# A. Carregar e explorar os dados

import pandas as pd

# 1.Ler o CSV usando pandas.

df = pd.read\_csv("dados1.csv")

print(df)

# 2. Verificar número de linhas/colunas e tipos de dados (.info()).

df.info()

# 3. Exibir as 5 primeiras linhas (.head()).

print(df.head(5))

# 4. Conferir se há valores nulos.

if df.isnull().values.any():

    print("Há valores nulos.")

else:

    print("Não há valores nulos.")

# B. Estatísticas Descritivas

import pandas as pd

import numpy as np

df = pd.read\_csv("dados1.csv")

# 1. Média, mediana, desvio-padrão e variância de:Passageiros, Distância (km), Ocupação (%), Receita (R$)

print("Estatísticas Descritivas:")

for col in ["Passageiros", "Distância (km)", "Ocupação (%)", "Receita (R$)"]:

    print(f"\n{col}:")

    print(f"Média: {df[col].mean()}")

    print(f"Mediana: {df[col].median()}")

    print(f"Desvio Padrão: {df[col].std()}")

    print(f"Variância: {df[col].var()}")

# 2.Calcular o percentil 25%, 50%, 75% da receita.

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

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

print(f"25%: {percentis[0.25]}")

print(f"50%: {percentis[0.5]}")

print(f"75%: {percentis[0.75]}")

# 3. Encontrar a companhia com maior receita total e com maior número de passageiros.

maior\_receita = df.loc[df["Receita (R$)"].idxmax()]

maior\_passageiros = df.loc[df["Passageiros"].idxmax()]

print("\nCompanhia com maior receita:")

print(maior\_receita)

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

print(maior\_passageiros)

# 4. Contagem de voos por companhia

contagem\_voos = df["Companhia"].value\_counts()

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

print(contagem\_voos)

# 5. Receita média por companhia e por aeroporto de origem

receita\_media = df.groupby("Companhia")["Receita (R$)"].mean()

print("\nReceita média por companhia:")

print(receita\_media)

receita\_media\_por\_aeroporto = df.groupby("Aeroporto Origem")["Receita (R$)"].mean()

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

print(receita\_media\_por\_aeroporto)

# C. Visualizações com Seaborn

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

#1. Histograma da distribuição de passageiros.

df = pd.read\_csv("dados1.csv")

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

sns.histplot(df["Passageiros"], bins=20, kde=True).set(title='Distribuição de Passageiros')

plt.show()

#2. Boxplot da ocupação (%) separada por companhia aérea.

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

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

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

plt.xticks(rotation=45)

plt.show()

#3. Gráfico de barras da receita média por companhia.

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

sns.barplot(x="Companhia", y="Receita (R$)", data=df, estimator=sum)

plt.title("Receita Total por Companhia")

plt.xticks(rotation=45)

plt.show()

#4. Scatterplot de distância x receita para verificar relação.

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

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

plt.title("Relação entre Distância e Receita")

plt.show()

#5. (Desafio) Heatmap de correlação entre variáveis numéricas (Passageiros, Distância (km), Ocupação (%), Receita (R$)).

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

correlation\_matrix = df[["Passageiros", "Distância (km)", "Ocupação (%)", "Receita (R$)"]].corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f")

plt.title("Heatmap de Correlação entre Variáveis Numéricas")

plt.show()

# D. Perguntas Analíticas

import pandas as pd

df = pd.read\_csv("dados1.csv")

#•  Qual companhia tem maior participação em número de voos?

participacao\_voos = df["Companhia"].value\_counts(normalize=True)

maior\_participacao = participacao\_voos.idxmax()

print(f"A companhia com maior participação em número de voos é: {maior\_participacao}")

#•  A distância influencia a receita?

correlacao = df[["Distância (km)", "Receita (R$)"]].corr().iloc[0, 1]

if correlacao > 0:

    print("A distância tem uma influência positiva na receita.")

elif correlacao < 0:

    print("A distância tem uma influência negativa na receita.")

else:

    print("Não há correlação entre distância e receita.")

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

ocupacao\_maior\_receita = df.loc[df["Receita (R$)"].idxmax(), "Ocupação (%)"]

print(f"A maior ocupação é de {ocupacao\_maior\_receita}%, mas isso não implica que todos os voos com alta ocupação tenham alta receita.")

#•  Quais aeroportos de origem concentram mais voos?

aeroportos\_concentracao = df["Aeroporto Origem"].value\_counts()

print("Aeroportos de origem com mais voos:")

print(aeroportos\_concentracao)

Gráfico, Histograma

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de mapa de árvore

O conteúdo gerado por IA pode estar incorreto.Gráfico, Gráfico de dispersão

O conteúdo gerado por IA pode estar incorreto.