Explicando o que faz a instrução para preparar

grouped = df.groupby(['Companhia', 'Aeroporto Origem', 'Aeroporto Destino']).agg({ 'Passageiros': 'sum', 'Receita (R$)': 'sum', 'Ocupação (%)': 'mean' }).reset\_index()

Vamos analisar linha por linha o que esse código faz. Ele está em **Python**, usando **pandas** para manipulação de dados.

grouped = df.groupby(['Companhia', 'Aeroporto Origem', 'Aeroporto Destino']).agg({

'Passageiros': 'sum',

'Receita (R$)': 'sum',

'Ocupação (%)': 'mean'

}).reset\_index()

**Passo a passo:**

1. **df.groupby([...])**
   * df é um DataFrame do pandas.
   * groupby(['Companhia', 'Aeroporto Origem', 'Aeroporto Destino']) agrupa os dados considerando **cada combinação única de Companhia, Aeroporto de Origem e Aeroporto de Destino**.
   * Exemplo: todas as linhas da Azul que vão de GIG para GRU serão agrupadas juntas.
2. **.agg({...})**
   * agg significa **agregação**. Aqui você diz como combinar os dados de cada grupo:
     + 'Passageiros': 'sum' → soma o total de passageiros de cada grupo.
     + 'Receita (R$)': 'sum' → soma a receita total de cada grupo.
     + 'Ocupação (%)': 'mean' → calcula a **média da ocupação** de cada grupo.
3. **.reset\_index()**
   * Depois do groupby, os grupos se tornam índices do DataFrame.
   * reset\_index() transforma esses índices em colunas normais novamente, deixando o DataFrame “limpo” para uso posterior.

**Resultado:**

Você terá um **novo DataFrame** (grouped) com estas colunas:

* Companhia
* Aeroporto Origem
* Aeroporto Destino
* Passageiros (total por grupo)
* Receita (R$) (total por grupo)
* Ocupação (%) (média por grupo)

Ele resume o **total de passageiros e receita**, e a **média da ocupação**, para cada rota de cada companhia.

O que faz a instrução:

grouped['Receita por Passageiro'] = grouped['Receita (R$)'] / grouped['Passageiros']

Essa linha cria uma **nova coluna** no DataFrame grouped chamada Receita por Passageiro. Vamos detalhar:

grouped['Receita por Passageiro'] = grouped['Receita (R$)'] / grouped['Passageiros']

**Passo a passo:**

1. **grouped['Receita (R$)']**
   * Seleciona a coluna de receita total do DataFrame grouped.
2. **grouped['Passageiros']**
   * Seleciona a coluna com o número total de passageiros por grupo.
3. **Divisão /**
   * Faz a **divisão elemento a elemento** (linha a linha) da receita total pelo número de passageiros.
   * Isso calcula quanto cada passageiro gerou de receita **em média**.
4. **grouped['Receita por Passageiro'] = ...**
   * Cria a nova coluna e armazena o resultado da divisão.

**Resultado:**

O DataFrame grouped terá agora uma coluna extra com o valor médio que cada passageiro contribuiu para a receita daquela rota de cada companhia.