

"ANÁLISE DE OCORRÊNCIAS AERONÁUTICAS NA AVIAÇÃO BRASILEIRA"

Robert Rossi Silva de Mesquita
Cientista de Dados

Teste Técnico Cientista de Dados - A3Data





OBJETIVO

- Explorar dados de ocorrências aeronáuticas disponibilizados pelo CENIPA (Centro de Investigação e Prevenção de Acidentes Aeronáuticos).
- Aplicar técnicas de ciência de dados para identificar padrões, insights para compreensão dos fatos relacionados a ocorrências aéreas.



DADOS UTILIZADOS

Fonte: CENIOA – Dados abertos do Governo Federal.

Tabelas utilizadas:

- **ocorrencia.csv;**
 - Dados gerais das ocorrências.
- **aeronave.csv;**
 - Dados gerais das aeronaves envolvidas.
- **fator_contribuinte.csv**
 - Fatores contribuintes das ocorrências.



TECNOLOGIAS UTILIZADAS

- Google Colab;
- Python;
 - Pandas, numpy, seaborn, matplotlib, scikit-learn, spacy, chardet.
- Github



ETAPAS DA EXPLORAÇÃO DOS DADOS

Tratamento dos Dados: Qualidade dos dados, limpeza, merge;

Análise Exploratória: Exploração visual dos dados, Análises temporais e espaciais.

Hipóteses e Estatísticas: Formulação de hipóteses e testes e validação.



ETAPAS DA EXPLORAÇÃO DOS DADOS

Tratamento dos Dados: Qualidade dos dados, limpeza, merge;

Análise Exploratória: Exploração visual dos dados, Análises temporais e espaciais.

Hipóteses e Estatísticas: Formulação de hipóteses e testes e validação.

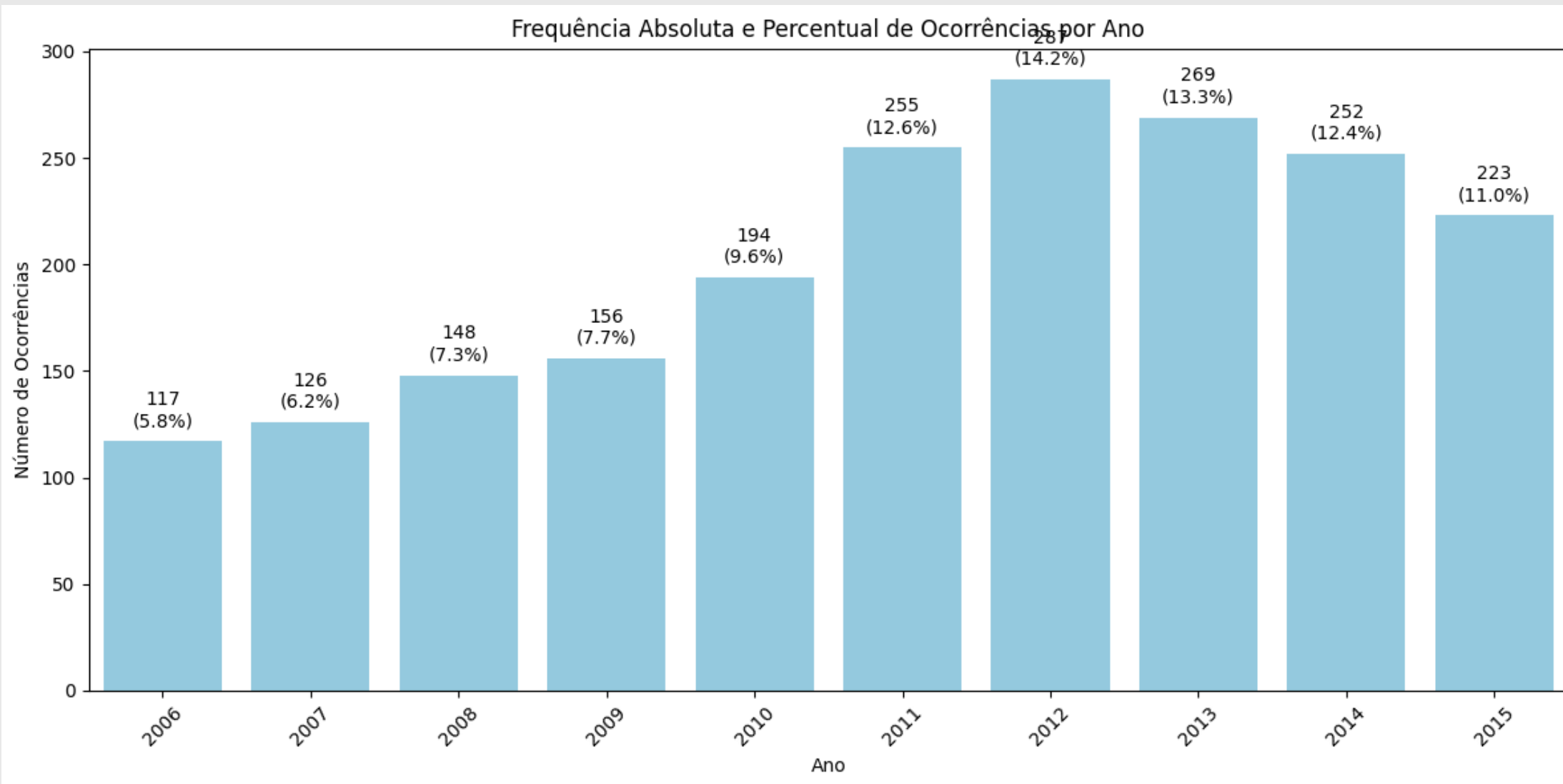
ETAPA 1- TRATAMENTO DOS DADOS

- Leitura e inspeção dos dados;
- Identificação de chaves primárias e estrangeiras;
- Padronização dos dados e imputação/exclusão de campos nulos;
- Padronização de Encoding de tabelas;
- Aplicação de técnica One-Hot Encoding em variável de interesse;
- Junção de tabelas.

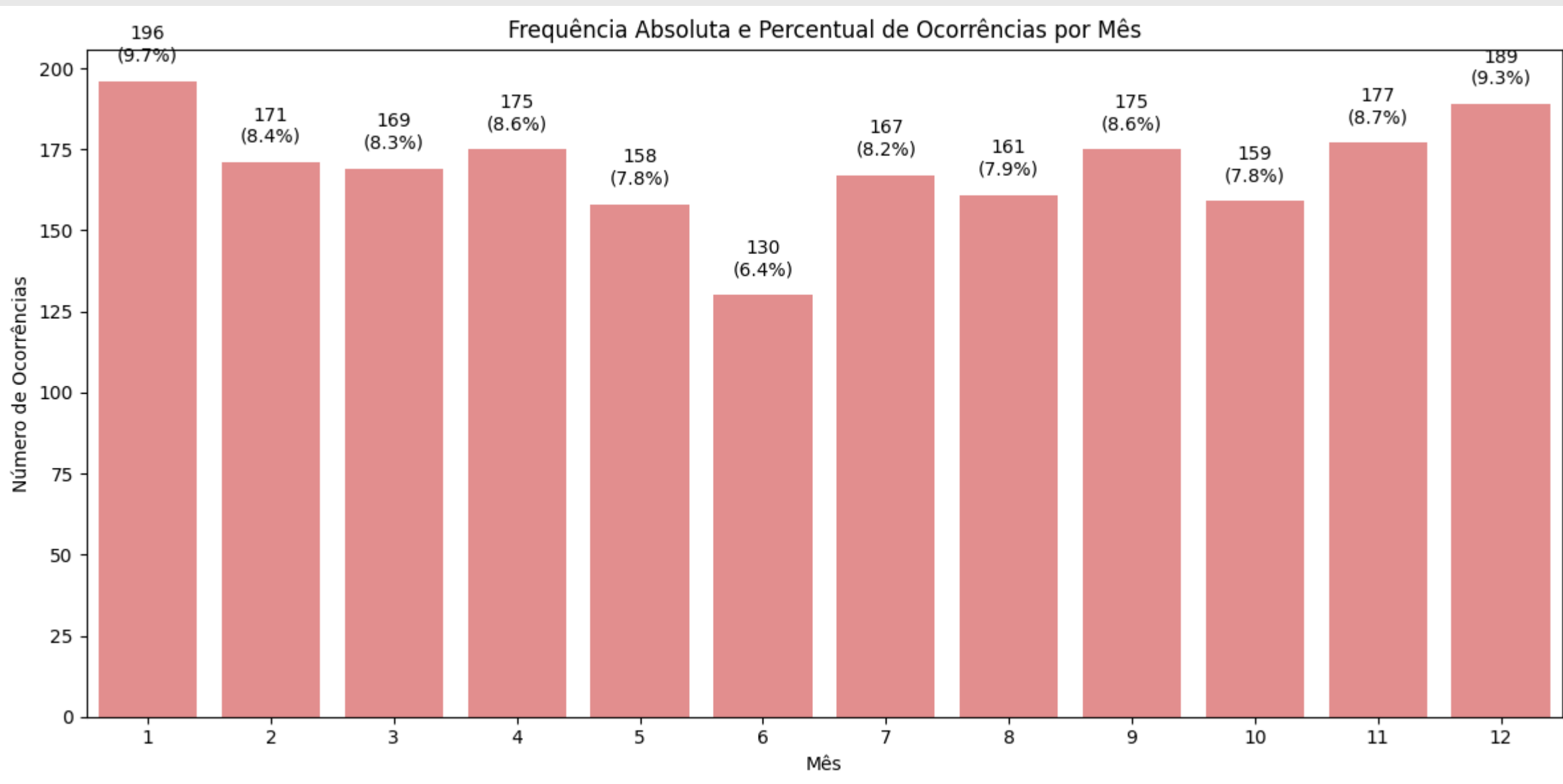
ETAPA 2- ANÁLISE EXPLORATÓRIA

- Análise temporal (ano, mês, hora);
- Análise de ocorrências por tipo de operação;
- Análise de ocorrências por fatores de área;
- Análise espacial-temporal de fatalidades, com destaque para os eventos com maior impacto.

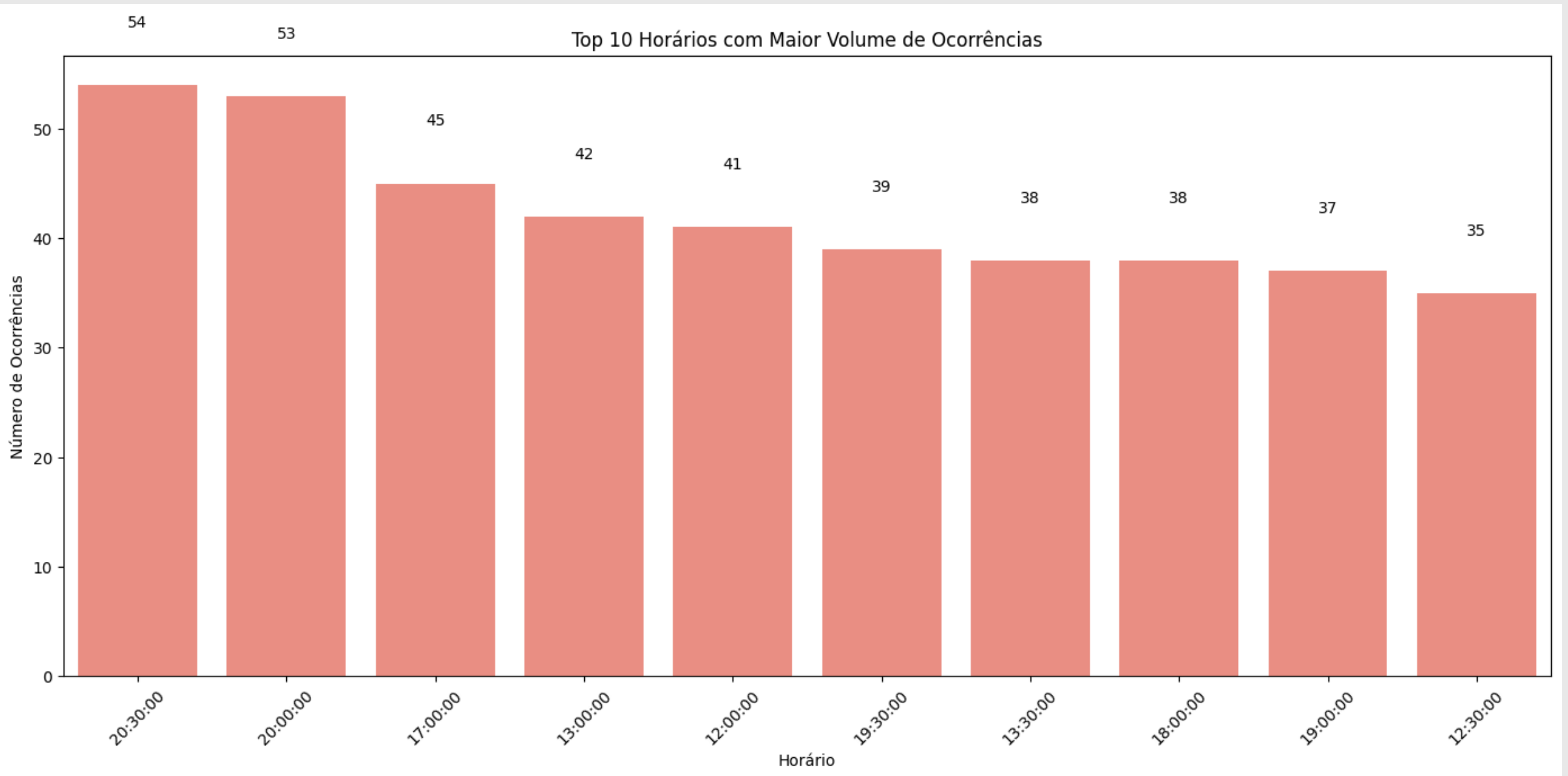
ETAPA 2 -Ocorrências por Ano



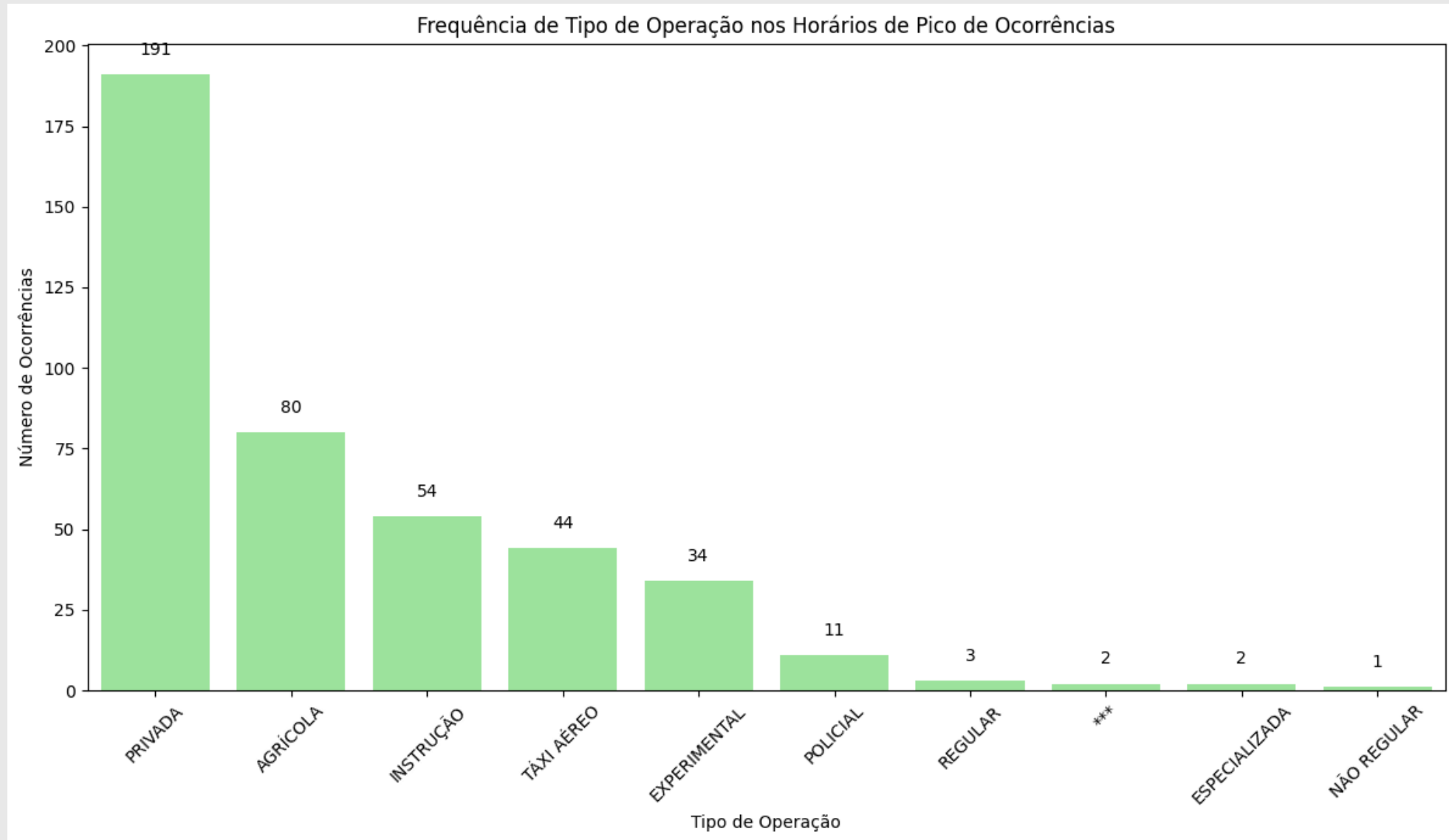
ETAPA 2 -Ocorrências por Mês



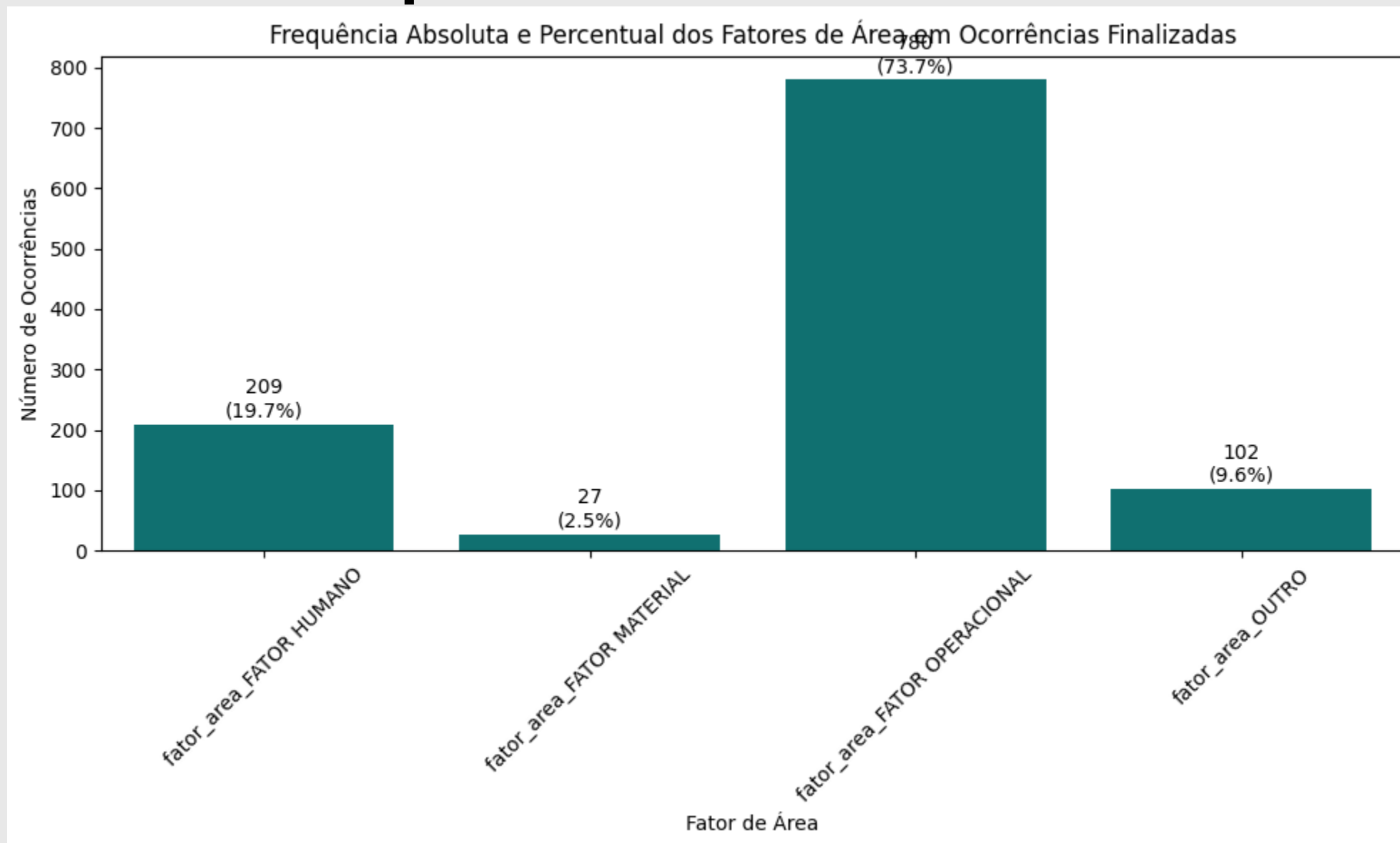
ETAPA 2 – Horários com PICO de ocorrências



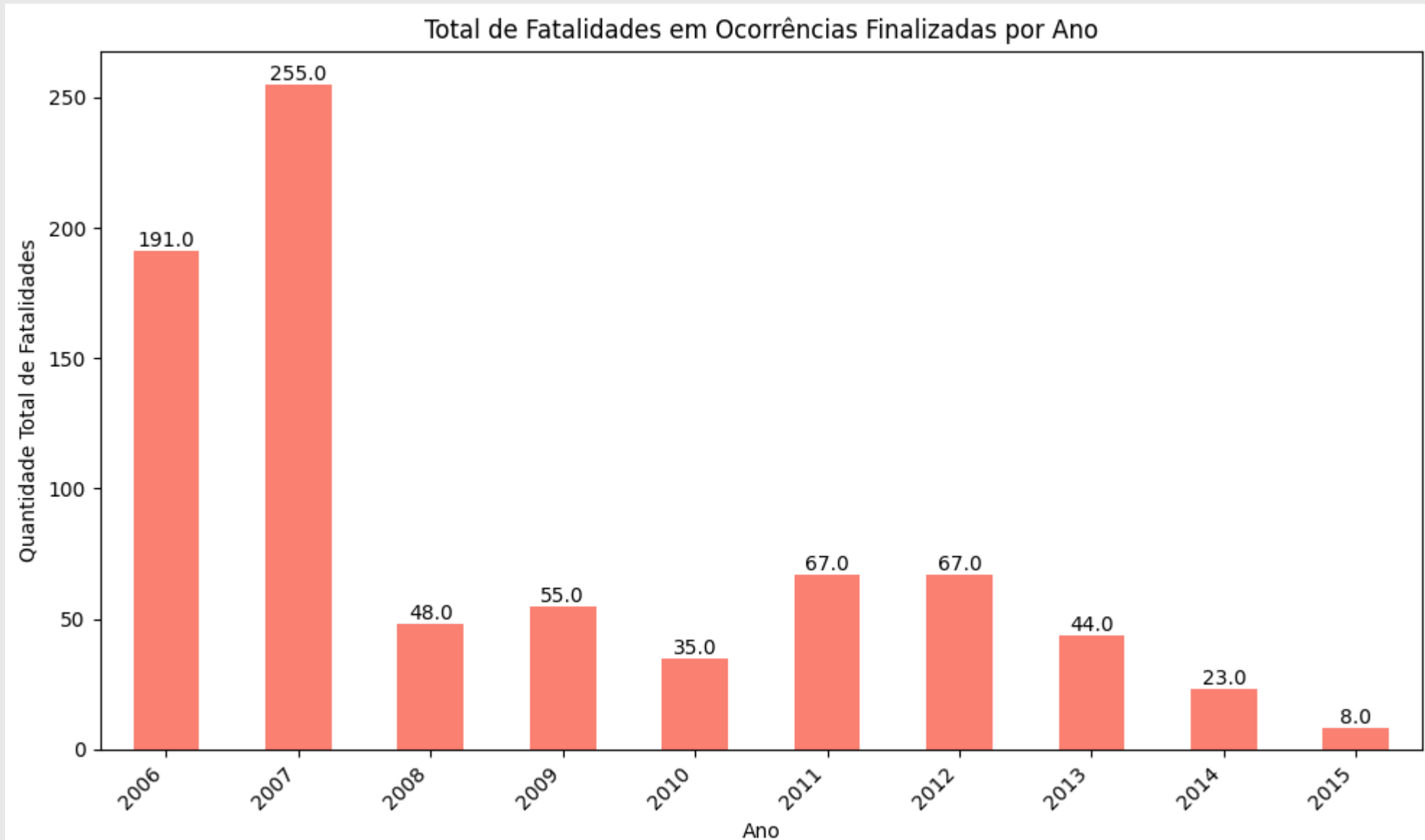
ETAPA 2 – Tipo de operação envolvida em ocorrências em horário de PICO



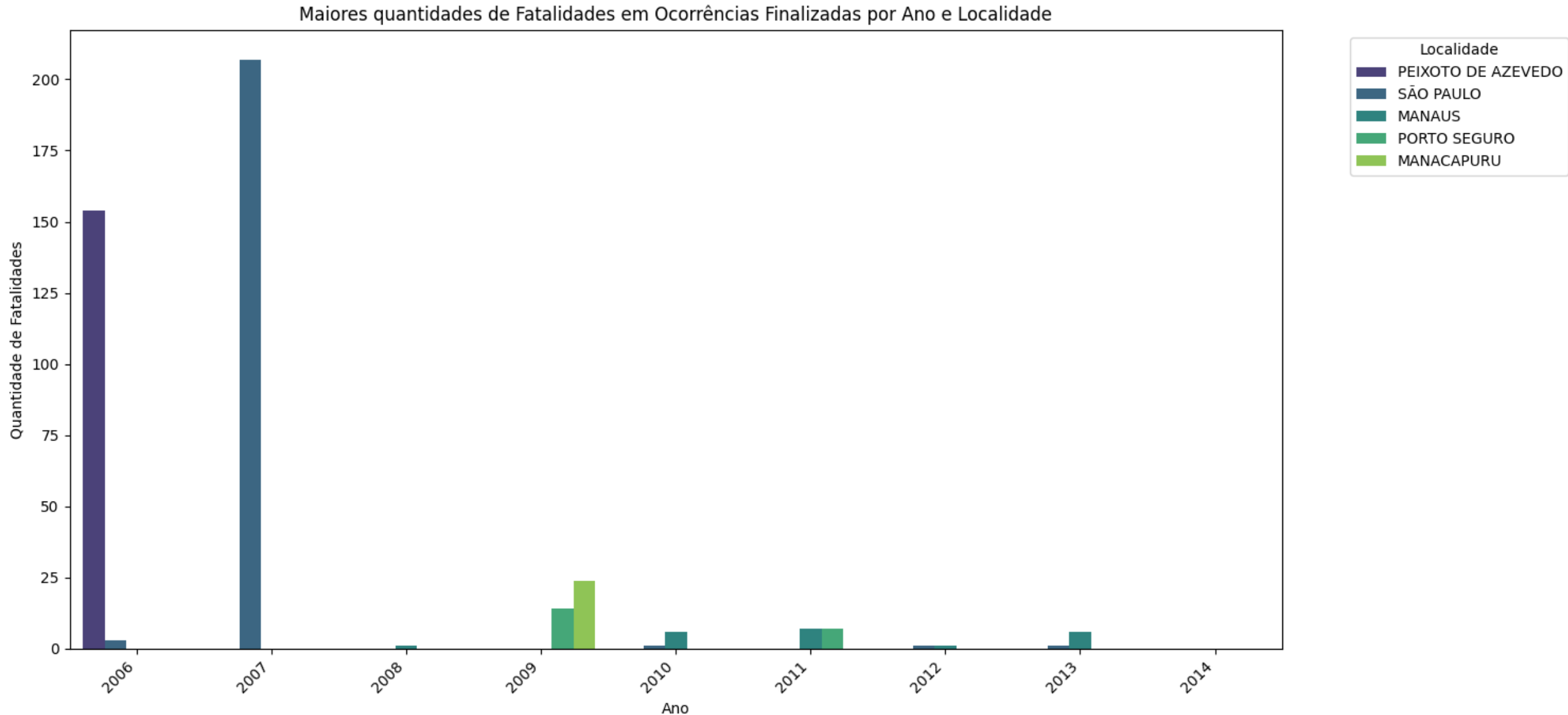
ETAPA 2 – Análise da frequência dos fatores contribuintes por área.



ETAPA 2 – Análise espacial-temporal de fatalidades.



ETAPA 2 – Análise espacial-temporal de fatalidades, com destaque para os eventos com maior impacto.



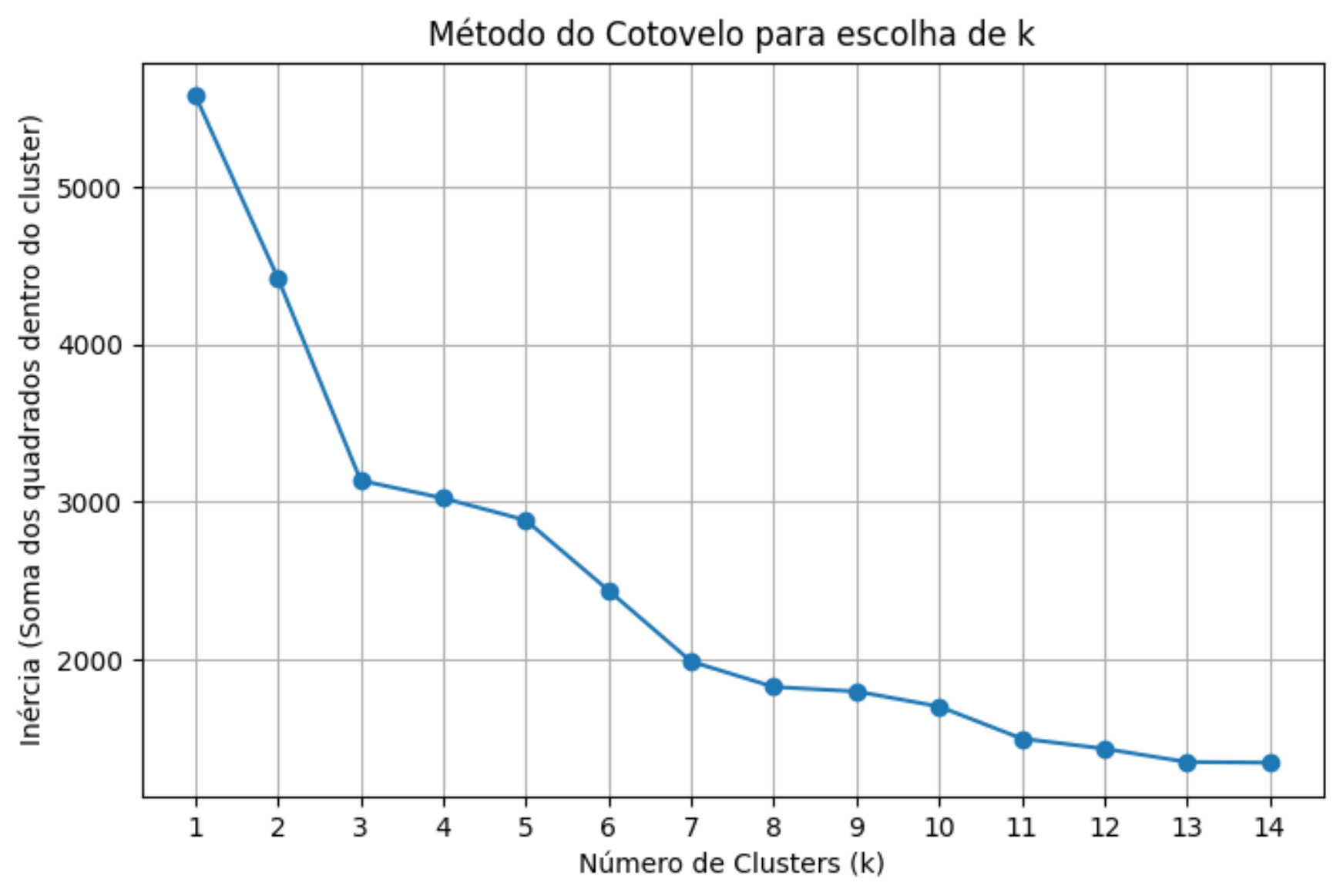
ETAPA 3- HIPÓTESE

- Hipótese:
 - Existem perfis distintos de ocorrências aeronáuticas que se agrupam segundo características da aeronave (como número de motores, peso e assentos), fatores contribuintes (humano, material, operacional e outros.) e fases de operação (pouso, decolagem, cruzeiro, manobra, etc).

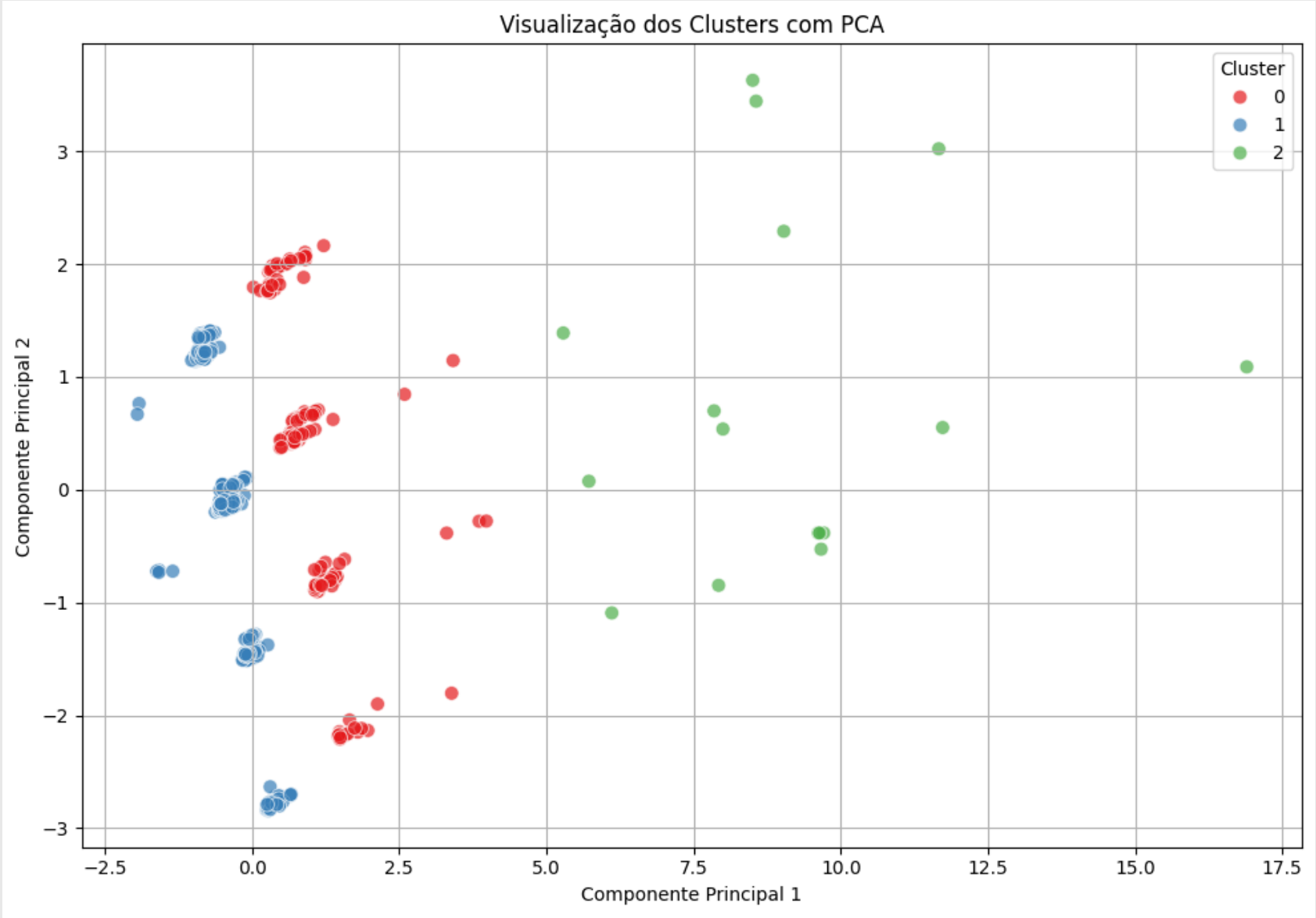
ETAPA 3- METODOLOGIA

- Método escolhido:
 - Clusterização com Kmeans/método Elbow;
 - Padronização do Z-score;
 - Variável categórica transformada por one-hot encoding;

ETAPA 3- RESULTADOS



ETAPA 3- RESULTADOS

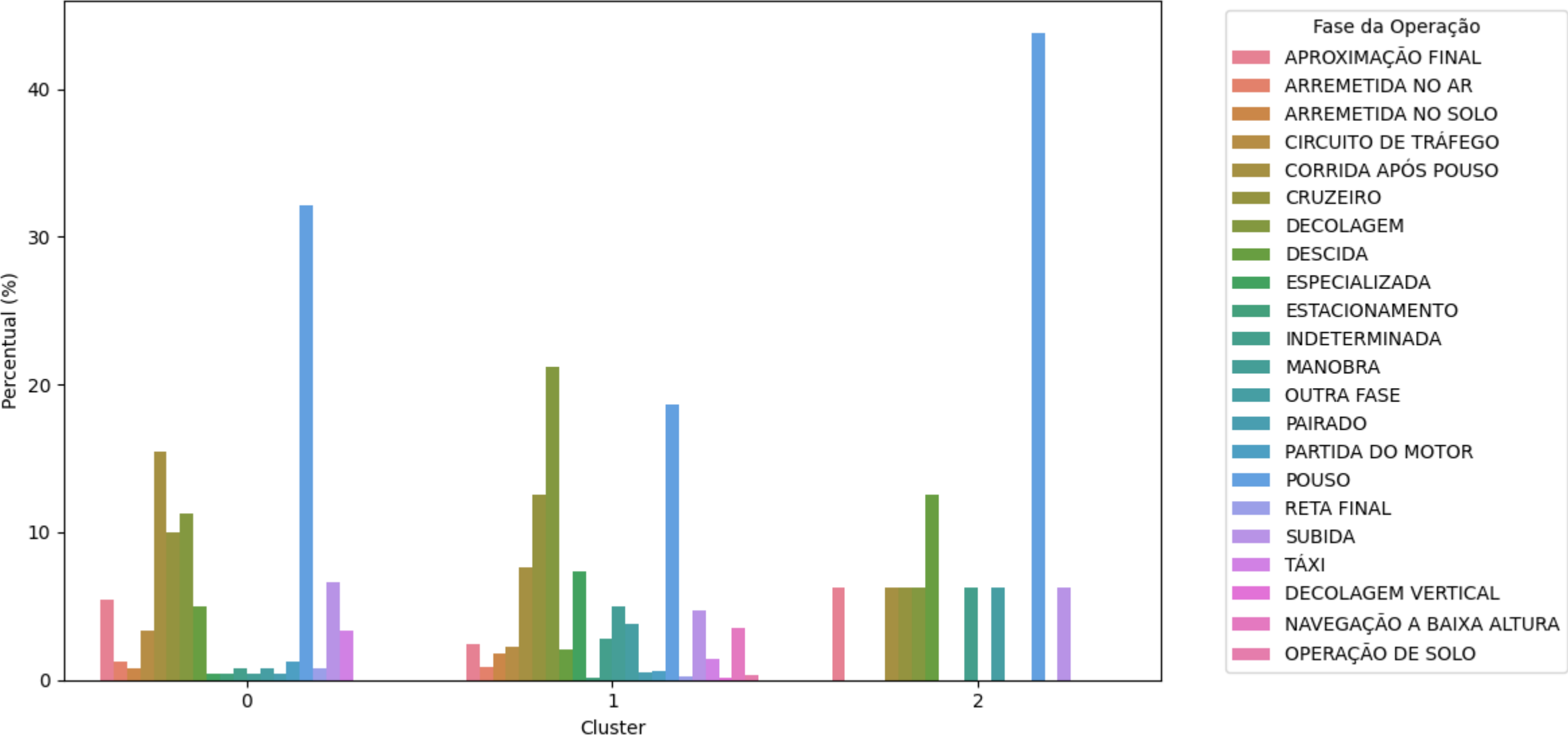


ETAPA 3- CLUSTERS

Grupos	Peso médio(Kg)	assentos	Dano médio	Fatores	Fase
Cluster 0	Aeronaves de pequeno porte				
	~3.676	~9	~1.9 (substancial)	Operacional (78%) Humano (27%)	Pouso(78%) corrida pós pouso (27%) Decolagem e cruzeiro
Cluster 1	Aeronaves muito leves				
	~1.393	~3.3	~2 (substancial)	Operacional (72%) Humano (17%)	Múltiplas
Cluster 2	Aviões comerciais/grande porte				
	~110.263	~135	~0.94(Leve)	Operacional(87%) Humano(50%)	Pouso(44%) Descida(12%)

ETAPA 3- CLUSTERS

Distribuição percentual das fases da operação por cluster



ETAPA 3- ANÁLISES ESTATÍSTICA

Método	Resultado
Silhouette Score	0.352
Calinski-Harabasz Index	405.5
Davies-Bouldin Index	1.1

- **Silhouette score:** Clusters moderadamente definidos, porém com sobreposição;
- **Calinski-Harabasz Index:** Relativamente alto, conseguiu segmentar bem;
- **Davies-Bouldin Index:** Existe proximidade entre os clusters, o que é esperado devido a perfis similares.

ETAPA 3- CONCLUSÃO

- A análise exploratória revelou padrões relevantes nas ocorrências aeronáuticas no Brasil, destacando a importância de entender os dados sob diferentes perspectivas.
- A aplicação de clusterização permitiu identificar três perfis distintos de acidentes, diretamente associados a características técnicas das aeronaves, aos fatores contribuintes e às fases do voo mais críticas.
- Essa segmentação evidenciou que diferentes perfis operacionais enfrentam riscos distintos, e que estratégias de prevenção genéricas podem ser ineficazes. Os resultados sugerem que ações personalizadas de mitigação, orientadas por dados, podem elevar significativamente a segurança operacional na aviação civil.