaela da Silva

Maria Eduarda Claro

Pedro Henrique Dias de Paula Santos

Tobias Perassi Alquezar

Professor: André Cassulino Araújo Souza

Disciplina: Ciência de Dados

17 de Junho de 2025

Sumário

1. Introdução	3
2. Referencial Teórico.....	3
3. Metodologia	3
4. Análise de Dados	3
5. Conclusão.....	3
6. Códigos do Colab.....	4
7. Referências.....	6

1. Introdução

Este projeto tem como objetivo realizar uma análise exploratória e estatística com base em dados simulados de uma locadora de filmes. Através da utilização de bibliotecas da linguagem Python, busca-se identificar padrões de comportamento dos clientes, preferências de gênero cinematográfico e relações entre variáveis como idade, avaliação dos filmes e quantidade de filmes alugados.

2. Referencial Teórico

Durante o projeto, foram aplicados os seguintes conceitos estatísticos: medidas de tendência central, medidas de dispersão, análise de correlação, testes de normalidade, análise de dados categóricos e visualização de dados. As bibliotecas utilizadas incluem pandas, numpy, matplotlib, seaborn e scipy, todas amplamente utilizadas na análise de dados em Python.

3. Metodologia

Foi criada uma base de dados simulada com 500 registros de clientes, contendo atributos como idade, gênero, avaliação de filmes, quantidade de filmes alugados, tempo médio de uso por mês e categoria de filmes favorita. Os dados foram gerados de forma aleatória, representando um cenário realista. Em seguida, foram aplicadas técnicas de limpeza, análise estatística e visualização.

4. Análise de Dados

As análises realizadas contemplaram:

- Cálculo da média, mediana e moda de variáveis quantitativas.
- Cálculo de variância, desvio padrão e amplitude.
- Análise de correlação entre idade e quantidade de filmes alugados.
- Verificação da normalidade por meio dos testes de Shapiro-Wilk e Anderson-Darling.
- Geração de gráficos como histogramas, boxplots, mapas de calor e gráficos de barras.
- Interpretação dos resultados obtidos.

5. Conclusão

A análise demonstrou que a maioria dos clientes da locadora tem entre 25 e 40 anos, com preferência por comédias e ação. Não foi observada forte correlação entre idade e número de filmes alugados. A variável idade mostrou distribuição próxima à normalidade, enquanto a quantidade de filmes não. A análise permitiu identificar padrões relevantes para tomada de decisões e melhorias no serviço.

6. Códigos do Colab

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats


np.random.seed(42)


n = 500


dados = pd.DataFrame({

    'idade': np.random.randint(18, 65, size=n),

    'genero': np.random.choice(['Masculino', 'Feminino'], size=n),

    'avaliacao_filme': np.random.randint(1, 6, size=n),

    'qtd_filmes_alugados': np.random.poisson(5, size=n),

    'tempo_medio_uso_mes': np.random.normal(6, 2, size=n).round(1),

    'categoria_favorita': np.random.choice(['Ação', 'Comédia', 'Terror', 'Romance',

'Drama'], size=n)

})


dados.head()


# 3 - Medidas de Tendência Central

print("Média de idade:", dados['idade'].mean())
```

```
print("Mediana:", dados['idade'].median())  
print("Moda da avaliação:", dados['avaliacao_filme'].mode()[0])
```

4 - Medidas de Dispersão

```
print("Desvio padrão de filmes alugados:", dados['qtd_filmes_alugados'].std())  
print("Variância:", dados['qtd_filmes_alugados'].var())  
print("Amplitude:", dados['qtd_filmes_alugados'].max() -  
      dados['qtd_filmes_alugados'].min())
```

8 - Visualizações com Gráficos Estatísticos

```
sns.histplot(dados['idade'], bins=10, kde=True)  
plt.title('Distribuição de Idade dos Clientes')  
plt.xlabel('Idade')  
plt.ylabel('Frequência')  
plt.show()
```

10 - Análise de Dados Categóricos

```
sns.boxplot(x='genero', y='tempo_medio_uso_mes', data=dados)  
plt.title('Tempo Médio de Uso por Gênero')  
plt.show()
```

5 - Teste de normalidade

```
shapiro_idade = stats.shapiro(dados['idade'])  
print('Shapiro-Wilk para Idade:', shapiro_idade)
```

6 - Correlação entre Variáveis

```
sns.heatmap(dados[['idade', 'avaliacao_filme', 'qtd_filmes_alugados',  
'tempo_medio_uso_mes']].corr(), annot=True, cmap='coolwarm')
```

```
plt.title('Correlação entre Variáveis Numéricas')
```

```
plt.show()
```

7. Referências

Pandas Documentation. <https://pandas.pydata.org/>

NumPy Documentation. <https://numpy.org/>

Matplotlib Documentation. <https://matplotlib.org/>

Seaborn Documentation. <https://seaborn.pydata.org/>

SciPy Documentation. <https://docs.scipy.org/>

8. Link do Colab

<https://colab.research.google.com/drive/1Wg7zaWxFjOW-XRRGaFB4m60w9HXkn6UI#scrollTo=9MqSfvmd68UV>