# Bibliotecas

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

#==== 1. A

# Ler CSV

dados = "UC4/dados1.csv"

df = pd.read\_csv(dados, parse\_dates=["Data"])

# Verificar dimensões e tipos de dados

print("Formato do dataset:", df.shape)

print("\nInformações do dataset:")

print(df.info())

# Visualizar primeiras linhas

print("\nPrimeiras linhas:")

print(df.head())

# Verificar valores nulos

print("\nValores nulos por coluna:")

print(df.isnull().sum())

#==== 1. B

# Estatísticas básicas com describe

print(df.describe())

# Colunas de interesse numérico

colunas\_numericas = ['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R$)']

# Média, mediana, desvio padrão e variância

for coluna in colunas\_numericas:

    print(f"\n{coluna}")

    print("Média:", df[coluna].mean())

    print("Mediana:", df[coluna].median())

    print("Desvio padrão:", df[coluna].std())

    print("Variância:", df[coluna].var())

# Percentis da Receita

print("\nPercentis da Receita (R$):")

print(df["Receita (R$)"].quantile([0.25, 0.5, 0.75]))

# Companhia com maior receita total

receita\_por\_companhia = df.groupby("Companhia")['Receita (R$)'].sum().sort\_values(ascending=False)

print("\nCompanhia com maior receita total:")

print(receita\_por\_companhia.head(1))

# Companhia com maior número de passageiros

passageiros\_por\_companhia = df.groupby("Companhia")['Passageiros'].sum().sort\_values(ascending=False)

print("\nCompanhia com maior número de passageiros:")

print(passageiros\_por\_companhia.head(1))

# Contagem de voos por companhia

voos\_por\_companhia = df['Companhia'].value\_counts()

print("\nNúmero de voos por companhia:")

print(voos\_por\_companhia)

# Receita média por companhia

print("\nReceita média por companhia:")

print(df.groupby("Companhia")['Receita (R$)'].mean().sort\_values(ascending=False))

# Receita média por aeroporto de origem

print("\nReceita média por aeroporto de origem:")

print(df.groupby("Aeroporto Origem")['Receita (R$)'].mean().sort\_values(ascending=False))

#==== 1. C

# Histograma da distribuição de passageiros

sns.histplot(df['Passageiros'], bins=30, kde=True)

plt.title("Distribuição de Passageiros")

plt.xlabel("Número de Passageiros")

plt.ylabel("Frequência")

plt.show()

# Boxplot da ocupação (%) por companhia

sns.boxplot(data=df, x='Companhia', y='Ocupação (%)')

plt.title("Boxplot da Ocupação por Companhia Aérea")

plt.xlabel("Companhia Aérea")

plt.ylabel("Ocupação (%)")

plt.show()

# Gráfico de barras da receita média por companhia

receita\_media = df.groupby("Companhia")['Receita (R$)'].mean().sort\_values(ascending=False).reset\_index()

sns.barplot(data=receita\_media, x='Companhia', y='Receita (R$)')

plt.title("Receita Média por Companhia")

plt.ylabel("Receita Média (R$)")

plt.xlabel("Companhia")

plt.show()

# Scatterplot de Distância x Receita

sns.scatterplot(data=df, x='Distância (km)', y='Receita (R$)', hue='Companhia')

plt.title("Distância x Receita por Companhia")

plt.xlabel("Distância (km)")

plt.ylabel("Receita (R$)")

plt.show()

# Heatmap de correlação entre variáveis numéricas

corr = df[['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R$)']].corr()

sns.heatmap(corr, annot=True, cmap="coolwarm")

plt.title("Mapa de Calor das Correlações")

plt.show()

#==== 1. D

print("\n\t============Respostas=============\n")

# Companhia com maior número de voos

print("Companhia com maior número de voos:", voos\_por\_companhia.idxmax())

# Correlação distância e receita

corr\_dist\_receita = df[['Distância (km)', 'Receita (R$)']].corr().iloc[0,1]

print("\nCorrelação Distância x Receita:", round(corr\_dist\_receita, 2))

# Correlação ocupação e receita

corr\_ocup\_receita = df[['Ocupação (%)', 'Receita (R$)']].corr().iloc[0,1]

print("Correlação Ocupação x Receita:", round(corr\_ocup\_receita, 2))

# Aeroportos de origem com mais voos

print("\nTop 5 aeroportos de origem com mais voos:")

print(df['Aeroporto Origem'].value\_counts().head(5))

# Respostas:

# 1. Qual companhia tem maior participação em número de voos?

#   R - A Azul é a companhia que mais aparece nos dados — foi a que teve o maior número de voos.

# 2. A distância influencia a receita?

#   R -  Até influencia um pouco, mas bem pouco mesmo. A correlação entre distância e receita foi de 0,11, o que é uma ligação fraca. Ou seja, voos mais longos tendem a gerar mais receita, mas não é uma regra forte. Outros fatores como o preço das passagens ou a ocupação dos voos acabam pesando mais.

# 3. Os voos com maior ocupação são necessariamente os de maior receita?

#   R -  Nem sempre. A correlação entre ocupação e receita também foi baixa (0,13), o que mostra que um voo cheio não garante uma receita alta. Às vezes o voo tá lotado, mas com passagens mais baratas. Então, ocupação ajuda, mas não é o único fator que influencia a receita.

# 4. Quais aeroportos de origem concentram mais voos?

#   R -

#       CGH (Congonhas) – 32 voos

#       GIG (Galeão) – 28 voos

#       SDU (Santos Dumont) – 22 voos

#       GRU (Guarulhos) – 20 voos

#       BSB (Brasília) – 18 voos

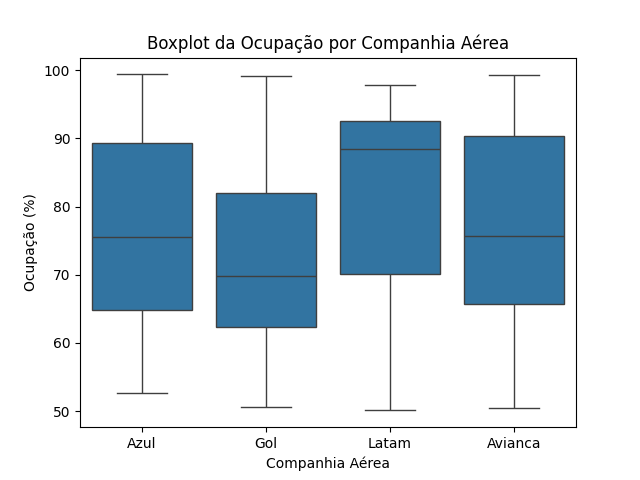
# GRÁFICOS:

1 - Histograma da distribuição de passageiros

Gráfico, Histograma

O conteúdo gerado por IA pode estar incorreto.

2 - Boxplot da ocupação (%) por companhia

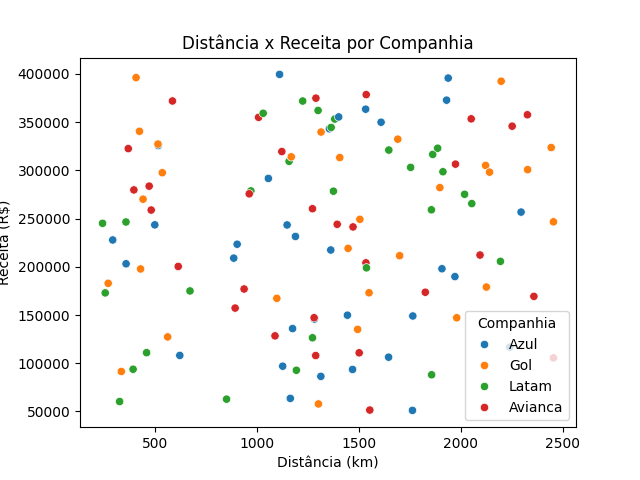


3 - Gráfico de barras da receita média por companhia

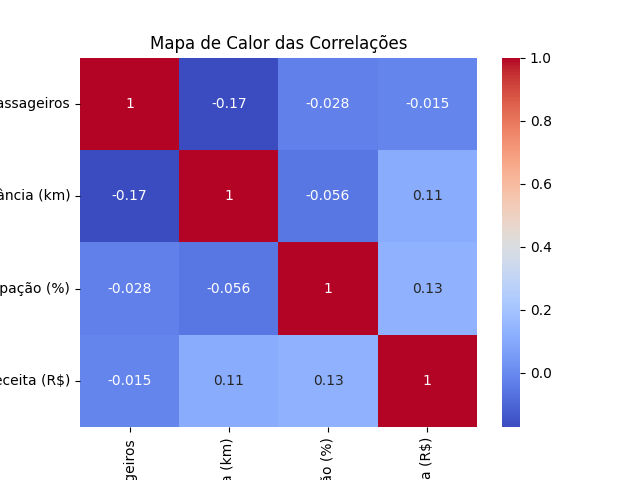
Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

4 - Scatterplot de Distância x Receita



5 - Heatmap de correlação entre variáveis numéricas



# **EXTENSÃO**

# Criar uma coluna com o mês/ano

df['Ano-Mês'] = df['Data'].dt.to\_period('M')

# Receita média por mês

receita\_media\_mes = df.groupby('Ano-Mês')['Receita (R$)'].mean().sort\_index()

print("\nReceita média por mês:")

print(receita\_media\_mes)

# Gráfico

receita\_media\_mes.index = receita\_media\_mes.index.astype(str)

plt.figure(figsize=(12, 6))

sns.lineplot(x=receita\_media\_mes.index, y=receita\_media\_mes.values, marker='o')

plt.title('Receita Média por Mês')

plt.xlabel('Ano-Mês')

plt.ylabel('Receita Média (R$)')

plt.tight\_layout()

plt.show()

# Ocupação média por companhia

ocupacao\_media = df.groupby('Companhia')['Ocupação (%)'].mean().sort\_values(ascending=False)

print("\nOcupação média por companhia:")

print(ocupacao\_media)

# Gráfico

plt.figure(figsize=(10, 6))

sns.barplot(x=ocupacao\_media.values, y=ocupacao\_media.index, palette='viridis')

plt.title('Ocupação Média por Companhia')

plt.xlabel('Ocupação Média (%)')

plt.ylabel('Companhia')

plt.tight\_layout()

plt.show()

