

## 1.Ler o CSV usando pandas.

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

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

4.Conferir se há valores nulos.

\*/ \*/

B. Estatísticas Descritivas

1. Calcular, usando pandas e numpy:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import statistics
```

2. Média, mediana, desvio-padrão e variância de:

```
print("\nestatística com describ")
print(df.describe())
estatística com describ
```

	Data	Passageiros	Distância (km)	Ocupação (%)	Receita (R\$)
count	120	120.000000	120.000000	120.000000	120.000000
mean	2023-03-01 12:00:00	170.566667	1318.662500	76.667500	236158.113333
min	2023-01-01 00:00:00	50.000000	240.200000	50.100000	51119.750000
25%	2023-01-30 18:00:00	95.000000	898.925000	64.925000	155383.922500
50%	2023-03-01 12:00:00	172.500000	1334.050000	77.750000	244611.075000
75%	2023-03-31 06:00:00	241.000000	1833.200000	90.325000	319910.972500
max	2023-04-30 00:00:00	298.000000	2455.300000	99.400000	399535.480000
std	NaN	75.905374	613.184727	14.886247	97011.399647

```
Passageiros
Distância (km)
Ocupação (%)
Receita (R$)
# estatísticas descritivas(media, desvio padrao, etc.)
print("\nestatística com describ")
print(df.describe())
campos = ['Passageiros','Distância (km)','Ocupação (%)','Receita (R$)']
for dados in campos:
    #print(f"\n{campos}")
    print("Média:", df[campos].mean())
    print("Mediana:", df[campos].median())
    print("Desvio padrão:", df[campos].std())
    print("Variância:", df[campos].var())

['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R$)']
Média campos: Passageiros      170.566667
Distância (km)      1318.662500
Ocupação (%)      76.667500
Receita (R$)      236158.113333
dtype: float64
Mediana campos: Passageiros      172.500
Distância (km)      1334.050
Ocupação (%)      77.750
Receita (R$)      244611.075
dtype: float64
Des padrão campos: Passageiros      75.905374
Distância (km)      613.184727
Ocupação (%)      14.886247
Receita (R$)      97011.399647
dtype: float64
Variância campos: Passageiros      5.761626e+03
Distância (km)      3.759955e+05
```

Ocupação (%) 2.216004e+02  
Receita (R\$) 9.411212e+09  
dtype: float64

['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R\$)']  
Média campos: Passageiros 170.566667  
Distância (km) 1318.662500  
Ocupação (%) 76.667500  
Receita (R\$) 236158.113333  
dtype: float64  
Mediana campos: Passageiros 172.500  
Distância (km) 1334.050  
Ocupação (%) 77.750  
Receita (R\$) 244611.075  
dtype: float64  
Des padrão campos: Passageiros 75.905374  
Distância (km) 613.184727  
Ocupação (%) 14.886247  
Receita (R\$) 97011.399647  
dtype: float64  
Variância campos: Passageiros 5.761626e+03  
Distância (km) 3.759955e+05  
Ocupação (%) 2.216004e+02  
Receita (R\$) 9.411212e+09  
dtype: float64

['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R\$)']  
Média campos: Passageiros 170.566667  
Distância (km) 1318.662500  
Ocupação (%) 76.667500  
Receita (R\$) 236158.113333  
dtype: float64  
Mediana campos: Passageiros 172.500  
Distância (km) 1334.050  
Ocupação (%) 77.750  
Receita (R\$) 244611.075  
dtype: float64  
Des padrão campos: Passageiros 75.905374  
Distância (km) 613.184727  
Ocupação (%) 14.886247  
Receita (R\$) 97011.399647  
dtype: float64  
Variância campos: Passageiros 5.761626e+03  
Distância (km) 3.759955e+05  
Ocupação (%) 2.216004e+02  
Receita (R\$) 9.411212e+09  
dtype: float64

['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R\$)']  
Média campos: Passageiros 170.566667  
Distância (km) 1318.662500  
Ocupação (%) 76.667500  
Receita (R\$) 236158.113333  
dtype: float64

```
Mediana campos: Passageiros          172.500
Distância (km)      1334.050
Ocupação (%)        77.750
Receita (R$)        244611.075
dtype: float64
Des padrão campos: Passageiros        75.905374
Distância (km)      613.184727
Ocupação (%)        14.886247
Receita (R$)        97011.399647
dtype: float64
Variância campos: Passageiros        5.761626e+03
Distância (km)      3.759955e+05
Ocupação (%)        2.216004e+02
Receita (R$)        9.411212e+09
```

```
3.Calcular o percentil 25%, 50%, 75% da receita.
print("\nPercentis (R$):")
print(df["Receita (R$)"].quantile([0.25, 0.5, 0.75]))
Percentis (R$):
0.25    155383.9225
0.50    244611.0750
0.75    319910.9725
Name: Receita (R$), dtype: float64
```

```
4.Encontrar a companhia com maior receita total e com maior número de passageiros.
maior_receita = df.groupby("Companhia")['Receita
(R$)'].sum().sort_values(ascending=False)
print("\nMaior receita total:")
print(maior_receita.head())
Maior receita total:
Companhia
Avianca    7275315.41
Gol        7219424.84
Azul       6944954.73
Latam      6899278.62
Name: Receita (R$), dtype: float64
```

```
5.Contagem de voos por companhia.
passageiros_comp =
df.groupby("Companhia")['Passageiros'].sum().sort_values(ascending=False)
print("\nMaior número de passageiros:")
print(passageiros_comp.head(1))
Maior número de passageiros:
Companhia
Avianca    5420
Name: Passageiros, dtype: int64
```

```
6.Receita média por companhia e por aeroporto de origem.
voos = df['Companhia'].value_counts()
print("\nNúmero de voos:")
print(voos)
Número de voos:
Companhia
Azul      32
```

Avianca	30
Gol	29
Latam	29

Name: count, dtype: int64

```
PS C:\Users\edson\OneDrive\Documents\GitHub\python IA\Matematica>
```

\*/\*

## C. Visualizações com Seaborn

### 1.Histograma da distribuição de passageiros.

```
plt.figure()
sns.histplot(df["Passageiros"], kde=True)
plt.title("Distribuição")
plt.xlabel("Total de Passageiros")
plt.show(block=False)
```

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

```
plt.figure()
sns.boxplot(x='Companhia', y='Ocupação (%)', data=df)
plt.title("Ocupação")
plt.xlabel("Companhia")
plt.show(block=False)
```

```
input("Pressione Enter para fechar tudo...")
```

```
plt.close('all')
```

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

```
receita_med = df.groupby("Companhia")['Receita (R$)'].mean().sort_values(ascending=False).reset_index()
sns.barplot(x="Companhia", y="Receita (R$)", data=receita_med, estimator="mean")
plt.title("Receita Média por Companhia ")
plt.ylabel("Receita (R$)")
plt.xlabel("Companhia")
plt.show(block=False)
```

4. Scatterplot de distância x receita para verificar relação.

```
plt.figure()
sns.scatterplot(x="Distância (km)", y="Receita (R$)", hue="Companhia", data=df)
plt.title("4. distância x receita")
plt.xlabel("Distância (km)")
plt.ylabel("Receita (R$)")
plt.show(block=False)
```

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

```
corr = df[['Passageiros', 'Distância (km)', 'Ocupação (%)', 'Receita (R$)']].corr()
sns.heatmap(corr, annot=True, cmap="coolwarm")
plt.title("5. Mapa de Calor das Correlações")
plt.show(block=False)
```

#### D. Perguntas Analíticas

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

```
voos = df['Companhia'].value counts()
```

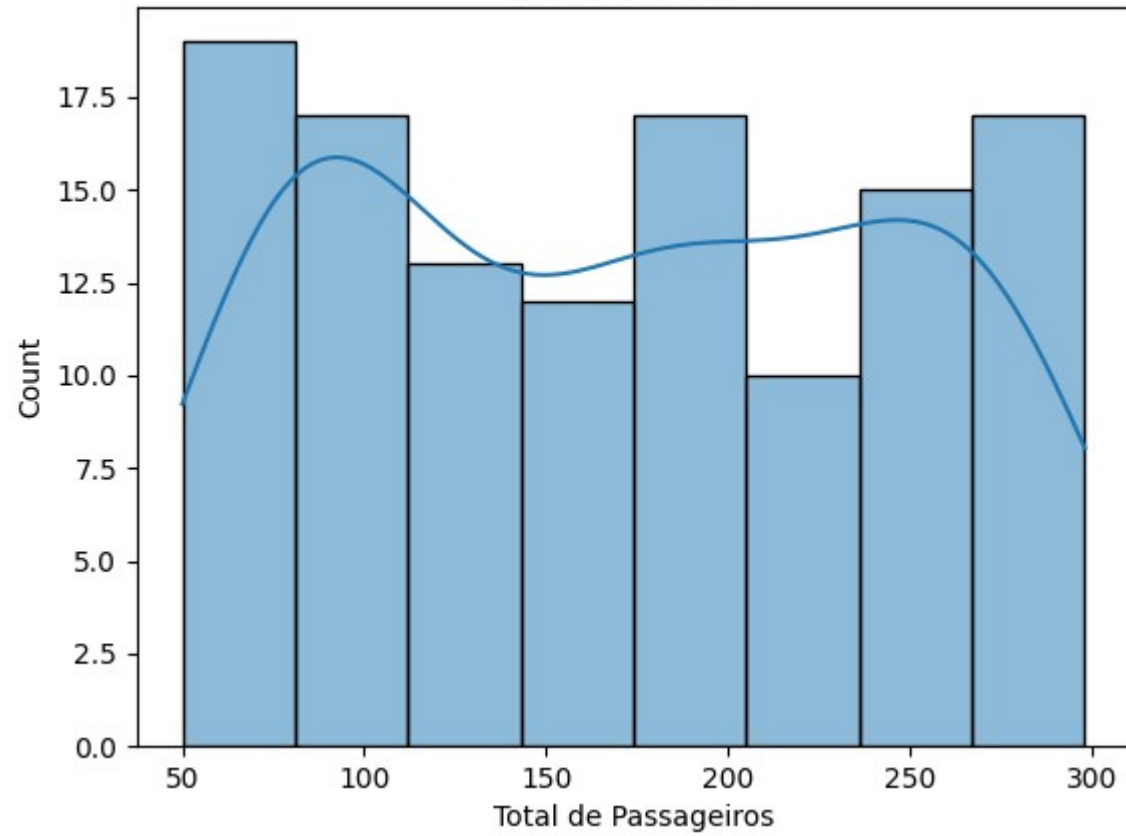
```
print (voos)
Companhia
Azul      32
```

```
2.A distância influencia a receita?
sim
Correlação Distância x Receita: 0.11
```

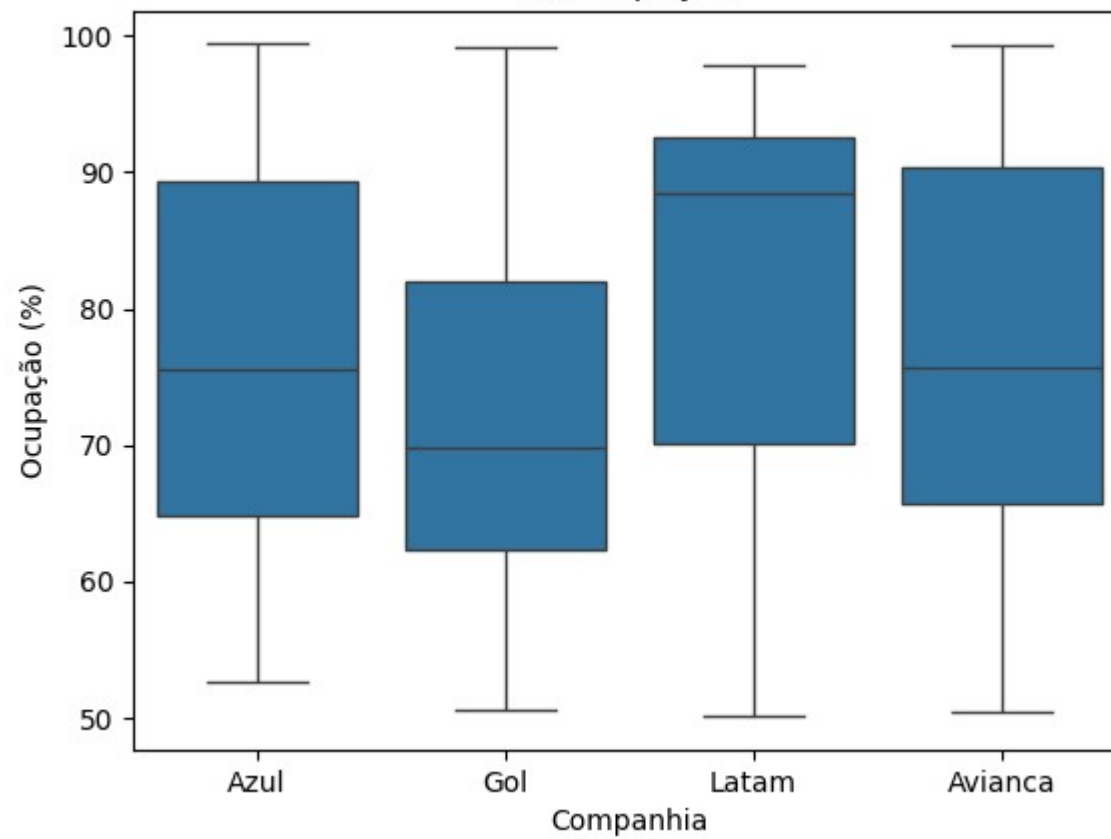
```
Os voos com maior ocupação são necessariamente os de maior receita?
sim
```

```
Quais aeroportos de origem concentram mais voos?
```

## 1. Distribuição

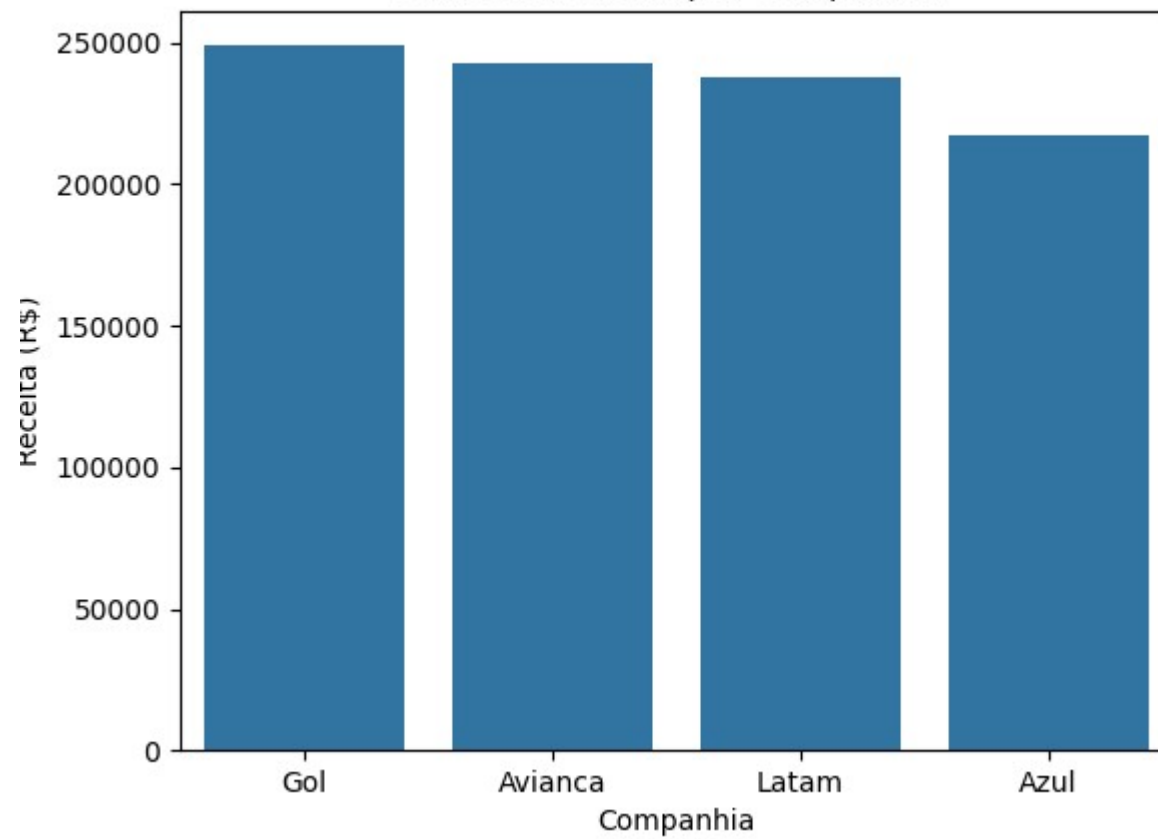


## 2. Ocupação

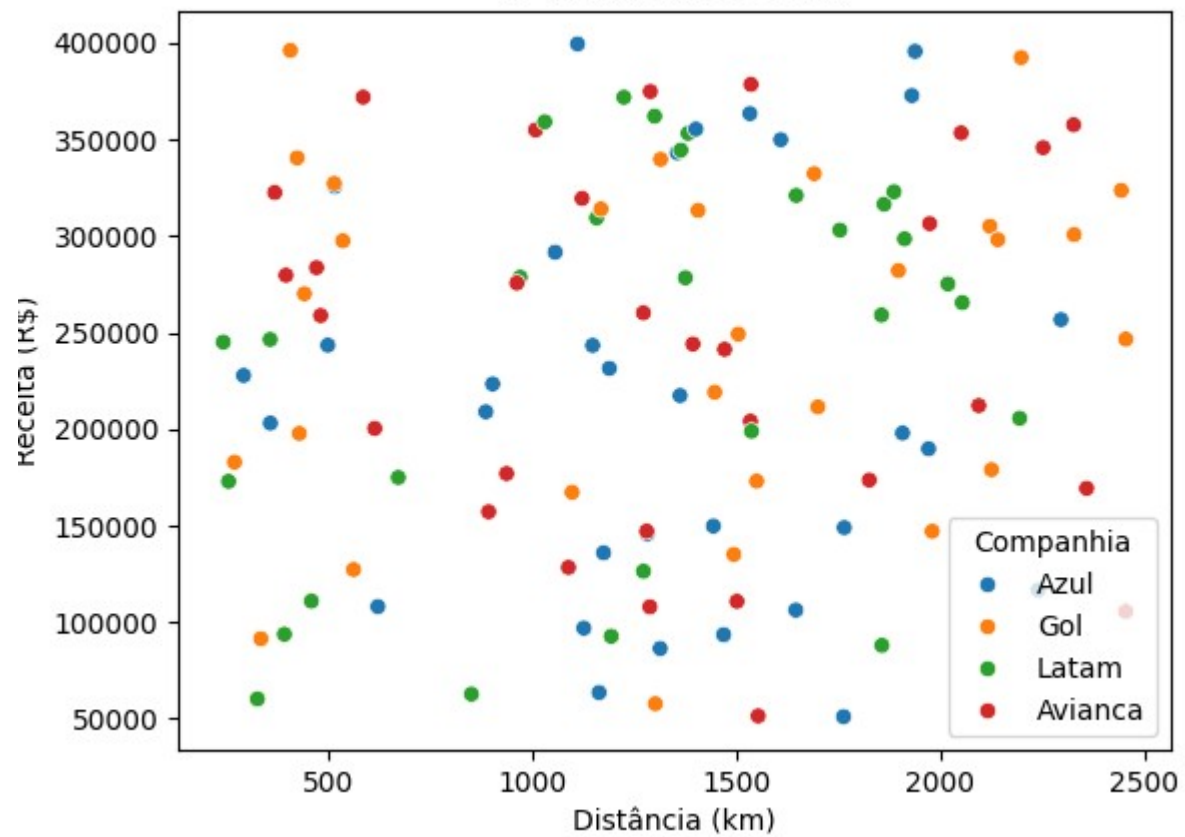




### 3. Receita Média por Companhia



4. distância x receita



5. Mapa de Calor das Correlações

